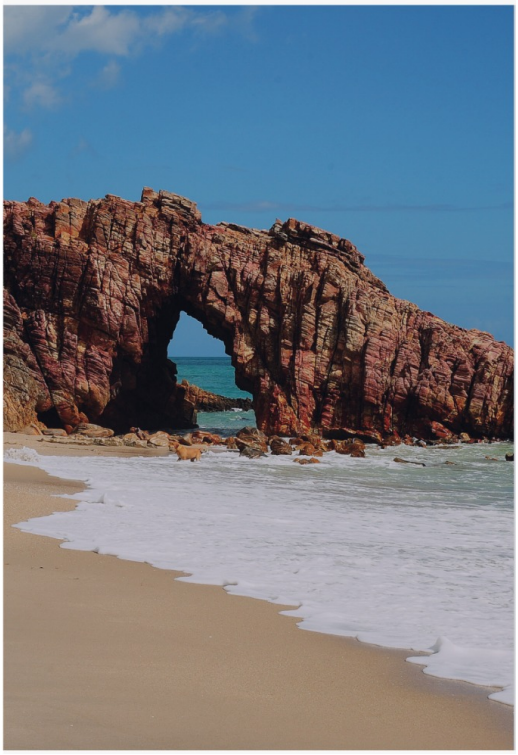
ADVANCE'2023

Fortaleza - Jericoacoara, Brazil
February, 2023



Proceedings of the ADVANCE'2023

https://lrsm.ibisc.univ-evry.fr/ADVANCE2023

ADVANCEs 2023

10th International Worshop on Advances in ICT Infrastructure and Services

Fortaleza - Jericoacoara, Brazil

Foreword

It is a great pleasure to present you ADVANCE'2023: the 10^{th} International Workshop on ADVANCEs in ICT Infrastructures and Services. This year, the event was held in Fortaleza and Jericoacoara (Brazil), returning to being an in-person event after two years of the COVID-19 pandemic. During the pandemic, ICT technologies played a key role in our societies, as the virus had dramatically limited our social face-to-face interactions. Communication technologies are helping us to mitigate constraints, and we envision that in the following years advances in networking, distributed services, and distributed infrastructures can also significantly help manage sanitary conditions. ADVANCE represents, therefore, a fantastic forum for discussion about how current ICT technologies and their research trends should evolve to meet our current and future emergency needs in the ICT domain. The focus of the ADVANCE series of workshops is to provide a forum for the publication, presentation, and discussion of relevant efforts of the worldwide scientific community, practitioners, researchers, and engineers from both academia and industry on the latest theoretical and technological advances in ICT.

After the successful organization of the 1st ADVANCE workshop in 2012 in Canoa Quebrada (Brazil) with the support of IFCE Aracati, the 2^{nd} edition was held in the city of Morro de Sao Paulo (Brazil). In 2013, with the support of IFCE, the 3^{rd} edition was held in Miami (USA). In 2014, with the support of IFU, the 4^{th} edition was held in Recife (Brazil). In 2015, with the support of UFPE, the 5^{th} edition was held in the city of Evry Val d'Essonne (France). In 2017, with the support of UEVE/Paris Saclay, the 6^{th} edition of the workshop was held in the beautiful city of Santiago de Chile (Chile), with the support of the Universidad De Chile (UC). The 7^{th} edition was held in Cape Verde Islands with the support of the Universidad de Cabo Verde. In 2020, the 8^{th} edition was held in the city of Cancun (Mexico) with the support of the Universidad del Caribe and the Universidad Autonoma de Yucatan. The 9^{th} edition was held online, with the support of the University of Zaragoza (Spain) and University College Cork (Ireland).

Fortunately, this 10^{th} edition of the workshop event is being held in-person in Fortaleza and Jericoacoara with the support of the Federal University of Ceará (UFC) and the Federal Institute of Education, Science and Technology of Ceará (IFCE-Fortaleza). ADVANCE'2023 consists of five main sessions, with three Invited Talks and four short professional courses. The technical sessions address current hot topics, including Mobile Computing and Offloading; SmartGrid, Smart Cities and Internet of Things; Cloud and Fog computing; Optical network; Network Slicing; Block-Chain technologies; UAV communication and VANETs; AI Applied to infrastructures and services; and Social data management. The five Technical Sessions consist of 8 full papers and 6 short papers. The first invited talk about "Performability Modeling: Complexity vs. Representativeness" was given by the invited speaker Prof. Paulo Romero Martin

Maciel (Federal University of Pernambuco - Brazil). The second invited talk, about "Intent-Based Networking", was given by the invited speaker Prof. Nazim Agoulmine (University of Evry - France). The third invited talk, about "Scaling Blockchain" was given by Prof Abdelhakim Hafid (University of Montreal - Canada).

We appreciate the presence of all participants, invited speakers, and the 44 authors that submitted their papers to ADVANCE'2023. We would also like to acknowledge the work of the 35 members of the Technical Program Committee for their hard work in reviewing the submissions. Finally, we thank our colleagues from UFC and IFCE for the organization and for making possible the celebration of ADVANCE'2023 in the beautiful state of Ceará in Brazil. Prof

Paulo Antonio Leal Rego, TPC Chair.

Table of contents

Technical Session 1 (Full Papers)

10th International Workshop on ADVANCEs in ICT Infrastructures and Services

Usability and security of electronic voting systems

Safia Aouragh*, Fanny Kalinowski *, Karima Boudaoud†
*UCA,Polytech'Nice Sophia
Email: safia_aouragh@hotmail.fr, kalinowski.fanny@orange.fr
†UCA, CNRS, I3S
Email: karima.boudaoud@univ-cotedazur.fr

***Abstract*—Nowadays, the interest towards electronic voting systems is increasing. However, in the existing e-voting systems several issues need to be resolved regarding usability and security. Is an e-voting system usable? Does it guarantee security? To answer these questions, we have conducted a comparative study of a non-exhaustive list of e-voting systems with a focus on usability and security properties. We have first identified the main tasks of an e-voting system and the usability and security properties that must be provided by a e-voting system. Then, we have analyzed what are the different tasks and usability and security properties provided by the selected e-voting systems and how they have been implemented, as well as their impact on the functioning of the system. Our study highlighted the strong relationship between security and usability. More specifically, we came to the conclusion that even if a system fulfills many security properties, its security can be fully guaranteed if it is usable. Finding the good tradeoff between security and usability increases the confidence of the users in these systems, which is an essential parameter.**

I. Introduction

Currently several e-voting systems have been developed. An e-voting system is a voting system that involves the use of electronic means. In general, two main types of electronic voting can be distinguished: electronic voting supervised by the physical presence of an electoral authority, such as electronic voting machines in polling stations or municipal offices; and electronic voting under the sole influence of the voter, not physically supervised by any electoral authority. For example, voting from a personal computer via the Internet or by mobile phone, including by SMS[1].

When designing e-voting systems, generally the developers have focussed more on providing an e-voting system that fulfils the voting process and offer a secure system[1] [2][3][4][5] than on the usability aspects. Some experiments and research have been done to try to find a good compromise between usability and security [5][6][7].

The objective of this paper is to present the results of a comparative study that we have conducted on a non-exhaustive list of e-voting systems with a focus on usability and security properties. We have first analysed how the existing systems work to extract and then define the main tasks and steps necessary for a e-voting system. Then, we have identified all the usability and security properties that must be provided by a e-voting system, after identifying all the properties provided by existing systems. Noticing that these properties were not defined and not evaluated in the same way by the studied systems, we have proposed a consensual definition for each property. Finally, we have analyzed what are the different tasks and properties of usability and security provided by the selected systems, and how they have been implemented to understand their impact on the functioning of the system.

The rest of the paper is organized as follows. In Section 2, we define the main tasks of an e-voting system and describe how these tasks have been implemented by the five e-voting systems that we have selected. In Section 3, we provide the list of usability and security properties that we have identified after analysing the existing e-voting systems and give a consensual definition. In Section 4, we analyse the usability and security properties provided by the five e-voting systems. Finally, we conclude this paper and provides an outlook on future work.

II. The tasks of the e-voting processes

An e-voting process includes several tasks, where a task is an activity associated with one or more steps in achieving an objective. In this study, we have identified and defined eight tasks:

- **Documentation**: at this step, the voters gather information about the parties running in the election and the process to prepare their vote. For the election organizers, this documentation phase consists of thinking about and comparing the methods to be adopted in order to set up the vote.
- **Identification**: it is necessary for the voters to identify themselves. This task allows them to establish their identity by declaring it with a unique identifier. More simply, the voters answer the question: "Who are you?".
- **Authentification**: After declaring their identity, the voters must authenticate themselves. The double authentication allows to prove that the voter is who she/he claims to be. In a client/server relationship, authentication can work both ways. The server needs to know who is actually accessing its site or information, but the client is equally legitimate in verifying that the server is the system it claims to be.
- **Making a choice**: after the authentication phase, the voters make a choice from among those available to them, such as the parties running in the election. This decision is unique: they can choose only one option, and they will not be able to reconsider their decision.

[1]E-Voting: International Developments and Lessons Learnt

- **Validation of a choice**: at this stage, the voters validate their choice.
- **Vote verification (voter side)**: after validating their choice, the voters can check that it has been added in the ballot and has been counted. This individual verification can also take place on the content of the vote.
- **Vote verification (assessor side):** the verification task is also performed by the voting organizers. The assessors can check the votes universally to make sure that the vote was correctly cast and that the count is correct.
- **Feedback sharing**: Finally, the voters are often invited to share their feedback via questionnaires or interviews.

In our study, we have analysed the implementation of these tasks for each selected voting system.

A. Prêt à Voter

Prêt à Voter is an e-voting system that has been the subject of an experiment piloted in May 2007 at Newcastle University. 105 volunteers were invited to vote for a donation to one of the Campus' institutions: Oxfam, Barnados and UNICEF [8].

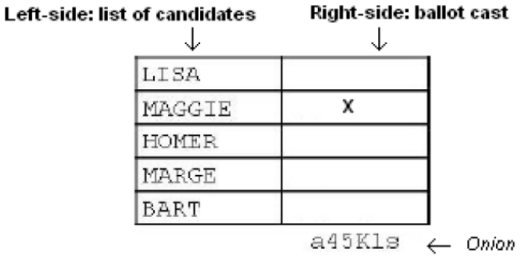


Fig. 1. System ballot of Prêt à Voter [8]

a) Documentation phase: Within the university, emails were sent to the mailing lists of each department. The volunteers present at the campus also encouraged participation in the charity vote.

b) Identification: A demographic questionnaire was distributed to each participant. Participants were asked to provide information about their age, educational background and past voting experience. This questionnaire replaced the presentation of identity document.

c) Authentication: This task was not available. This is probably due to the experimental aspect of the study.

d) Making a choice: As we can see it on Fig. 1, on the left side of the ballot, the names of the charitable organizations are listed. The order in which the names appeared is random. The right side of the ballot is reserved for the voter's choice. The voters have to make a cross next to the charity their choose to support.

e) Validation of the choice: After making a choice, the left part of the ballot, containing the order of the candidates is destroyed. Voters separate the ballot in two parts and keep only the right part, containing their choice and the onion. The onion encapsulating the order of the list of candidates is encrypted on several layers with the public keys of different officials and representatives of the parties. To decrypt the onion, and thus determine the value of the vote, the officials have to collaborate. The right side of the ballot is scanned into the system by an assessor.

f) Vote verification (voter): Voters receive a receipt on which is indicated the boxes they have checked, the encrypted onion and a serial number. The receipt allows voters to ensure that the machine has recorded their choice. With the serial number, they are able to track their vote online. On the Web Bulletin Board, their number appears if the ballot has been counted.

g) Vote verification (assessor): The successive layers of encryption of the onion make the counting of the election collaborative.

h) Sharing feedback: The participant completetes a SUS questionnaire.

B. Scratch Card

Scratch Card is a variation of Prêt à Voter. [9] proposes an incremental improvement approach to the manual voting system used in the UK to ensure the secrecy of the vote with an analysis focused on the technical-social dimension.

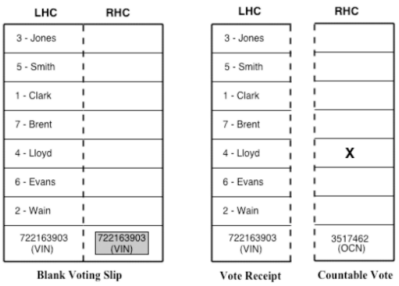


Fig. 2. Scratch Card System paper ballot before (left) and after (right) it has been counted [9]

a) Documentation phase: This phase is not specified by the authors.

b) Identification: Voters give their name and the assessor makes sure they are registered on the electoral list.

c) Authentication: This phase is not specified by the authors. Probably, the ID card is requested to prove the identity.

d) Making a choice: The ballot is similar to the *Prêt à voter* ballot (see Fig.2), except that at the bottom of each column a unique voting identification number is printed - VIN (Voting Identification Number). The voters can choose the ballot of their choice. The names of the candidates are printed on the left column (LHC), their order varies on each ballot. This portion of the ballot serves as a receipt to be kept after voting. The right-hand column (RHC) is used to collect the candidates' choice and is carried forward into the vote counting process. The RHC is actually a "scratch card" containing a small rectangle of opaque coating that initially masks a pre-printed code. This code (OCN) identifies the order in which the candidates' names were printed on the left-hand column. The VIN copy at the foot of this HCR is printed on this opaque coating.

e) Validation of the choice: Voters separate the ballot in two parts and keep only the LHC. This does not give any information about the content of the vote. Voters place the RHC in a closed box. Even before scratching the VIN on the RHC and revealing the OCN, the assessors publish the VIN at the bottom of the RHC on the web.

f) Vote verification (voter): Using the VIN on the LHC, the voters can verify that their vote has been counted on the Web Bulletin Board. The Web Bulletin Board is an online, publicly accessible display of the various registered votes that have been encrypted. The voters can verify that their VIN number is displayed. Also, the voters have the right to select any ballot. They are encouraged to select more than one, to ensure that the order of the candidates is random. Also, on the ballots having the list of candidates in the same order, the voters can scratch off the VIN on the RHC. By doing so, they can check that the OCNs are indeed identical, i.e. they do match this list order.

g) Vote verification (assessor): The counting is done by the officials at the counting center and must of course be supervised. The assessors must not count damaged ballots, i.e. ballots whose VIN has been scratched off, revealing the order of the candidates. If the ballot has not been damaged, the assessors scrape the VIN to reveal the OCN. Until the VIN is scratched off, it can be used as evidence that a vote was cast and not subsequently lost. To avoid any attempt to undermine the anonymity of the vote by recording VIN-OCN pairs, the RHCs are shuffled. Associating a sequence of OCNs with any recorded sequence of VINs that had previously obscured them is more difficult.

h) Sharing feedback: No post-vote interviews or questionnaires were conducted.

C. Code Voting

Code Voting is the subject of a study that proposes three approaches for vote registration [6]. The study involves 18 participants.

a) Documentation phase: Code Voting explores three approaches to record the vote: a manual approach, an approach using a QR-code and an approach using palpable objects. Unlike manual voting where voters enter their choice directly, here voters enter a code that represents their choice. The sheet containing voters and their associated code is distributed before the elections. The associated code is adapted to each approach. A series of numbers, a QR-code or a palpable object.

1) Manual approach:

b) Identification: A default login and password are provided on the code sheet distributed in advance. These parameters are not ”specific” to the voter.

c) Authentication: Not specified by the authors. In this experiment, authentication is meaningless because voters identify themselves with default data.

d) Making a choice: Voters first enter the serial number of the code sheet. This allows matching the code number with a voter. Then, they enter the code corresponding to their choice.

e) Validation of the choice: Voters confirm their choice. Their vote is automatically sent to the electronic ballot box.

f) Vote verification (voter): The electronic ballot box sends an acknowledgement code. If this code matches the one on the code sheet, the participants have confirmation that their vote has been recorded.

g) Vote verification (assessor): This task is not specified by the authors.

h) Sharing feedback: The participants complete an SUS questionnaire and a UEQ questionnaire.

2) QR-code approach: Only the task *Making a choice* differs from the manual approach.

d) Making a choice: The voters scan the QR code which encodes the serial number of the code sheet. This allows matching the code number with the voter. Then, they scan the QR code corresponding to their choice.

3) Approach with palpable object: Only the task *Making a choice* differs from the manual approach.

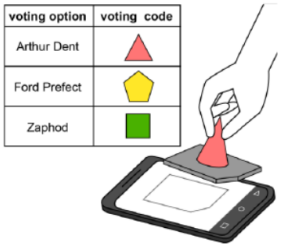


Fig. 3. Modality of palpable object approach of the Code Voting system [6]

d) Making a choice: Voters place first the item corresponding to the serial number of the code card on a touch screen (see Fig. 3). The system detects it and decodes the serial number. Then, voters put the item corresponding to their choice on the touch screen. The system detects it, decodes the choice and records it.

D. Blockchain-based e-voting system

We have also studied a fully decentralized e-voting system based on Blockchain and using smart contracts to address security aspects [5].

a) Documentation phase: The list of voters contains at least their name, their national identification number and their fingerprint. Alternatives to fingerprinting exist: sending a pin code to the voter’s number and the voter must provide the pin code to verify themselves, or using a grooming finger if the person does not have a thumb.

As input of the genesis block (parent block of the first block of the chain), the organizers provide the list of eligible voters consisting of the fingerprint and an associated binary value, candidates, date and time of the beginning of the vote, date and time of the end of the vote. The role of the organizer stops here. The program code is previously integrated into the blockchain according to the concept of smart contract. When the starting date and time are reached, a function is called

invoking the election procedure to begin and the corresponding activities are performed.

Based on the number of voters, they are grouped randomly. Each group has a separate schedule, with a starting and ending dates and times for voting. Voting duration (represented by a boolean flag) is adjusted so that no voter suffers from a network slowdown or failure. Each group is notified by message and email. Once the time is up, the voters of the group can no longer access the vote.

b) Identification: Voters give their public key. The program checks whether the voter is part of the group whose turn is to vote (flag = true), and is present on the eligibility list.

c) Authentication: Voters give their private key, which is their own fingerprint. The program converts it into binary and checks if it matches with the list of the genesis block. The SHA-256 hash of this binary value is used as the only representation of the voter in the block.

d) Making a choice: The list of candidates is represented as logos associated with binary values. Once the candidate is chosen, the ballot is created and contains the hash of the binary fingerprint and the choice string. The choice string consists of the candidate's choice hidden in other randomly generated values. The random string consists of randomly generated 0/1 values. A block is created containing the ballot, and another sibling block is created that consists of the voters hash (binary fingerprint), the reference number of the broadcast block, its own reference number, and the opening value of the choice.

e) Validation of the choice: When voters validate their choice, they broadcast the block that contains the bulletin in the chain.

f) Vote Verification (voter): Once all voters voted, all sibling blocks are broadcasted one by one sequentially. The even-numbered nodes start calculating the result by referring to the blocks and extract the candidate's choice for each block. Here, all nodes are expected to get the same result, as no blocks are discarded unnecessarily in between and the blockchain does not support any changes.

g) Vote Verification (assessor): After validation, the peer nodes begin the proof of work on the sister block. The peer that wins the proof of work will be the first to verify that the voter did not vote earlier and that the ballot is in the correct format. After all the checks, the block containing the ballot is added to the blockchain and other peer nodes check and update their chains. The majority is taken into account. If the majority does not agree, the block is rejected.

h) Sharing feedback: No post-vote interviews or questionnaires were conducted.

E. Benaloh Challenge

The Benaloh Challenge is not a voting system but a technique to support the verification of voting intention which is widely implemented in e-voting systems. The experiment chosen [7] presents a realistic scenario. It has been used for the elections of the Federal Parliament of Germany in 2017.

The purpose of this experiment is to verify the usability of the verification process. To do this, the screens are recorded. An intention card, indicating for whom to vote is given to voters, to preserve their privacy and avoid revealing real voting intentions. They will be then able to verifiy if the recorded vote corresponds to the intention on the card, using the site or the mobile. Three verification approaches are proposed: manual, automatic and mobile.

a) Documentation phase: When starting the users are informed about the purpose of the study and that their actions will be recorded. They are asked to sign a consent form. Then, they are asked to fill out a demographic questionnaire: age, gender, occupation, previous voting experience. If the individual is under the age of majority, she/he must provide permission from a legal guardian to participate.

b) Identification: Before starting the experiment, the participants received the information and materials necessary to conduct the verification approach. All the participants had a falsified letter sent by the election authority containing login credentials and a blank space to use for writing down the verification code. The mobile outreach group received a slightly different version mentioning the verification device. In addition, a smartphone was provided with the application pre-installed.

c) Authentication: Not specified by the authors. In this experiment, authentication is meaningless since the credentials are false.

d) Making a choice: Participants select an option and send to the system, which encrypts it. During the encryption, the system generates a random value that acts as a salt to individualize each vote. This avoids that two identical information provide the same fingerprint when hashed. A verification code is generated from the hash of the customer's choice and the salt. At this stage, the participants can either vote by validating the encrypted vote or proceed to a verification. As the verification supported by the Benaloh Challenge is not compatible with the secrecy of the vote, the verified vote cannot be taken into account and must therefore be discarded.

e) Validation of the choice (voter): The participants validate their encrypted choice.

f) Validation of the choice (assessor): The verification is done with the help of a verifier, which is a software that is either present on the voting device (for the manual and automatic approach) or on an auxiliary device like a smartphone (for the mobile approach).

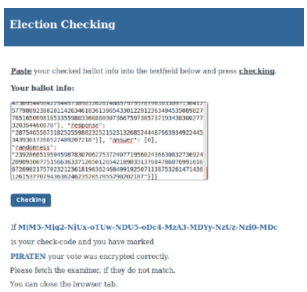


Fig. 4. Manual approach to verification in the Benaloh challenge [7]

1) Manual and automatic approaches: The system generates the data to be checked (option + random value). The participant clicks on the option *Check* (See Fig. 4 and 5.) and chooses the verification entity from a list. The data to be checked is automatically sent to the verification entity. The verification entity compiles a new verification code from the *hash* of the sent data and displays it. The participant compares the original verification code, written on her/his letter, with the newly created one. If the two codes match, the participant has confirmation that her/his choice was correctly encrypted.

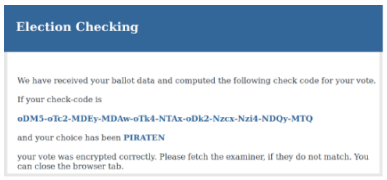


Fig. 5. Automatic approach to verification in the Benaloh challenge [7]

2) Mobile Approach: Instead of writing the verification code on a paper, the participant scans a QR-code that represents the verification code, using a mobile device such as a smartphone. The verification code is then transfered to the mobile device. By clicking on *Verify*, the participant is redirected to a second QR-code, which contains the verification data. She/he scans this QR-code and the mobile device uses the data to recalculate a verification code. This verification code is then automatically compared to the previously scanned code. Even, if participants do not have to compare the two results, they must verify that the mobile has encrypted the right choice. The mobile device displays the name of the candidate and participants must confirm that they have voted for the displayed candidate. (See Fig. 6.) If this is the case, participants have the confirmation that their choice has been correctly encrypted.

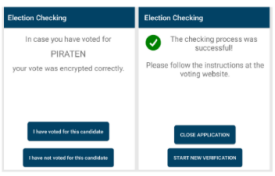


Fig. 6. Mobile approach to verification in the Benaloh challenge [7]

g) Vote Verification (assessor): Not specified by the authors.

h) Sharing feedback: Once the participants declare that they have finished, they are asked to complete the SUS questionnaire. Open-ended questions are included to collect their impressions on the different approaches (problems at the verification phase, whether they were really going to use it, how often). Each question has a space to allow the participants to justify their answers. The participant can also ask questions at the end of the questionnaire.

III. Properties

We have identified all the usability and security properties characterizing the e-voting systems that we have studied. When studying these systems, we have seen that the definitions vary according to the articles (i.e. authors), the e-voting systems and the evaluation methods. A same property can be defined and evaluated differently. Therefore, we had to propose a unified definition for each property and discuss the evaluation methods. In a way, we wanted to create our reference document (i.e. a kind of repository) regarding the usability and security properties cahracterizing e-voting systems. This repository takes into account the sub-properties and adjacent properties. The definitions provided are based on what we found in the litterature.

After definig the usability and security properties, we could conduct the comparative analysis of the e-voting systems.

A. Usability

According to ISO 9241-11[2], usability is the degree to which a system, product or service can be used by specified users to achieve defined goals with effectiveness (performance), efficiency (utility) and satisfaction, in a specified context of use. Usability is thus declined according to three properties, namely:

- **Utility or efficiency:** Is the task feasible? With what means? The interface must meet the needs of the users. Usefulness can be evaluated by doing a task analysis.
- **Performance or efficiency** : Performance is related to the efficiency property defined as the accuracy and degree of completion with which the user achieves specified objectives. Thus, the error rate also impacts the performance property.
- **Satisfaction** : The degree to which the user's physical, cognitive, and emotional reactions resulting from the use of a system, product, or service meet the user's needs and expectations. Although subjective, satisfaction can be evaluated based on user feedback. Different evaluation methods exist, such as the SUS (System Usability Scale) test, the implementation of the Thinking Out Loud protocol and behavioral observation.

Depending on the objectives of the analysis, usability is evaluated according to several evaluation methods, namely:

- *Inspection:* heuristic evaluation allows, through usability guidelines, such as those defined by Nielsen, to find and solve interface problems. The inspection can also be done through the verification of compliance with usability rules and recommendations. ISO standards can be used for standardized inspections.
- *User tests:* The Thinking Aloud protocol asks participants to say anything that comes to their mind as they complete a task. This may include what they are looking at,

[2]https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:fr

thinking, doing, and feeling. Observers are asked to take notes on what participants say and do, without attempting to interpret their actions and words. In particular, they should note where they are having difficulty. The test sessions are mostly recorded on audio and video. This observation phase allows the developers to analyze afterwards what the participants did and how they reacted.

- *The survey:* The demographic questionnaire (age, past experience, social category, etc.) allows to target the groups of users to be tested. Coupled or not with a pre-interview, the participants' background sheds light on the way they will apprehend and use the system. The SUS (Brooke's System Usability Scale) questionnaire is used to measure usability[10]. The SUS uses a Likert scale ranging from "strongly disagree" to "strongly agree" to allow the participants to assess the usability of the system they have tested. The participants must respond quickly to capture the feeling, and if they do not know what to answer, they must respond in the middle of the scale. Although subjective, the SUS score can show extreme correlations or disagreements. The UEQ (User Experience Questionnaire, Laugwitz et al) test measures user experience.

B. Acceptability

Nielsen [11] distinguishes between practical acceptability and social acceptability. Practical acceptability emphasizes the relationship between the proposed functionalities and the ease of use. It therefore includes usability, ergonomic criteria and the interaction model. Social acceptability includes the users' impressions, attitudes and social and normative constraints leading to the choice or support of the use of a given technology.

C. Trust

According to ISO/IEC 10181-1:1996[3], Entity X is said to trust Entity Y for a set of activities if and only if Entity X assumes that Entity Y will behave in a certain way with respect to the activities. In the context of e-voting, the degree of trust is based on security, accuracy, privacy and verifiability. This property is considered critical because it can influence the intention to vote.

D. Consistency

According to Schneiderman, consistency occurs when consistent sequences of actions are required in similar situations[4]. Nielsen adds that users should not have to wonder if different words, situations, or actions mean the same thing[5]. In security terms, consistency would be the fact that for the same sample to be analyzed, the same result is obtained.

E. Accuracy

ISO 5725-1:1994[6] defines accuracy as the correctness and precision of measurement results and methods. All votes cast must be counted.

F. User experience

For Donald Norman, the user experience (UX) is "the responses and perceptions of a person that result from using or anticipating the use of a product, service or system"[7]. In the selected e-voting systems, three properties emerged:

- **Stimulation:** is a hedonic attribute of a product, which can lead to new impressions, opportunities and ideas.
- **Attractiveness :** the Magnus Revang's User Experience wheel[8] presents attractiveness as the combination of several design elements (interface and graphics). Attractiveness in the user experience would then depend on the placement of elements, typography, colors and contrasts, media used and graphic elements.
- **The novelty :** it is estimated to be a period of three months. In this temporary range, the user shows curiosity and discovers the environment.

G. Security

The main security properties are:

- **Integrity :** guarantees that the elements considered are accurate and complete. Integrity guarantees the accuracy of the information. This property ensures that the content of the vote has not been altered from its original intent, nor destroyed in an unauthorized manner (either incidentally or intentionally). We have identified two sub-properties of integrity. Namely:
  - **Transparency :** e-voting systems are often compared to "black boxes". However, it must be possible to audit them (thanks to an analysis conducted by experts and with the release of the source code for example). According to a BeVoting study, a system is transparent if there are audit possibilities adapted to any voter. Everyone must be able to verify that the election was conducted correctly. This audit must allow voters to determine how their vote was composed, and how this vote will be taken into account in the election result.
  - **Verification :** According to He [12] and Riera [13], a system is verifiable if each voter can ensure that her/his vote was included in the final tally. Sako [14] distinguishes between individual and universal verification. In individual verification, the participants must be able to verify whether or not their message reached its destination, but cannot determine whether this is true for other voters. In universal verification, everyone can independently verify that all votes

[3]https://urlz.fr/eVqV
[4]https://bit.ly/3EFmHll
[5]https://www.nngroup.com/articles/ten-usability-heuristics/
[6]https://urlz.fr/eVsd
[7]https://www.usabilis.com/definition-ux-experience-utilisateur-user-experience/
[8]Magnus Revang's User Experience wheel: https://urlz.fr/eVpO

have been correctly counted. We would add that the verification process must ensure that the system has encrypted the vote according to the voter's intent. In this way, no one can falsify the result of the vote [15].

- **Confidentiality :** only authorized persons have access to the elements considered. Confidentiality ensures secrecy. In this context, no one should be able to link a voter to a vote, and the audit should not alter the secrecy of the vote. The privacy of the vote must be preserved during the election as well as afterwards, and over a long period of time. We have identified two sub-properties of confidentiality. Namely:
  - **Privacy :** It is a multidisciplinary concept, with multiple perspectives depending on the type of actors. Westin [16] defines it as the control of the communication of personal data, Altman [17] as the control of interpersonal boundaries. The CNIL defines personal data as "any information relating to a natural person who can be identified, directly or indirectly"[9]. According to this definition, the vote is a personal data, and must therefore remain secret.
  - **Anonymity:** excluding the identity of the voter.

- **Availability:** it is a question of guaranteeing access to a service or a resource. The system must be able to perform a function under predefined conditions of time or performance. An underlying property has been defined, namely :
  - **Scalability :** The ability of an IT device to adapt to demand. It must be able to maintain its functionality and performance in case of high demand.

- **Non repudiation :** it guarantees that a transaction cannot be denied. One sub-property has been identified, namely:
  - **Tracability:** guarantees that accesses and attempts to access the elements considered are traced and that these traces are kept and exploitable. Everything must be recorded in the activity logs, so that it can be traced back in a history.

- **Authentication :** guarantees the origin of an element. This property allows to be sure that the sender is the real one. It ensures the identity of a user. An underlying property has been identified, namely :
  - **Unicity :** each voter has a unique identifier that she/he can attest by authenticating herself/himself. This property guarantees that the voters can vote only once and the ballot is valid only once.

[9]https://www.cnil.fr/fr/definition/donnee-personnelle

IV. Analysis of the voting systems from the repository

A. Prêt à Voter

Prêt à voter was developed to enable vote verification and election audits. It combines a paper trail (for recount) and a sophisticated cryptographic process to ensure secrecy and verification. Usability was evaluated using the SUS questionnaire. Among the 105 voting participants, 70 completed the SUS test.

In terms of effectiveness, 75.8% of participants understood the voting instructions. Only half of the participants understood the value of discarding the left side of the ballot, which contains the list of candidates. Those who did not discard the left side of the ballot may have jeopardized the secrecy of their vote, and therefore their privacy. Indeed, if an individual recovers the left part and attaches the right part kept by the candidate before the latter scans it, she/he can reconstitute the ballot. It is interesting to note that voters did not have much difficulty using the Web Bulletin Board, despite the fact that verification was a completely new experience for them. So when the usability guide is clear, security is enhanced.

The scan that records the content of the vote can be compromised by the behavior of the participants. If the participant withdraws the ballot too soon, the system may incorrectly record the check mark and/or the onion containing the order of the list of candidates. An incorrectly scanned onion renders the ballot null and void because if no match is found, it cannot be deciphered and therefore the order of the candidates is lost. Also, if the cross on the ballot is misshapen, or the expected time during the scan is not considered by the voter, the system may misinterpret its intent. It would undermine the integrity of the vote, against the will of the system.

The SUS score was 68.5. 63.3% of participants were reassured that the marks on the receipt matched those on their report card. This verification step not only provided security, but also improved participant comfort. 40.5% of participants felt confident using the voting procedure. The authors acknowledge that the usability of the *Prêt a voter* is lower than comparable voting systems in the United States. As a complement to the SUS test, the UTAUT model was used to assess acceptability and confidence. Among the 105 voting participants, 53 responded to this survey. 36.2% of participants would agree to use *Prêt a voter* in national or local elections, but 56.9% would prefer to use it for other types of elections. Acceptance is therefore above average when the stakes are lower. To explain this, the authors found a positive correlation between the average value of acceptance and security. In terms of safety, opinions are very divided. 41.5% think the system is vulnerable or very vulnerable to attack, 41.5% think it is somewhat not vulnerable and 17% think it is not vulnerable.

Considering the misunderstanding problems encountered, usability and security could be significantly improved by adding information on the voting procedure.

B. Scratch Card

One of the major challenges of e-voting systems is their adoption by the general public. Indeed, despite the benefits of e-voting systems, the confidence is these systems is decisive for their acceptability. In order to understand what is behind the already established trust in manual voting systems in the UK, the authors analyzed several elements. These include the close supervision of voting and counting by the electoral authorities, the fact that the system is based on physical evidence, kept under seal with the possibility of re-examination, and the simplicity of the process and its steps, enhanced by the several years of experience. The reticence towards e-voting systems is understandable, as thanks to the electronic devices the authorities do not have to supervise the election process, the votes are recorded in a dematerialized way and it is possible to transmit the votes and access to them remotely. The process and its steps are becoming more complex, the volatility of digital data is feared and the verification tasks are being rushed. These changes and new methods obscure the perception of e-voting systems, impede their acceptability and hinder their usability. The ability to review votes in the event of allegations of irregularity in the UK represents a privacy vulnerability as ballots are stamped with a discrete identification code linked to the polling station. What might be perceived as a problem is in fact known but generally accepted. The assumption made in the face of this paradox is that linking a code to an individual is a manually non-trivial task. This paradox shows the importance of the belief in the non-subversive character and the robustness of the system. This is possible only through the trust and transparency of the e-voting system. Thus, it seems crucial to the authors to tend towards these two essential characteristics - the non-subversive aspect and a degree of comprehension accessible to the larger number of users - while preserving usability.

The design of Scratch Card is part of an incremental improvement approach to the manual voting system used in the UK. Partially automated, the e-voting systems are, according to the authors, likely to preserve and/or gain the level of trust already given to manual voting systems. Thus, the use of paper ballots as the norm is retained. In particular, the authors aim to improve the secrecy of the vote, the accuracy and the overall efficiency of the system, rather than innovating the voting medium.

The secrecy of the vote is preserved, until the counting of the votes is done by the scratch card. The scratch card system is a widespread system and the general public is generally familiar with it. Since the secrecy of the vote is ensured, the properties of privacy and resistance to coercion are also enhanced.

Various aspects of the system are being changed and can be automated to speed up the voting process and improve the efficiency of the system. The use of voting machines to record and/or count votes is being considered, due to the large number of votes. To maintain and preserve the confidence of this proposal, two schemes are considered. In the first case, a single voting machine would be present at the counting center. The ballots would be transported in paper format to the counting center where they would be scratched before being scanned for counting by the voting machine. In the second case, each voting center would have a voting machine that would electronically transmit the counts to the counting center. Automated vote counting would eliminate human errors and thus increase the accuracy of the system. In addition, if the recording of the votes is entrusted to a machine, voters would be able to verify that their vote has been counted.

C. Code Voting

Code Voting [6] is the subject of a study that proposes 3 approaches to experiment different ways of recording a vote. This study focused on usability only, security was not addressed. Usability was evaluated through a SUS questionnaire.

All the participants were able to successfully register their vote. The level of efficiency is therefore 100 percent. Although the study sample is not representative of a population with only 18 participants, this indicator is promising for the usability of the three approaches. The QR-code approach had the highest SUS score (84.02), followed by the palpable object approach (78.61) and the manual approach (61.25). To understand the reasons of this score, the authors conducted a UEQ test where the allocation of points ranged from -3 to +3 depending on whether the property was rated as very poor or very good. The following properties were assessed: novelty, stimulation, dependence, reliability, effectiveness, clarity and attraction.

The manual approach had the lowest scores. The evaluation of its innovative character was negative. Its attraction and stimulation score was relatively low, and can be explained by its daunting nature. The participant had to enter multiple codes by hand. 5.5% of the participants were in favour of this approach. They felt it was safer because they were in control.

In contrast, the palpable object approach was recognized for its innovation and stimulation. 38.8% of participants were in favor of this approach. 4 people appreciated its intuitive nature and 3 felt pleasure in voting. Palpable objects are seen as a good alternative to the manual and QR-code for the elderly and visually impaired. Nevertheless, the study was not conducted on such profiles. Therefore, usability needs to be evaluated in more detail on these groups.

Overall, the QR-code approach was evaluated positively. Its clarity and effectiveness received the highest scores of the three approaches. 55.5% of participants were in favor of this approach. The familiarity of the QR-code allowed them to quickly get the hang of the system, and to feel relatively comfortable. 5 people justified this choice by the fact that the objects have a more important manufacturing and distribution cost than the QR-Code. Indeed, the devices must be custom-made for the election. To be recognized on touch screens, a particular recognition technology must be developed in a conductive material. As a result, it will be more difficult to implement this approach on a large scale. The scalability of this approach is compromised.

D. Blockchain

The blockchain [5] is a distributed database. It is public, meaning that all peer nodes in the network can access it. The records ensure traceability. Also, each node has the same data records. This is called consistency. Each activity is transparent; peer nodes can verify and validate it. Each transaction or activity in a block is verified. If the majority of peer nodes do not approve it, the action will not be entered into the registry. Thus, there is both individual and universal verification. Since all nodes have the same records, they must provide the same result as to the outcome of the vote. Non-repudiation is enforced. The majority of the group takes precedence over the individuality of the results. It is also a protection against DoS (Denial Of Service) attacks: all nodes have copies, there is no loss of information.

The blockchain is decentralized. The reduction of the third party is an interesting avenue as the participation of the third party can have a vulnerable effect on the procedure. The smart contract, in the form of code, is executed automatically. It establishes the terms of the contract between the two parties. During the execution of the smart contract, all validation steps are recorded in order to secure all data. This prevents the data from being modified or deleted afterwards. Once a data has been inserted, it is very difficult to falsify it. Dishonest miners must modify the previously broadcast block to insert themselves into the chain, and these modifications must be approved by the other miners on the network. The integrity of the vote is then preserved. Also, attempts to ”double spend” are difficult. Proof of work requires a lot of computing power and energy to generate fingerprints that uniquely identify the blocks. If a node tries to vote twice, its second vote will be rejected as its fingerprint already exists in the booklet.

Finally, the blockchain guarantees confidentiality. The identity of the voter is recorded by the system as the hash of the fingerprint converted into binary. Privacy is preserved, in the sense that personal information is not broadcasted on the network. The voter is anonymous and her/his hashed public key is broadcasted and attests her/his identity. The voting slot is randomly generated and then allocated to groups when needed. It is more difficult to plan blackmail attacks or to try to manipulate intentions. Attackers do not know which individuals make up the next group to vote, nor when they will vote.

E. Benaloh challenge

The Benaloh challenge study explores three approaches to conduct a voting intention audit [7]. Since the voter is particularly active in this challenge, usability is a crucial parameter. In fact, the usability of the manual, automatic and mobile approaches was evaluated. Overall, the majority of participants were able to successfully verify their vote. 61.3% successfully completed the manual verification. The automatic and mobile approach had the same completion score of 81.25%. However, the experience of 5 participants was not included in the post-vote analysis either because they dropped out of the experiment or because they encountered technical difficulties preventing the necessary data collection.

The results of the study have shown that the automatic approach was not more effective than the mobile approach, and vice versa. In the manual approach, participants reported difficulties in understanding what they had to copy and paste and what verification data was displayed. Either they misunderstood the instruction and thought they had to copy and paste the verification code, or they did not understand what to copy and paste at all. The interface lacked some information: the status of the verification was not indicated, the instructions were not clear, and error handling was not taken into account. Several readings were required to understand the instructions, resulting in wrong actions that were not detected by the system.

The time required for a successful verification was recorded, with a starting time defined as *t* when the participant were pressing the verification button, and the ending time defined as *t'* when the participant were redirected to the system.

On average, the manual approach took three times longer than the automatic approach. Those who did the check with a smartphone were on average twice as fast as those in the manual approach. However, the mobile approach took a little longer, as the user had more actions to satisfy. QR-code scanning time was characterized by wide variations, ranging from 2 to 15 seconds per QR-code. Some participants found this waiting time too long, others gave up. A common usability issue highlighted by the authors was participants' motivation to verify the voting system. Those who thought the verification was too complex to understand or too time consuming did not want to proceed with the verification. There was a counter-intuitive aspect, participants were checking the voting system and not their personal vote which had to remain secret. If we want to check the voting system efficiently, we need to be able to test it several times and quickly. This is why the speed of execution is also important.

Regarding satisfaction, the automatic approach obtained the highest SUS score with 79.4 points. The mobile and manual approach had a similar score, respectively 75.8 and 75.4. In both the manual and automatic approach, usability issues related to the verification code were reported. Consisting of 43 characters, including both numbers and letters, the sequence was time consuming to copy and compare. It was easy to see that verification errors could occur. If the participants do not notice a mismatch, they may believe that their vote did not derive from the original intention and miss a fraud, or conversely, think that they vote was manipulated. The security of the vote is then compromised, and more particularly its integrity. A careless mistake done by the participant compromises the accuracy of the information. Moreover, QR-codes are limited in terms of character capacity. Thus, the encrypted voting data, too important, cannot be satisfied with a single QR-code. To facilitate the usability of the mobile approach, it would be relevant to consider reducing the number of characters to be able to use a single QR-code. On the other hand, since security also depends on the size of the verification

data sequence, this option must be subject to a risk analysis. Indeed, if the size of the sequence is reduced, brute force attacks aiming at establishing a correspondence between the vote and its verification code will be faster and have more chances to succeed. It would be interesting to analyze the size of the verification data sequence in such a way that the risk of compromising the security is acceptable. This problem is an example of the importance of having a trade-off between security and usability. An alternative to this scalability problem is the one used in Estonia, where the QR-code redirects to a link where the verification data are located.

V. Conclusion

End-to-end verifiable e-voting systems facilitate the verification of the integrity of individual votes during the election process. More specifically, end-to-end verification methods allow voters to confirm that their votes have not been manipulated by the client. Verification can be done in two ways.

- Verification based on the vote cast; ensuring that the voting system has encrypted the vote corresponding to the voter's intention.
- Verification based on the recorded vote; ensuring that the vote recorded by the voting system corresponds to the vote cast, and that it is correctly included in the election result

The common finding of the e-voting system experiments studied is that the approaches using QR-code are the most successful. To explain this, we hypothesize that the widespread use of this technology makes it more acceptable and usable. In the Benaloh Challenge [7], we saw that human errors in verifying the verification code is aborted, as the integrity of the vote is checked by the system. Unlike manual approaches, QR-code technology offloads the user by automating a few steps that seem burdensome, whether in terms of time, comprehension, or stimulation. While trusting this technology is a lead for usability, QR-code technology is limited in terms of security. Blockchain [5], on the other hand, appears to address many of the trust issues with e-voting systems identified by the authors of the Scratch Card study [9]. Although there is no physical evidence in blockchain, this distributed database concept in which all nodes have a copy of the information prevents information loss. Better yet, blockchain provides the traceability property that is essential to ensure non-repudiation. However, blockchain is still an abstract concept. In terms of usability, this can be problematic as it contributes to the obscure perception of the cryptography supported by e-voting system technologies. In addition, traditional voting systems were usually delegating the responsibility of managing the vote to an election authority whereas blockchain proposes a system that is decentralized. This can be not well perceived by the users and potential voters who can have a certain reluctance since the responsibility of the good functionning of the vote is no longer delegated to a single entity (to a trusted third party) but to individuals.

To increase the adoption of electronic voting systems, it is important to take into account both usability and security and find the best compromise between them. As future work, we plan to improve an existing e-voting system focussing on providing the best tradeoff between usability and security based on the properties and tasks that we have defined in the context of this study.

Through our analysis we have also identified a third inextricably linked criterion related to voter education. Indeed, voters' perception regarding the act and process of voting, as well as the relationship to electoral participation or democracy is considerable in the adoption of a new system. We believe that this is an area that should be explored in future work.

References

[1] Ameen, Zinah J Mohammed. Secure Electronic Voting Application Based on Face recognition and Ciphering. Journal of Computer Science Applications and Information Technology, s. d., 2018

[2] Adida, Ben, et Ronald L. Rivest. Scratch Vote: Self-Contained Paper-Based Cryptographic Voting. In Proceedings of the 5th ACM Workshop on Privacy in Electronic Society - WPES '06, 29. Alexandria, Virginia, USA: ACM Press, 2006. https://doi.org/10.1145/1179601.1179607.

[3] Chaum, D. Secret-Ballot Receipts: True Voter-Verifiable Elections . IEEE Security Privacy Magazine 2, no 1 (janvier 2004): 38-47. https://doi.org/10.1109/MSECP.2004.1264852.

[4] Weldemariam, Komminist, Adolfo Villafiorita, et Andrea Mattioli. Assessing Procedural Risks and Threats in E-Voting: Challenges and an Approach. In E-Voting and Identity, edited by Ammar Alkassar and Melanie Volkamer, 4896:38-49. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. https://doi.org/10.1007/978-3-540-77493-8_4.

[5] Sadia, Kazi, Md. Masuduzzaman, Rajib Kumar Paul, et Anik Islam. Blockchain-Based Secure E-Voting with the Assistance of Smart Contract. In Proceedings of the IETE International Conference on Blockchain Technology (IC-BCT 2019) Blockchain Technologies. Mumbai, India, 2020. https://doi.org/10.1007/978-981-15-4542-9_14.

[6] Marky, Karola, Martin Schmitz, Felix Lange, et Max Mühlhäuser. Usability of Code Voting Modalities . In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 1-6. Glasgow Scotland Uk: ACM, 2019. https://doi.org/10.1145/3290607.3312971.

[7] Marky, Karola, Oksana Kulyk, Karen Renaud, et Melanie Volkamer. What Did I Really Vote For? On the Usability of Verifiable E-Voting Schemes . In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, 1-13. Montreal QC, Canada: ACM Press, 2018. https://doi.org/10.1145/3173574.3173750.

[8] Marco Winckler, Regina Bernhaupt, Philippe Palanque, David Lundin, Kieran Leach, Peter Ryan, Eugneio Alberdi, Lorenzo Strigini. Assessing the usability of open verifiable e-voting systems: a trial with the system Prêt à Voter . https://www.researchgate.net/publication/228559244_Assessing_the_usability_of_open_verifiable_e-voting_systems_A_trial_with_the_system_Pret_a_Voter

[9] Randell, Brian, et Peter Y A Ryan. Voting Technologies and Trust. In Proceedings of the Formal Aspects in Security and Trust, Third International Workshop, FAST 2005, Newcastle upon Tyne, UK, July 18-19, 2005

[10] Brooke, John. SUS - A Quick and Dirty Usability Scale , s. d., 7.

[11] Nielsen, J. (1993) Usability Engineering. Academic Press, Inc., Harcourt Brace Company, San Diego, USA.

[12] He Q., Su Z. (1998) “A New Practical Secure e-Voting Scheme”, IFIP/SEC’98, Austrian Computer Society, Austria, pp. 196-205.

[13] Riera A., Borrell J., Rifa J. (1998) “An Uncoercible Verifiable Electronic Voting Protocol”, IFIP/SEC’98, Austrian Computer Society, Austria, pp. 206-215.

[14] Sako K., Kilian J. (1995) “Receipt-Free Mix-Type Voting Scheme: A Practical Solution to the Implementation of A Voting Booth”, EUROCRYPT’95, Malo, France, pp. 393-403.

[15] Fujioka, Atsushi, Tatsuaki Okamoto, and Kazuo Ohta, *A Practical Secret Voting Scheme for Large Scale Elections*, s. d., 2.

[16] Westin Alan F. Privacy and Freedom. New York: Atheneum Press, 1967

[17] Altman, I. The environment and social behavior: Privacy, personal space, territoryand crowding. Monterey, CA.: Brooks/Cole, 1975.

Evaluating the use of the Homomorphic Algorithm on Computational Offloading

Francisco A. A. Gomes[1], Filipe Fernandes S. B. de Matos[1], Paulo A. L. Rego[2], Fernando A. M. Trinta[2], and José N. de Souza[2]

[1] Federal University of Ceará (UFC), Crateús, CE, Brazil
almada@crateus.ufc.br, and filipe.fernandes@crateus.ufc.br
[2] Group of Computer Networks, Software Engineering and Systems (GREat)
Federal University of Ceará (UFC), Fortaleza, CE, Brazil
pauloalr@ufc.br, fernando.trinta@dc.ufc.br, and neuman@ufc.br

Abstract

Mobile accounts for almost half of the web traffic worldwide. However, these devices still have computational and energy limitations. Mobile Cloud Computing (MCC) tackles problems like this by migrating tasks and data to remote cloud environments. This technique is known as offloading. However, during this procedure, data is transmitted on the network without protection, which is unfeasible for applications with confidential data that cannot be exposed without security. This work propose a module to ensure the security of data transmitted during computational offloading and that it is easily expandable to any cryptographic algorithms. This work evaluates the performance of computational offloading when adopting Homomorphic cryptography during data migration. The results showed that depending on the operation performed, the time in the offloading process increases up to 900 times.

1 Introduction

The last few years have shown great popularization and growth in the use of mobile devices in society (e.g., smartphones, tablets, and smartwatches). These devices have more and more processing capacity and an increasing number of embedded sensors (e.g., temperature, luminosity, accelerometer, gyroscope), which enable the sensing of environmental data so that they can be interpreted and processed by various applications. However, the mobility provided by mobile devices brings limitations to their use. Furthermore, while using applications like these, users often enter sensitive information that could be exposed to various types of cyber attacks.

To get around the issue of limited resources on mobile devices, research has promoted the integration of mobile devices and cloud resources, creating the research area known today as Mobile Cloud Computing (MCC) [8]. Among the topics addressed in MCC, the most prominent is the use of the *offloading* technique to mitigate processing and storage problems on mobile devices by migrating tasks and data to remote cloud environments [3, 18]. As it is a shared medium, it is essential that any data transmitted during offloading be protected, especially those containing sensitive user information, such as bank details and medical exam results.

One of the methods adopted for data protection in these environments is encryption. Encryption makes the data unreadable at the sender, sends it across the network, and from that nonsense data retrieves the original data at the receiver. Usually, this is performed by encryption algorithms, which are basically programs that implement the idea of a chosen encryption method [4]. In addition to the classic techniques, such as symmetric (AES) and asymmetric (RSA) encryption, homomorphic encryption also guarantees data privacy[1][10]. Furthermore,

this kind of encryption allows operating directly on the encrypted data, while other techniques require decrypting the data before operating on it.

Using homomorphic encryption during the offloading can benefit the client, as it allows tasks or submitted data to be processed by the server without access to the original content. On the other hand, adopting homomorphic encryption requires a computational effort from the transmitter, which can be a problem for mobile devices and their limited resources.

Thus, this paper evaluates the impact of using homomorphic cryptography during the computational offloading performed by mobile applications. For this, we developed an extensible security module to add encryption algorithms to assist the developer in building secure applications. The results showed that depending on the operation performed, the time in the offloading process increases up to 900 times when using homomorphic encryption. The most significant contribution of this work is to present to developers the cost-benefit of using this type of encryption and, in this way, help them decide when to adopt it.

The rest of this article is organized as follows: Section 2 deals with the theoretical foundation and presents the fundamental concepts of the research. Section 3 presents related works. Section 4 presents an overview of the proposal, architecture, established definitions, and an application developed from the proposed solution. Section 5 presents the results of experiments carried out to show the solution's performance in terms of total time, communication, encryption, and execution. Finally, Section 6 concludes the article and exposes possible future works.

2 Background

This Section presents the main concepts, definitions, and characteristics related to the areas on which base this research. Thus, we discuss Mobile Cloud Computing, Offloading, and Homomorphic Encryption concepts.

2.1 Mobile Cloud Computing

MCC's main objective is to mitigate the problems of mobile computing (*e.g.*, energy, processing, and storage limitations). There are many definitions for this paradigm. According to [21], MCC is a computational paradigm that exploits the advantages of Cloud Computing to mitigate the problems of mobile computing and integrates these two areas. According to [15], MCC was created based on the concepts of Cloud and Mobile Computing and aimed to allow applications that require more sophisticated computing resources to run on different mobile devices, providing a good experience for users. Among the various techniques associated with MCC, the most cited in the literature is offloading [17]

2.2 Offloading

According to a research conducted by [8], mobile devices use the offloading technique to reduce energy consumption and improve computational performance by migrating processing and data to equipment with greater computational power and storage. The offloading operation differs from traditional Client-Server architecture's Request/Response mechanism. In the Request/Response mechanism, the servers are always responsible for processing tasks transmitted by the client. In the offloading operation, the client can process tasks if there is no Internet connection or if the mobile device does not benefit from delegating the task computation to the server [22].

Lin et al.[18] indicate two types of offloading operations: computational and data. Computational offloading is an operation that delegates the processing performed on the mobile device to another execution environment (Cloud), aiming to prolong battery life and increase computational capacity. Data offloading aims to extend the mobile device's storage capacity, sending the data to a machine with greater storage capacity.

In addition to the Cloud, mobile devices use other remote devices as offloading targets, such as cloudlets and other mobile devices. According to [24], cloudlets are server instances allocated on the same network as the clients and can handle offloading requests. Thus, when cloudlets are used, offloading remains close to client devices and generates advantages such as higher speed rates and lower latency rates [5].

2.3 Security Techniques

Due to the data migration required when offloading a mobile device to a remote environment, this data can travel on the network without any protection. Thus, there must be security in this migration. Traditional symmetric or asymmetric encryption methods are adopted to provide security in cloud environments. However, processing the data related to offloading must be decrypted on the server that has access to the original data, leaving it vulnerable [11].

2.3.1 Homomorphic Encryption

The homomorphic encryption method allows data to be encrypted and sent to the server. Even so, operations can be performed on that data without needing to decrypt it until it returns to the sender. It is not necessary to have access to the original text to manipulate it [2]. In addition to the operations of encrypting and decrypting data, this technique uses addition and multiplication operations.

Homomorphic encryption systems can be divided into fully homomorphic and partially homomorphic. Drozdowski et al. [7] state that a completely homomorphic system is a system that supports any number of addition and multiplication operations on the data. This first type of system was proposed in [12], but it proved inefficient in terms of processing time. Therefore, most systems with homomorphic properties are partially homomorphic systems [25]. Unlike the one mentioned above, this type of system is computationally practical but comes with the cost of supporting only limited mathematical operations on encrypted data. A partial solution mainly contains two operations: additive or multiplicative homomorphic encryption schemes. In partial, Paillier cryptosystems support addition, and ElGamal cryptosystems support multiplication.

3 Related Works

This Section presents the works related to the present proposal. These works consist of security solutions in the computer offloading the mobile device to a remote environment.

Gomes et al.[14] presented an analytical study on the impact of encryption algorithms on the performance of computational offloading performed by mobile applications. The work developed a security module that guarantees the confidentiality and integrity of data trafficked in offloading and added a framework that supports offloading [13]. Such a module has two components: one on the mobile side and one on the cloud. Such components adopt a hybrid encryption approach, using symmetric and asymmetric encryption algorithms to transmit information over the network. Therefore, the data object being transferred during offloading is encrypted; thus,

the privacy of the mobile device's data and, consequently, of its user is preserved. A notable weakness of this work is security in the offloading process because when the data arrives at the server, it needs to decrypt to operate on it. At this point, the server has access to the plain data, which can often be personal information that should not be visible.

Liu et al.[19] presented an implementation that adopts Steganography techniques in the computational offloading of images from mobile devices to remote servers. The main objective of this work is to provide data security and examine the energy consumption spent in this process. Through these techniques, offloading hides the data to be sent in an image and, on the server side, processes it in this hidden format. In addition, the authors propose an image recovery method based on the block data hiding method. Unfortunately, the authors did not evaluate the performance of the Steganography technique in offloading in terms of the total time of the operation.

Ren et al.[23] claim that data privacy concerns are increasingly affecting the Internet of things (IoT) and artificial intelligence (AI) applications, in which it is very challenging to protect the privacy of the underlying data. In recent, the advancements in the performances of homomorphic encryption have made it possible to help protect sensitive and personal data in IoT applications using homomorphic encryption-based schemes. This paper proposed a practical homomorphic encryption scheme that can enable data users in IoT systems to securely operate data over encrypted data, which can effectively protect the privacy of key data in the system. Furthermore, experiments were carried out to verify the encryption and decryption time and homomorphic operations, presenting a result that has little impact on offloading but does not show the amount of data in which these experiments were carried out. Thus, it is noted that a more detailed analysis was not carried out on the impact of this type of encryption at the processing level for the various scenarios with a large amount of data, which is what happens in a mobile cloud computing scenario.

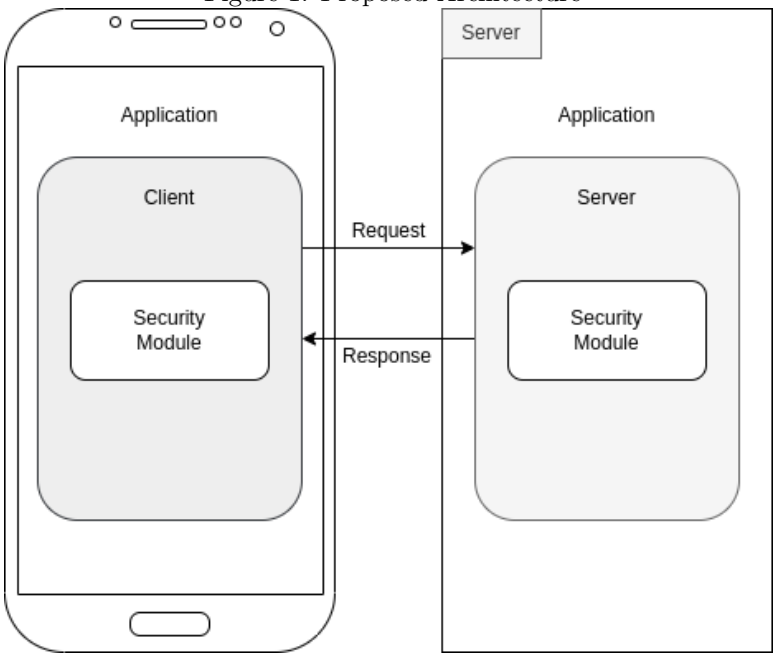
4 Proposal

Previous sections have pointed out relevant issues in the areas of mobile computing (*e.g.*, limited power, storage, and processing) and Mobile Cloud Computing (*e.g.*, lack of security during offloading). Thus, a solution was created that helps develop applications that use computational offloading to mitigate the problem of scarce resources on mobile devices. In addition, such a solution implements security techniques for the privacy of the data transferred during the procedure. Paillier's algorithm was adopted for the homomorphic cryptography process, the main public-key algorithm in which homomorphism is applied to its encryption and addition operations are performed.

4.1 Architecture

Figure 1 presents the proposed architecture. A security module was implemented to have encryption algorithms available to the application developer. The idea is that this module is used to ensure the security of data transmitted during computational offloading and that it is easily expandable to other cryptographic algorithms. The module is composed of two parts: a client and a server. At the beginning of computational offloading, the application invokes the client module to encrypt the data. Next, the encrypted data is sent to the server, which performs the desired computation directly on them in the case of homomorphic encryption. The result is returned to the client once processing is complete. Finally, upon receiving the encrypted result, the client module decrypts it and delivers the raw result to the application.

Figure 1: Proposed Architecture



The modules were developed using the Java programming language. To use homomorphic cryptography, we adopted the external library Javallier[1], which provides a set of methods that encrypt data and compute operations of addition and multiplication of numbers. When the server receives the encrypted data from the client, it can perform different procedures depending on the technique used for data security (in principle, it has the implementation of homomorphic encryption), and the server performs the necessary processing on the encrypted data itself. The result of the operation, of course, will also be encrypted data that will be promptly sent to the mobile application.

4.2 LoadBench

LoadBench[2] is a mobile application developed from the proposed solution that adopts the concept of benchmarking, where it is possible to measure the processing time of the requested/performed operation. The application is divided into a client and a server (allocated, respectively, on a mobile device and a Cloudlet). *LoadBench* supports the offloading of tasks using the homomorphic encryption technique. The tasks computed by *LoadBench* are three mathematical operations: 1) factorial of an integer (from 50 to 200) chosen by the user; 2) sum and 3) multiplication of square matrices, with dimensions of 50x50, 100x100, ..., 1000x1000 (defined by the user) and composed of random values of type *Integer* generated by the application.

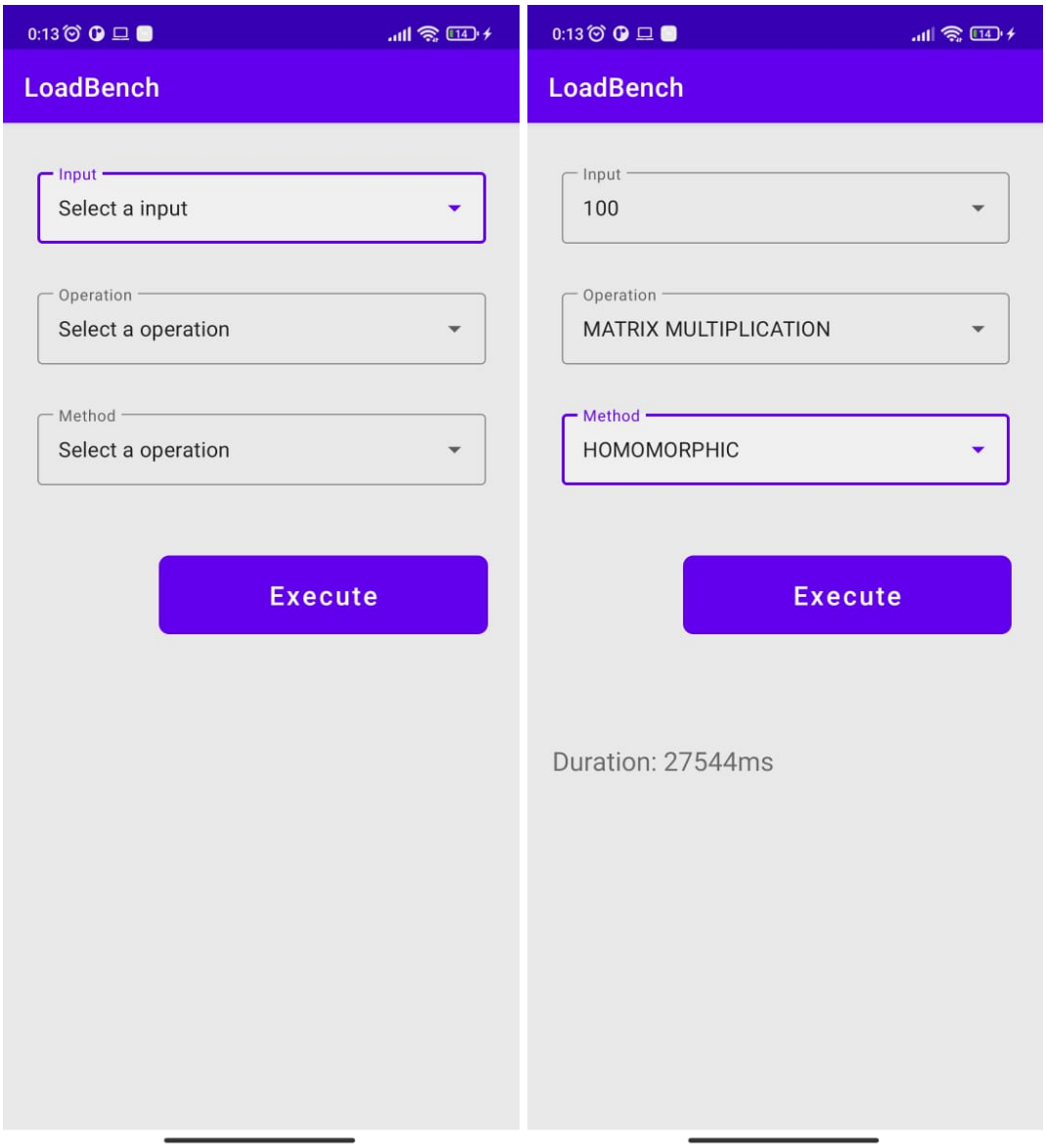
The operation of the application and the communication between its components occurs as follows: once the mobile application is started (Figure 2.a), the user chooses the application's input data (matrix dimensions or operand factorial), the desired operation (factorial of a number, addition or multiplication of matrices) and whether to use the homomorphic data

[1] https://github.com/n1analytics/javallier

[2] https://github.com/henrique010/tcc-implementation

protection method or not (unsafe mode). After, the user requests the execution of the operation by clicking the *Execute* button. If the mobile device establishes a connection with the remote server, offloading is performed according to the previously chosen mode. Otherwise, the application processes the task on the device. At the end of the computation, the application displays the time required to compute, locally or remotely, the task (Figure 2.b).

Figure 2: Application Screens of LoadBench



6

5 Experiments and Results

To evaluate the impacts caused by the adoption of homomorphic encryption techniques during the computational *offloading* of mobile devices, the application *LoadBench* and a server were used to perform the processing requested by the mobile application. This application was used to perform tests, evaluating the total time of *offloading*, communication, and execution of the operation by the server.

5.1 Execution Environment Description

For the experiment, a *smartphone* and a *laptop* were used. The *smartphone* has the following features: Xiaomi Redmi Note 10; Qualcomm Snapdragon 678 2.2Ghz Octa-Core 64-bit processor; 4GB of main memory; 64GB of internal memory; and Android 11.0.0 (Google API 30). The *laptop*, which acts as a *cloudlet*, has the following settings: Acer Aspire A31542G; AMD Ryzen 5 3500u processor; 8GB DDR4 RAM; 256 SSD; Video Card Radeon 540x 2GB; and Ubuntu 20.04 LTS.

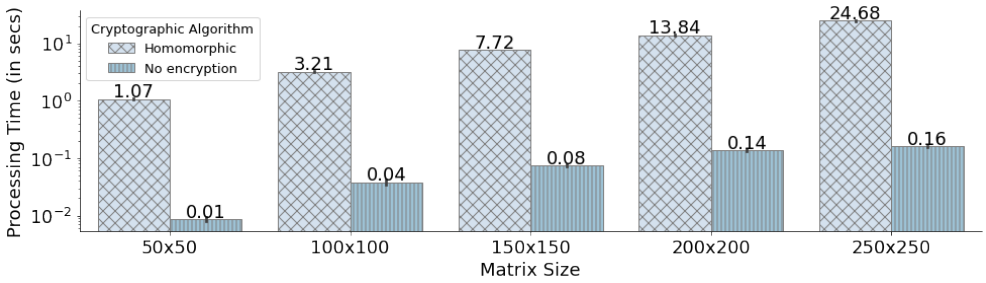
5.2 Description of the experiment

For the present research, the application *LoadBench* was adopted, and the following metrics were evaluated: 1) the processing time of offloading (total); 2) communication time (consists of upload and download time of data), and 3) execution time (time taken by the server to compute the task). For the experiment, 5 ranges of dimensions of rows and columns were empirically chosen for the matrices. Initially, starting with dimensions that always vary by fifty in the difference of rows and columns, such as 50x50, 100x100, 150x150, 200x200, and 250x250. Due to homomorphic encryption presenting RAM memory limitations when encrypting matrices with dimensions greater than 500x500, dimensions of this type were disregarded in the test. Regarding the mathematical operation, we have chosen matrix sum and multiplication operations because of their high computational complexity ($\theta(n^2)$ and $\theta(n^3)$, respectively). We run each scenario thirty times because, according to the Central Limit Theorem (TCL), when you have a sufficiently large sample, the probability distribution of the sample mean can be approximated by a normal distribution [9].

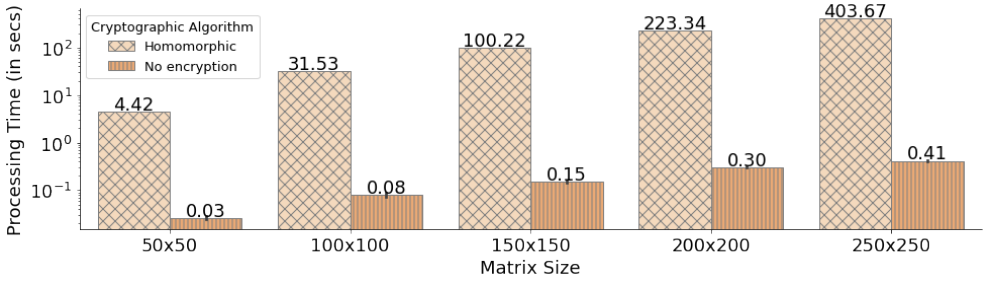
5.3 Results Obtained

Figure 3 presents the average processing time spent with offloading for the Homomorphic encryption algorithm used in the solution. Thus, in the following paragraphs, the results of using this encryption will be compared with the non-use. When comparing the results, we noted that the processing time is higher when adopting homomorphic encryption for all scenarios evaluated. We already expected this behavior, as the application's security layer was added, and it consumed more time to encrypt data before transmission. Notably, when performing the matrix multiplication operation, we noted that the increase in offloading time was 147.3 times for 50x50 matrices and 984.56 times for 250x250 matrices. We observed a similar result to the matrix addition operation, where the offloading using homomorphic encryption was 107 times (50x50) and 154.25 (250x250) slower than without encryption. These results indicate that, even for small arrays, this type of encryption is not suitable for using offloading for mobile applications.

Figure 4 presents the results related to the communication time, execution and encryption in the *offloading* processing. In this case, the No Encryption case (Figures 4(a) and (c)) has the



(c) Sum of Matrices



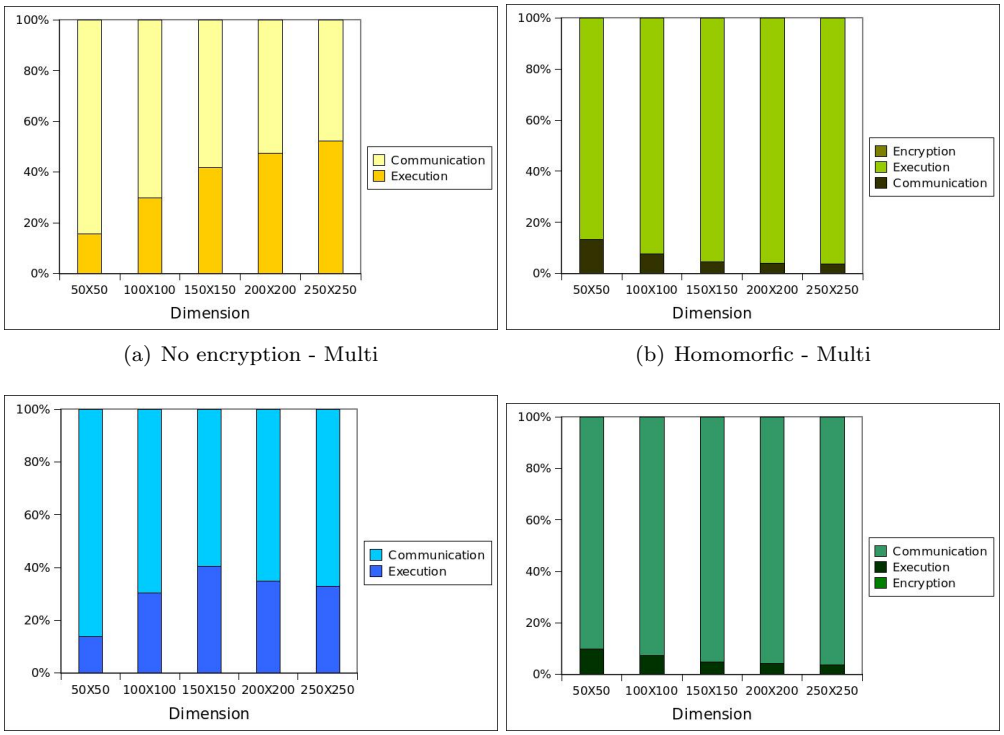
(d) Multiplication of Matrices

Figure 3: Results related to processing time metric (on logarithmic scale)

processing time divided only between communication and execution since no type of encryption is performed. As we observed in previous results when the Homomorphic encryption technique was adopted, most of the total offloading was dedicated to computing the operation on the server (Figure 4(b) and (d)). In addition, regardless of the technique used, as the matrix entries increase, the time spent to compute the operation significantly influences the total time spent in the process. It is also possible to notice that the inverse happens for the encryption time because, as the entries increase, the time spent with encryption proves less influential. From all that has been exposed, it is possible to conclude that using the homomorphic encryption technique will bring benefits related to data security. However, if the problem to be solved has a high complexity (multiplication), it will be a very inefficient process as that the data object to be trafficked has an increasing size. Suppose the problem to be solved is less complex (addition of matrices). In that case, the process takes 43 times more time, which confirms the inefficiency of this type of encryption for mobile devices, but that depending on the method to be performed, this impact it becomes smaller.

5.4 Statistical Data Analysis

Until now, we based our processing time analysis on the average values of the thirty repetitions performed in each scenario. However, there is no guarantee that these average values represent the results related to them well. Thus, we decided to apply a statistical test to assess whether the mean values are good representations of the results and, consequently, to reinforce the observations presented so far. Initially, we divided the results obtained into five groups. Each group contained the results of the two types of offloading performed (with homomorphic encryption



(a) No encryption - Multi

(b) Homomorfic - Multi

(c) No encryption - Add

(d) Homomorfic - Add

Figure 4: Percentage of processing time dedicated to Communication, Execution and Encryption operations

and without encryption) for each array size (50x50 to 250x250). As each group consisted of two subgroups of unpaired samples, we conducted the necessary tests to assess whether or not it would be feasible to apply the T-Student test. By applying the Bartlett test, we verified that the sample variances were not equal for all matrix sizes, which makes using the T-Student test unfeasible. Because of this, we chose the Mann-Whitney test [20]. In the Mann-Whitney test, the null hypothesis (H0) indicates no difference between the evaluated means, i.e., rejecting H0 means that the means differ. Table 1 shows the results of the Mann-Whitney test. We noticed that the null hypothesis was rejected for all matrix sizes. Thus, we conclude that the means shown in Figure 3 are statistically relevant and summarize the behavior of each method of offloading in each scenario.

Matrix Size	**50x50**	**100x100**	**150x150**	**200x200**	**250x250**
Sum of Matrices					
Mann-Whitney	$2.80 \cdot 10^{-11}$	$2.98 \cdot 10^{-11}$	$2.98 \cdot 10^{-11}$	$3.00 \cdot 10^{-11}$	$2.97 \cdot 10^{-11}$
Reject H0?	✓	✓	✓	✓	✓
Multiplication of Matrices					
Mann-Whitney	$2.95 \cdot 10^{-11}$	$3.00 \cdot 10^{-11}$	$3.00 \cdot 10^{-11}$	$3.00 \cdot 10^{-11}$	$2.99 \cdot 10^{-11}$
Reject H0?	✓	✓	✓	✓	✓

Table 1: Mann-Whitney test results related to the processing time metric

5.5 Discussions

The results were quite unfavorable to adopting homomorphic cryptography in a computational offloading scenario. Although some works in the literature have shown that conventional cryptographic algorithms are feasible in computational offloading[14], they still demand that the data be decrypted before task processing effectively, which can be an opportunity to access raw data. Therefore, by overcoming this disadvantage, homomorphic cryptography makes computational offloading even more secure, although significantly slower. This section presents ideas to speed up computing a task using homomorphic cryptography faster, maintaining data privacy.

The first method is caching to map input and output data. The server machine can save the results of operations performed in memory and link them to the received operands. Thus, when receiving a new offloading request with the same entry, the server does not need to compute the same operation again. Instead, it just returns the result saved in memory. Another approach would be to use a more efficient programming language to compute the task. In this case, the developer would need to use the [6] computational offloading multi-language approach. Although more laborious in terms of development, the technique could significantly reduce the processing time of the task and make offloading faster than local processing if the developer chooses the most appropriate server language. A third mechanism would be to improve the hardware of the server machine that will process the task submitted in offloading. The cloudlet used in our experiments has a good hardware configuration for a personal computer. However, it is still weak compared to powerful servers allocated in the Cloud and/or Fog. Therefore, using more powerful server machines can significantly accelerate task processing using homomorphic cryptography, especially when the task is computationally more complex and involves larger input parameters. Even parallel computing can be used to improve the performance of homomorphic encryption. The work [16] introduces a generic method to perform arithmetic operations on encrypted matrices using a homomorphic system and presents the using matrix operations in parallel.

6 Conclusion and Future Works

The present work presented an implementation of a client-server architecture focused on providing security using a homomorphic encryption algorithm for the offloading operation used in the scenario of applications for mobile devices, in addition to having an extensible solution for any encryption algorithm. The purpose is to present a possibility for developers of the *Android* platform and the Java programming language to safely use offloading resources from the implemented algorithms. As a result, it was shown that the module's implementation with the homomorphic cryptography algorithm is evaluated in terms of offloading time, which proved inefficient for the mobile application scenario. A limitation was observed when performing addition and multiplication operations using homomorphic cryptography for matrices with dimensions of 500x500 or greater due to the lack of resources related to the device's main memory when using this technique.

In future works, the following improvements are proposed in this work: *(i)* Implement a Java Annotation, which receives parameters (*e.g.*, security level, required performance) and, from them, defines the cryptographic algorithm to be used; and *(ii)* Analyze the energy consumption of offloading when performed using the proposed solution.

7 Acknowledgments

This work is partially supported by INES[3], CNPq grant 465614/2014-0, FACEPE grants APQ-0399-1.03/17 and APQ/0388-1.03/14, CAPES grant 88887.136410/2017-00

References

[1] Abbas Acar, Hidayet Aksu, A. Selcuk Uluagac, and Mauro Conti. A survey on homomorphic encryption schemes: Theory and implementation. *ACM Comput. Surv.*, 51(4), jul 2018.

[2] Abbas Acar, Hidayet Aksu, A Selcuk Uluagac, and Mauro Conti. A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys (Csur)*, 51(4):1–35, 2018.

[3] Ahmed Aliyu, Abdul Hanan Abdullah, Omprakash Kaiwartya, Syed Hamid Hussain Madni, Usman Mohammed Joda, Abubakar Ado, and Muhammad Tayyab. Mobile cloud computing: taxonomy and challenges. *Journal of Computer Networks and Communications*, 2020, 2020.

[4] Steve Burnett and Stephen Paine. *RSA Security's official guide to cryptography*. McGraw-Hill, Inc., 2001.

[5] Costa et al. Mpos: A multiplatform offloading system. 2015.

[6] Filipe F. S. B. de Matos, Paulo A. L. Rego, and Fernando A. M. Trinta. An empirical study about the adoption of multi-language technique in computation offloading in a mobile cloud computing scenario. In *Proceedings of the 11th International Conference on Cloud Computing and Services Science - CLOSER,*, pages 207–214. INSTICC, SciTePress, 2021.

[7] P. Drozdowski, N. Buchmann, C. Rathgeb, M. Margraf, and C. Busch. On the application of homomorphic encryption to face identification. In *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2019.

[8] Niroshinie Fernando, Seng W Loke, and Wenny Rahayu. Mobile cloud computing: A survey. *Future generation computer systems*, 29(1):84–106, 2013.

[9] Hans Fischer. *A history of the central limit theorem: From classical to modern probability theory*. Springer Science & Business Media, 2010.

[10] Caroline Fontaine and Fabien Galand. A survey of homomorphic encryption for nonspecialists. *EURASIP Journal on Information Security*, 2007:1–10, 2007.

[11] Joffre Gavinho Filho, Gabriel P Silva, and Claudio Miceli. A public key compression method for fully homomorphic encryption using genetic algorithms. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1991–1998. IEEE, 2016.

[12] Craig GENTRY. Fully homomorphic encryption using ideal lattices. *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 169–178, 2009.

[13] Francisco AA Gomes, Paulo AL Rego, Lincoln Rocha, José N de Souza, and Fernando Trinta. Caos: A context acquisition and offloading system. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 1, pages 957–966. IEEE, 2017.

[14] Francisco AA Gomes, Paulo AL Rego, Fernando Antonio Mota Trinta, Windson Viana, Francisco Airton Silva, José AF de Macêdo, and José N de Souza. A study about the impact of encryption support on a mobile cloud computing framework. In *CLOSER*, pages 400–407, 2020.

[15] Dijiang Huang and Huijun Wu. Mobile cloud computing taxonomy. In *Mobile Cloud Computing*, pages 5–29. Elsevier, 2018.

[16] Xiaoqian Jiang, Miran Kim, Kristin Lauter, and Yongsoo Song. Secure outsourced matrix computation and application to neural networks. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security*, pages 1209–1222, 2018.

[3] www.ines.org.br

[17] Karthik Kumar, Jibang Liu, Yung-Hsiang Lu, and Bharat Bhargava. A survey of computation offloading for mobile systems. *Mobile networks and Applications*, 18(1):129–140, 2013.

[18] Hai Lin, Sherali Zeadally, Zhihong Chen, Houda Labiod, and Lusheng Wang. A survey on computation offloading modeling for edge computing. *Journal of Network and Computer Applications*, 169:102781, 2020.

[19] Jibang Liu, Karthik Kumar, and Yung-Hsiang Lu. Tradeoff between energy savings and privacy protection in computation offloading. pages 213–218, 2010.

[20] H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1):50 – 60, 1947.

[21] Shahryar Shafique Qureshi, Toufeeq Ahmad, Khalid Rafique, et al. Mobile cloud computing as future for mobile applications-implementation methods and challenging issues. In *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, pages 467–471. IEEE, 2011.

[22] Paulo AL Rego, Philipp B Costa, Emanuel F Coutinho, Lincoln S Rocha, Fernando AM Trinta, and Jose N de Souza. Performing computation offloading on multiple platforms. *Computer Communications*, 105:1–13, 2017.

[23] Wang Ren, Xin Tong, Jing Du, Na Wang, Shan Cang Li, Geyong Min, Zhiwei Zhao, and Ali Kashif Bashir. Privacy-preserving using homomorphic encryption in mobile iot systems. *Computer Communications*, 165:105–111, 2021.

[24] Mahadev Satyanarayanan, Paramvir Bahl, Ramón Caceres, and Nigel Davies. The case for vm-based cloudlets in mobile computing. *IEEE pervasive Computing*, 8(4):14–23, 2009.

[25] Leandro V Silva, Pedro Barbosa, Rodolfo Marinho, and Andrey Brito. Security and privacy aware data aggregation on cloud computing. *Journal of Internet Services and Applications*, 9(1):1–13, 2018.

12

SmartGrid Improvements on Photovoltaic Systems by Analyzing Energy and Weather Measures*

Vagner Henrique de Souza[1], Wendell Rodrigues[2], Matheus Lima[3], Cláudio Marques de Sá Medeiros[4], and Edilson Mineiro Sá Júnior[5]

[1] PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
vagner@lit.ifce.edu.br
https://orcid.org/0000-0002-0928-1609
[2] PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
wendell@ifce.edu.br
https://orcid.org/0000-0001-6140-3610
[3] PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
matheuslimaufc@gmail.com
[4] PPGCC - Federal Institute of Ceara, Fortaleza, Brazil
claudiosa@ifce.edu.br
[5] IFCE - Federal Institute of Ceara, Fortaleza, Brazil
edilson.mineiro@gmail.com

Abstract

This paper presents an approach for power generation analysis in photovoltaic systems. The objective is to identify the qualitative impact on the energy produced according to fault detection and prediction based on the electrical and weather measurement dataset. Currently, machine and deep learning have been used effectively to recognize patterns in solar panel measurements. Dirt, breakage and shading can greatly reduce the designed efficiency of solar power plants. Thus, increasing the quantity and quality of the failure measurement dataset can significantly improve reliability and its performance.

Contents

*This is a full research paper submitted to ADVANCE 2023 in the area of interest: *Verticales Industries, Usage and Societal Impacts.*

1 Introduction

Smart grid [5][6] is a system that consists of, by using digital solutions, monitoring, controlling and analyzing in bidirectional power flow among all types of generating units like Thermal power station, hydropower generation, solar power plant, nuclear power station, etc. and end-use consumers. It is used to monitor and control the power flow automatically, through a master station with the help of technology. Smart grid technology enables the controlling of components of the power system in generation, transmission, distribution system. It enables monitoring and controlling the power consumption at consumers' premises by the automation system and provides proper control in the distribution system. It also analyzes the power consumption so that it provides proper control in the generation system and transmission automatically. Smart grid can be applied on all layers of energy handling. Regarding energy producing, solar panel systems provide an interesting power source.

Energy produced by solar panel systems is proliferating fast in recent years. This is due, mainly, to the reduced cost and the benefits relied to environment. However, the performance of those systems is impacted by many issues. Ancuta and Cepisca in [1] present fault analysis for photovoltaic (PV) panels that decrease their result according to plant design.

Thus, this work addresses the external factors that directly impact the power loss. For this, we analyze some industries that use solar plants and enumerate the main issues regarding failures. Most of them are related to vandalism and dust, even bird dust. As the dataset for this problem is almost non-existent, we experimentally collected measurement data from a small-scale solar panel plant from our institution and applied some pattern recognition methods to provide a real-time solution to improve the system performance.

2 Photovoltaic Systems

Krauter[2] defines Photovoltaics(PV) as a solar electric power generated via the direct conversion of solar radiation into electricity. Currently, this conversion can achieve around 26% of a theoretical 33% of the incoming solar irradiance.

According to Sumathi et al [3], in 2010, among the world's total power, the contribution of solar PV was a very small fraction. However a rapid growth is observed in the installation of solar PV and would continue to increase in the upcoming years. In the New Policies Scenario, power generated from solar PV in 2035 is over 26-times that of 2010, expanding from 32 Twh to 846 Twh. The energy from installed solar PV would increase from 67 GW in 2011 to 600 GW in 2035, on account of decreasing expenses and government aids.

Thus, the industry and other productive sectors of society have high competitive gains adopting this technology.

Briefly, a single PV device is known as a cell. An individual PV cell is usually small, typically producing about 1 or 2 watts of power. These cells are made of different semiconductor materials and are often less than the thickness of four human hairs. In order to withstand the outdoors for many years, cells are sandwiched between protective materials in a combination of glass and/or plastics.

PV modules and arrays are just one part of a PV system. Systems also include support structures, along with the components, such as inverters, that take the direct-current (DC) electricity produced by modules and convert it to the alternating-current (AC) electricity used to power electrical devices.

3 Fault Detection and Prediction Methods

An effective way to detect pattern changes is comparing a large number of measures in different days in different weather conditions. Usually, voltage and current have small changes if we have same solar irradiance. Hence, an experimental survey was implemented in order to acquire electrical measures from PV in several time and weather conditions. Figure 1 shows the elements of this additional testbed for small-scale analysis adding external interference such as dirties or shadows. During energy production to supply a small set of loads (lamps and sockets), three measures are made: voltage, current (so power), and solar irradiance from an external metering sensor. Current and voltage come from the installed inverter and all these measures are stored as a time-series data. Data are processed using some python scripts in order to be used as input information to pattern recognizing methods. However, in order to validate small-scale data on larger power plants, we have used data acquired from commercial inverters and meteorological station.

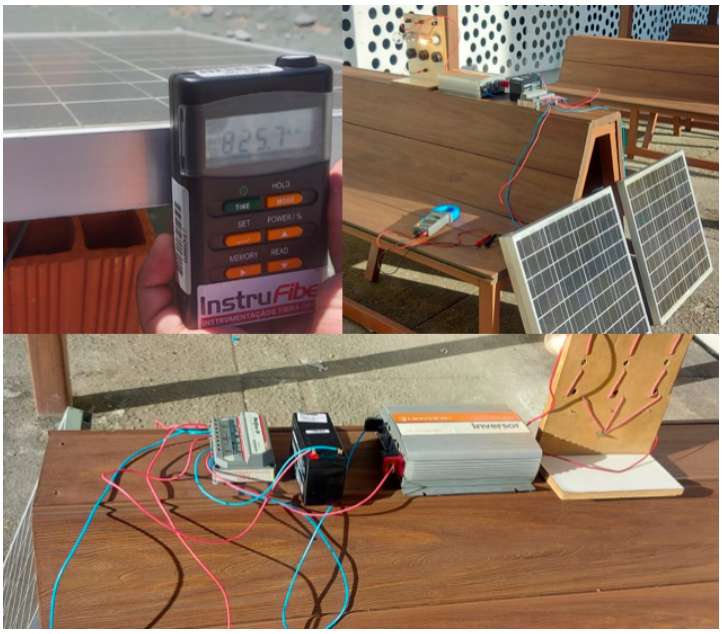


Figure 1: Experimental Dataset Capture. Source: authors

This problem fits typically in regression methods. Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the *outcome variable*, or *target*) and one or more independent variables (often called *features*, *predictors*, or *covariates*). The objective of regression in machine learning is to predict continuous values such as generated power.

3.1 Light Gradient Boosting Machine

Next sections show how interesting is Light Gradient Boosting Machine (LightGBM) in results. LightGBM [7] extends the Gradient Boosting Regressor (GBR) [9] algorithm by adding a type of automatic feature selection as well as focusing on boosting examples with larger gradients.

This can result in a dramatic speedup of training and improved predictive performance. Later is preferable for our approach.

4 PV Measurement Dataset

The data were collected over 6 months in northeastern Brazil during 2021, where the sun is available most of the time. For training and validation procedures, we have acquired data from a commercial inverter connected to some solar panel strings and, also, a weather station. Data were matched in time in order to link measures. Inverters can handle multiple MPPT modules and, thus, allows better layout scenarios.

4.1 Measures

There are many output measures from data source. However, many of them can be calculated from the other columns. Pre-analysis shows non-relevant features and enlightens others such as solar irradiance and temperature. In fact, as the panels transform sunlight into energy, irradiance plays a key role in production. However, it cannot be disregarded that the temperature and even the wind speed (which helps to cool the panels) directly contribute to the best performance in power generation. Thus, dataset is provided as shown in Table 1 with 35336 rows and 6 columns.

DateTime	Irrad	StTemperature	WindSpeed	PnTemperature	Power
2021-05-01 05:22:00	0	31.3	0.0	29	0
2021-05-01 05:23:00	0	31.3	0.0	31	0
...	...	...	...	...	...

Table 1: Cross Data Measures from Inverters and Weather Station

5 Results

Dataset was split in a twofold subset. First one has 24735 rows for training and a second one has 10601 rows for validation test. The prediction is based on the regression in search of the generated power according to the characteristics obtained in the field.

5.1 Comparing Methods

Table 2 presents the regression methods comparison ordered by R^2, the coefficient of determination indicating the proportion of variation between generated power and its prediction.

Although the training time for the Light Gradient Boosting Machine (lightgbm) method is the longest, we prefer to use it as the best result knowing that the training time does not have much impact on the final application. Moreover, it is always possible to decrease that time by computing on GPUs.

Still observing the results, it is important to note that other methods such as Random Forests (rf) [8] show expressive gains. The low training time has a good R^2 value and presents a better result for the accuracy indicated in the Mean Absolute Percentage Error (MAPE). This method is an excellent candidate when training time is a critical factor in implementing the application.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE	TT (Sec)
lightgbm	2953.3167	3.523394e+07	5933.3738	**0.8705**	0.5873	0.1750	300
gbr	3076.8711	3.620912e+07	6014.6031	0.8670	0.6267	0.2085	2.186
rf	2821.3143	3.858068e+07	6209.9684	0.8582	0.4255	**0.1496**	9.808
knn	3200.9721	4.239244e+07	6508.1509	0.8442	0.4262	0.1769	231
lr	4094.1461	4.443212e+07	6662.7248	0.8367	0.8606	0.4469	403
lasso	4092.0608	4.443848e+07	6663.1764	0.8367	0.8591	0.4479	62
ridge	4093.5858	4.443253e+07	6662.7510	0.8367	0.8601	0.4469	36
lar	4094.5139	4.443441e+07	6662.8956	0.8367	0.8635	0.4468	41
br	4092.0380	4.443597e+07	6662.9925	0.8367	0.8584	0.4474	77
llar	4120.2396	4.489454e+07	6696.9883	0.8351	0.8407	0.4629	40
et	2879.3797	4.660053e+07	6825.7398	0.8288	0.4145	0.1501	9.893
en	4542.0858	5.224133e+07	7223.4427	0.8081	0.9578	0.7955	84
omp	4672.4580	5.351067e+07	7310.9689	0.8034	0.8998	0.5791	27
huber	3771.4987	5.446469e+07	7374.9502	0.7999	0.6986	0.2994	**1.293**
ada	5317.3509	5.663545e+07	7524.5133	0.7919	0.7790	0.4537	635
dt	3354.4411	6.235039e+07	7894.1402	0.7708	0.4479	0.1737	186
par	4750.7526	7.142720e+07	8354.1221	0.7377	0.7240	0.3689	179
dummy	14337.4690	2.723538e+08	16502.6821	-0.0008	1.6260	4.1296	29

Table 2: Regression Methods Comparison :: gbr(gradient boosting), rf(random forests), knn(k-nearest neighbour), lr(logistic regression),lasso(least absolute shrinkage and selection operator),lar(least-angle regression),br(bayesian ridge),llar(lasso Least angle regression),et(extra tree),en(elastic net),omp(orthogonal machine porsuit),ada(adaboost regressor),dt(decision tree),par(passive aggressive regressor)

5.2 Training Score for lightGBM

A learning curve shows the validation and training score of the estimator for varying numbers of training samples. It allows us to find out how much we benefit from adding more training data and whether the estimator suffers more from a variance error or a bias error. In Figure 2, both the cross validation score and the training score converge to a value that is quite low with increasing size of the training instances.

5.3 Feature Importance

The *feature importance* will return an array where each element of it is a feature of our model. It will tell you, in proportions, how important that feature is to the model, where the higher the value, the more important the feature is to the model.

As expected, irradiance is the most important feature in the model (see Figura 3). However, it is not a direct function of the generation and, in cases where there are external interferences such as dirt on the panels, this feature may indicate an anomalous pattern in the generation.

5.4 Prediction Error for lightGBM Regressor

A prediction error plots the actual targets (generated power on panels) from the dataset against the predicted values generated by our model. This allows us to see how much variance is in the model. The regression model have a identity line well fitting when comparing against the 45 degree line, where the prediction exactly matches the model (Figure 4).

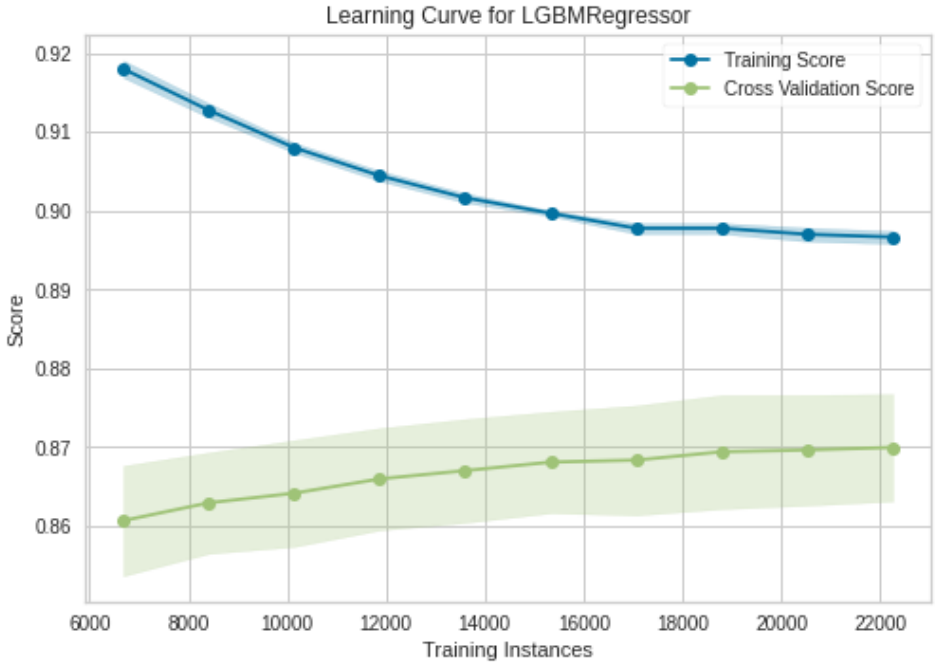


Figure 2: Learning Curve for lightGBM. Source: from experimental results

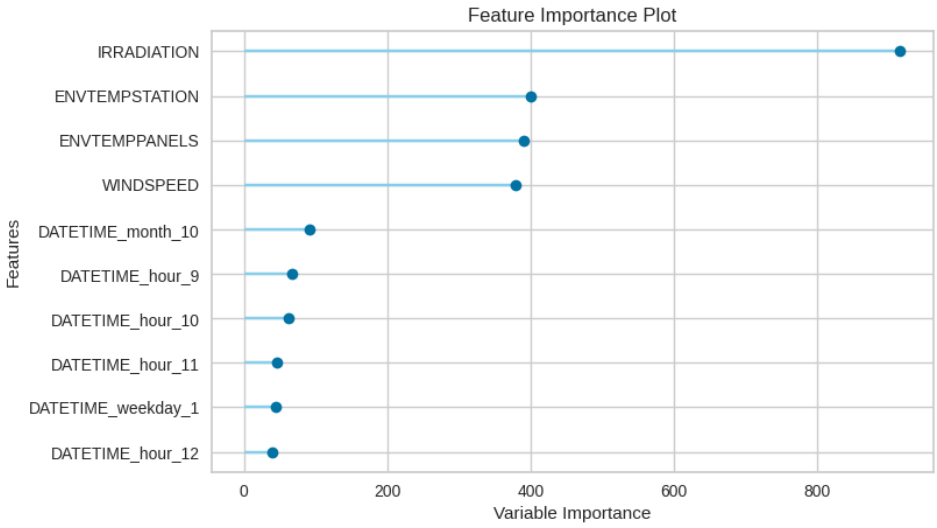


Figure 3: Feature Importance. Source: from experimental results

6

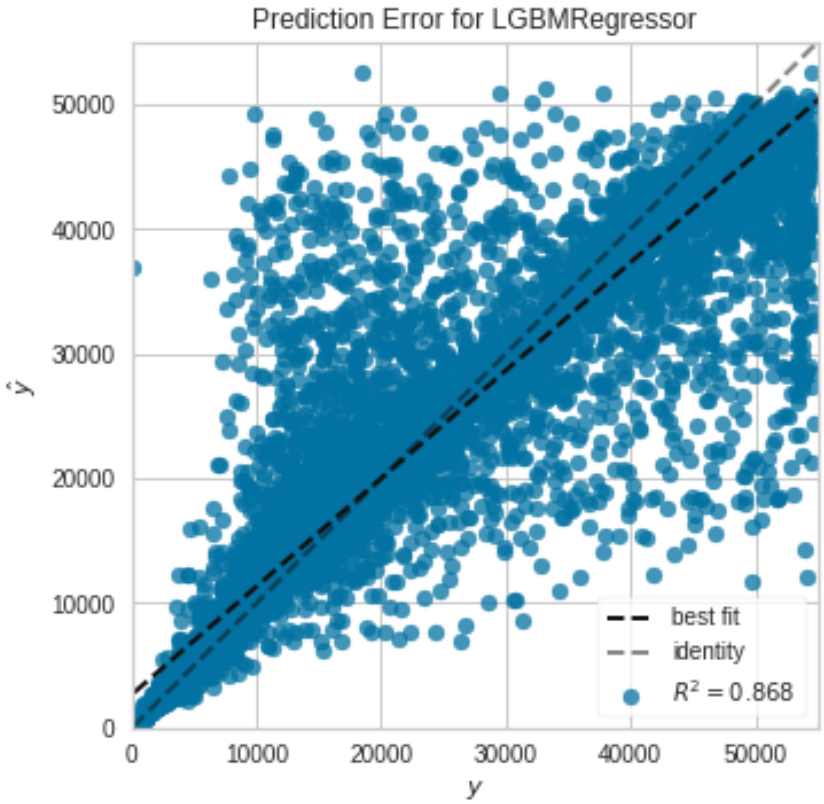


Figure 4: Prediction Error Plot. Source: from experimental results

6 Conclusion and Perspectives

This work presented a proposal to analyze a multi-source dataset to guarantee improvements in the performance of photovoltaic systems. Although the work was based on computational artificial intelligence methods of machine learning, this was not the only focus. In fact, the scarcity of datasets to train these methods has been one of the biggest problems. However, the dataset must be restructured into several data sources in addition to measurements of electrical and meteorological quantities. It is an important perspective to make an association of conventional and thermal image data to improve the assertiveness of the results. With this, the performance of photovoltaic systems is monitored in real time and it will be possible to anticipate problems. As seen in results, several machine learning methods can meet the prediction requirements. Among the methods, lightGBM stands out as it presents the best results in R^2 for our dataset. The generated model can be easily deployed on low cost hardware in order to analyze solar power plants in real-time. That allows increasing efficiency of energy generation. Nevertheless, there is still room for testing with other algorithms and when the set of features (evaluated variables such as individual MPPTs) are in a more complex definition, the use of deep learning as a tool for the identification of anomaly patterns becomes valid.

Acknowledgment

This research was supported by FUNCAP and CAPES. We thank our colleagues from Telematics and Industry Department of IFCE who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper.

References

[1] F. Ancuta and C. Cepisca, "Fault analysis possibilities for PV panels," Proceedings of the 2011 3rd International Youth Conference on Energetics (IYCE), 2011, pp. 1-5.

[2] S. C. W. Krauter, Solar Electric Power Generation, Springer, New York, NY, USA, 2006.

[3] S. Sumathi, L. A. Kumar, and P. Surekha, Solar PV and Wind Energy Conversion Systems, Springer, Switzerland, 2015.

[4] Alsharif, Mohammed H., Anabi Hilary, Kelechi, Khalid, Yahya, and Shehzad Ashraf, Chaudhry. "Machine Learning Algorithms for Smart Data Analysis in Internet of Things Environment: Taxonomies and Research Trends".Symmetry 12, no.1 (2020).

[5] Qiang Sun, Xubo Ge, Lin Liu, Xin Xu, Yibin Zhang, Ruixin Niu, Yuan Zeng, "Review of Smart Grid Comprehensive Assessment Systems", Energy Procedia, Volume 12, 2011, Pages 219-229, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2011.10.031.

[6] Paul, Riya and Sen, Anirban and Dey, Basudeb, "Study of Application of Smart Grid on Solar Energy" (November 25, 2019). 2nd International Conference on Non-Conventional Energy: Nanotechnology & Nanomaterials for Energy & Environment (ICNNEE) 2019, http://dx.doi.org/10.2139/ssrn.3492860

[7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: a highly efficient gradient boosting decision tree." In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 3149–3157.

[8] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[9] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001;1189–232.

8

Technical Session 2 (Short Papers)

Aerial Base Station Allocation Strategy

Wilson R. S. Silva[1]*and Diego. L. Cardoso[2]†

[1] Instituto Federal do Pará, Vigia de Nazaré, Pará, Brazil
rogerio.soares@ifpa.edu.br
[2] Universidade Federal do Pará, Belém, Pará, Brazil
diego@ufpa.br

Abstract

Unmanned aerial vehicles (UAV) serving as aerial base stations (ABSs) can be deployed to provide wireless connectivity to ground devices in events of increased network demand, point-failure in existing infrastructure, or disasters. However, positioning, signal coverage is still a challenge for this technology. In this work we propose a UAVbs allocation strategy to provide wireless signal to dynamically distributed users. Experimental results show that the proposed using swarm algorithm design maximizing the number of connected devices inthe UAV-BS.

1 Introduction

Aerial Base Stations (UAVs) serving as aerial base stations (UAV-BSs) can be deployed to provide wireless connectivity to ground devices. The adjustable altitude and mobility of UAVs make them suitable candidates for flexible deployment as aerial base stations (UAV-BSs) [3]

UAVs are one of the supplements of 6G. We can use drones in 6G as mobile hot spots during disaster situations, when the ground base stations (BSs) or even the electrical infrastructure are damaged and when the users are out of service, according to work [1].

Figure 1: UAV coverage wireless.

Recent popularity of UAVs and all of these potential applications, integrating UAVs in mobile networks is still a challenging topic. In this context, several issues still remain, such as determining the optimal deployment location of multiple UAVs, designing their optimal trajectory [2].

*Designed and implemented the class style
†Did numerous tests and provided a lot of suggestions

10th International Workshop on ADVANCEs in ICT Infrastructures and Services

In this paper we propose a solution to determine the optimal placement of multiple UAVs scenario. Based on that, a simulation scenario is built, in which an urban area is considered and user traffic demands are generated. we propose the utilization swarm algorithm to find the optimal placement of multiple UAVs BS.

The remainder of this paper is as follows. Section II introduces the proposed System Model. Section III shows the simulation results, while Section IV concludes the paper.

2 Proposed System Model

2.1 UAV Base Station

According to work [4] the UAV is carrying a directional antenna, and it has a coverage wireless. This is illustrated in Fig. 1, where is the directivity angle, H is the UAV height and R is the radius of coverage.

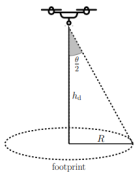


Figure 2: UAV coverage wireless [4].

The model as which considers free space path loss between the user and the UAV, which depends on whether or not there is line of sight (LOS) in the link between the UE and the UAV. Where f is the carrier frequency, d is the link distance, c is the light speed, as shown in work [4].

In the work of [5], he considers that there is a transmission between a user allocated and a drone. Then each user's data rate is that found from.

2.2 Problem formulation

UAVs are equipped with antennas capable of providing network access to mobile users. According to the height of the UAV, it will have a certain cellular coverage area. The problem is finding the ideal height, coverage radius and power to provide access to mobile users on the ground. This is an optimization problem. Based on this, the following cost function is proposed approach .

$$\min \sum_{u=1}^{u}(h) \tag{1}$$

$$\text{subject to: } p < Power \tag{2}$$

$$r < Radius \tag{3}$$

Where U is sum the of UAVs bases stations. P is the signal Power. h is the Height of the uav and r is comunication Radius. To provide access to the network for users, the UAVs must remain at a certain height, signal strength and communication radius.

10th International Workshop on ADVANCEs in ICT Infrastructures and Services

2.3 Allocation Strategy

To solve the above optimization problem, which consists of finding the ideal height, power and radius of the UAVS BS. The proposed algorithm consists of:
1. user cluster using kmeans;
2. for each cluster (i)
3. optimize uav height, power and radius
4. end for
5. plot the solution

3 Simulation and results

In our experiments, We consider 5 UAVs BS with backhall dedicated. Assume that there are 500 stationary users in the specified geographical area of 500x500 distributed in a uniform. The simulation are performed in Matlab®2022.

The figure below shows the allocation of users in the chosen scenario. There is a central BS that provides the backhall for the UAVs. Each UAV BS provides wireless signal to distributed users.

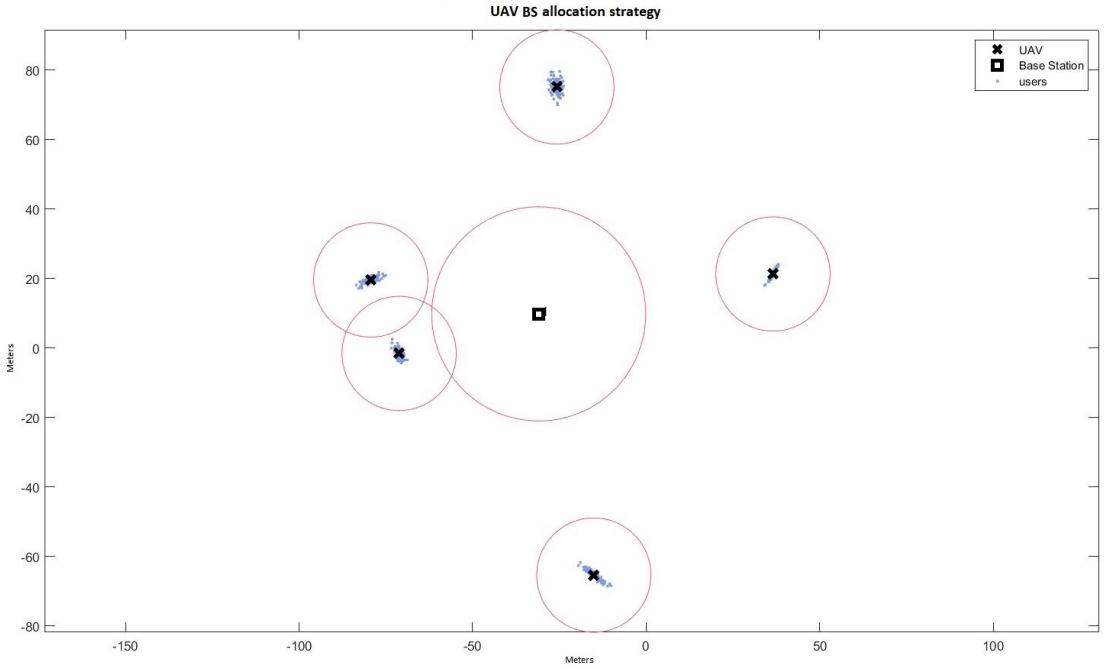


Figure 3: Allocation the UAVs BS.

The a Figure below shows performance metrics the algorithm PSO used to allocation UAV BS of the users. Note that the power of the uav increases according to the coverage radius of the wireless signal.

10th International Workshop on ADVANCEs in ICT Infrastructures and Services

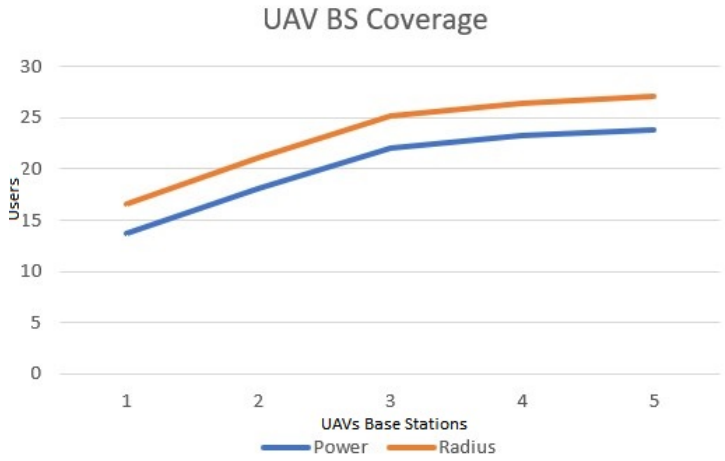


Figure 4: Power UAV BS coverage wireless.

Each group is being served according to the QoS of the users. Realize that more than 98% of users are being served by the wireless signal per cluster served. This is shown in the Figure below.

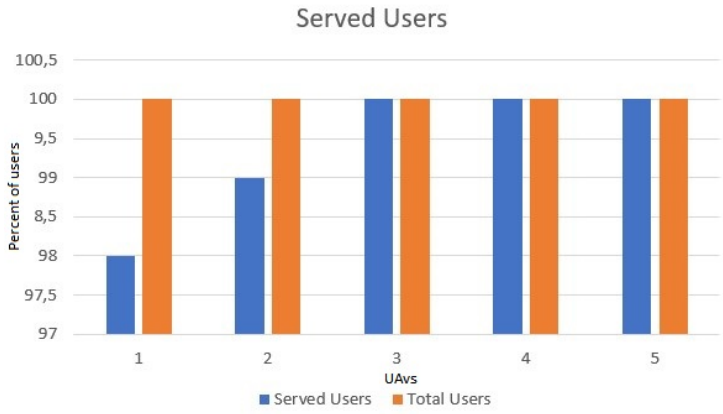


Figure 5: Percent of users served by each UAV BS.

4 Conclusion

In this paper, the problem the allocation optimal for multiple UAVs BS in is studied. We formulate the problem which finds the optimal placements, radius and power for multiple UAVs BS using swarm algorithm. This propost, allocation more than 98 percent of users in the chosen scenario. In future, we will investigate others algorithm to optimal of UAVs to cover the movable user in the area remotes.

4

10th International Workshop on ADVANCEs in ICT Infrastructures and Services

References

[1] Malkawi W, Shakhatreh H, and Musa A (2021). 3D placements of drones in a millimeter-wave network to maximize the lifetime of wireless devices. International Journal of Advanced and Applied Sciences, 8(11): 119-128

[2] M. Ozturk, J. P. B. Nadas, P. H. V. Klaine, S. Hussain and M. A. Imran, "Clustering Based UAV Base Station Positioning for Enhanced Network Capacity," 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), 2020, pp. 1-6, doi: 10.1109/AECT47998.2020.9194188.

[3] B. Omoniwa, B. Galkin and I. Dusparic, "Energy-aware optimization of UAV base stations placement via decentralized multi-agent Q-learning," 2022 IEEE 19th Annual Consumer Communications Networking Conference (CCNC), 2022, pp. 216-222, doi: 10.1109/CCNC49033.2022.9700536.

[4] M. Ozturk, J. P. B. Nadas, P. H. V. Klaine, S. Hussain and M. A. Imran, "Clustering Based UAV Base Station Positioning for Enhanced Network Capacity," 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), 2020, pp. 1-6, doi: 10.1109/AECT47998.2020.9194188.

[5] Wa'ed Malkawi *, Hazim Shakhatreh, Ahmed Musa.3D placements of drones in a millimeter-wave network to maximize the lifetime of wireless devices. Volume 8, Issue 11 (November 2021), Pages: 119-128.

[6] T. M. Shami, D. Grace and A. Burr, "Load Balancing and Control Using Particle Swarm Optimisation in 5G Heterogeneous Networks," 2018 European Conference on Networks and Communications (EuCNC), 2018, pp. 1-9, doi: 10.1109/EuCNC.2018.8442519.

Publication Agreement for Papers Accepted for Publication in ADVANCE '2020 Proceedings

Please fill in this form and submit it in a pdf format (with your final version of the paper) using your easychair account:

Paper No.: 1290

TITLE OF PAPER: UAVBS Allocation Strategy Using Swarm Algorithm

AUTHOR(S): Wilson Rogério Soares e Silva

PUBLICATION TITLE: Proceedings of ADVANCE Conference

By submitting your work for publication in the Proceedings of **Sixth International Workshop on ADVANCEs in ICT (ADVANCE)**, you have agreed, **by default**, to the following terms and conditions:

(a) The author and coauthor(s) must complete this form and submit it via easychair **along with the camera-ready**. If the paper is a work made for hire and the author or coauthor(s) do not have the authority to sign the paper, the owner of the paper must also sign the form.

(b) You grant the publisher of ADVANCE Proceedings a right to publish your work and to distribute the published work according to the relevant laws and regulations governing publications.

(c) You must be either the owner of the work or have obtained permission from the owner of the work (e.g. your employer) to publish the work. You have agreed to take the responsibility to solve any disputes that may arise anytime in the future with regard to the ownership of the work and the right of publishing the work, including penalty and compensation payment.

(d) If excerpts from copyrighted works are included in your paper, you must have obtained written permission from the copyright owners and show credit to the sources in your paper. You have agreed to take the responsibility to solve any disputes that may arise anytime in the future with regard to copyright materials that you have used in your work.

If any of the above terms and conditions does not apply to your work or you do not agree with any of these provisions, please inform the **ADVANCE Workshop** Organizing Committee immediately at the following address:

E-mail: advance2020@easychair.org

Corresponding Author's

Signature: _____ **Title/Employer:** _____
(if not signed by author)

Print Name: WILSON ROGÉRIO SOARES E SILVA

As the form is being signed by the corresponding author on behalf of the others, the following additional statement must be accepted and signed by the author signing for the co-authors: "I represent and warrant that I am authorised to execute this transfer of copyright on behalf of all the authors of the article named above."

Date: 11/01/2023 **Corresponding Author's Signature:** _____

The use of Machine Learning for predicting the sentence given by the population at a crime scene

Keyvilânia V. V. de Freitas[1], Dr. C. Alexandre R. Fernandes[1], Dr. James Ferreira M. Jr.[1] and Gabriela A. dos Santos[1]
[1] Universidade Federal do Ceará, Sobral, Brazil.
keyvilania@alu.ufc.br, alexandrefernandes@ufc.br, james.mourajr@unilab.edu.br, gabriela.albuquerque@alu.ufc.br

Abstract

Violence and authoritarianism are present factors in most societies, as well as prejudice and discrimination. Social variables, such as race, profession, and place of residence, are important factors that interfere with the portrait of each member of society and may override the truth and how judicial laws are applied to them. The main goal of this work is to apply machine learning techniques to predict the verdict given by the population to the accused at a crime scene. The database was created with the social information collected from 1102 participants from 21 different states of Brazil, in the year 2020. The data used are part of the "Brasil 2020 Decide" survey, which has assertions about social conditions and authoritarian behavior related to economic conditions, color and race. In the described crime scenario, the SVM (Support Vector Machine) algorithms achieved an average accuracy of 68% in predicting the sentence for the accused.

1 Introduction

Brazilian society lives with the recurrent fear caused by crime and violence. Organizations that disseminate hate, prejudice, and intolerance have been gaining more and more followers (Fórum Brasileiro de Segurança Pública, 2017).In this context, we live in a period where the constant validation of democratic institutions in Brazil is needed.

The growth of social media also brings visibility to cases of intolerance and prejudice suffered, for example, by the lower-class population, making evident the neglect and disengagement of part of society (Lavor F., et al., 2018). The feeling of helplessness and invalidity felt by the population contribute to increase the number of supporters of groups with authoritarian discourses in search of quick solutions (Lavor F., et al., 2018). These aspects influence the opinion of individuals, such as, for example, the sentence they would give in a criminal case. In this scenario, the popular verdict often differs from the judicial one because it is distorted by the conditions experienced by each individual in society.

The work aims to apply machine learning techniques to predict the verdict given by the population at a crime scene. The database used was assembled with information from the social context of 1102 participants from 21 different states of Brazil, mostly from the states of Ceará and Rio Grande do Sul, in the year 2020. The data used are part of the "Brasil 2020 Decides" survey which has assertions about social conditions, authoritarian behavior related to economic conditions, and authoritarian behavior related to color/race differences. The questions were intended to measure the predisposition to authoritarianism and assess how demographic and social variables affect these predispositions. The indices of the F scale, RWA, modern racism index, and class prejudice scale, among others, are also obtained from the survey (BRASIL 2020 DECIDE).

This work will determine the effectiveness of the SVM (Support Vector Machine) as a tool for judgment. The main objective is to obtain a model to predict the sentence of an accused at a crime scene, based on data collected through the form. To improve the efficiency of the SVM, the linear, polynomial, and RBF kernels were tested. The predictive model is based on a classification system that follows the degree of knowledge discovery in the database (KDD) which allows a better exploration of databases for using the SVM. The accuracy obtained in this work presents an average success rate of 60% reaching 68% of effectiveness.

The remainder of the work is divided as follows. In topic 2, the papers related to this article are presented. In topic 3, the theoretical framework for the measurement of authoritarian positions and the influence of social variables is addressed. In topic 4, the methods used in the work are presented. In section 5, the results are presented and discussed. In section 6, the conclusion is addressed.

2 Measuring Authoritarian Positions and the Influence of Social Variables on Violence

In this section, the theoretical aspects will be presented, namely the indices of support for authoritarian positions and the influence of social class and color/race variables.

2.1 Measuring Support for Authoritarian Positions

At the beginning of the 20th century, with the development of totalitarian regimes such as Nazism, Fascism, and Communism, lines of study on authoritarianism gained more emphasis, as well as its relationship with prejudice, support for authoritarian behavior, and the way that social variables facilitated the development of ideologies that led to World War II. (Barros, Torres, & Pereira, 2009). In this way, the study of the measurement of authoritarian tendency became of great importance.

In 1950 Theodor Adorno developed the F scale, one of the main ways to measure predisposition to authoritarianism. This scale, developed along with other psychologists, was designed to implicitly measure adherence to authoritarianism through a set of statements about everyday events. The first versions of the form addressed ideas related to Nazism, ethnocentrism, and political-economic conservatism to capture pre-fascist tendencies. The assertions are divided into nine strands: Conventionalism; Authoritarian Submission; Authoritarian Aggression; Anti-intracepction; Superstition and Stereotypy; Power and Toughness; Destructiveness and Cynicism; Projectivity; and Sex (Adorno, 1950).

From three of the strands used by Theodor Adorno, in 1981 Robert Altemeyer developed the theory of extreme right authoritarianism. To measure the level of authoritarianism based on his theory, Robert A. created the Right-Wing Authoritarianism – RWA scale. Influence of Social Variables on Violence

According to Atlas da Violência, there is an increase in homicide rates, especially for black men and women (Atlas da Violência 2018). The black and poor population are those with the highest rates of mental disorders, victimization by homicide and other violent crimes, mortality, among others (Cerqueira, et al., 2018). In this way, intersectional reflection becomes necessary, within the scope of

public security, regarding the characteristics of race and classes that have an emphasis on social inequality indices, interfering with adequate access to education, justice, health, employment, etc. (Rocha, 2015).

In “A Naturilazação da Pobreza”, it is reported that the exclusion and impoverishment of certain social strata are highlighted by the relationship between the accumulation or scarcity of wealth with the virtues and imperfections of each individual. Likewise, about race, there is an attempt to mask racial tensions based on the idea that the intense crossing of races resulted in peaceful racial interactions in society (Lavor F., et al., 2018). Thus, it is observed how the social variables interfere in the portrait of each member before society, being able to overlap truth and justice.

3 Method

This section describes the methods used in this work, being divided into database and classification.

3.1 Database

To carry out the prediction model, the database of the survey “Brasil 2020 Decide” was used. This database was constructed by researchers of the postgraduate program of the Universidade Federal do Ceará – UFC in partnership with the Universidade da Integração Internacional da Lusofonia Afro Brasileira – UNILAB. The procedure carried out in the research follows the rules established by Resolution 466/12 of the “Conselho Nacional de Saúde” (BRASIL 2020 DECIDE, 2020). The research results are restricted to the institutions responsible for it's development. In the form used for the research, a crime scenario is described, in which there is an accused and a victim with informed social class and employment relationship. Along with that information, are presented two images of two people with different physical features. At the end of the survey the participant must give a verdict on whether to convict or acquit the accused. The prediction model proposed in this paper aims to predict the sentence given by the population in these scenarios (Acquit or Condemn). For this, it is necessary to take into account the values that each individual carries with them, such as beliefs, level of satisfaction with the government, income, prejudices, as well as their socio-economic situation. In this way, the assertions of the survey that had the most relevant characteristics were selected to compose the base. The chosen questions were: 1, 30, 32, 33, 35, 37, 38, 41, 44, 59, 75, 87, 88, 97, 100, 106, 112 e 118 (form in annex). With the questions and the indexes extracted from the original form, a database with a size of 13KB was set up and later applied in the Matlab software.

3.2 Prediction model

For the prediction model, the KNN, SVM and MLP algorithms were tested, from which the SVM classifier was selected because of it's higher accuracy values and shorter processing time. O support vector machine (SVM) é um dos classificadores mais eficientes para problemas de elevada dimensionalidade e problemas com dados não lineares que estão sujeitos a informações que podem diminuir a eficácia do algoritmo. The Support Vector Machine (SVM) is one of the most efficient classifiers for high-dimensional problems and problems with non-linear data that rely on information that can decrease the effectiveness of the algorithm.

The prediction model follows the steps of the KDD method: (1) data selection, (2) data pre-processing, (3) data transformation, (4) data mining, and (5) interpretation. In the first step, attributes that had no relevance for the classification were discarded and an exclusion question was applied to remove samples that could impair the effectiveness of the algorithm. In the second stage, irregularities

were corrected and missing values of age were filled in with the average age per label. Information about the victim and the accused are encoded in binary variables. After pre-processing, the database had 1102 samples, 1 binary output, and 15 attributes, namely: The assertion about the crime scenario described; the participant's age and gender; the profile of the accused; the profile of the victim; social class, income, and education of the participant; the state where he/she resides; the modern racism index; the RWA index; the F scale index; the social dominance index; and the class prejudice index. In the third stage, data normalization was performed using the z-score method. The kfold technique with k=10 was also applied to split the training and test data. In the fourth step, the database is applied to the algorithm for performing the classification, of the sentence determined for the crime scene described in item 4.1. Finally, the results are presented and analyzed in the fifth step, which will be shown in the next section. The charts shown present the results obtained according to the parameters used.

4 Results and Discussion

In this section, the results and tests carried out are discussed. Using the previously mentioned database, the results were obtained by applying kfold with k=10, which consists of dividing the database into 10 groups where each group is used for testing while the rest is used for training, alternating the groups until all have been used as a test. The z-score normalization was also applied, where the difference between the data values and the population mean, about the standard deviation, is calculated for each attribute.

4.1 Unnormalized data

The results obtained with the SVM for the different types of kernels without z-score normalization are shown in Table 1. The results show that in the tests carried out for the linear kernel, the algorithm obtained a success rate of 67%, which is the highest rate.

The database has attributes of different magnitudes such as age, modern racism index, RWA index, F scale index, social dominance index, and class prejudice index. Therefore, it is necessary to normalize the data so that they are all represented in the same scale.

4.2 Normalized data

The results obtained with the SVM for the different types of the kernel with the application of the z-score and the PCA are shown in Table 2. The results show that the highest average accuracy is given for the polynomial kernel with a rate of 68, 20%, reaching a maximum accuracy of 74.54%. The lowest average accuracy was obtained with the RBF kernel, but this kernel obtained a maximum accuracy of 75.45%. Note that the difference between the algorithms before and after normalization is approximately 7%, for the RBF and Polynomial kernels. However, the averages of success for all the types of kernels after normalization are very close. Despite the improvement in accuracy levels, this rate still is a medium value, indicating that these factors influence decision-making in the population but there are others that were not disclosed in the database.

Kernel Types	Average Accuracy
Linear	67,00%
RBF	59,27%
Polynomial	59,53%

Table 1: The average accuracy rate for different kernel types.

Kernel Types	Average Accuracy
Linear	68,09%
RBF	67,55%
Polynomial	68,20%

Table 2: The average accuracy rate for the different kernel types with the standardized

5 Conclusion and Future Work

This work sought to analyze the judgment and position of people in a crime scenario, based on the socioeconomic and racial data of those present at the crime scene. Data were collected through the survey “Brasil 2020 Decide”, which gathered social and racial information from 1102 participants from 21 different states of Brazil. It's possible to see that the accuracy of the algorithms for the normalized data, being this the best case, have results very close to each other: approximately 68%. As a way of seeking to improve this accuracy, for future work we will apply this database in the XGBoost and CNN (Convolution Neural Network) algorithms. The obtained results will be applied in studies in the area of psychology related to the behavior of society.

References

Adorno, T. W. (1950). The Authoritarian Personality.

Barros, T., Torres, A. R., & Pereira, C. (2009, Abril). Autoritarismo e Adesão a Sistemas de Valores Psicosociais. *PSICO - UFSP*, pp. 47-57.

Carlisle, D. (2010, April). *graphicx: Enhanced support for graphics.* Retrieved from http://www.ctan.org/tex-archive/ help/Catalogue/entries/graphicx.html

Cerqueira, D., Lima, R., Bueno, S., Neme, C., Ferreira, H., & Coelho, D. (2018). Atlas da Violência 2018. *IPEA.*

Fernandes, L. (2022). APRENDIZAGEM DE MÁQUINA PARA PREVISÃO DE PREDISPOSIÇÃO AO MEDO DO CRIME. *Repositório UFC*, p. 47.

Fórum Brasileiro de Segurança Pública. (2017). Medo da Violência e Apoio ao autoritarismo no Brasil. *Revista Brasileira de Segurança Pública*, p. 39.

Gioia, T., & Barros, J. (2021). ASSOCIAÇÃO ENTRE VARIÁVEIS SOCIOECONÔMICAS E A OCORRÊNCIA DE DENGUE NO ESTADO DE GOIÁS: UMA ANÁLISE A PARTIR DE ALGORITMOS DE MACHINE LEARNING. *Revista Brasileira de Geográfia Médica e da Saúde.*

Lavor F., T. L., Almeida S., D. S., Martino J., P., Malherme B., V. N., Ferreira M. J., J., & Sergio L., R. (2018). Análises Interseccionais a Partir da Raça e da Classe: Medo do Crime e Autoritarismo. *Psicologia: Ciência e Profissão*, 223-237.

PROGRAMA DE PÓS GRADUAÇÃO DA UFC E DA UNILAB. (2020). *BRASIL 2020 DECIDE.*

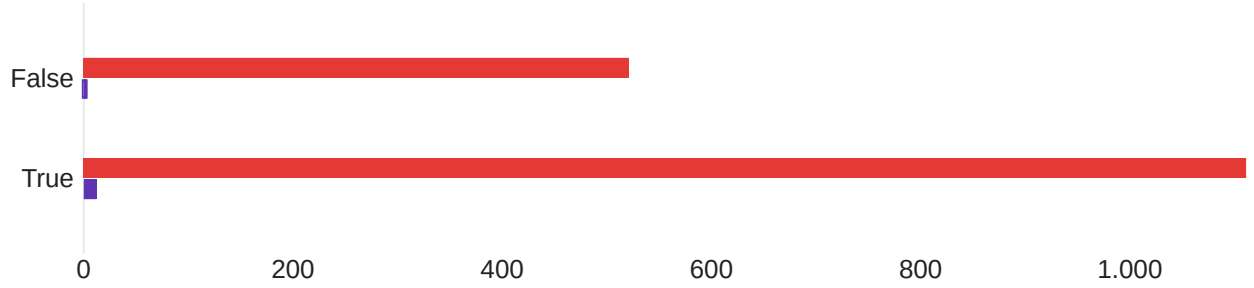
Rocha, E. (2015). O negro no mundo dos ricos: um estudo sobre a disparidade racial de riqueza no Brasil com os dados do censo demográfico de 2010. *Universidade de Brasilia*.

Sarriés, G., Trivellato, G., & Furlan, G. (n.d.). Sistema de Avaliação Ponderada da Multinacionalidade da Agricultura: Machine Learning, Índices de Sustentabilidade Ambiental e Serviços Ecossistêmicos. *4th National Workshop on UI Green Metric World University Rankings for Brazilian Universities*.

ANEXO

Q1 - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO Prezado (a) colaborador (a), Você é convidado (a) a participar da pesquisa "Brasil 2020 Decide", que tem como finalidade conhecer como as pessoas tomam decisões em um julgamento, comparando a percepção de estudantes matriculados em cursos de graduação em Direito e com população em geral. Esta investigação é coordenada pelo pesquisador Prof. Dr. James Ferreira Moura Junior, do Programa de Pós Graduação em Psicologia da UFC e da UNILAB, e pelo pesquisador Prof. Dr. Angelo Brandelli Costa, do Programa de Pós-Graduação em Psicologia da PUCRS. Para tanto é necessário que você preencha os questionários, com duração aproximada de 20 minutos. 1. PARTICIPANTES DA PESQUISA: Pessoas maiores de 18 anos da população geral e estudantes de cursos de Graduação em Direito. Participarão da pesquisa aqueles que, ao serem convidados a colaborar, voluntariamente concordem. 2. ENVOLVIMENTO NA PESQUISA: Ao participar deste estudo, você deve responder a um questionário online envolvendo perguntas sobre aspectos psicológicos individuais, bem como um cenário simulado de julgamento. Você tem a liberdade de se recusar a participar e pode ainda deixar de responder em qualquer momento da pesquisa, sem nenhum prejuízo. Sempre que quiser, você poderá pedir mais informações sobre a pesquisa. Para isso, poderá entrar em contato com o coordenador da pesquisa. 3. RISCOS E DESCONFORTOS: A participação nesta pesquisa não traz complicações; talvez, apenas, algum constrangimento que algumas pessoas sentem quando estão fornecendo informações sobre si mesmas. Os procedimentos utilizados nesta pesquisa seguem as normas estabelecidas pela Resolução 466/12 do Conselho Nacional de Saúde, e não oferecem risco à sua integridade física, psíquica e moral. Nenhum dos procedimentos utilizados oferece riscos à sua dignidade. 4.

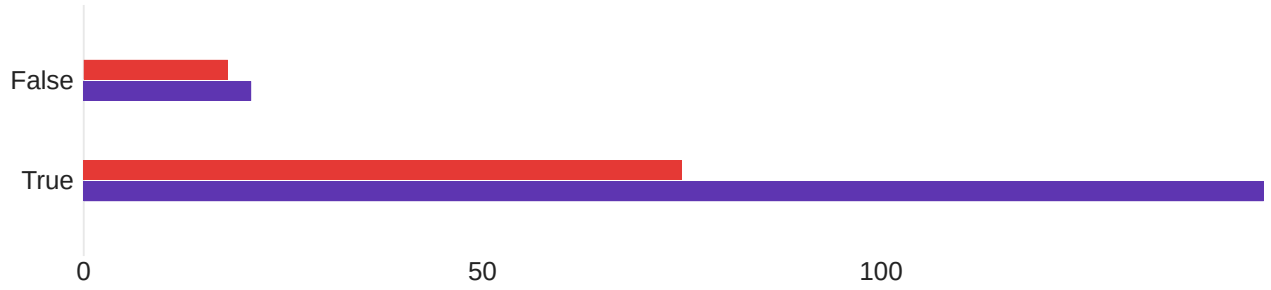
CONFIDENCIALIDADE DA PESQUISA: Todas as informações coletadas neste estudo são estritamente confidenciais. Apenas os membros do grupo de pesquisa terão conhecimento das respostas, e seu nome não será usado em nenhum momento. Todos os dados serão analisados em conjunto, garantindo o caráter anônimo das informações. Os resultados poderão ser utilizados em eventos e publicações científicas. 5. BENEFÍCIOS: Ao participar desta pesquisa, você não deverá ter nenhum benefício direto. Entretanto, espera-se que a mesma forneça dados importantes acerca de possíveis fatores contribuintes para a explicação de comportamentos sociais e para a construção de guidelines para estabelecimento de decisões em julgamento no Brasil. Eu fui informado(a) dos objetivos da pesquisa acima de maneira clara e detalhada. Recebi informações a respeito da pesquisa e esclareci minhas dúvidas. Sei que em qualquer momento poderei solicitar novas informações e modificar minha decisão de participação se assim eu o desejar. Quaisquer dúvidas relativas a esta pesquisa poderão ser esclarecidas pelos pesquisador responsável James Ferreira Moura Junior, fone (085) 997767070, ou pela entidade responsável o Comitê de Ética em Pesquisa da Universidade Federal do Ceará (CEP-UFC) em (85) 33668344/46, Rua Coronel Nunes de Melo, 1000 - Rodolfo Teófilo, de segunda a sexta-feira das 8h às 12h. O Comitê de Ética é um órgão independente constituído de profissionais das diferentes áreas do conhecimento e membros da comunidade. Sua responsabilidade é garantir a proteção dos direitos, a segurança e o bem-estar dos participantes por meio da revisão e da aprovação do estudo, entre outras ações. Se você concordar em participar deste estudo, você apertará o botão logo abaixo de que concorda em voluntariamente participar da pesquisa.



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	1	1	0	0	521
True	1	2	1	0	0	1123

Campo	False	True
Estou de acordo e concordo em participar da pesquisa	521	1.111
Não estou de acordo e não concordo em participar da pesquisa	0	12
Total	521	1123

Q1 - VEJA COM ATENÇÃO a notícia abaixo sobre um caso de homicídio! Todas as informações, dados e imagens dos casos são fictícios. INSTRUÇÕES. Abaixo temos o caso judicial referente à notícia anterior e você é um dos membros do júri. Considere o relatório resumido, que traz os principais elementos levantados ao longo do processo, e tome sua decisão. O restante do Júri já se pronunciou. Mesmo com poucos elementos, é preciso que você se posicione. FRANCISCO RAIMUNDO SILVA, brasileiro, desempregado, 33 anos, morador de uma favela do Rio de Janeiro, foi denunciado pelo crime de HOMICÍDIO SIMPLES DOLOSO. A vítima era JOSÉ ANTÔNIO LIMEIRA, brasileiro, 36 anos, ajudante de caminhoneiro, que habita em uma comunidade dessa mesma cidade. Considere que Homicídio Simples Doloso é matar alguém com a intenção ou assumindo o risco do que se está fazendo. Resumo do caso: A vítima volta para casa e ao descer do ônibus foi apunhalado e acabou morrendo. Acontece que na noite anterior, envolveu-se numa briga de bar com o acusado. As testemunhas afirmaram que a discussão foi feia entre eles. Que depois do acontecido todos foram embora e o bar fechou. A acusação alegou que ninguém teria motivos para matar a vítima, que era um homem trabalhador e que a morte só poderia estar ligada à briga anterior. A defesa afirmou que a vítima tinha um passado duvidoso e envolvimento com o crime. E que o acusado não teve nenhuma participação no fato ocorrido. Nada a acrescentar, conclusos os autos para julgamento. Indique qual sua decisão final, marcando UM dos vereditos:

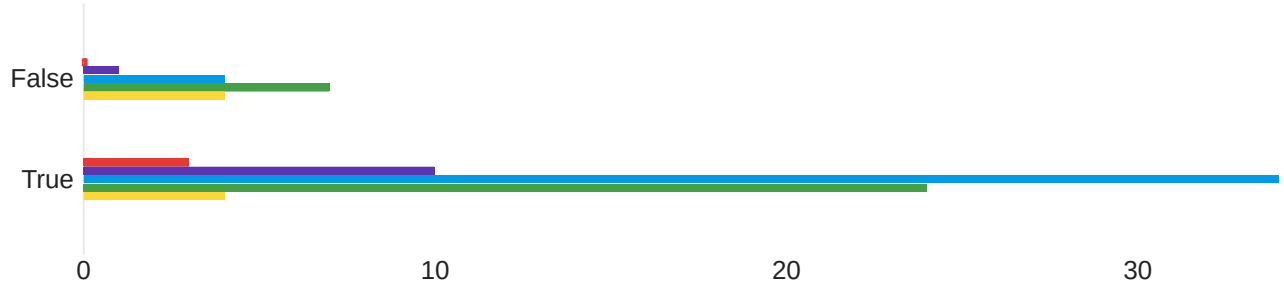


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	2	0	0	39
True	1	2	2	0	0	223

Campo	False	True
Condenar o réu pelo cometimento de Homicídio Simples Doloso.	18	75
Absolver o réu pelo cometimento de Homicídio Simples Doloso.	21	148
Total	39	223

10

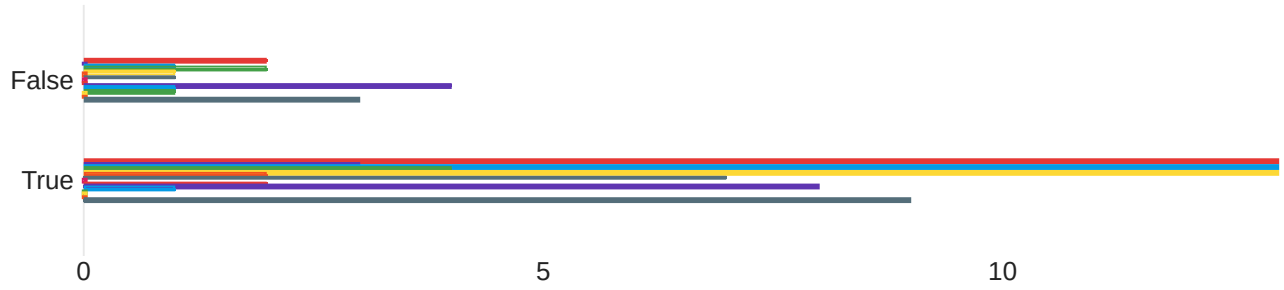
Q2 - Qual o grau da pena que o réu merece receber?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	34	38	36	2	3	16
True	32	38	36	2	4	75

Campo	False	True
Mínima	0	3
Leve	1	10
Moderada	4	34
Forte	7	24
Máxima	4	4
Total	16	75

Q3 - Se a pena para esse crime varia de 6 (seis) a 20 (vinte) anos, a qual tempo de sentença você condena o acusado?

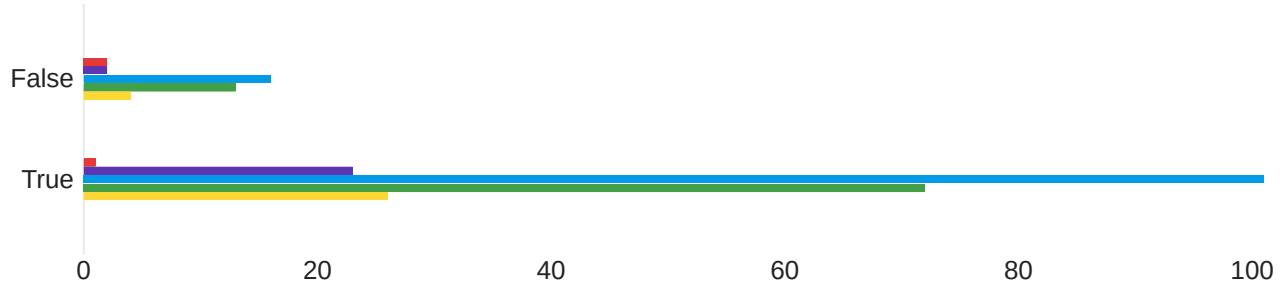


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	23	40	33	5	27	16
True	23	40	32	6	36	75

Campo	False	True
6 anos	2	13
7 anos	0	3
8 anos	1	13
9 anos	2	4
10 anos	1	13
11 anos	0	2
12 anos	1	7
13 anos	0	0
14 anos	0	2
15 anos	4	8
16 anos	1	1

17 anos	1	0
18 anos	0	0
19 anos	0	0
20 anos	3	9
Total	16	75

Q4 - O quanto você confia na sua decisão de sentença?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	5	3	1	1	37
True	1	5	3	1	1	223

Campo	False	True
Nada	2	1
Pouco	2	23
Médio	16	101
Muito	13	72
Totalmente	4	26
Total	37	223

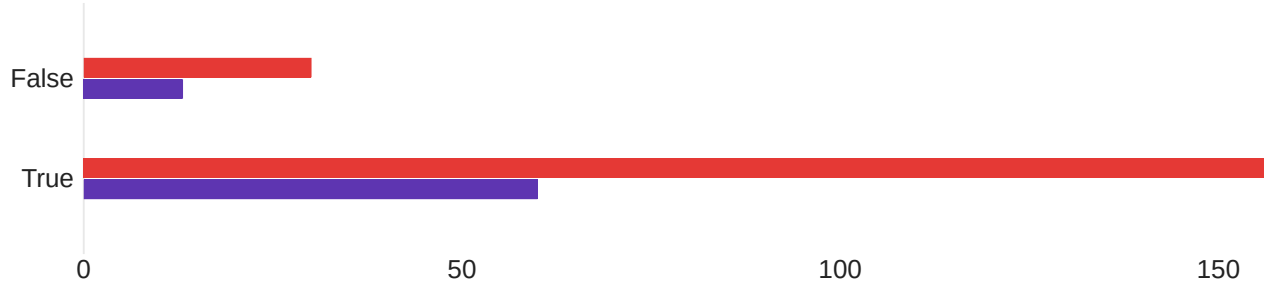
Q5 - O quanto você acredita que o réu estava no controle de suas ações?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	10	7	2	5	16
True	1	10	7	2	4	75

Q6 - Na sua opinião, qual a probabilidade de o réu se envolver em um crime?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	4	10	6	2	2	16
True	2	10	7	2	4	75

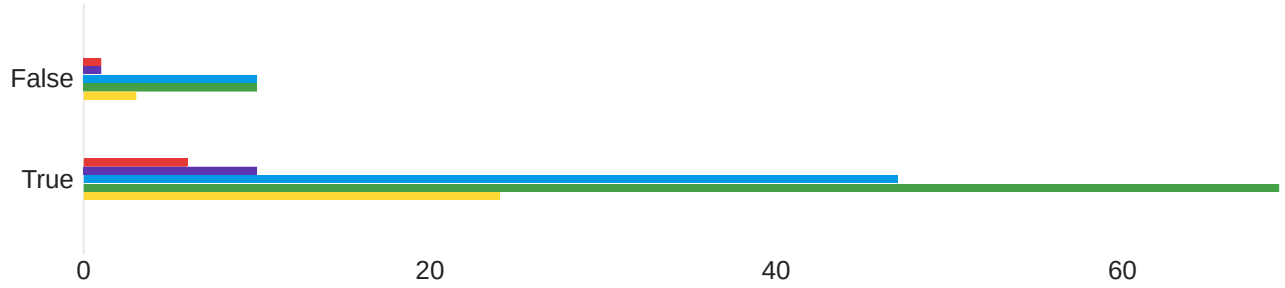
Q100 - VEJA COM ATENÇÃO a notícia abaixo sobre um caso de homicídio! Todas as informações, dados e imagens dos casos são fictícios. INSTRUÇÕES. Abaixo temos o caso judicial referente à notícia anterior e você é um dos membros do júri. Considere o relatório resumido, que traz os principais elementos levantados ao longo do processo, e tome sua decisão. O restante do Júri já se pronunciou. Mesmo com poucos elementos, é preciso que você se posicione. EIKE CASTELO BRAZ TERCEIRO, brasileiro, 34 anos, gerente de banco, que reside em condomínio na cidade do Rio de Janeiro foi denunciado pelo crime de HOMICÍDIO SIMPLES DOLOSO. A vítima era JOSÉ ANTÔNIO LIMEIRA, brasileiro, 36 anos, ajudante de caminhoneiro, que habita em uma comunidade dessa mesma cidade. Considere que Homicídio Simples Doloso é matar alguém com a intenção ou assumindo o risco do que se está fazendo. Resumo do caso: A vítima volta para casa e ao descer do ônibus foi apunhalado e acabou morrendo. Acontece que na noite anterior, envolveu-se numa briga de bar com o acusado. As testemunhas afirmaram que a discussão foi feia entre eles. Que depois do acontecido todos foram embora e o bar fechou. A acusação alegou que ninguém teria motivos para matar a vítima, que era um homem trabalhador e que a morte só poderia estar ligada à briga anterior. A defesa afirmou que a vítima tinha um passado duvidoso e envolvimento com o crime. E que o acusado não teve nenhuma participação no fato ocorrido. Nada a acrescentar, conclusos os autos para julgamento. Indique qual sua decisão final, marcando UM dos vereditos:



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	1	0	0	43
True	1	2	1	0	0	216

Campo	False	True
Condenar o réu pelo cometimento de Homicídio Simples Doloso.	30	156
Absolver o réu pelo cometimento de Homicídio Simples Doloso.	13	60
Total	43	216

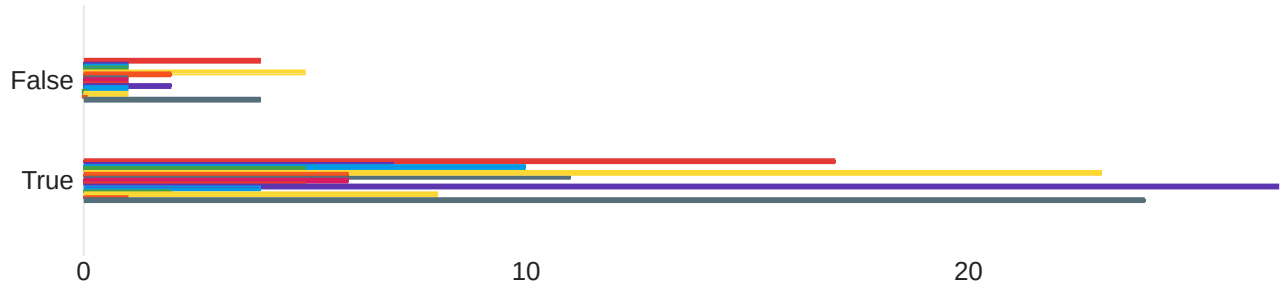
Q101 - Qual o grau da pena que o réu merece receber?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	32	38	36	2	4	25
True	32	38	36	2	4	156

Campo	False	True
Mínima	1	6
Leve	1	10
Moderada	10	47
Forte	10	69
Máxima	3	24
Total	25	156

Q102 - Se a pena para esse crime varia de 6 (seis) a 20 (vinte) anos, a qual tempo de sentença você condena o acusado?



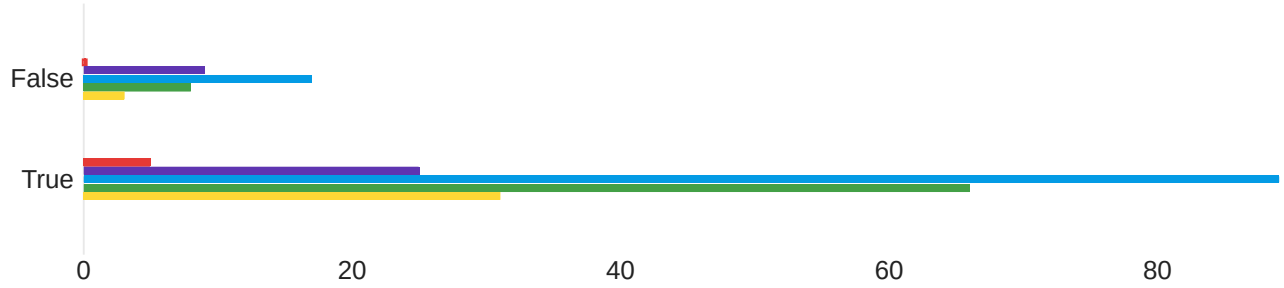
Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	23	40	31	5	29	25
True	23	40	32	5	25	156

Campo	False	True
6 anos	4	17
7 anos	1	7
8 anos	1	10
9 anos	1	5
10 anos	5	23
11 anos	2	6
12 anos	1	11
13 anos	1	6
14 anos	1	5
15 anos	2	27
16 anos	1	4

20

17 anos	0	2
18 anos	1	8
19 anos	0	1
20 anos	4	24
Total	25	156

Q103 - O quanto você confia na sua decisão de sentença?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	2	5	3	1	1	37
True	1	5	3	1	1	216

Campo	False	True
Nada	0	5
Pouco	9	25
Médio	17	89
Muito	8	66
Totalmente	3	31
Total	37	216

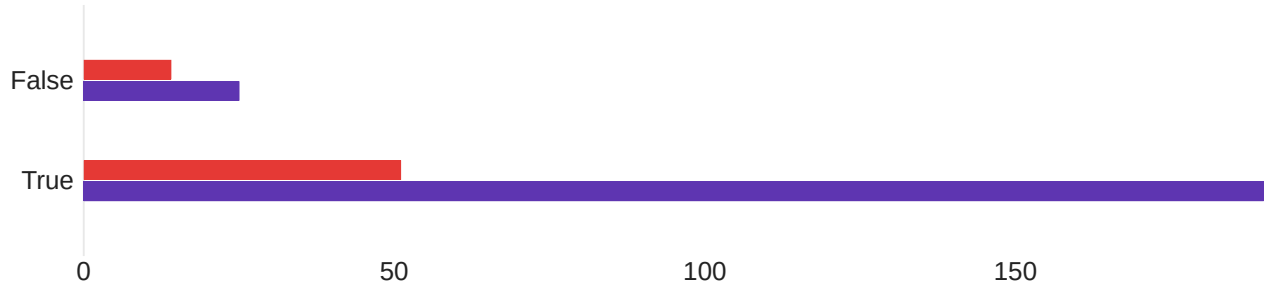
Q104 - O quanto você acredita que o réu estava no controle de suas ações?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	5	10	8	2	3	25
True	1	10	8	2	5	156

Q105 - Na sua opinião, qual a probabilidade de o réu se envolver em um crime?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	2	10	7	2	5	25
True	2	10	7	2	5	156

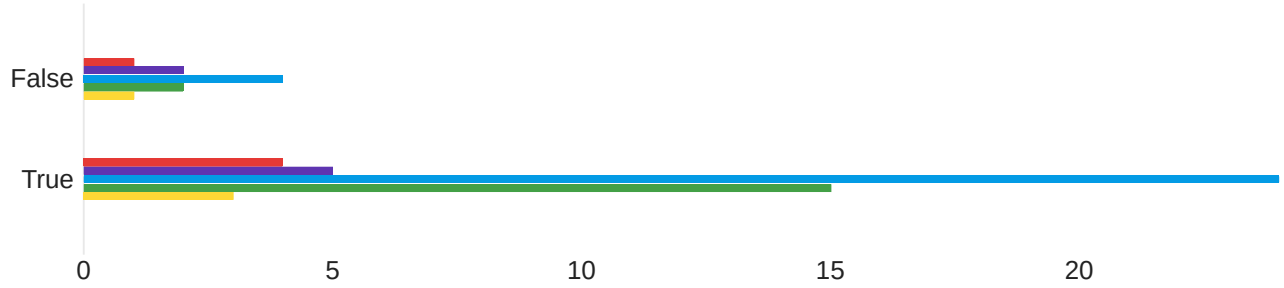
Q106 - VEJA COM ATENÇÃO a notícia abaixo sobre um caso de homicídio! Todas as informações, dados e imagens dos casos são fictícios. INSTRUÇÕES. Abaixo temos o caso judicial referente à notícia anterior e você é um dos membros do júri. Considere o relatório resumido, que traz os principais elementos levantados ao longo do processo, e tome sua decisão. O restante do Júri já se pronunciou. Mesmo com poucos elementos, é preciso que você se posicione. FRANCISCO RAIMUNDO SILVA, brasileiro, desempregado, 33 anos, morador de uma favela do Rio de Janeiro, foi denunciado pelo crime de HOMICÍDIO SIMPLES DOLOSO. A vítima era MARCEL SCHMISK BÔAS, brasileiro, empresário, 35 anos, morador de um bairro na zona nobre da mesma cidade. Considere que Homicídio Simples Doloso é matar alguém com a intenção ou assumindo o risco do que se está fazendo. Resumo do caso: A vítima volta para casa e ao descer do ônibus foi apunhalado e acabou morrendo. Acontece que na noite anterior, envolveu-se numa briga de bar com o acusado. As testemunhas afirmaram que a discussão foi feia entre eles. Que depois do acontecido todos foram embora e o bar fechou. A acusação alegou que ninguém teria motivos para matar a vítima, que era um homem trabalhador e que a morte só poderia estar ligada à briga anterior. A defesa afirmou que a vítima tinha um passado duvidoso e envolvimento com o crime. E que o acusado não teve nenhuma participação no fato ocorrido. Nada a acrescentar, conclusos os autos para julgamento. Indique qual sua decisão final, marcando UM dos vereditos:



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	2	0	0	39
True	1	2	2	0	0	241

Campo	False	True
Condenar o réu pelo cometimento de Homicídio Simples Doloso.	14	51
Absolver o réu pelo cometimento de Homicídio Simples Doloso.	25	190
Total	39	241

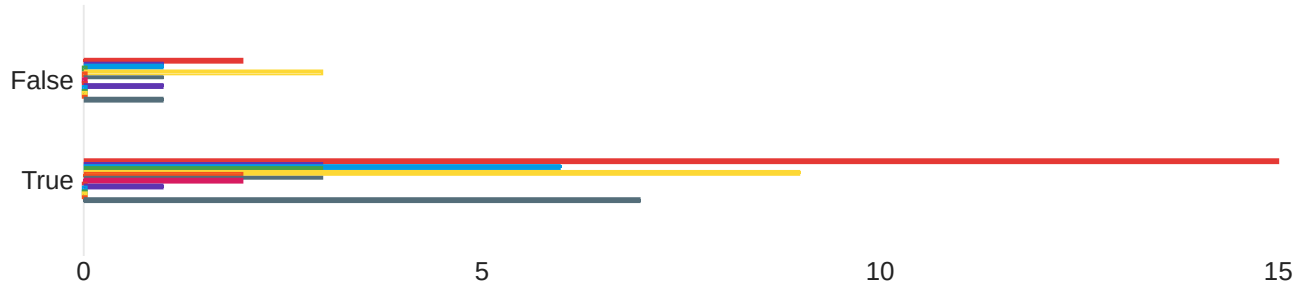
Q107 - Qual o grau da pena que o réu merece receber?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	32	38	35	2	4	10
True	32	38	35	2	4	51

Campo	False	True
Mínima	1	4
Leve	2	5
Moderada	4	24
Forte	2	15
Máxima	1	3
Total	10	51

Q108 - Se a pena para esse crime varia de 6 (seis) a 20 (vinte) anos, a qual tempo de sentença você condena o acusado?

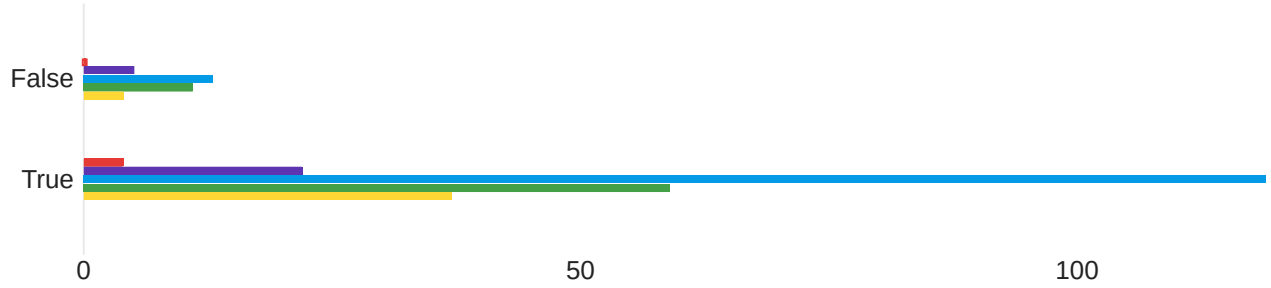


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	23	39	30	6	32	10
True	23	40	30	6	42	51

Campo	False	True
6 anos	2	15
7 anos	1	3
8 anos	1	6
9 anos	0	3
10 anos	3	9
11 anos	0	2
12 anos	1	3
13 anos	0	2
14 anos	0	0
15 anos	1	1
16 anos	0	0

17 anos	0	0
18 anos	0	0
19 anos	0	0
20 anos	1	7
Total	10	51

Q109 - O quanto você confia na sua decisão de sentença?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	2	5	3	1	1	33
True	1	5	3	1	1	241

Campo	False	True
Nada	0	4
Pouco	5	22
Médio	13	119
Muito	11	59
Totalmente	4	37
Total	33	241

30

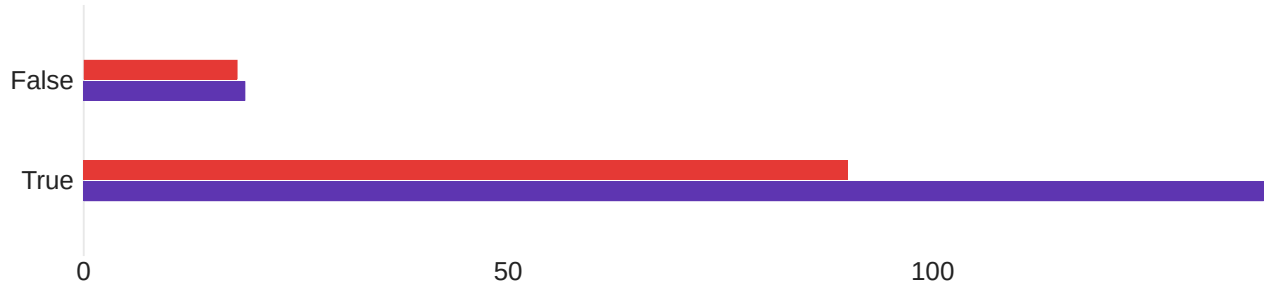
Q110 - O quanto você acredita que o réu estava no controle de suas ações?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	3	10	7	2	5	10
True	3	10	7	2	4	51

Q111 - Na sua opinião, qual a probabilidade de o réu se envolver em um crime?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	4	10	7	2	3	10
True	1	10	7	2	4	51

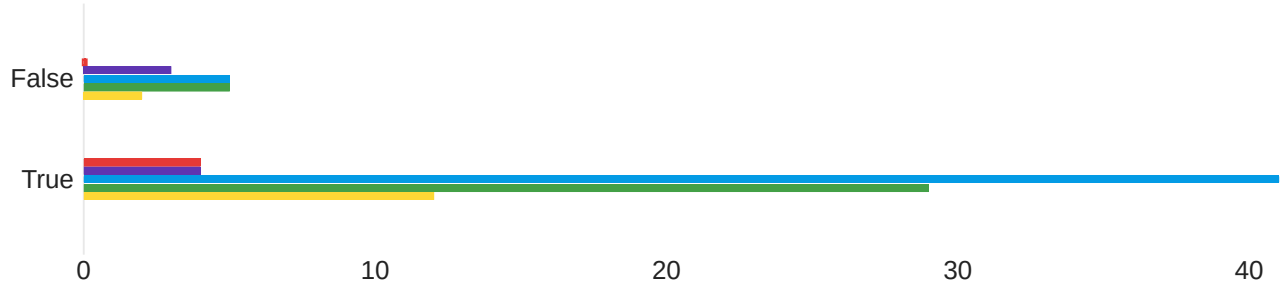
Q112 - VEJA COM ATENÇÃO a notícia abaixo sobre um caso de homicídio! Todas as informações, dados e imagens dos casos são fictícios. INSTRUÇÕES. Abaixo temos o caso judicial referente à notícia anterior e você é um dos membros do júri. Considere o relatório resumido, que traz os principais elementos levantados ao longo do processo, e tome sua decisão. O restante do Júri já se pronunciou. Mesmo com poucos elementos, é preciso que você se posicione. EIKE CASTELO BRAZ TERCEIRO, brasileiro, 34 anos, gerente de banco, que reside em condomínio na cidade do Rio de Janeiro foi denunciado pelo crime de HOMICÍDIO SIMPLES DOLOSO. A vítima era MARCEL SCHMISK BÔAS, brasileiro, empresário, 35 anos, morador de um bairro na zona nobre da mesma cidade. Considere que Homicídio Simples Doloso é matar alguém com a intenção ou assumindo o risco do que se está fazendo. Resumo do caso: A vítima volta para casa e ao descer do ônibus foi apunhalado e acabou morrendo. Acontece que na noite anterior, envolveu-se numa briga de bar com o acusado. As testemunhas afirmaram que a discussão foi feia entre eles. Que depois do acontecido todos foram embora e o bar fechou. A acusação alegou que ninguém teria motivos para matar a vítima, que era um homem trabalhador e que a morte só poderia estar ligada à briga anterior. A defesa afirmou que a vítima tinha um passado duvidoso e envolvimento com o crime. E que o acusado não teve nenhuma participação no fato ocorrido. Nada a acrescentar, conclusos os autos para julgamento. Indique qual sua decisão final, marcando UM dos vereditos:



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	2	0	0	37
True	1	2	2	0	0	229

Campo	False	True
Condenar o réu pelo cometimento de Homicídio Simples Doloso.	18	90
Absolver o réu pelo cometimento de Homicídio Simples Doloso.	19	139
Total	37	229

Q113 - Qual o grau da pena que o réu merece receber?

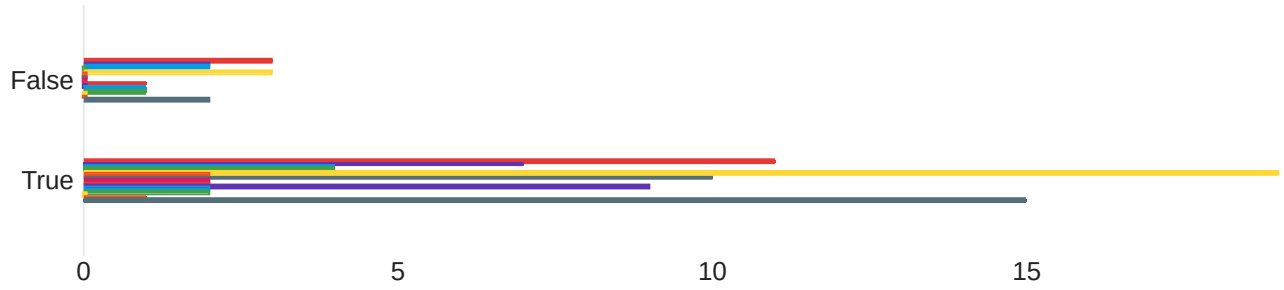


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	34	38	36	2	3	15
True	32	38	36	2	4	90

Campo	False	True
Mínima	0	4
Leve	3	4
Moderada	5	41
Forte	5	29
Máxima	2	12
Total	15	90

35

Q114 - Se a pena para esse crime varia de 6 (seis) a 20 (vinte) anos, a qual tempo de sentença você condena o acusado?

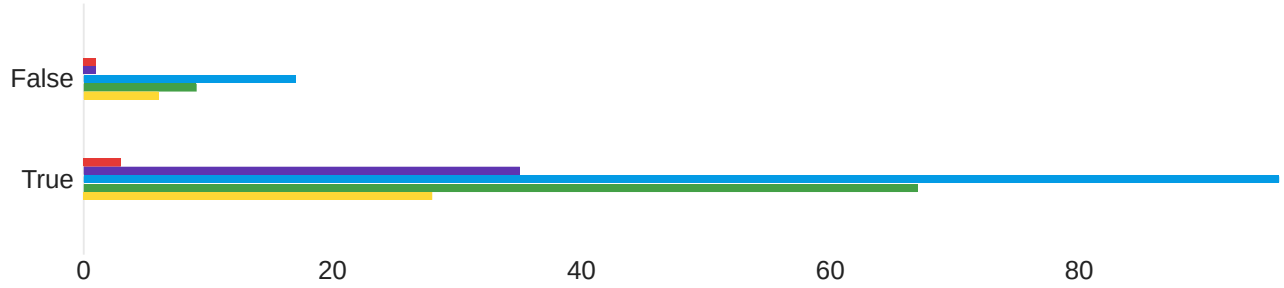


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	23	39	32	6	36	15
True	23	40	31	5	29	90

Campo	False	True
6 anos	3	11
7 anos	2	7
8 anos	2	4
9 anos	0	4
10 anos	3	19
11 anos	0	2
12 anos	0	10
13 anos	0	2
14 anos	1	2
15 anos	0	9
16 anos	1	2

17 anos	1	2
18 anos	0	0
19 anos	0	1
20 anos	2	15
Total	15	90

Q115 - O quanto você confia na sua decisão de sentença?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	5	4	1	1	34
True	1	5	3	1	1	229

Campo	False	True
Nada	1	3
Pouco	1	35
Médio	17	96
Muito	9	67
Totalmente	6	28
Total	34	229

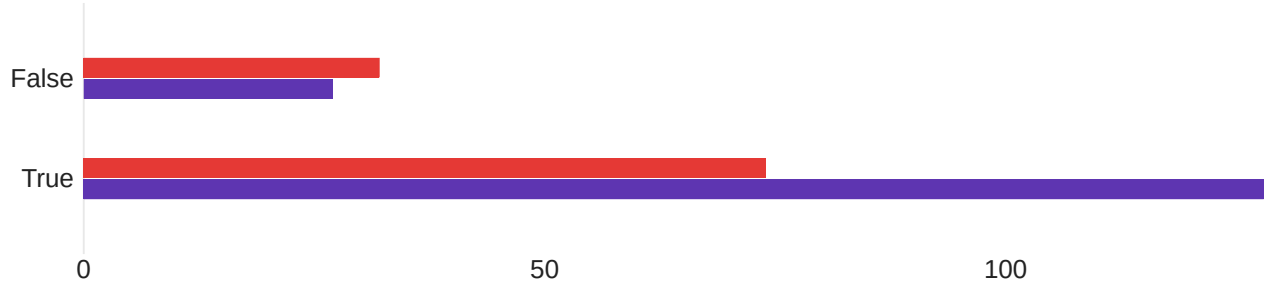
Q116 - O quanto você acredita que o réu estava no controle de suas ações?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	2	10	7	3	7	15
True	1	10	8	2	5	90

Q117 - Na sua opinião, qual a probabilidade de o réu se envolver em um crime?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	10	7	2	6	15
True	1	10	7	2	4	90

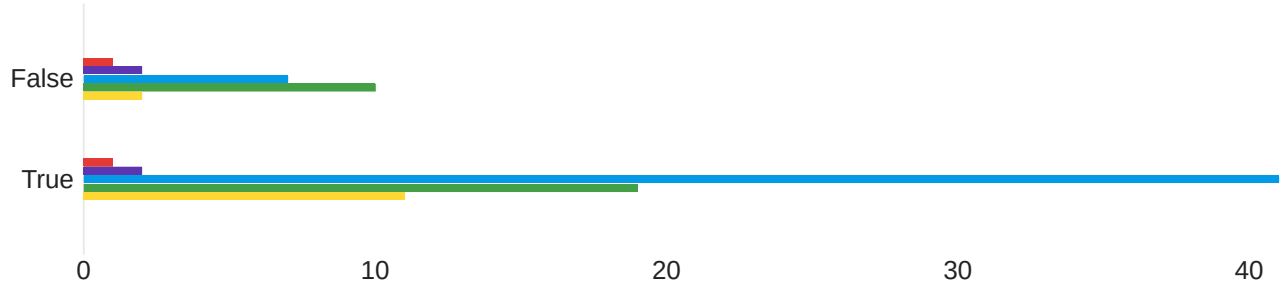
Q118 - VEJA COM ATENÇÃO a notícia abaixo sobre um caso de homicídio! Todas as informações, dados e imagens dos casos são fictícios. INSTRUÇÕES. Abaixo temos o caso judicial referente à notícia anterior e você é um dos membros do júri. Considere o relatório resumido, que traz os principais elementos levantados ao longo do processo, e tome sua decisão. O restante do Júri já se pronunciou. Mesmo com poucos elementos, é preciso que você se posicione. A, brasileiro, 34 anos, que reside na cidade do Rio de Janeiro foi denunciado pelo crime de HOMICÍDIO SIMPLES DOLOSO. A vítima era B, brasileiro, 35 anos, morador da mesma cidade. Considere que Homicídio Simples Doloso é matar alguém com a intenção ou assumindo o risco do que se está fazendo. Resumo do caso: A vítima volta para casa e ao descer do ônibus foi apunhalado e acabou morrendo. Acontece que na noite anterior, envolveu-se numa briga de bar com o acusado. As testemunhas afirmaram que a discussão foi feia entre eles. Que depois do acontecido todos foram embora e o bar fechou. A acusação alegou que ninguém teria motivos para matar a vítima, que era um homem trabalhador e que a morte só poderia estar ligada à briga anterior. A defesa afirmou que a vítima tinha um passado duvidoso e envolvimento com o crime. E que o acusado não teve nenhuma participação no fato ocorrido. Nada a acrescentar, conclusos os autos para julgamento. Indique qual sua decisão final, marcando UM dos vereditos:



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	1	0	0	59
True	1	2	2	0	0	202

Campo	False	True
Condenar o réu pelo cometimento de Homicídio Simples Doloso.	32	74
Absolver o réu pelo cometimento de Homicídio Simples Doloso.	27	128
Total	59	202

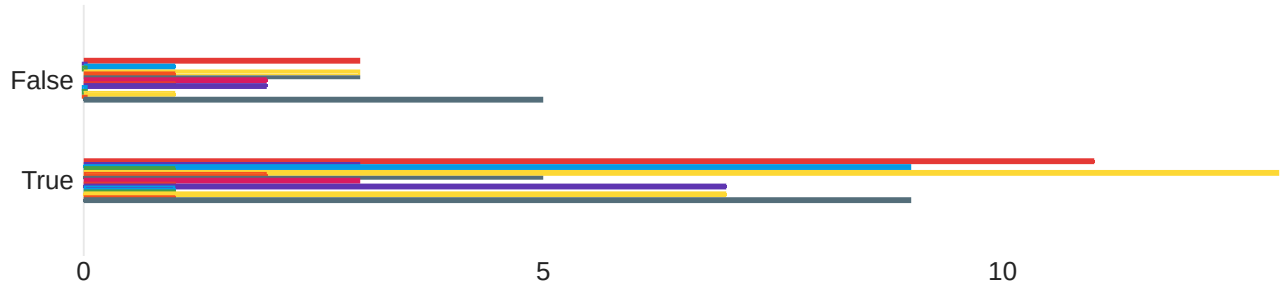
Q119 - Qual o grau da pena que o réu merece receber?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	32	38	36	2	4	22
True	32	38	35	2	3	74

Campo	False	True
Mínima	1	1
Leve	2	2
Moderada	7	41
Forte	10	19
Máxima	2	11
Total	22	74

Q120 - Se a pena para esse crime varia de 6 (seis) a 20 (vinte) anos, a qual tempo de sentença você condena o acusado?

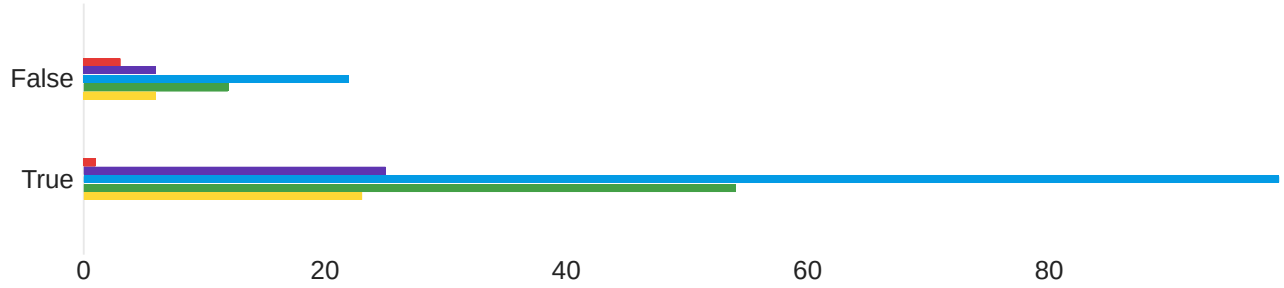


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	23	39	31	5	24	22
True	23	40	32	5	30	74

Campo	False	True
6 anos	3	11
7 anos	0	3
8 anos	1	9
9 anos	0	1
10 anos	3	13
11 anos	1	2
12 anos	3	5
13 anos	2	3
14 anos	1	1
15 anos	2	7
16 anos	0	1

17 anos	0	1
18 anos	1	7
19 anos	0	1
20 anos	5	9
Total	22	74

Q121 - O quanto você confia na sua decisão de sentença?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	5	3	1	1	49
True	1	5	3	1	1	202

Campo	False	True
Nada	3	1
Pouco	6	25
Médio	22	99
Muito	12	54
Totalmente	6	23
Total	49	202

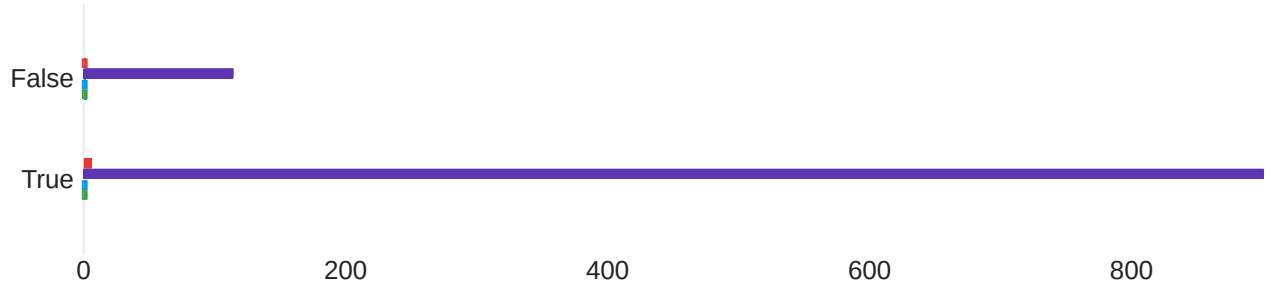
Q122 - O quanto você acredita que o réu estava no controle de suas ações?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	4	10	8	2	3	22
True	1	10	7	2	6	74

Q123 - Na sua opinião, qual a probabilidade de o réu se envolver em um crime?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	5	10	8	2	3	22
True	3	10	7	2	4	74

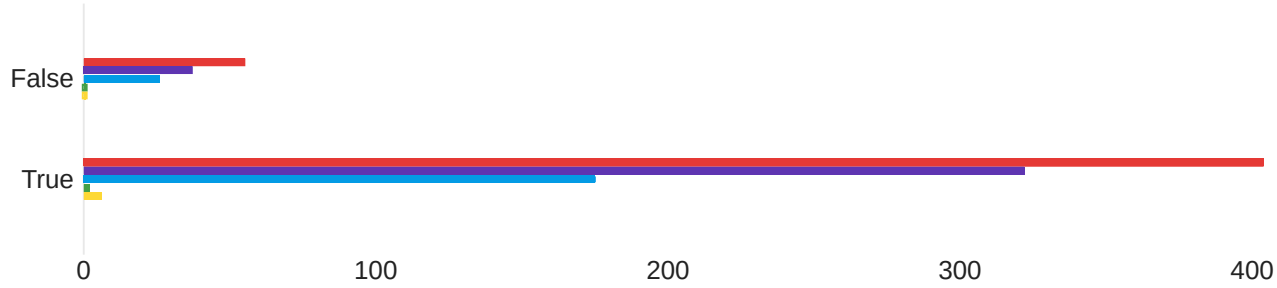
Q33 - Qual foi o fato descrito no cenário?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	6	4	0	0	118
True	1	5	4	0	0	909

Campo	False	True
Uma pessoa foi vítima de estelionato.	1	6
Uma pessoa foi morta/assassinada.	114	901
Uma pessoa sofreu um acidente de carro.	1	2
Uma pessoa sofreu uma queda.	2	0
Total	118	909

Q30 - Qual você considera que era a cor/raça do acusado?



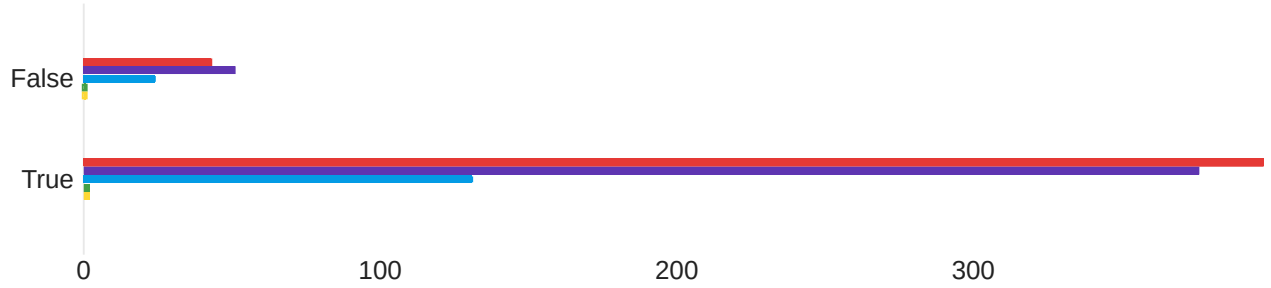
Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	5	3	2	3	118
True	1	7	3	2	3	909

Campo	False	True
Branco	55	404
Preto	37	322
Pardo	26	175
Indígena	0	2
Amarelo	0	6
Total	118	909

Q31 - Qual você considera que era a classe do acusado?

Nenhum dado encontrado - talvez seus filtros estejam restritos demais!

Q32 - Qual você considera que era a cor/raça da vítima?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	7	4	3	7	118
True	1	9	4	3	7	909

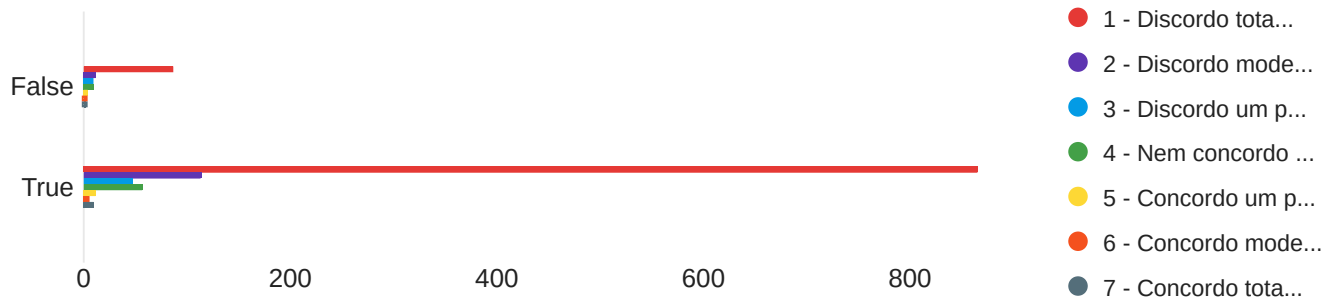
Campo	False	True
Branco	43	398
Preto	51	376
Pardo	24	131
Indígena	0	2
Amarelo	0	2
Total	118	909

Q34 - Qual você considera que era a classe da vítima?

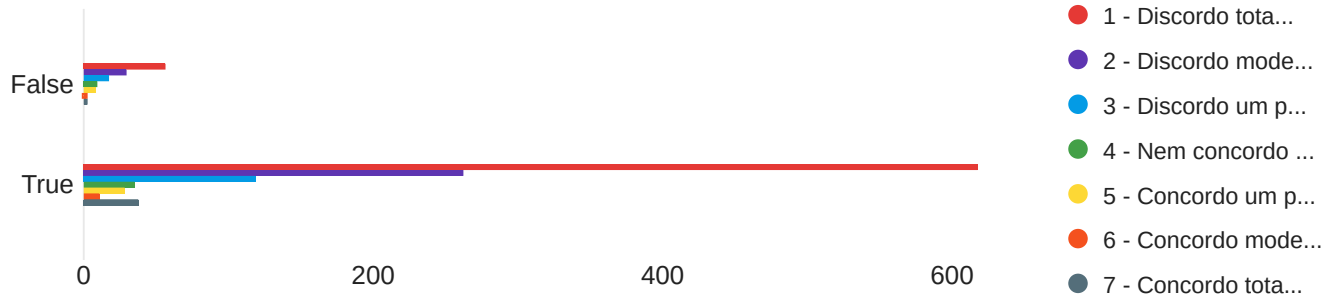
Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	10	5	2	5	118
True	0	10	5	3	7	909

Q35 - Por favor, leia atentamente cada um dos itens abaixo a respeito da população negra e em seguida, indique seu grau de concordância com cada um deles. Por gentileza responda todos, utilizando a escala abaixo, selecione o que melhor representa sua resposta.

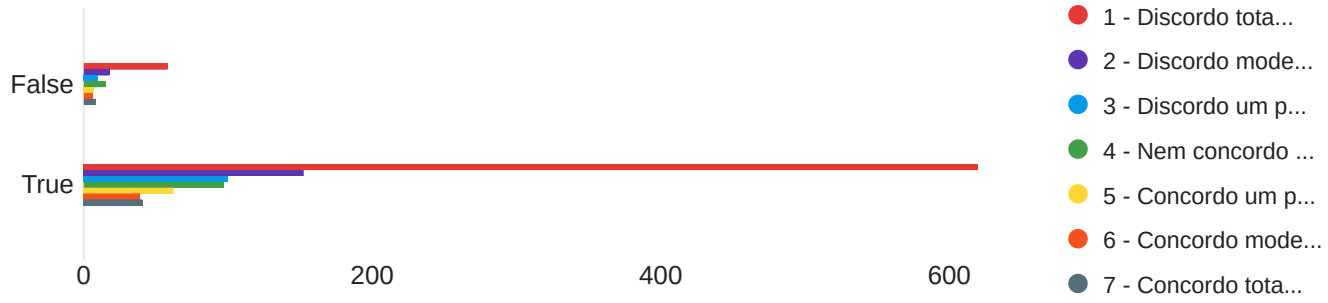
Eles têm conseguido mais do que merecem.



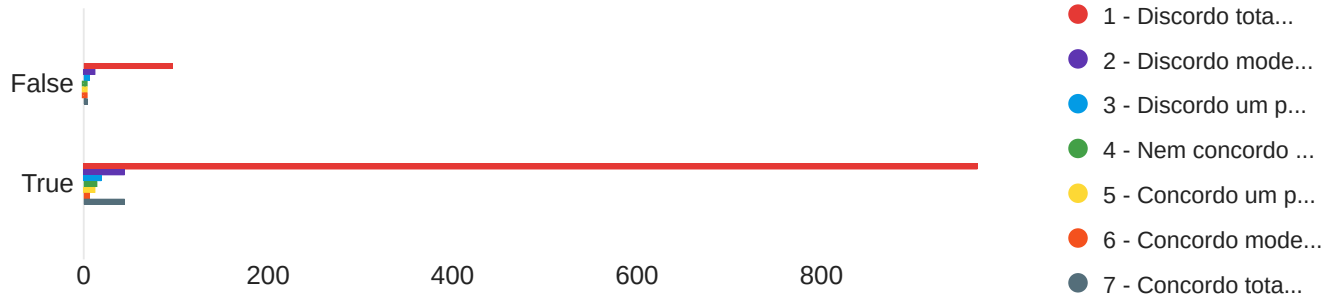
Eles recebem muito respeito e consideração.



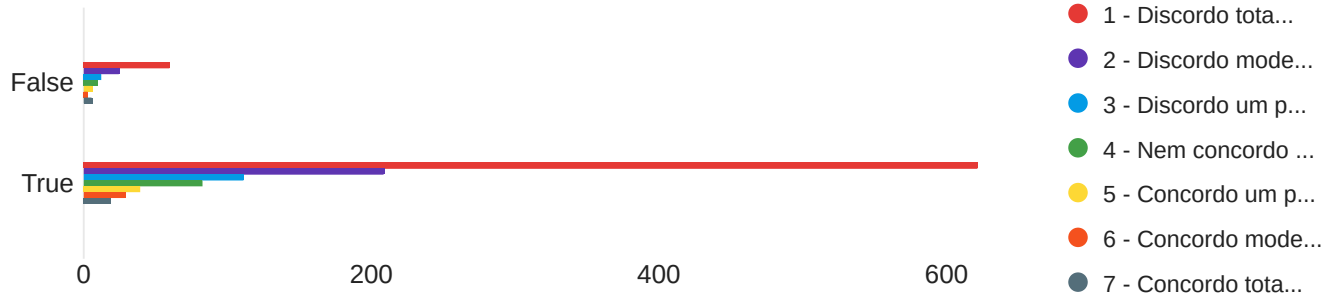
Eles são muito exigentes em relação aos seus direitos.



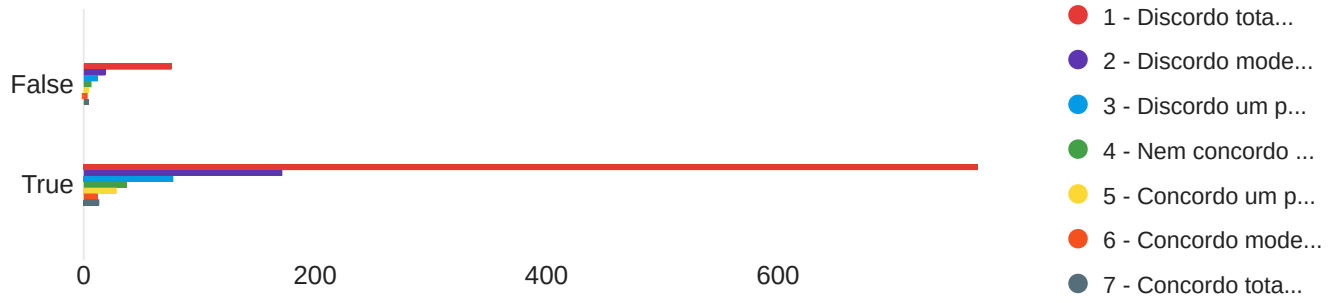
A discriminação não é um problema do Brasil.



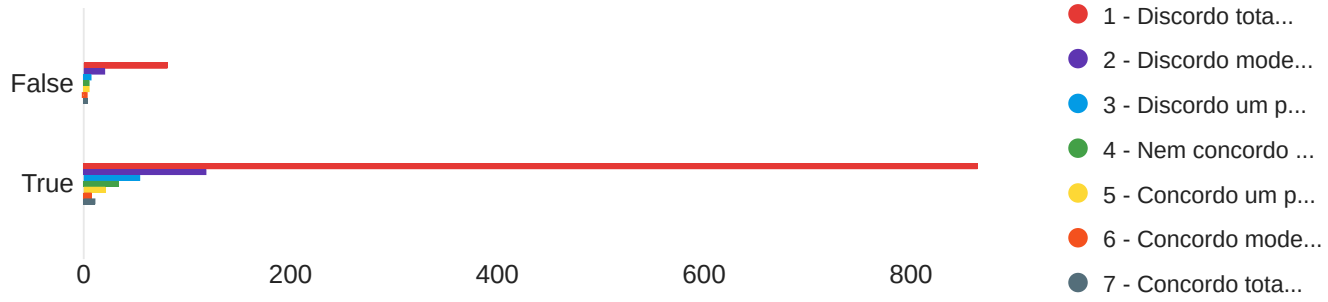
Eles têm muita influência política.



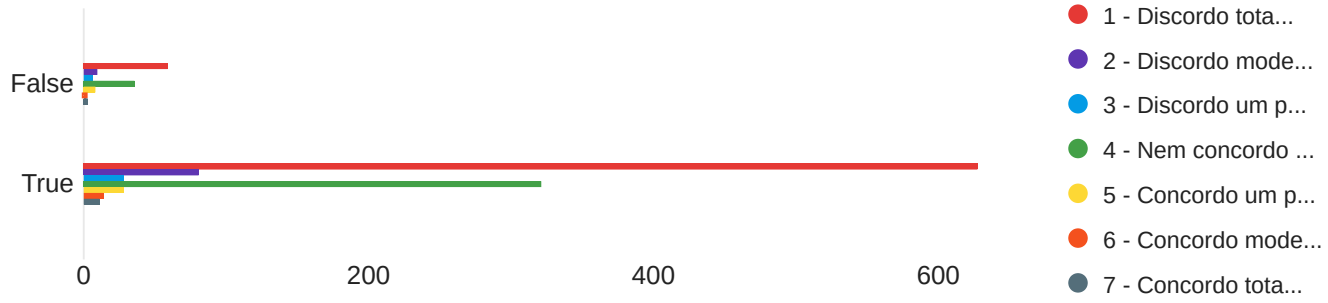
Eles não necessitam de ajuda, apenas devem se esforçar.



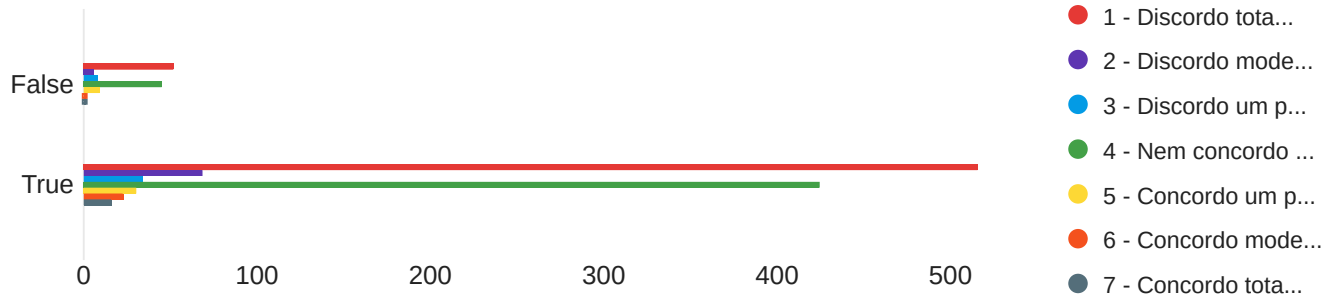
Eles devem superar o preconceito sem apoio como aconteceu com outros grupos...



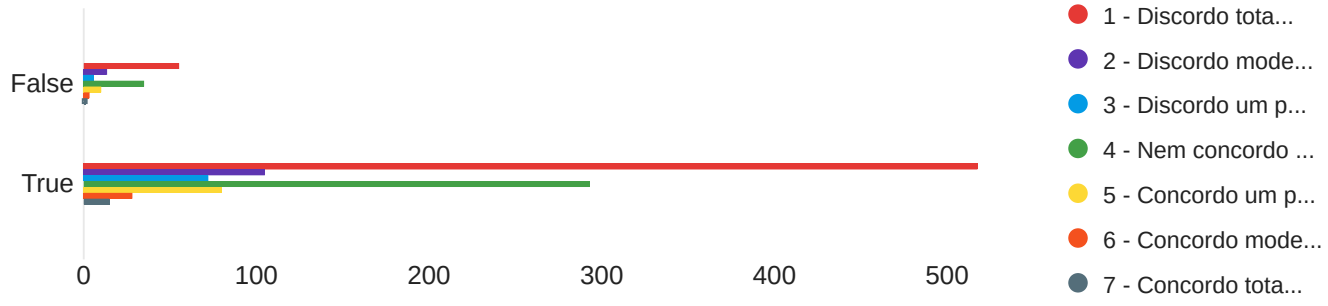
Eles são mais habilidosos em trabalhos manuais.



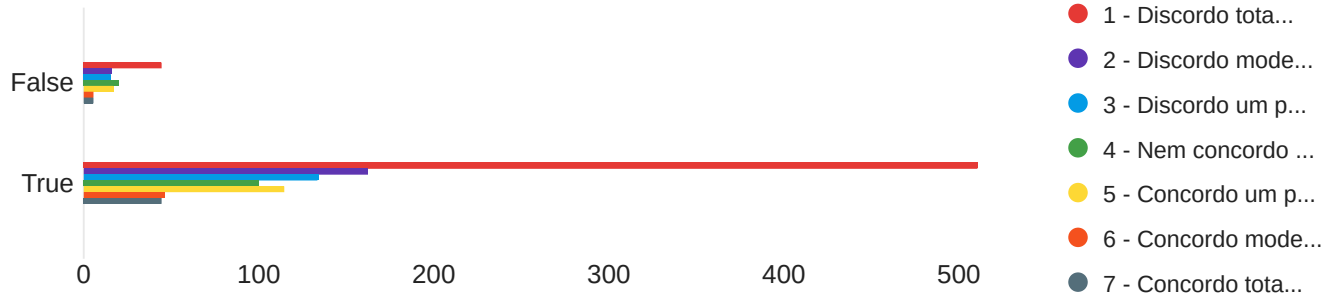
Possuem maior habilidade culinária.



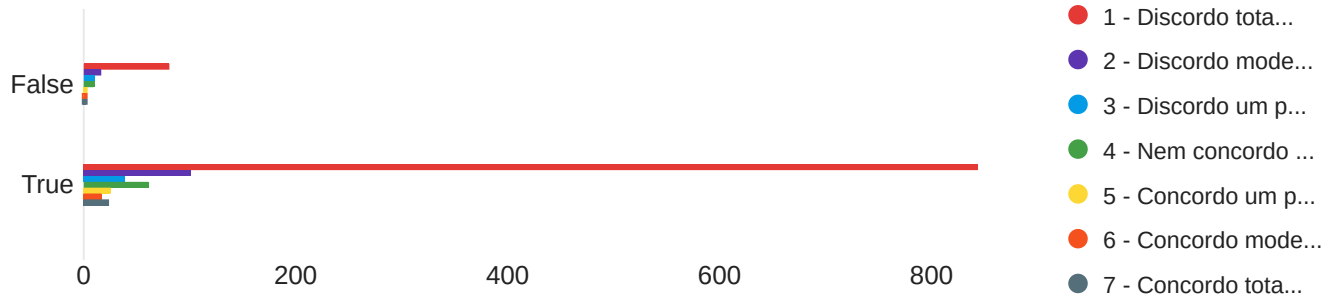
Estão em moda suas danças pela sensualidade que expressam.



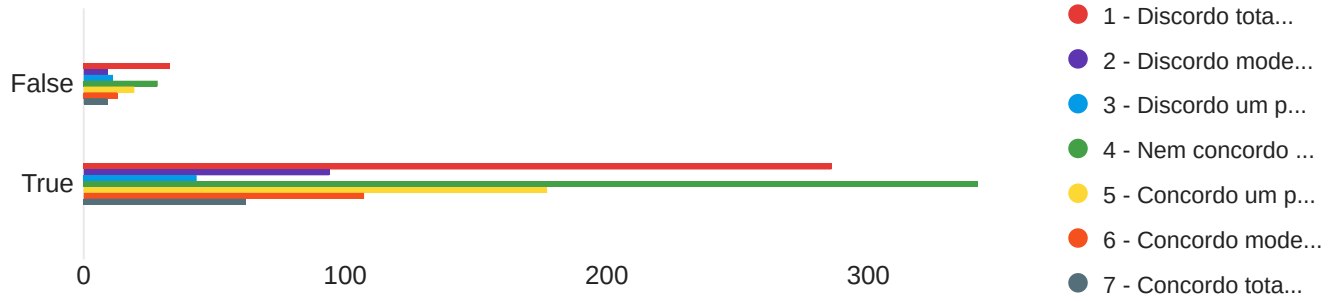
Tem-se dada demasiada importância aos seus movimentos de protesto.



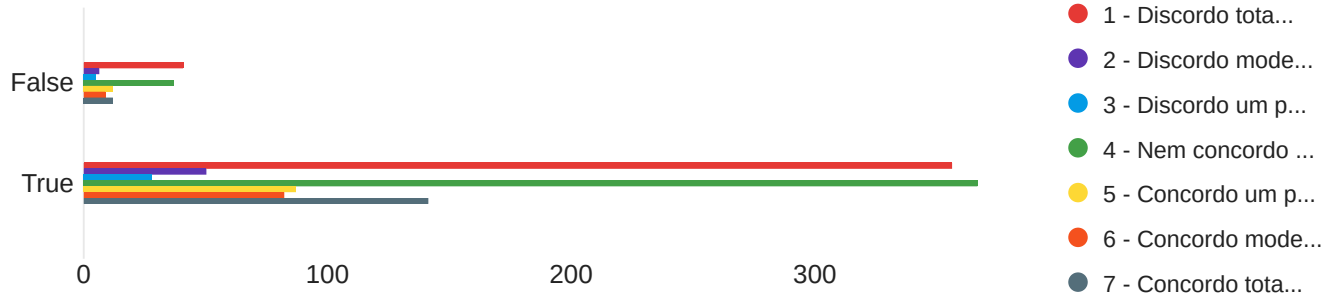
Parece pouco prudente dar importância às suas queixas.



Apresentam melhor desempenho em modalidades esportivas.



Possuem uma beleza diferente.



False

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Eles têm conseguido mais do que merecem.	36	42	37	1	1	122
Eles recebem muito respeito e consideração.	36	42	37	1	2	122
Eles são muito exigentes em relação aos seus direitos.	36	42	38	2	4	122
A discriminação não é um problema do Brasil.	36	42	37	1	2	122
Eles têm muita influência política.	36	42	37	2	3	122
Eles não necessitam de ajuda, apenas devem se esforçar.	36	42	37	1	2	122
Eles devem superar o preconceito sem apoio como aconteceu com outros grupos.	36	42	37	1	2	122
Eles são mais habilidosos em trabalhos manuais.	36	42	38	2	3	122
Possuem maior habilidade culinária.	36	42	38	2	3	122
Estão em moda suas danças pela sensualidade que expressam.	36	41	38	2	3	122
Tem-se dada demasiada importância aos seus movimentos de protesto.	36	42	38	2	3	122
Parece pouco prudente dar importância às suas queixas.	36	42	37	1	2	122
Apresentam melhor desempenho em modalidades esportivas.	36	42	39	2	4	122
Possuem uma beleza diferente.	36	42	38	2	4	122

True

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Eles têm conseguido mais do que merecem.	36	42	36	1	1	1111

Eles recebem muito respeito e consideração.	36	42	37	1	2	1111
Eles são muito exigentes em relação aos seus direitos.	36	42	37	2	3	1111
A discriminação não é um problema do Brasil.	36	42	36	1	2	1111
Eles têm muita influência política.	36	42	37	1	2	1111
Eles não necessitam de ajuda, apenas devem se esforçar.	36	42	37	1	1	1111
Eles devem superar o preconceito sem apoio como aconteceu com outros grupos.	36	42	36	1	1	1111
Eles são mais habilidosos em trabalhos manuais.	36	42	37	2	2	1111
Possuem maior habilidade culinária.	36	42	38	2	3	1111
Estão em moda suas danças pela sensualidade que expressam.	36	42	38	2	3	1111
Tem-se dada demasiada importância aos seus movimentos de protesto.	36	42	38	2	3	1111
Parece pouco prudente dar importância às suas queixas.	36	42	37	1	2	1111
Apresentam melhor desempenho em modalidades esportivas.	36	42	39	2	4	1111
Possuem uma beleza diferente.	36	42	39	2	4	1111

Eles têm conseguido mais do que merecem.

Campo	False	True	Total
1 - Discordo totalmente	86	866	952
2 - Discordo moderadamente	12	114	126
3 - Discordo um pouco	9	47	56
4 - Nem concordo nem discordo	10	57	67

5 - Concordo um pouco	4	12	16
6 - Concordo moderadamente	0	5	5
7 - Concordo totalmente	1	10	11

Eles recebem muito respeito e consideração.

Campo	False	True	Total
1 - Discordo totalmente	56	618	674
2 - Discordo moderadamente	29	262	291
3 - Discordo um pouco	17	119	136
4 - Nem concordo nem discordo	9	35	44
5 - Concordo um pouco	8	28	36
6 - Concordo moderadamente	1	11	12
7 - Concordo totalmente	2	38	40

Eles são muito exigentes em relação aos seus direitos.

Campo	False	True	Total
1 - Discordo totalmente	58	620	678
2 - Discordo moderadamente	18	152	170
3 - Discordo um pouco	10	100	110
4 - Nem concordo nem discordo	15	97	112
5 - Concordo um pouco	7	62	69
6 - Concordo moderadamente	6	39	45
7 - Concordo totalmente	8	41	49

A discriminação não é um problema do Brasil.

Campo	False	True	Total
1 - Discordo totalmente	96	970	1066
2 - Discordo moderadamente	12	45	57
3 - Discordo um pouco	6	20	26
4 - Nem concordo nem discordo	1	14	15
5 - Concordo um pouco	3	12	15
6 - Concordo moderadamente	0	6	6
7 - Concordo totalmente	4	44	48

Eles têm muita influência política.

Campo	False	True	Total
1 - Discordo totalmente	60	622	682
2 - Discordo moderadamente	25	209	234
3 - Discordo um pouco	12	111	123
4 - Nem concordo nem discordo	10	82	92
5 - Concordo um pouco	6	39	45
6 - Concordo moderadamente	3	29	32
7 - Concordo totalmente	6	19	25

Eles não necessitam de ajuda, apenas devem se esforçar.

Campo	False	True	Total
1 - Discordo totalmente	76	773	849

2 - Discordo moderadamente	19	171	190
3 - Discordo um pouco	12	77	89
4 - Nem concordo nem discordo	6	37	43
5 - Concordo um pouco	4	28	32
6 - Concordo moderadamente	1	12	13
7 - Concordo totalmente	4	13	17

Eles devem superar o preconceito sem apoio como aconteceu com outros grupos...

Campo	False	True	Total
1 - Discordo totalmente	81	865	946
2 - Discordo moderadamente	20	118	138
3 - Discordo um pouco	7	54	61
4 - Nem concordo nem discordo	5	34	39
5 - Concordo um pouco	5	21	26
6 - Concordo moderadamente	0	8	8
7 - Concordo totalmente	4	11	15

Eles são mais habilidosos em trabalhos manuais.

Campo	False	True	Total
1 - Discordo totalmente	59	628	687
2 - Discordo moderadamente	9	81	90
3 - Discordo um pouco	6	28	34
4 - Nem concordo nem discordo	36	321	357

5 - Concordo um pouco	8	28	36
6 - Concordo moderadamente	1	14	15
7 - Concordo totalmente	3	11	14

Possuem maior habilidade culinária.

Campo	False	True	Total
1 - Discordo totalmente	52	516	568
2 - Discordo moderadamente	6	68	74
3 - Discordo um pouco	8	34	42
4 - Nem concordo nem discordo	45	424	469
5 - Concordo um pouco	9	30	39
6 - Concordo moderadamente	1	23	24
7 - Concordo totalmente	1	16	17

Estão em moda suas danças pela sensualidade que expressam.

Campo	False	True	Total
1 - Discordo totalmente	55	518	573
2 - Discordo moderadamente	13	105	118
3 - Discordo um pouco	6	72	78
4 - Nem concordo nem discordo	35	293	328
5 - Concordo um pouco	10	80	90
6 - Concordo moderadamente	3	28	31
7 - Concordo totalmente	0	15	15

Tem-se dada demasiada importância aos seus movimentos de protesto.

Campo	False	True	Total
1 - Discordo totalmente	44	511	555
2 - Discordo moderadamente	16	162	178
3 - Discordo um pouco	15	134	149
4 - Nem concordo nem discordo	20	100	120
5 - Concordo um pouco	17	114	131
6 - Concordo moderadamente	5	46	51
7 - Concordo totalmente	5	44	49

Parece pouco prudente dar importância às suas queixas.

Campo	False	True	Total
1 - Discordo totalmente	80	844	924
2 - Discordo moderadamente	16	101	117
3 - Discordo um pouco	10	39	49
4 - Nem concordo nem discordo	10	61	71
5 - Concordo um pouco	3	25	28
6 - Concordo moderadamente	1	17	18
7 - Concordo totalmente	2	24	26

Apresentam melhor desempenho em modalidades esportivas.

Campo	False	True	Total
1 - Discordo totalmente	33	286	319

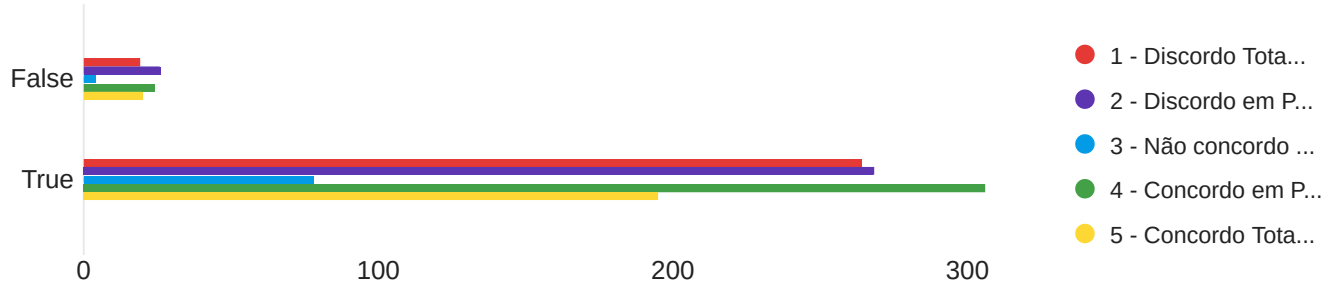
2 - Discordo moderadamente	9	94	103
3 - Discordo um pouco	11	43	54
4 - Nem concordo nem discordo	28	342	370
5 - Concordo um pouco	19	177	196
6 - Concordo moderadamente	13	107	120
7 - Concordo totalmente	9	62	71

Possuem uma beleza diferente.

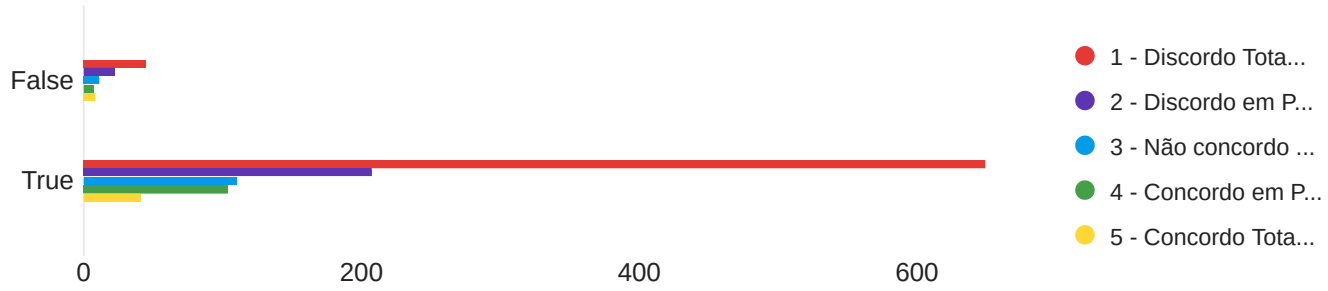
Campo	False	True	Total
1 - Discordo totalmente	41	356	397
2 - Discordo moderadamente	6	50	56
3 - Discordo um pouco	5	28	33
4 - Nem concordo nem discordo	37	367	404
5 - Concordo um pouco	12	87	99
6 - Concordo moderadamente	9	82	91
7 - Concordo totalmente	12	141	153

Q37 - Agora este questionário busca entender as suas opiniões sobre diversos temas sociais. Por favor, marque o quanto você concorda com as afirmações abaixo:

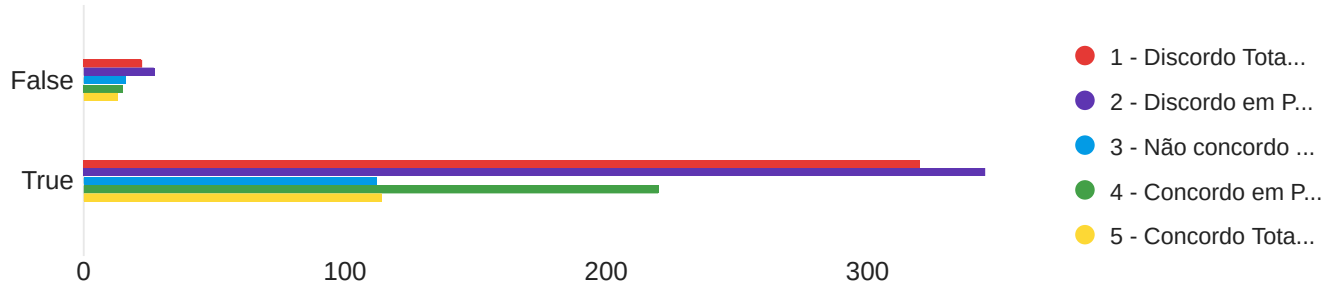
Do jeito que as coisas estão indo nesse país, serão necessárias medidas sev...



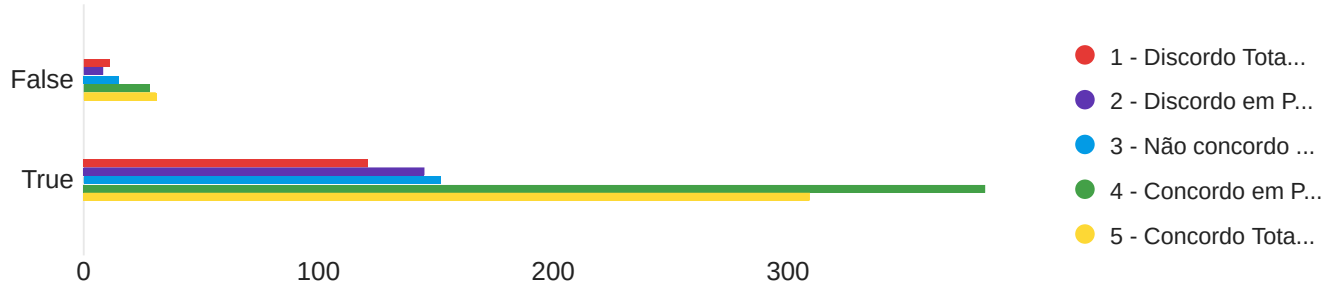
A situação do nosso país está ficando tão séria que ações firmes seriam jus...



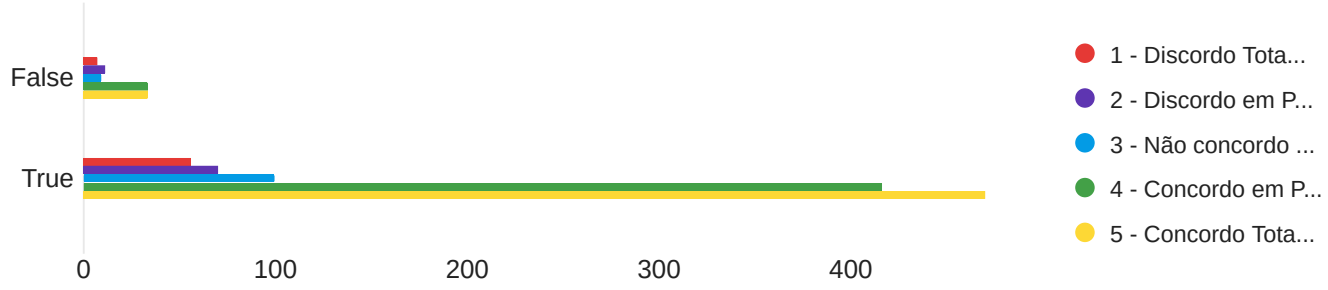
Ser gentil com criminosos só os encoraja a tirar proveito de sua fraqueza,...



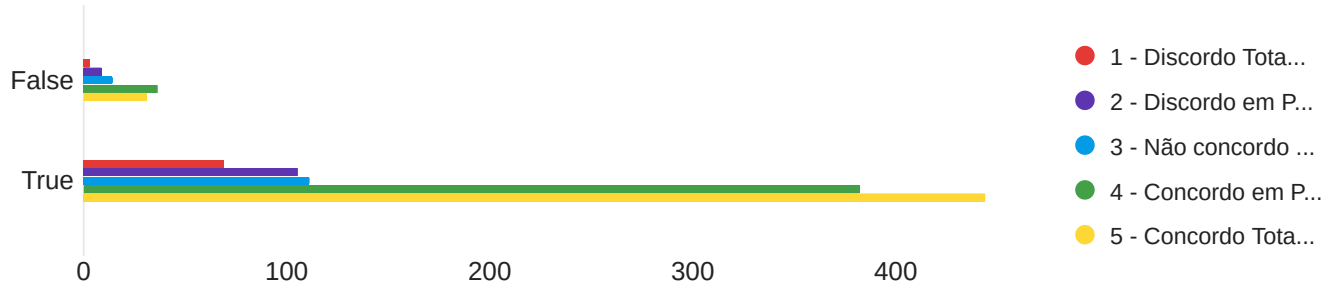
Quanto maior o número de pessoas preparadas para desafiar o governo, melhor...



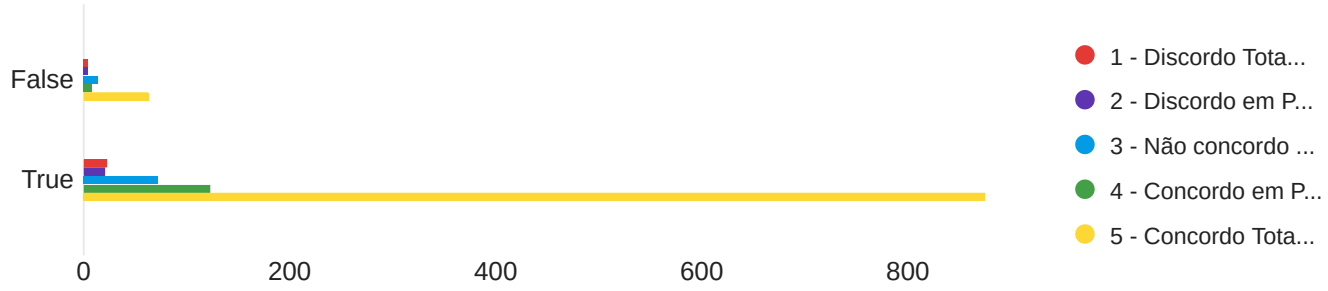
Quanto maior o número de pessoas preparadas para protestar contra o governo...



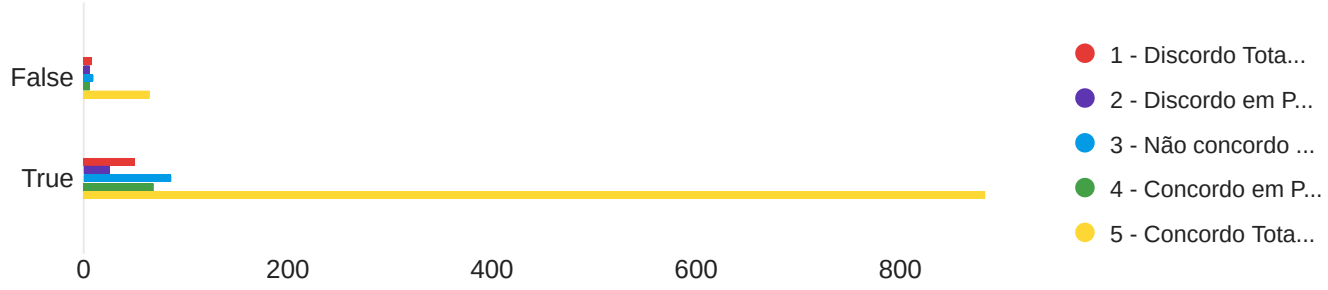
Quanto maior o número de pessoas preparadas para criticar as autoridades, m...



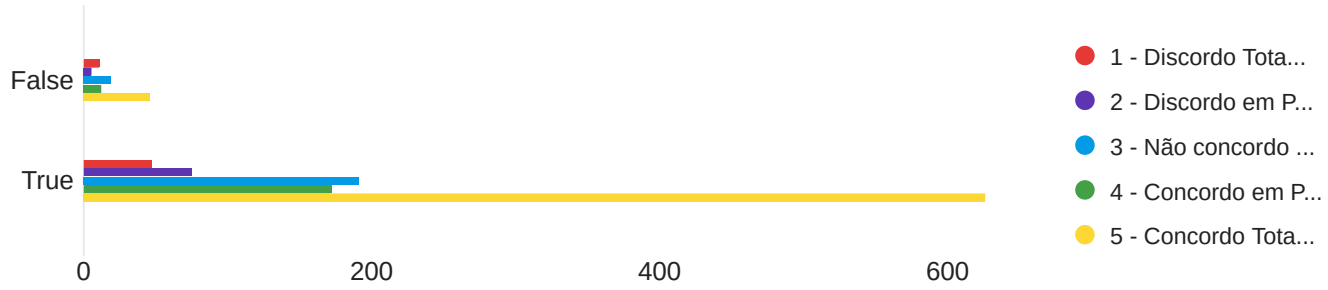
As pessoas deveriam ter as suas próprias preferências sexuais, mesmo se iss...



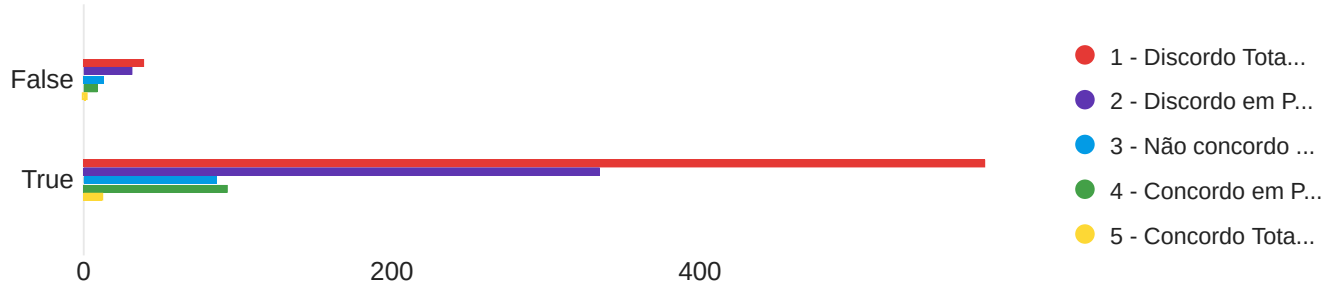
Não há nada de errado com sexo antes do casamento.



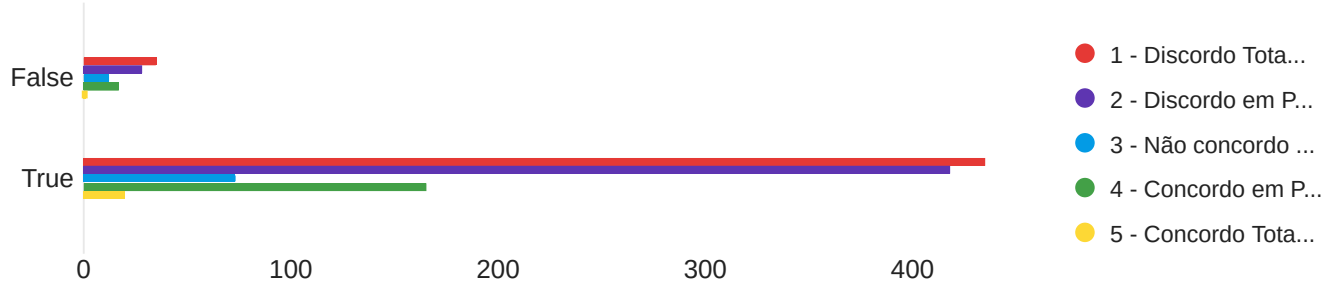
Não tem absolutamente nada errado com praias de nudismo.



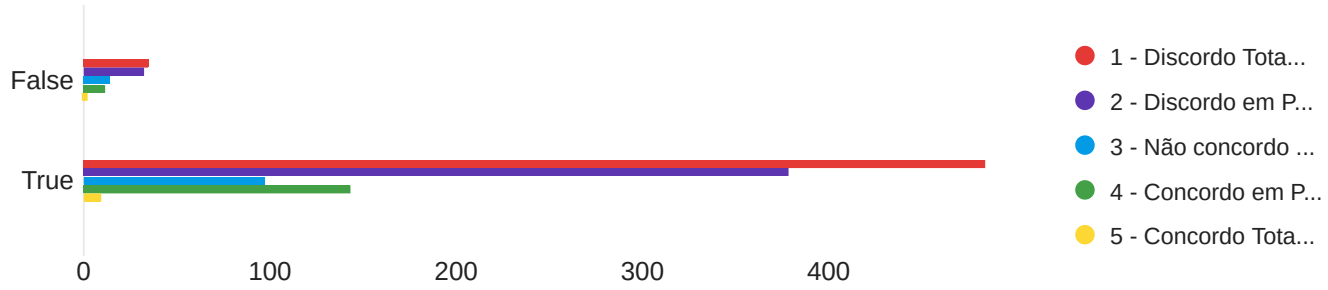
O segredo para uma boa vida é a obediência àqueles que estão no controle.



As autoridades devem ser obedecidas porque elas estão na melhor posição par...



Nosso país será melhor se obedecermos nossos líderes.



False

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Do jeito que as coisas estão indo nesse país, serão necessárias medidas severas para endireitar os meliantes, os criminosos e os pervertidos.	1	5	3	1	2	93
A situação do nosso país está ficando tão séria que ações firmes seriam justificadas se eliminassem os desordeiros e nos levassem de volta ao nosso verdadeiro caminho.	1	5	2	1	2	93
Ser gentil com criminosos só os encoraja a tirar proveito de sua fraqueza, sendo melhor agir de maneira firme e dura com eles.	1	5	3	1	2	93
Quanto maior o número de pessoas preparadas para desafiar o governo, melhor para a sociedade.	1	5	4	1	2	93
Quanto maior o número de pessoas preparadas para protestar contra o governo, melhor para a sociedade.	1	5	4	1	2	93
Quanto maior o número de pessoas preparadas para criticar as autoridades, melhor para a sociedade.	1	5	4	1	1	93
As pessoas deveriam ter as suas próprias preferências sexuais, mesmo se isso torná-las diferentes do resto da sociedade.	1	5	4	1	1	93

Não há nada de errado com sexo antes do casamento.	1	5	4	1	2	93
Não tem absolutamente nada errado com praias de nudismo.	1	5	4	1	2	93
O segredo para uma boa vida é a obediência àqueles que estão no controle.	1	5	2	1	1	93
As autoridades devem ser obedecidas porque elas estão na melhor posição para saber o que é bom para o país.	1	5	2	1	1	93
Nosso país será melhor se obedecermos nossos líderes.	1	5	2	1	1	93

True

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Do jeito que as coisas estão indo nesse país, serão necessárias medidas severas para endireitar os meliantes, os criminosos e os pervertidos.	1	5	3	1	2	1111
A situação do nosso país está ficando tão séria que ações firmes seriam justificadas se eliminassem os desordeiros e nos levassem de volta ao nosso verdadeiro caminho.	1	5	2	1	1	1111
Ser gentil com criminosos só os encoraja a tirar proveito de sua fraqueza, sendo melhor agir de maneira firme e dura com eles.	1	5	3	1	2	1111
Quanto maior o número de pessoas preparadas para desafiar o governo, melhor para a sociedade.	1	5	4	1	2	1111
Quanto maior o número de pessoas preparadas para protestar contra o governo, melhor para a sociedade.	1	5	4	1	1	1111
Quanto maior o número de pessoas preparadas para criticar as autoridades, melhor para a sociedade.	1	5	4	1	1	1111

As pessoas deveriam ter as suas próprias preferências sexuais, mesmo se isso torná-las diferentes do resto da sociedade.	1	5	5	1	1	1111
Não há nada de errado com sexo antes do casamento.	1	5	5	1	1	1111
Não tem absolutamente nada errado com praias de nudismo.	1	5	4	1	1	1111
O segredo para uma boa vida é a obediência àqueles que estão no controle.	1	5	2	1	1	1111
As autoridades devem ser obedecidas porque elas estão na melhor posição para saber o que é bom para o país.	1	5	2	1	1	1111
Nosso país será melhor se obedecermos nossos líderes.	1	5	2	1	1	1111

Do jeito que as coisas estão indo nesse país, serão necessárias medidas sev...

Campo	False	True	Total
1 - Discordo Totalmente	19	264	283
2 - Discordo em Parte	26	268	294
3 - Não concordo nem discordo	4	78	82
4 - Concordo em Parte	24	306	330
5 - Concordo Totalmente	20	195	215

A situação do nosso país está ficando tão séria que ações firmes seriam jus...

Campo	False	True	Total
1 - Discordo Totalmente	45	649	694
2 - Discordo em Parte	22	207	229

3 - Não concordo nem discordo	11	110	121
4 - Concordo em Parte	7	104	111
5 - Concordo Totalmente	8	41	49

Ser gentil com criminosos só os encoraja a tirar proveito de sua fraqueza,...

Campo	False	True	Total
1 - Discordo Totalmente	22	320	342
2 - Discordo em Parte	27	345	372
3 - Não concordo nem discordo	16	112	128
4 - Concordo em Parte	15	220	235
5 - Concordo Totalmente	13	114	127

Quanto maior o número de pessoas preparadas para desafiar o governo, melhor...

Campo	False	True	Total
1 - Discordo Totalmente	11	121	132
2 - Discordo em Parte	8	145	153
3 - Não concordo nem discordo	15	152	167
4 - Concordo em Parte	28	384	412
5 - Concordo Totalmente	31	309	340

Quanto maior o número de pessoas preparadas para protestar contra o governo...

Campo	False	True	Total
1 - Discordo Totalmente	7	56	63

2 - Discordo em Parte	11	70	81
3 - Não concordo nem discordo	9	99	108
4 - Concordo em Parte	33	416	449
5 - Concordo Totalmente	33	470	503

Quanto maior o número de pessoas preparadas para criticar as autoridades, m...

Campo	False	True	Total
1 - Discordo Totalmente	3	69	72
2 - Discordo em Parte	9	105	114
3 - Não concordo nem discordo	14	111	125
4 - Concordo em Parte	36	382	418
5 - Concordo Totalmente	31	444	475

As pessoas deveriam ter as suas próprias preferências sexuais, mesmo se iss...

Campo	False	True	Total
1 - Discordo Totalmente	4	22	26
2 - Discordo em Parte	4	20	24
3 - Não concordo nem discordo	14	72	86
4 - Concordo em Parte	8	122	130
5 - Concordo Totalmente	63	875	938

Não há nada de errado com sexo antes do casamento.

Campo	False	True	Total

1 - Discordo Totalmente	8	50	58
2 - Discordo em Parte	6	25	31
3 - Não concordo nem discordo	9	85	94
4 - Concordo em Parte	6	68	74
5 - Concordo Totalmente	64	883	947

Não tem absolutamente nada errado com praias de nudismo.

Campo	False	True	Total
1 - Discordo Totalmente	11	47	58
2 - Discordo em Parte	5	75	80
3 - Não concordo nem discordo	19	191	210
4 - Concordo em Parte	12	172	184
5 - Concordo Totalmente	46	626	672

O segredo para uma boa vida é a obediência àqueles que estão no controle.

Campo	False	True	Total
1 - Discordo Totalmente	39	585	624
2 - Discordo em Parte	31	335	366
3 - Não concordo nem discordo	13	86	99
4 - Concordo em Parte	9	93	102
5 - Concordo Totalmente	1	12	13

As autoridades devem ser obedecidas porque elas estão na melhor posição par...

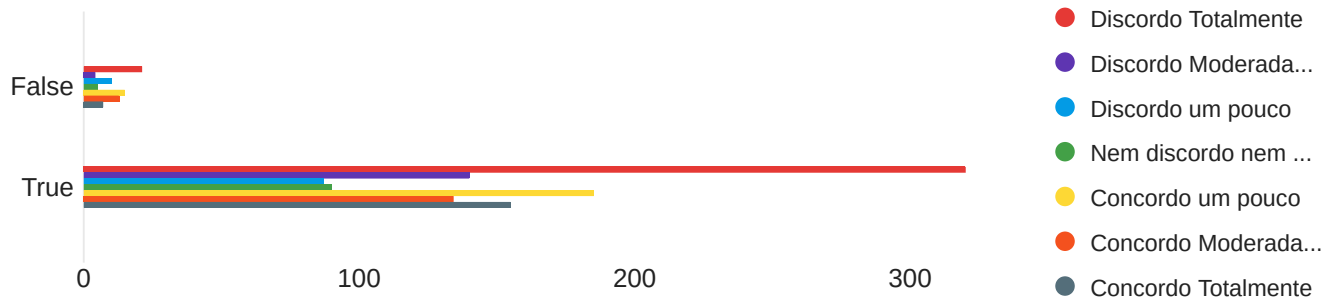
Campo	False	True	Total
1 - Discordo Totalmente	35	435	470
2 - Discordo em Parte	28	418	446
3 - Não concordo nem discordo	12	73	85
4 - Concordo em Parte	17	165	182
5 - Concordo Totalmente	1	20	21

Nosso país será melhor se obedecermos nossos líderes.

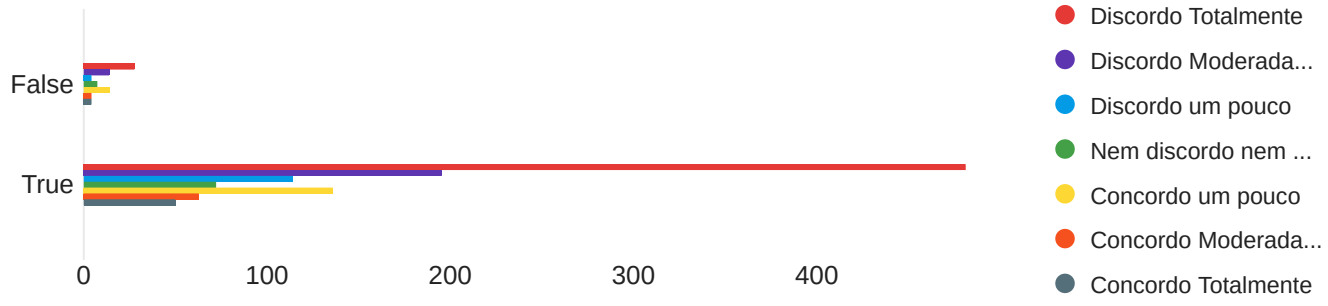
Campo	False	True	Total
1 - Discordo Totalmente	35	484	519
2 - Discordo em Parte	32	378	410
3 - Não concordo nem discordo	14	97	111
4 - Concordo em Parte	11	143	154
5 - Concordo Totalmente	1	9	10

Q97 - A seguir encontram-se algumas afirmações com as quais você pode concordar ou não. Considerando que não existe resposta certa ou errada, leia atentamente cada uma delas e utilize a escala para indicar em que medida você discorda ou concorda.

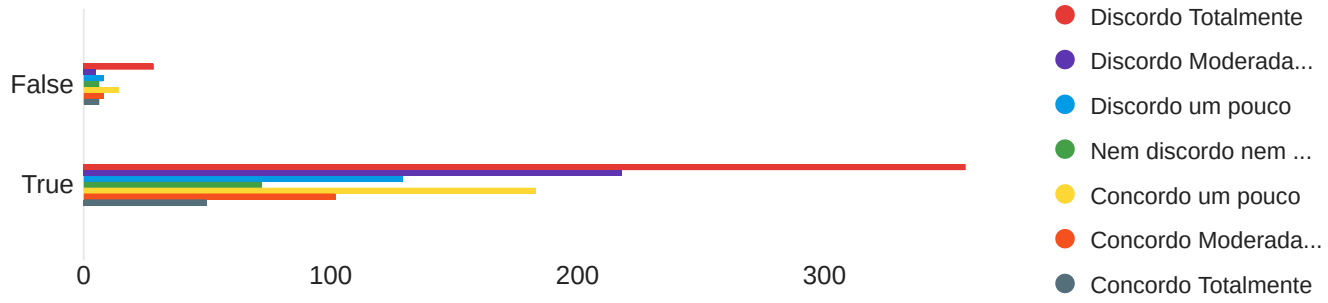
O que este país necessita, principalmente, antes de leis ou planos político...



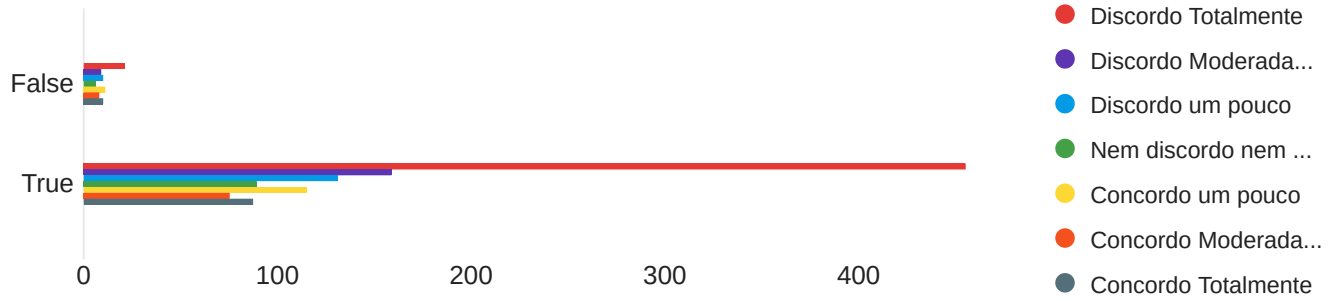
A maioria de nossos problemas sociais estaria resolvida se pudéssemos nos l...



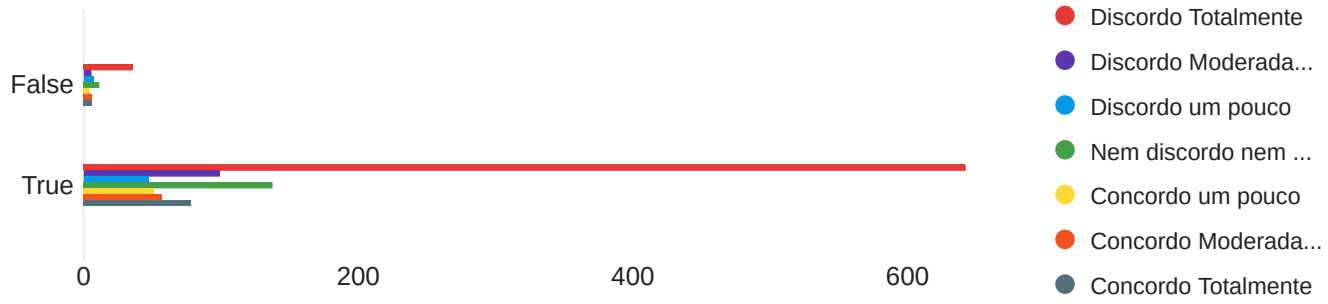
A obediência e o respeito à autoridade são as principais virtudes que devem...



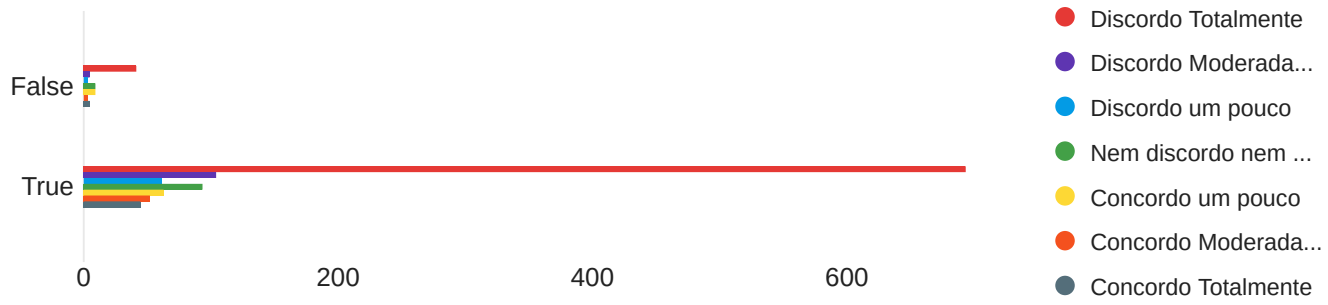
Se falássemos menos e trabalhássemos mais, todos estaríamos melhor.



Todos devemos ter fé absoluta em um poder sobrenatural (divino), cujas deci...



O policial é um guerreiro de Deus para impor a ordem e proteger as pessoas...



False

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
O que este país necessita, principalmente, antes de leis ou planos políticos, é de alguns líderes valentes, incansáveis e dedicados em quem o povo possa depositar a sua fé.	1	7	4	2	5	75
A maioria de nossos problemas sociais estaria resolvida se pudéssemos nos livrar das pessoas imorais, dos marginais e dos pervertidos.	1	7	3	2	4	75
A obediência e o respeito à autoridade são as principais virtudes que devemos ensinar as nossas crianças.	1	7	3	2	5	75
Se falássemos menos e trabalhássemos mais, todos estaríamos melhor.	1	7	4	2	5	75
Todos devemos ter fé absoluta em um poder sobrenatural (divino), cujas decisões devemos acatar.	1	7	3	2	4	75
O policial é um guerreiro de Deus para impor a ordem e proteger as pessoas de bem.	1	7	3	2	4	75

True

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
O que este país necessita, principalmente, antes de leis ou planos políticos, é de alguns líderes valentes, incansáveis e dedicados em quem o povo possa depositar a sua fé.	1	7	4	2	5	1111
A maioria de nossos problemas sociais estaria resolvida se pudéssemos nos livrar das pessoas imorais, dos marginais e dos pervertidos.	1	7	3	2	4	1111
A obediência e o respeito à autoridade são as principais virtudes que devemos ensinar as nossas crianças.	1	7	3	2	4	1111
Se falássemos menos e trabalhássemos mais, todos estaríamos melhor.	1	7	3	2	4	1111
Todos devemos ter fé absoluta em um poder sobrenatural (divino), cujas decisões devemos acatar.	1	7	2	2	4	1111
O policial é um guerreiro de Deus para impor a ordem e proteger as pessoas de bem.	1	7	2	2	3	1111

O que este país necessita, principalmente, antes de leis ou planos político...

Campo	False	True	Total
Discordo Totalmente	21	320	341
Discordo Moderadamente	4	140	144
Discordo um pouco	10	87	97
Nem discordo nem concordo	5	90	95
Concordo um pouco	15	185	200
Concordo Moderadamente	13	134	147
Concordo Totalmente	7	155	162

A maioria de nossos problemas sociais estaria resolvida se pudéssemos nos l...

Campo	False	True	Total
Discordo Totalmente	28	481	509
Discordo Moderadamente	14	195	209
Discordo um pouco	4	114	118
Nem discordo nem concordo	7	72	79
Concordo um pouco	14	136	150
Concordo Moderadamente	4	63	67
Concordo Totalmente	4	50	54

A obediência e o respeito à autoridade são as principais virtudes que devem...

Campo	False	True	Total
Discordo Totalmente	28	357	385
Discordo Moderadamente	5	218	223
Discordo um pouco	8	129	137
Nem discordo nem concordo	6	72	78
Concordo um pouco	14	183	197
Concordo Moderadamente	8	102	110
Concordo Totalmente	6	50	56

Se falássemos menos e trabalhássemos mais, todos estaríamos melhor.

Campo	False	True	Total
Discordo Totalmente	21	455	476

Discordo Moderadamente	9	159	168
Discordo um pouco	10	131	141
Nem discordo nem concordo	6	89	95
Concordo um pouco	11	115	126
Concordo Moderadamente	8	75	83
Concordo Totalmente	10	87	97

Todos devemos ter fé absoluta em um poder sobrenatural (divino), cujas deci...

Campo	False	True	Total
Discordo Totalmente	36	642	678
Discordo Moderadamente	5	99	104
Discordo um pouco	7	47	54
Nem discordo nem concordo	11	137	148
Concordo um pouco	4	51	55
Concordo Moderadamente	6	57	63
Concordo Totalmente	6	78	84

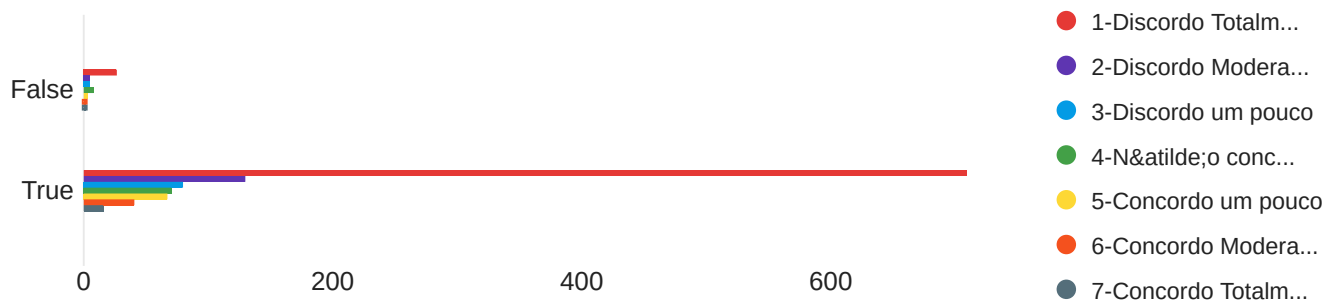
O policial é um guerreiro de Deus para impor a ordem e proteger as pessoas...

Campo	False	True	Total
Discordo Totalmente	41	693	734
Discordo Moderadamente	5	104	109
Discordo um pouco	3	61	64
Nem discordo nem concordo	9	93	102

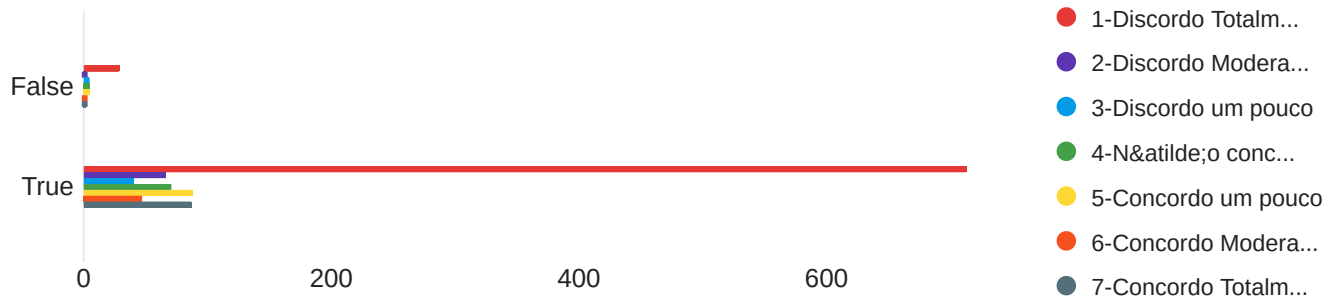
Concordo um pouco	9	63	72
Concordo Moderadamente	3	52	55
Concordo Totalmente	5	45	50

Q41 - Por favor marque o quanto você concorda ou discorda de cada ideia abaixo indicando um número de 1 a 7. Você pode responder rapidamente, a sua primeira impressão geralmente é a melhor.

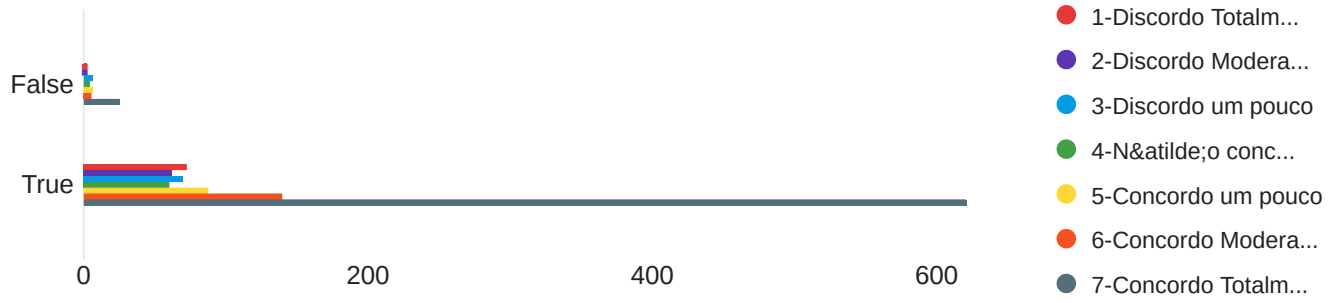
Uma sociedade ideal exige que alguns grupos estejam em posições superiores...



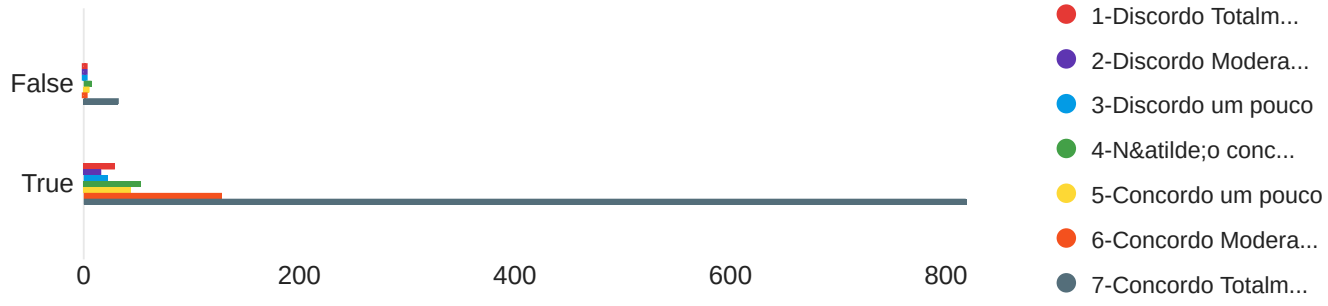
Alguns grupos de pessoas são simplesmente inferiores a outros na sociedade



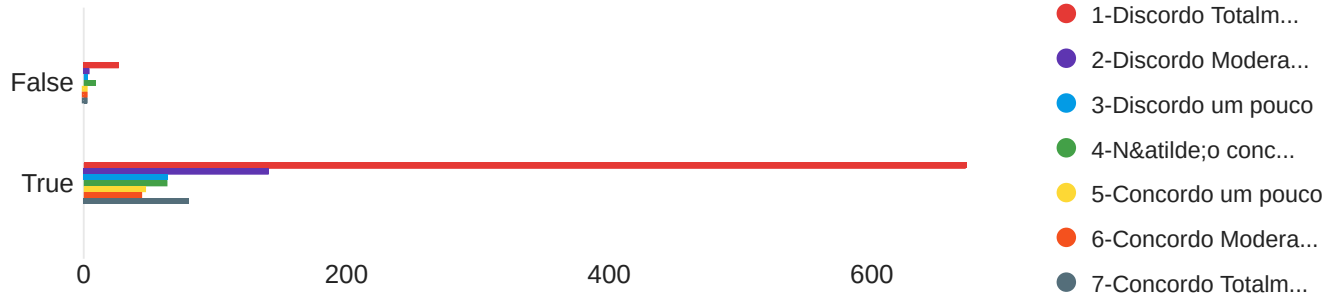
Nenhum grupo deveria ser dominante na sociedade



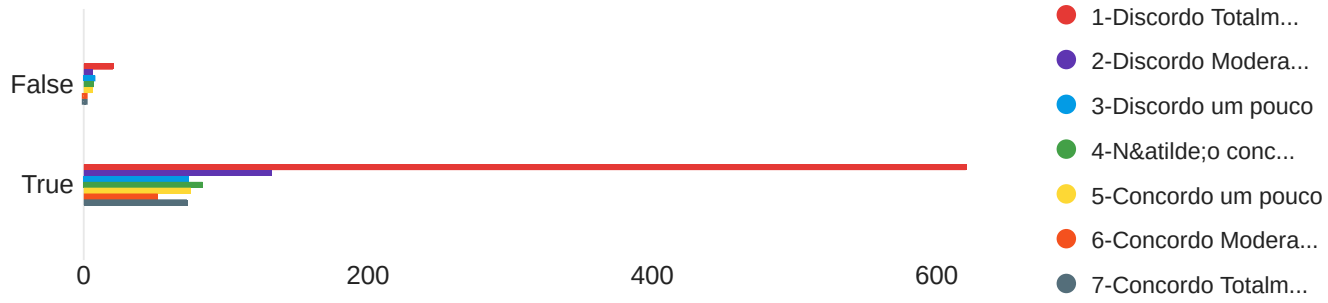
Grupos em posições inferiores na sociedade merecem tanto quanto grupos que...



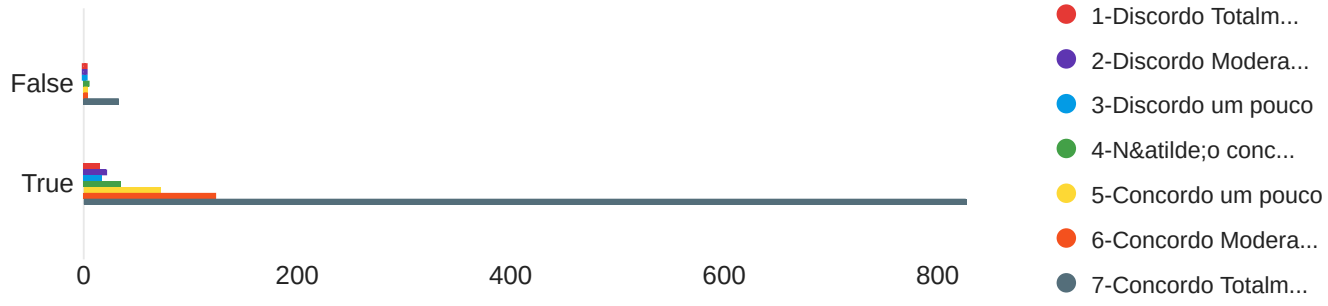
Nosso objetivo principal não deveria ser a igualdade entre os grupos



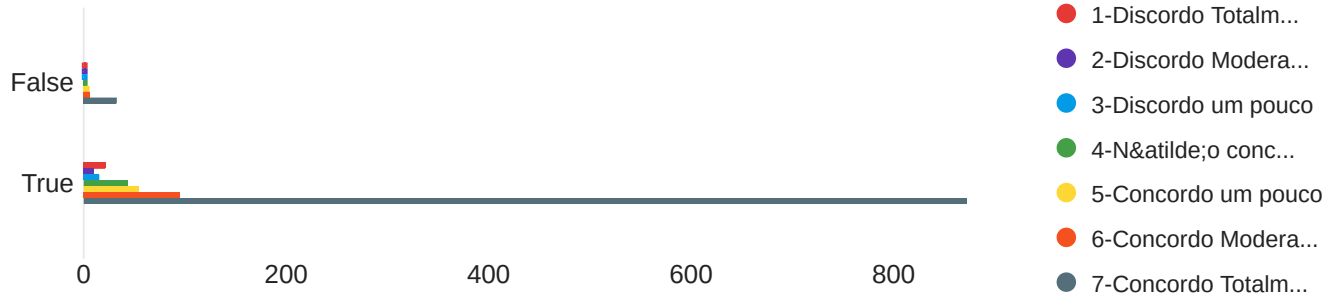
É injusto tentar fazer com que os grupos sejam iguais



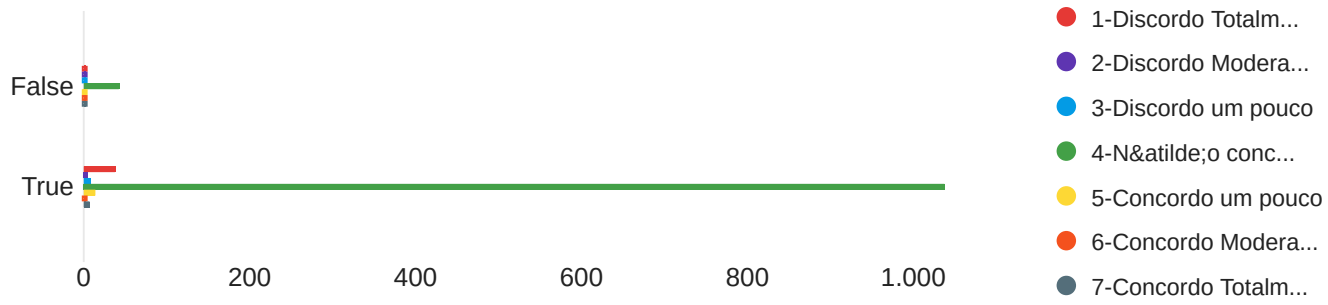
Deveríamos fazer o possível para assegurar condições iguais para os diferen...



Nós deveríamos trabalhar para oferecer a todos os grupos chances iguais de...



Se você está lendo aqui, por favor, marque "4-Não concordo nem discordo".



False

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Uma sociedade ideal exige que alguns grupos estejam em posições superiores e outros estejam em posições inferiores na sociedade	1	7	2	2	3	49
Alguns grupos de pessoas são simplesmente inferiores a outros na sociedade	1	7	2	2	4	49
Nenhum grupo deveria ser dominante na sociedade	1	7	6	2	3	49
Grupos em posições inferiores na sociedade merecem tanto quanto grupos que estão em posições superiores	1	7	6	2	3	49
Nosso objetivo principal não deveria ser a igualdade entre os grupos	1	7	2	2	3	49
É injusto tentar fazer com que os grupos sejam iguais	1	6	2	2	2	49
Deveríamos fazer o possível para assegurar condições iguais para os diferentes grupos	1	7	6	1	2	49
Nós deveríamos trabalhar para oferecer a todos os grupos chances iguais de serem bem-sucedidos	1	7	6	1	2	49
Se você está lendo aqui, por favor, marque "4-Não concordo nem discordo".	1	5	4	1	0	49

True

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Uma sociedade ideal exige que alguns grupos estejam em posições superiores e outros estejam em posições inferiores na sociedade	1	7	2	2	2	1111
Alguns grupos de pessoas são simplesmente inferiores a outros na sociedade	1	7	2	2	4	1111
Nenhum grupo deveria ser dominante na sociedade	1	7	6	2	4	1111
Grupos em posições inferiores na sociedade merecem tanto quanto grupos que estão em posições superiores	1	7	6	1	2	1111
Nosso objetivo principal não deveria ser a igualdade entre os grupos	1	7	2	2	4	1111
É injusto tentar fazer com que os grupos sejam iguais	1	7	2	2	4	1111
Deveríamos fazer o possível para assegurar condições iguais para os diferentes grupos	1	7	6	1	2	1111
Nós deveríamos trabalhar para oferecer a todos os grupos chances iguais de serem bem-sucedidos	1	7	6	1	2	1111
Se você está lendo aqui, por favor, marque "4-Não concordo nem discordo".	1	7	4	1	0	1111

Uma sociedade ideal exige que alguns grupos estejam em posições superiores...

Campo	False	True	Total
1-Discordo Totalmente	26	709	735
2-Discordo Moderadamente	5	129	134
3-Discordo um pouco	5	79	84

4-Não concordo nem discordo	8	71	79
5-Concordo um pouco	3	67	70
6-Concordo Moderadamente	1	40	41
7-Concordo Totalmente	1	16	17

Alguns grupos de pessoas são simplesmente inferiores a outros na sociedade

Campo	False	True	Total
1-Discordo Totalmente	29	713	742
2-Discordo Moderadamente	2	66	68
3-Discordo um pouco	4	40	44
4-Não concordo nem discordo	5	70	75
5-Concordo um pouco	5	88	93
6-Concordo Moderadamente	2	47	49
7-Concordo Totalmente	2	87	89

Nenhum grupo deveria ser dominante na sociedade

Campo	False	True	Total
1-Discordo Totalmente	2	72	74
2-Discordo Moderadamente	1	62	63
3-Discordo um pouco	6	70	76
4-Não concordo nem discordo	4	60	64
5-Concordo um pouco	6	87	93
6-Concordo Moderadamente	5	139	144

7-Concordo Totalmente	25	621	646

Grupos em posições inferiores na sociedade merecem tanto quanto grupos que...

Campo	False	True	Total
1-Discordo Totalmente	1	29	30
2-Discordo Moderadamente	1	16	17
3-Discordo um pouco	2	22	24
4-Não concordo nem discordo	7	53	60
5-Concordo um pouco	5	44	49
6-Concordo Moderadamente	1	128	129
7-Concordo Totalmente	32	819	851

Nosso objetivo principal não deveria ser a igualdade entre os grupos

Campo	False	True	Total
1-Discordo Totalmente	27	672	699
2-Discordo Moderadamente	4	141	145
3-Discordo um pouco	3	64	67
4-Não concordo nem discordo	9	63	72
5-Concordo um pouco	2	47	49
6-Concordo Moderadamente	2	44	46
7-Concordo Totalmente	2	80	82

É injusto tentar fazer com que os grupos sejam iguais

Campo	False	True	Total
1-Discordo Totalmente	21	621	642
2-Discordo Moderadamente	6	132	138
3-Discordo um pouco	8	74	82
4-Não concordo nem discordo	7	84	91
5-Concordo um pouco	6	75	81
6-Concordo Moderadamente	1	52	53
7-Concordo Totalmente	0	73	73

Deveríamos fazer o possível para assegurar condições iguais para os diferen...

Campo	False	True	Total
1-Discordo Totalmente	1	15	16
2-Discordo Moderadamente	0	21	21
3-Discordo um pouco	2	17	19
4-Não concordo nem discordo	5	35	40
5-Concordo um pouco	4	72	76
6-Concordo Moderadamente	4	124	128
7-Concordo Totalmente	33	827	860

Nós deveríamos trabalhar para oferecer a todos os grupos chances iguais de...

Campo	False	True	Total
1-Discordo Totalmente	1	22	23
2-Discordo Moderadamente	1	10	11

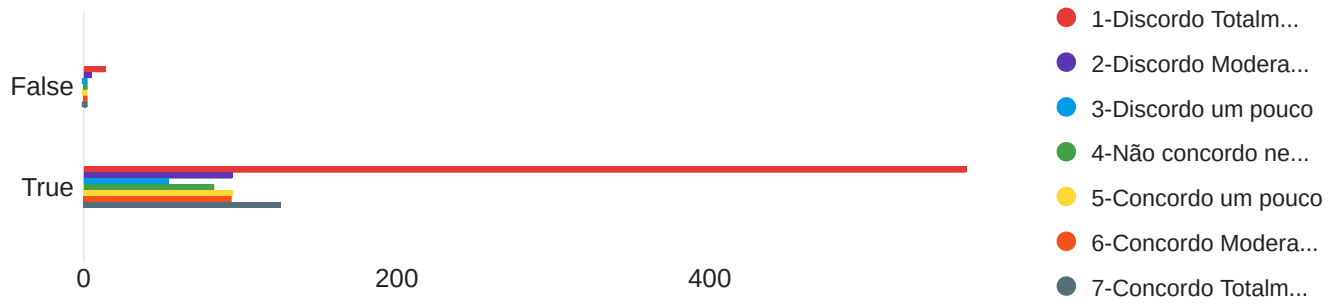
3-Discordo um pouco	1	15	16
4-Não concordo nem discordo	3	43	46
5-Concordo um pouco	5	54	59
6-Concordo Moderadamente	6	95	101
7-Concordo Totalmente	32	872	904

Se você está lendo aqui, por favor, marque "4-Não concordo nem discordo".

Campo	False	True	Total
1-Discordo Totalmente	1	38	39
2-Discordo Moderadamente	2	5	7
3-Discordo um pouco	0	8	8
4-Não concordo nem discordo	43	1.038	1081
5-Concordo um pouco	3	13	16
6-Concordo Moderadamente	0	2	2
7-Concordo Totalmente	0	7	7

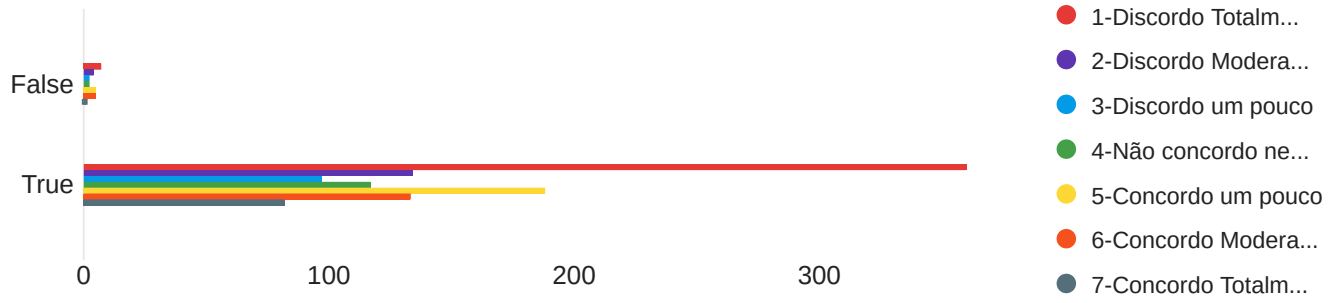
Q38 - Por favor, leia atentamente cada um dos itens abaixo a respeito dos indivíduos pobres e em seguida, indique seu grau de concordância com cada um deles. Por gentileza responda todos, utilizando a escala abaixo, selecione o que melhor representa sua respostas.

Os pobres pertencem a uma classe social menos dotada e isso explica porque...

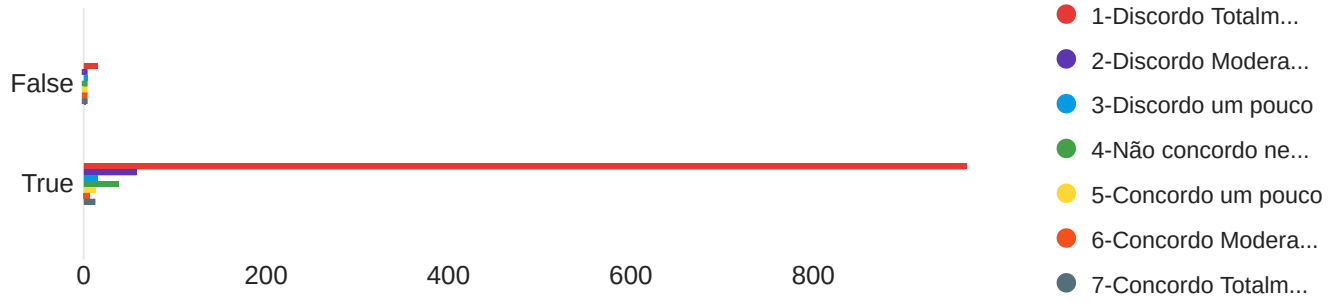


Os pobres transmitem aos filhos valores e aptidões diferentes dos necessári...

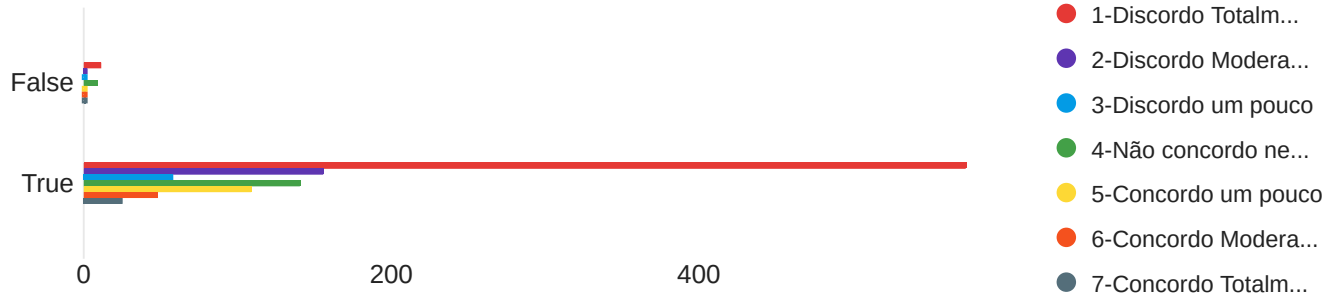
Comparados com os ricos, os pobres são muito diferentes nos valores que ens...



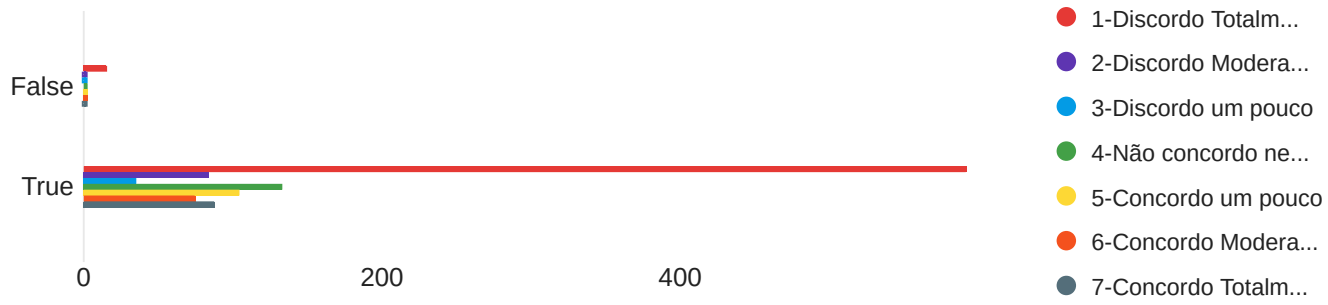
Me incomodaria se um familiar próximo casasse com uma pessoa pobre.



Comparado com os ricos, os pobres são muito diferentes nos valores e compor...



No que diz respeito à honestidade, os ricos e os pobres são muito diferente...



False

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
Os pobres pertencem a uma classe social menos dotada e isso explica porque não estão numa situação tão boa quanto os ricos.	1	7	2	2	3	25
Os pobres transmitem aos filhos valores e aptidões diferentes dos necessários para ser bem-sucedido na sociedade brasileira.	1	5	2	2	2	25
Comparados com os ricos, os pobres são muito diferentes nos valores que ensinam aos filhos.	1	6	3	2	4	25
Me incomodaria se um familiar próximo casasse com uma pessoa pobre.	1	5	2	1	2	25
Comparado com os ricos, os pobres são muito diferentes nos valores e comportamentos sexuais.	1	6	3	2	3	25
No que diz respeito à honestidade, os ricos e os pobres são muito diferentes.	1	7	3	2	5	25

True

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas

Os pobres pertencem a uma classe social menos dotada e isso explica porque não estão numa situação tão boa quanto os ricos.	1	7	3	2	5	1111
Os pobres transmitem aos filhos valores e aptidões diferentes dos necessários para ser bem-sucedido na sociedade brasileira.	1	7	2	2	4	1111
Comparados com os ricos, os pobres são muito diferentes nos valores que ensinam aos filhos.	1	7	3	2	4	1111
Me incomodaria se um familiar próximo casasse com uma pessoa pobre.	1	7	1	1	1	1111
Comparado com os ricos, os pobres são muito diferentes nos valores e comportamentos sexuais.	1	7	2	2	3	1111
No que diz respeito à honestidade, os ricos e os pobres são muito diferentes.	1	7	3	2	4	1111

Os pobres pertencem a uma classe social menos dotada e isso explica porque...

Campo	False	True	Total
1-Discordo Totalmente	14	564	578
2-Discordo Moderadamente	5	95	100
3-Discordo um pouco	0	54	54
4-Não concordo nem discordo	2	83	85
5-Concordo um pouco	1	95	96
6-Concordo Moderadamente	2	94	96
7-Concordo Totalmente	1	126	127

Os pobres transmitem aos filhos valores e aptidões diferentes dos necessári...

Campo	False	True	Total

1-Discordo Totalmente	14	589	603
2-Discordo Moderadamente	2	116	118
3-Discordo um pouco	1	81	82
4-Não concordo nem discordo	5	104	109
5-Concordo um pouco	3	105	108
6-Concordo Moderadamente	0	74	74
7-Concordo Totalmente	0	42	42

Comparados com os ricos, os pobres são muito diferentes nos valores que ens...

Campo	False	True	Total
1-Discordo Totalmente	7	360	367
2-Discordo Moderadamente	4	134	138
3-Discordo um pouco	2	97	99
4-Não concordo nem discordo	2	117	119
5-Concordo um pouco	5	188	193
6-Concordo Moderadamente	5	133	138
7-Concordo Totalmente	0	82	82

Me incomodaria se um familiar próximo casasse com uma pessoa pobre.

Campo	False	True	Total
1-Discordo Totalmente	15	968	983
2-Discordo Moderadamente	2	58	60
3-Discordo um pouco	4	15	19
4-Não concordo nem discordo	3	38	41

5-Concordo um pouco	1	13	14
6-Concordo Moderadamente	0	7	7
7-Concordo Totalmente	0	12	12

Comparado com os ricos, os pobres são muito diferentes nos valores e compor...

Campo	False	True	Total
1-Discordo Totalmente	11	574	585
2-Discordo Moderadamente	2	156	158
3-Discordo um pouco	1	58	59
4-Não concordo nem discordo	9	141	150
5-Concordo um pouco	1	109	110
6-Concordo Moderadamente	1	48	49
7-Concordo Totalmente	0	25	25

No que diz respeito à honestidade, os ricos e os pobres são muito diferente...

Campo	False	True	Total
1-Discordo Totalmente	15	592	607
2-Discordo Moderadamente	1	84	85
3-Discordo um pouco	0	35	35
4-Não concordo nem discordo	2	133	135
5-Concordo um pouco	3	104	107
6-Concordo Moderadamente	3	75	78
7-Concordo Totalmente	1	88	89

Q84 - Utilize a escala de + 4 a – 4 para indicar, em relação às qualificações descritas abaixo, o quão próximo você considera a atuação do judiciário (a atuação de juízes e de tribunais para a garantia de direitos e da justiça na sociedade): Negativa/Positiva

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	-3	4	0	2	4	19
True	-4	4	0	2	5	1111

Q85 - Injusta/Justa

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	-4	3	-1	2	4	19
True	-4	4	0	2	5	1111

Q86 - Ruim/Boa

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	-3	3	0	2	3	19
True	-4	4	0	2	4	1111

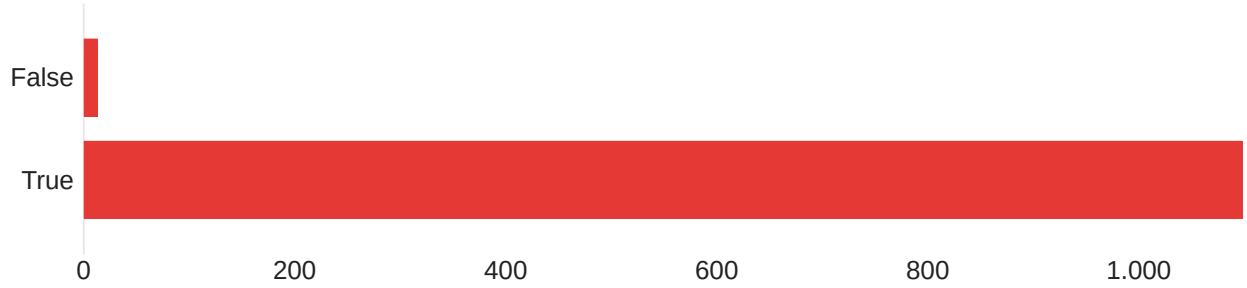
Q87 - Indesejável/Desejável

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	-3	4	0	2	4	19
True	-4	4	0	2	5	1111

Q89 - Duvidosa/Confiável

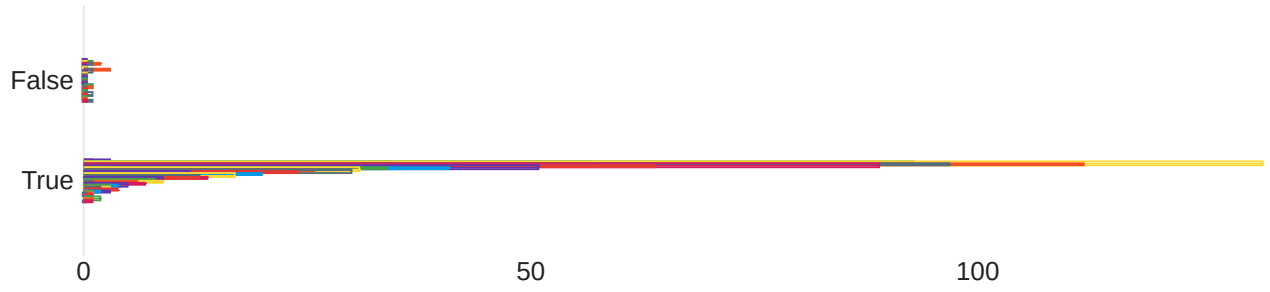
Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	-4	2	-1	2	5	19
True	-4	4	-1	2	5	1111

Q59 - Qual sua idade?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	4	4	4	0	0	13
True	4	4	4	0	0	1098

Campo	False	True
Idade em anos	13	1.098
Total	13	1098



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	19	99999	7735	26634	709390462	13

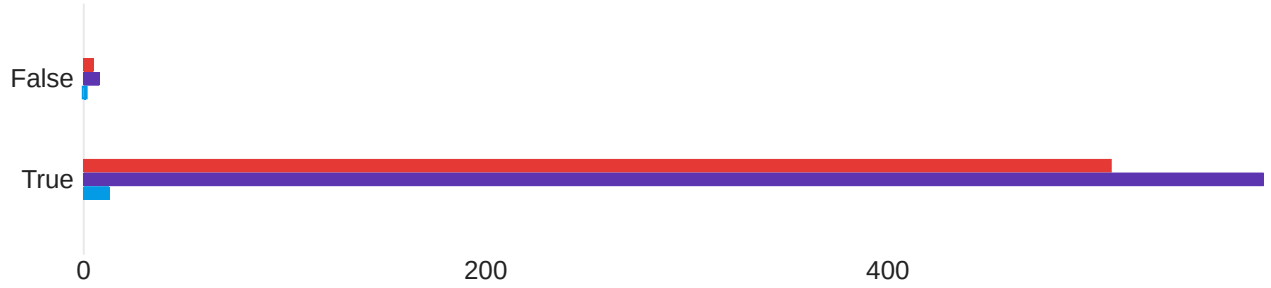
True	16	13081976	11955	394614	155720587650	1098

Campo	False	True
16	0	1
17	0	3
18	0	57
19	1	93
20	0	132
21	2	112
22	1	97
23	0	89
24	0	64
25	0	51
26	0	41
27	0	34
28	0	31
29	3	26
30	1	30
31	0	19
32	0	24
33	0	12
34	0	20
35	0	5
36	0	17
37	0	5

38	0	13
39	0	14
40	0	13
41	0	9
42	0	7
43	0	8
44	0	9
45	0	6
46	0	1
47	0	7
48	0	5
49	0	5
50	1	4
51	1	3
52	0	3
53	1	2
54	0	2
55	0	3
56	0	4
57	0	3
58	0	2
61	0	1
67	0	1
75	0	1
201	1	0

1976	0	1
1994	0	1
1995	0	1
1997	0	1
1998	0	2
2000	0	1
2001	0	1
99999	1	0
13081976	0	1
Total	13	1098

Q75 - Como você se reconhece?

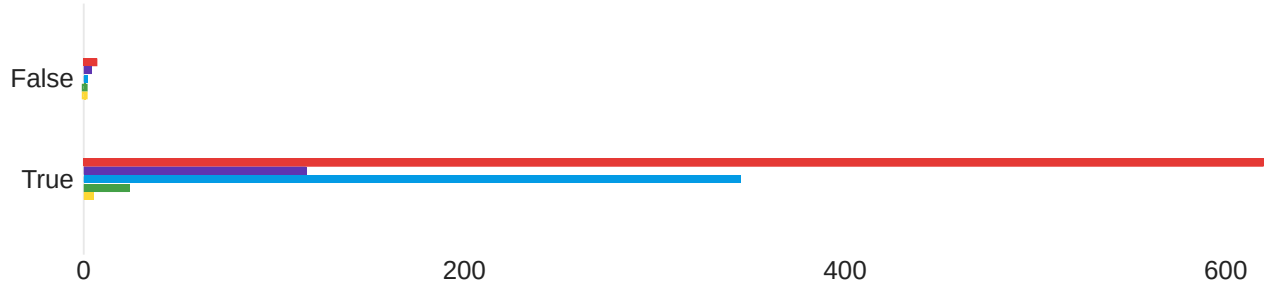


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	3	2	1	1	13
True	1	6	2	1	1	1111

Campo	False	True
Mulher	5	511
Homem	8	587
Outro, por favor especificar:	0	13
Total	13	1111

Outro, por favor especificar: – Texto

Q44 - Você se identifica como?

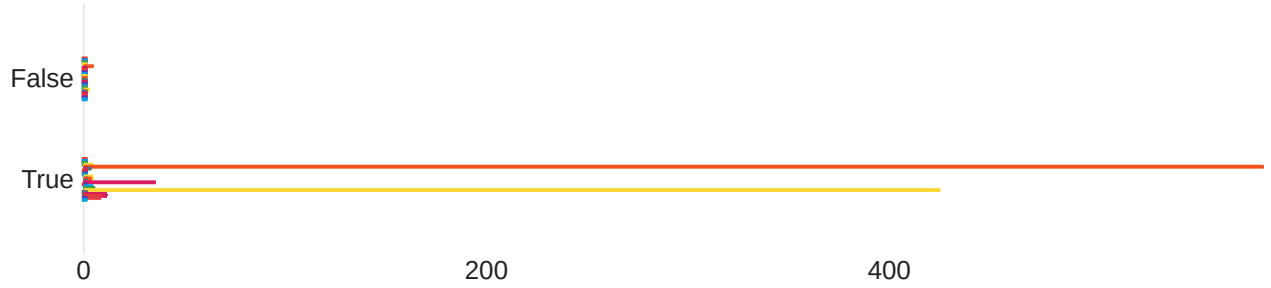


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	3	2	1	1	13
True	1	5	2	1	1	1111

Campo	False	True
Branco	7	620
Preto	4	117
Pardo	2	345
Amarelo	0	24
Outro	0	5
Total	13	1111

Outro – Texto

Q40 - Em qual Estado você mora?

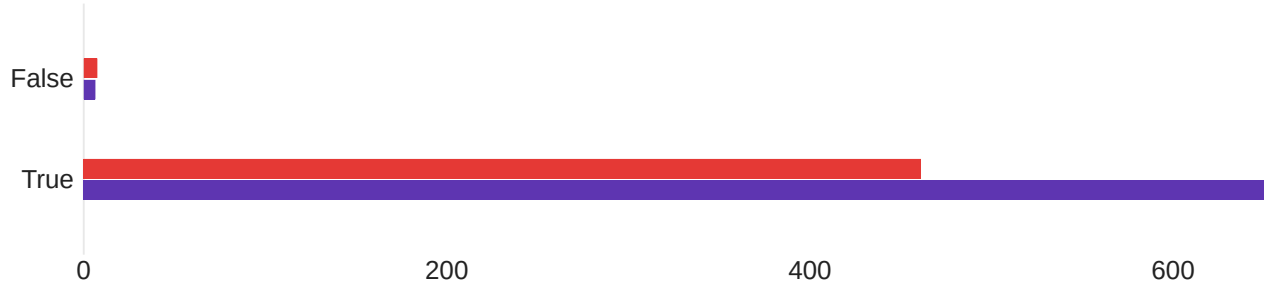


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	6	25	15	7	45	13
True	9	30	17	7	54	1111

Campo	False	True
Acre	0	0
Alagoas	1	0
Amapá	0	0
Amazonas	0	0
Bahia	1	5
Ceará	5	586
Distrito Federal	0	4
Espírito Santo	0	2
Goiás	0	1
Maranhão	0	2
Mato Grosso	0	1
Mato Grosso do Sul	0	2

Minas Gerais	1	5
Pará	1	4
Paraíba	0	2
Paraná	1	36
Pernambuco	0	1
Piauí	0	2
Rio de Janeiro	0	5
Rio Grande do Norte	0	6
Rio Grande do Sul	3	425
Rondônia	0	0
Roraima	0	0
Santa Catarina	0	12
São Paulo	0	9
Sergipe	0	1
Tocantins	0	0
Total	13	1111

Q42 - Você mora na capital do estado?



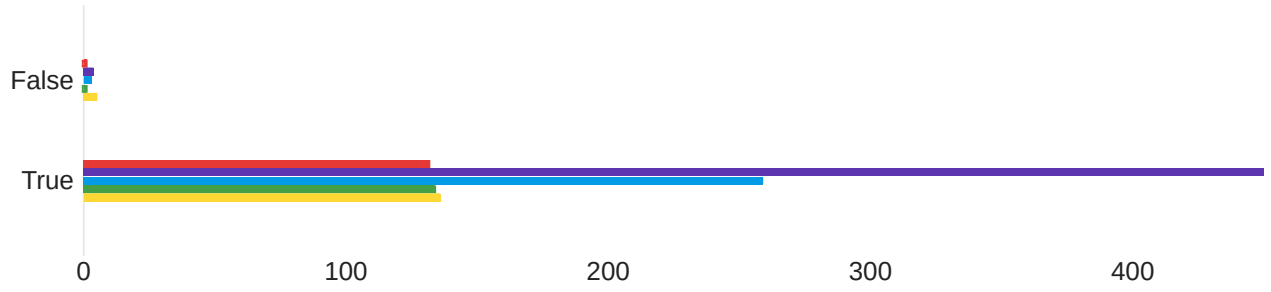
Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	1	0	0	13
True	1	2	2	0	0	1111

Campo	False	True
Sim	7	461
Não	6	650
Total	13	1111

Q86 - Em comparação às pessoas do seu país, você se considera da:

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	3	8	5	2	3	13
True	1	9	4	2	3	1111

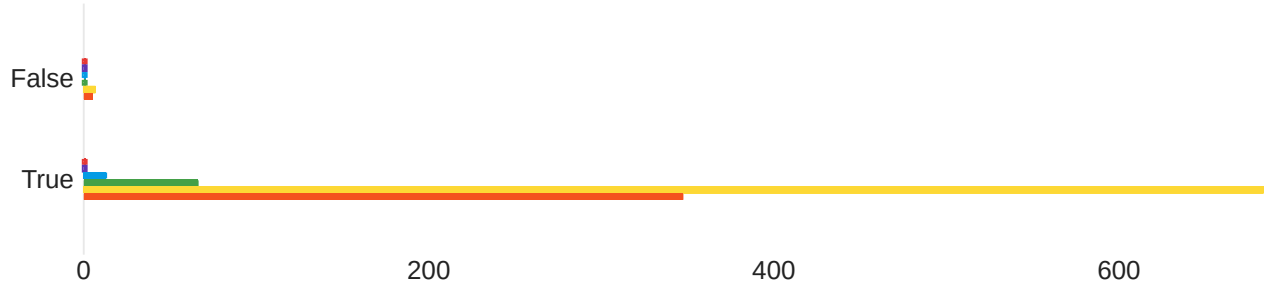
Q87 - Atualmente, qual é a média de renda mensal familiar da sua casa?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	7	5	2	3	13
True	1	7	4	2	3	1111

Campo	False	True
Menos de 1 salário mínimo (até R$1.045,00).	1	132
Entre 1 e 3 salários mínimos (de R$1.045,01 a R$3.135,00).	4	450
Entre 3 e 6 salários mínimos (de R$3.135,01 a R$6.270,00).	3	259
Entre 6 e 10 salários mínimos (de R$6.270,01 a R$10.450,00).	0	134
Mais do que 10 salários mínimos (acima de R10.450,00).	5	136
Total	13	1111

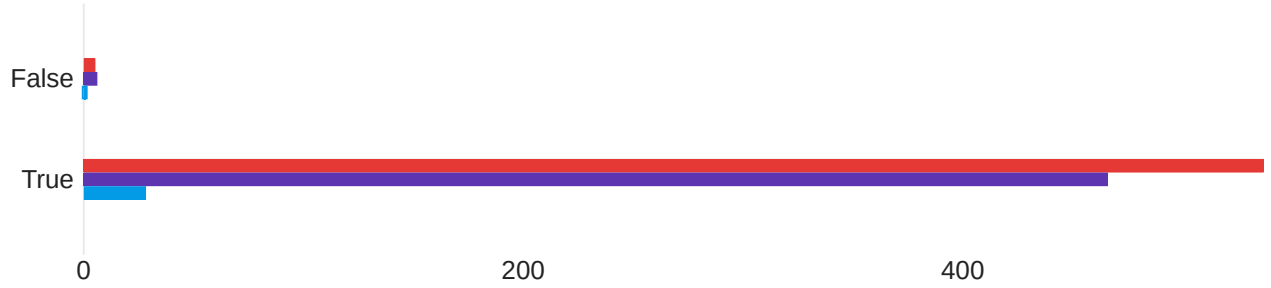
Q88 - Qual a sua escolaridade?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	4	8	7	1	1	13
True	4	8	7	1	0	1111

Campo	False	True
Clique Analfabeto / Fundamental I incompleto.	0	0
Fundamental I completo / Fundamental II Incompleto.	1	1
Fundamental completo/Médio Incompleto.	0	13
Médio completo.	0	66
Ensino Superior (universitário) incompleto.	7	684
Ensino Superior (universitário) completo.	5	347
Total	13	1111

Q95 - A faculdade/universidade que você estuda/estudou é/foi:

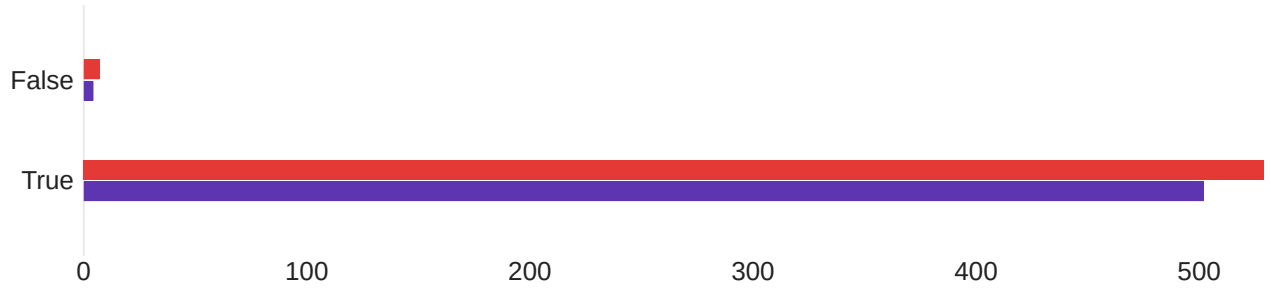


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	4	3	1	2	11
True	1	5	2	2	2	1031

Campo	False	True
Privada	5	537
Pública	6	466
Outros:	0	28
Total	11	1031

Outros: – Texto

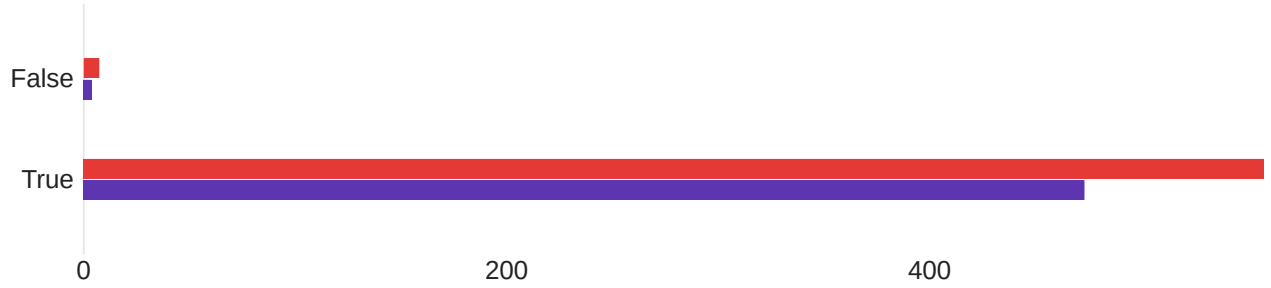
Q96 - A faculdade/universidade que você estuda/estudou é localizada em:



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	4	2	1	2	11
True	1	4	2	1	2	1031

Campo	False	True
Capital	7	529
Interior	4	502
Total	11	1031

Q89 - Você é estudante ou bacharel em Direito?



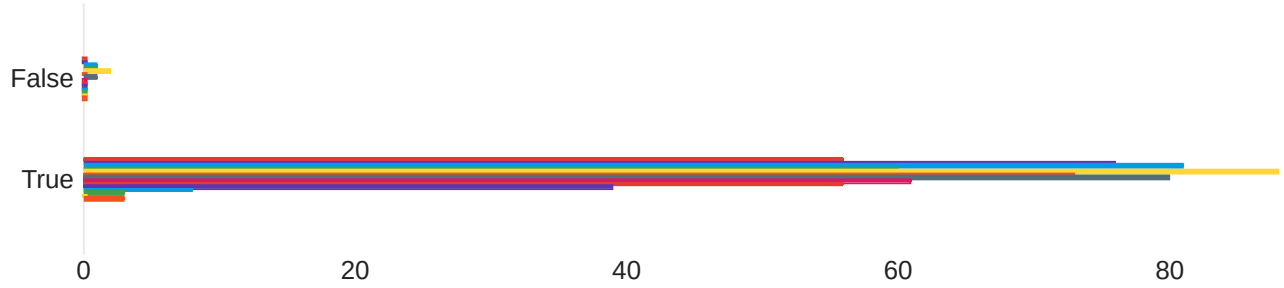
Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	3	2	1	1	11
True	1	3	2	1	1	1031

Campo	False	True
Sim.	7	558
Não.	4	473
Total	11	1031

Q72 - Qual a sua área de estudo ou formação no ensino superior?

Qual a sua área de estudo ou formação no ensino superior?

Q70 - Em qual semestre da graduação você está?

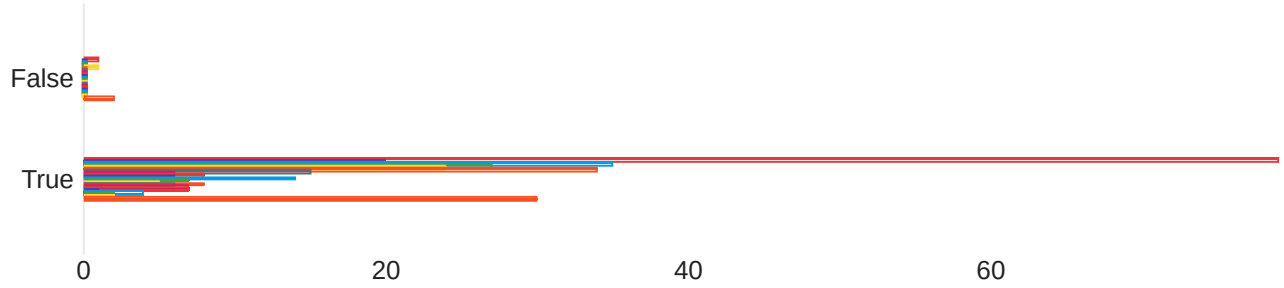


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	3	7	5	1	2	5
True	1	14	5	3	8	684

Campo	False	True
1	0	56
2	0	76
3	1	81
4	1	60
5	2	88
6	0	73
7	1	80
8	0	61
9	0	56
10	0	39
11	0	8
12	0	3

13	0	0
14	0	3
Total	5	684

Q71 - Quantos anos fazem que você concluiu a graduação?

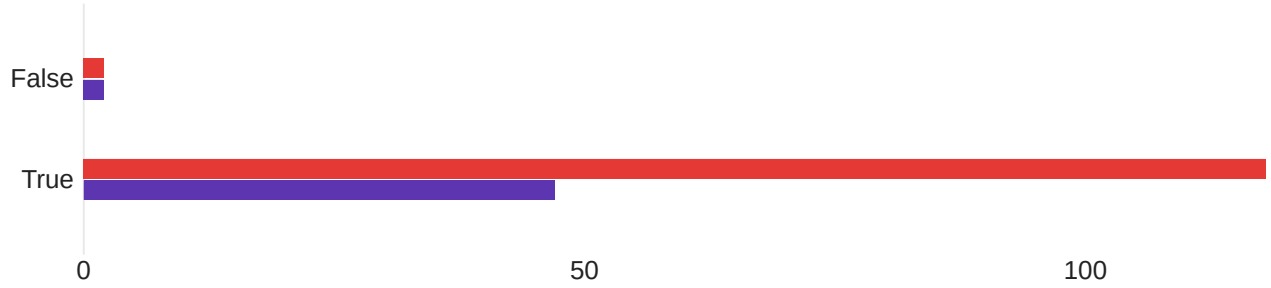


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	22	13	10	92	4
True	1	22	7	7	44	347

Campo	False	True
Menos que 1.	1	79
1	0	20
2	0	35
3	0	27
4	1	24
5	0	34
6	0	15
7	0	6
8	0	8
9	0	6
10	0	14
11	0	7

12	0	5
13	0	8
14	0	6
15	0	7
16	0	7
17	0	1
18	0	4
19	0	2
20	0	2
20 ou mais.	2	30
Total	4	347

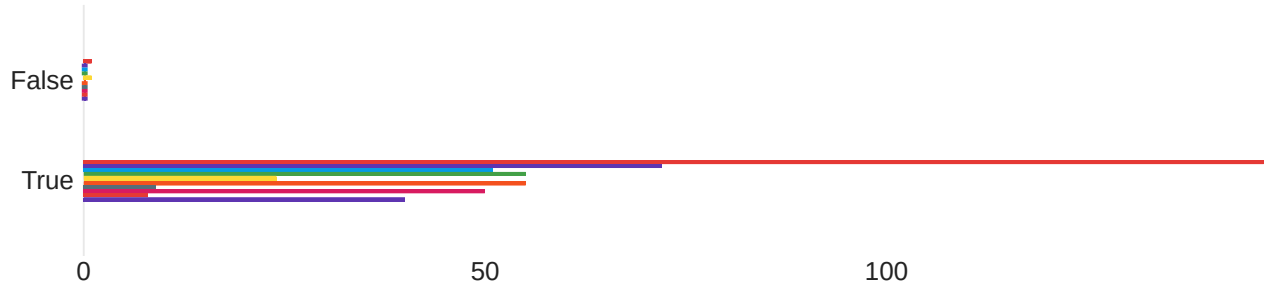
Q74 - Você atua na área do Direito?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	2	2	1	0	4
True	1	2	1	0	0	165

Campo	False	True
Sim.	2	118
Não.	2	47
Total	4	165

Q91 - Em que carreira você atua ou pretende, principalmente, atuar como jurista?

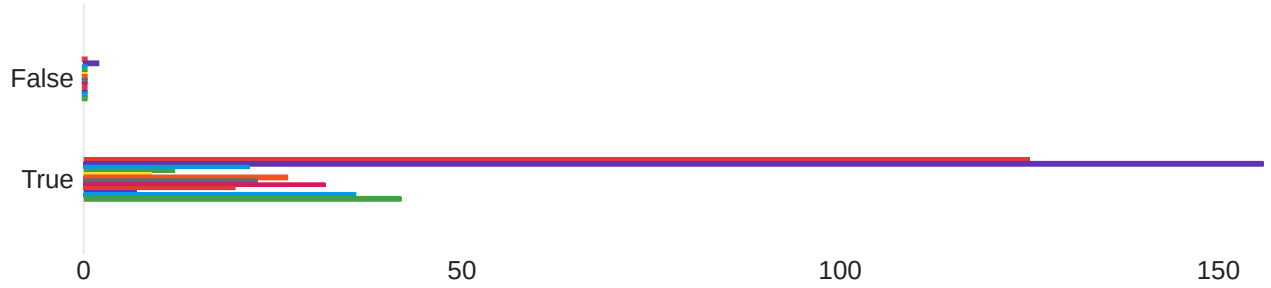


Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	7	4	3	9	2
True	1	12	5	4	13	511

Campo	False	True
Advogado	1	147
Juiz	0	72
Promotor	0	51
Defensor	0	55
Procurador	1	24
Delegado	0	55
Consultor jurídico	0	9
Professor	0	50
Em escritório de advocacia, mas não advogando	0	8
Outros:	0	40
Total	2	511

Outros: – Texto

Q92 - Em que área você, predominantemente, atua ou pretende atuar?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	4	4	4	0	0	2
True	1	14	6	4	19	511

Campo	False	True
Direito Civil.	0	125
Direito Penal.	2	156
Direito comercial ou empresarial.	0	22
Direito do consumidor.	0	12
Direito da tecnologia da informação.	0	9
Direito tributário.	0	27
Direito administrativo.	0	23
Direito trabalhista.	0	32
Direito previdenciário.	0	20
Direito Ambiental.	0	7
Direitos Humanos.	0	36
Outros:	0	42

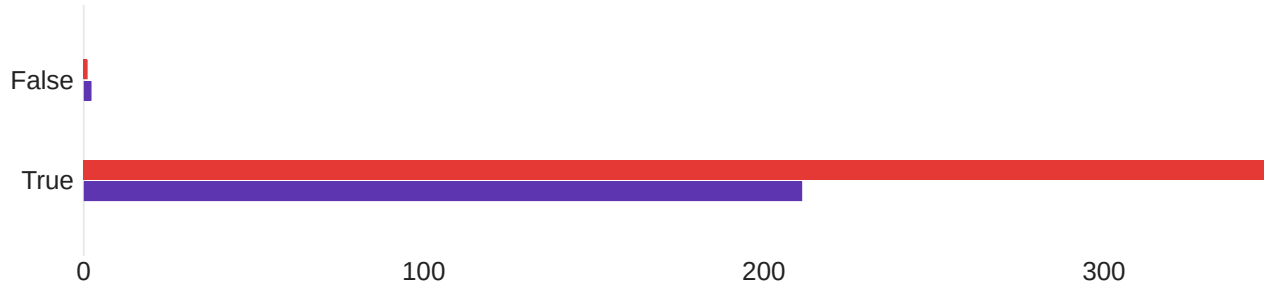
Total	2	511

Outros: – Texto

Q78 - Considerando sua trajetória no direito, o quão próximo você se sente do Direito Penal?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	3	10	8	3	11	3
True	1	10	6	3	7	558

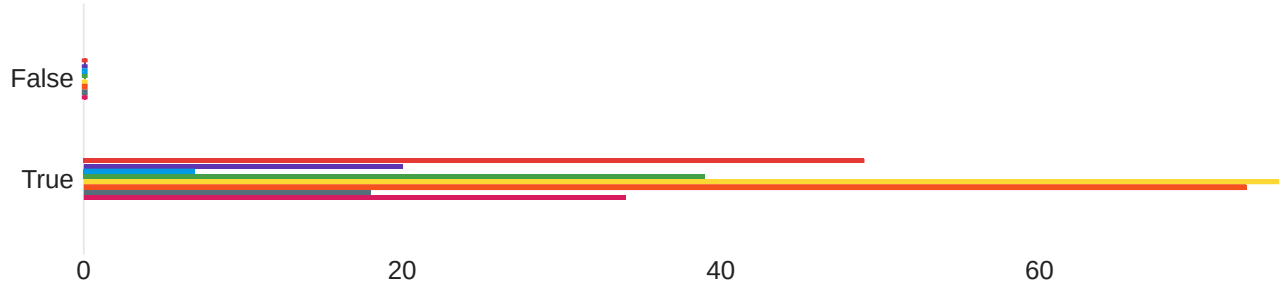
Q79 - Alguém de seu círculo familiar mais próximo exerce ou exerceu algum cargo jurídico?



Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	1	4	3	1	2	3
True	1	4	2	1	2	558

Campo	False	True
Não.	1	347
Sim.	2	211
Total	3	558

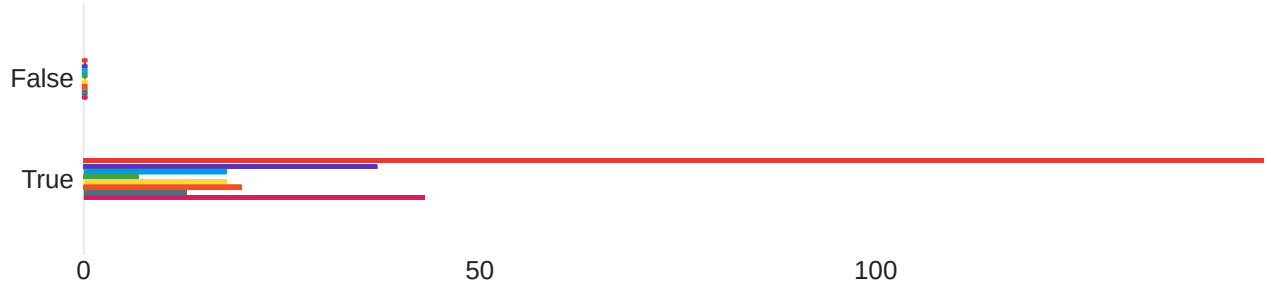
Q80 - Qual(is) familiar(es)?



Campo	False	True
Pai.	0	49
Mãe.	0	20
Padrasto/madrasta.	0	7
Irmão/irmã.	0	39
Tio/tia.	0	75
Primo/prima.	0	73
Avô/avó.	0	18
Outros:	0	34
Total	0	315

Selecionar um campo

Q81 - Qual(is) cargo(s)?



Campo	False	True
Advogado.	0	149
Juiz.	0	37
Promotor.	0	18
Defensor.	0	7
Consultor jurídico.	0	18
Professor.	0	20
Em escritório de advocacia, mas não advogando.	0	13
Outros:	0	43
Total	0	305

Selecionar um campo

Selecionar um campo

Q82 - Em que medida você se considera politicamente alinhado:

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	0	0	0	0	0	0
True	1	9	4	2	4	1111

Q92 - Você acha que o preconceito racial está mais ligado à cor da pele ou à classe social dos negros?

Campo	Mínimo	Máximo	Meio	Desvio padrão	Variância	Respostas
False	0	0	0	0	0	0
True	1	5	3	1	1	1111

Q93 - Você tem algum comentário, crítica ou sugestão?

Você tem algum comentário, crítica ou sugestão?

Q94 - Se desejar contribuir com etapas futuras dessa pesquisa deixe aqui seu email:

Se desejar contribuir com etapas futuras dessa pesquisa deixe aqui seu email:

pedrovr9544@gmail.com

Publication Agreement for Papers Accepted for Publication in ADVANCE '2023 Proceedings

Please fill in this form and submit it in a pdf format (with your final version of the paper) using your easychair account:

Paper No.: 1970

TITLE OF PAPER: The use of Machine Learning for predicting the sentence given by the population at a crime scene.

AUTHOR(S): Keyvilânia Vieira V. de Freitas

PUBLICATION TITLE: Proceedings of ADVANCE Conference

By submitting your work for publication in the Proceedings of **Sixth International Workshop on ADVANCEs in ICT (ADVANCE)**, you have agreed, **by default**, to the following terms and conditions:

(a) The author and coauthor(s) must complete this form and submit it via easychair **along with the camera-ready**. If the paper is a work made for hire and the author or coauthor(s) do not have the authority to sign the paper, the owner of the paper must also sign the form.

(b) You grant the publisher of ADVANCE Proceedings a right to publish your work and to distribute the published work according to the relevant laws and regulations governing publications.

(c) You must be either the owner of the work or have obtained permission from the owner of the work (e.g. your employer) to publish the work. You have agreed to take the responsibility to solve any disputes that may arise anytime in the future with regard to the ownership of the work and the right of publishing the work, including penalty and compensation payment.

(d) If excerpts from copyrighted works are included in your paper, you must have obtained written permission from the copyright owners and show credit to the sources in your paper. You have agreed to take the responsibility to solve any disputes that may arise anytime in the future with regard to copyright materials that you have used in your work.

If any of the above terms and conditions does not apply to your work or you do not agree with any of these provisions, please inform the **ADVANCE Workshop** Organizing Committee immediately at the following address:

E-mail: advance2020@easychair.org

Corresponding Author's

Signature: Keyvilânia Vieira Vemancio de Freitas **Title/Employer:** _____
(if not signed by author)

Print Name: Keyvilânia

As the form is being signed by the corresponding author on behalf of the others, the following additional statement must be accepted and signed by the author signing for the co-authors: "I represent and warrant that I am authorized to execute this transfer of copyright on behalf of all the authors of the article named above."

Date: 20/01/2023 Corresponding Author's Signature: Keyvilânia Vieira Vemancio de Freitas

Technical Session 3 (Full Papers)

Impacts of GPS module on energy consumption and machine-learning based battery lifetime estimation

André Teixeira de Aquino[1]*, José Ailton Leão Barboza Júnior[1]†, Nícolas de Araújo Moreira[1]‡, and Paulo Peixoto Praça[1]§

Federal University of Ceara, Fortaleza, Brazil
andret.aquino@gmail.com, ailton.junior@ufersa.edu.br, nicolas.araujom@gmail.com, paulopp@dee.ufc.br

Abstract

The maintenance of a Wireless Sensor Network may represent a logistic challenge. Battery replacement on applications involving huge numbers of sensors spread over a wide and distant area may be expensive and difficult. So, good planning of the maintenance schedule is necessary. This work discusses the impact of the GPS module for bovine tracking on farms on current consumption and its estimation. It compares Long-Short Term Memory (LSTM) networks and Decision Trees with AdaBoost to estimate this consumption. The results show that the activation of the GPS module increases 114.36% the current consumption and Decision Trees with AdaBoost using 300 estimators and with a depth equal to 20 outperforms LTSM with Root Mean Square (RMS) error of 0.00015.
Keywords: Machine Learning, Internet of Things, Wireless Sensor Networks

1 Introduction

The Internet of Things (IoT) is the connection between different types of sensors, devices, objects, machines, vehicles, and buildings using Wireless Sensor Networks (WSNs). IoT is a key concept in the evolution of communication, industry, and agriculture, for example, [19, 21, 4].

The nodes of this network are self-organized and decentralized intelligent sensors able to communicate with each other through radio links [10, 1]. WSN are characterized by complexity and heterogeneity: it integrates different analog and digital blocks, such as Analog-Digital Converters (ADCs), Digital Signal Processors (DSPs), and Radio-Frequency Circuits [10]. Usually, the communication of WSNs is multi-path asymmetric, presenting a many-to-one data flow, i.e., the nodes send data to a monitoring station [10].

Wireless Sensor Networks (WSNs) have been used widely in many scientific, industrial, and commercial applications, such as ecological control (forest fire detection, pollution control), intelligent houses, etc., acquiring environmental parameters (e.g. temperature, illumination, etc.) [10]. WSN must be low cost and have strong energy constraints: depending on the scenario, it is quite difficult to recharge or replace the battery, as for example, applications with a huge number of nodes spread over a large area (forest or farm monitoring, for instance) or applications in distant and dangerous locations (such as volcano monitoring) [10].

The uneven use of livestock pastures has a significant impact on productivity, the ecosystem, biodiversity, and function. The use of collars with Global Positioning System (GPS) allows for obtaining a large amount of data about the position and activity of the animals, giving

*Responsible for experimental results
†Responsible for experimental results
‡Co-Advisor, and responsible for theoretical results
§Advisor

new possibilities to researchers [12]. Similarly to the previously described scenarios, one of the limitations related to using these devices is their autonomy, more precisely, the battery lifetime. Battery replacement or recharging impacts the logistics of animal handling. So, it is important to have an estimate of battery lifetime to have a more adequate maintenance schedule for these networks. A second relevant point is related to the way the data is accessed. The current work is inserted in the context of a project of a collar for a location using satellite, data acquisition, and transmission using LoRa for remote data access for real-time surveying of pastures.

This paper has discussed the impact of the GPS module, to track bovines on farms, activation on current consumption, and two machine-learning-based estimation techniques are used to estimate the current consumed and then compared. The remainder of this paper is as follows: the next section (2) gives an overview of existing works related to this research. The following section (3) presents the two machine-learning techniques used: Long-Short Term Memory (LSTM) Networks and Decision Trees with AdaBoost. Section 4 presents the used hardware. Then, section 5 presents and discusses the results of measurements and estimation.

2 Literature Review and State of the Art

As discussed before, IoT has applications in a wide variety of areas. A considerable number of researches discusses also how to increase the battery lifetime and study factors that affect it, such as the impact of interference on energy consumption. Applications of IoT in precision irrigation for agriculture are explored in [8, 4]. The author in [15] explores the use of IoT in wildlife tracking.

The author in [16] presents a platform named SYNERGIE to measure the energy consumption on nodes of WSNs and quantify the impact of interference on energy consumption. It measures the consumption of microcontroller, memory, sensor, and radio-frequency circuit at a rate of 1,500 samples per second through General Purpose Input-Output (GPIO) lines and a resistor interface. The measurements are transmitted to a computer through a serial link. This platform is based on commercially available low-cost and low-power components: AT-mega328p microcontroller, which contains an embedded 10-bit Analog-to-Digital-Converter (ADC) and five operational amplifiers XCT1086 able to acquire up to five independent measurements. A similar power meter system is presented in [23]: the authors present a real-time power metering system for a wireless sensor network called Nemo, a noninvasive and plug-and-play device that can be easily installed without wires. Nemo has a circuit that dynamically adjusts the resistance of shunt resistors according to the current load (shunt resistor switch). Nemo presents a high measurement accuracy (average measurement error of 1.34%) on TelosB devices.

The author in [6], in addition to the introduction of SYNERGIE platform, also explores the problem of energy efficiency and discusses a model of energy consumption on WSN nodes, presenting software that allows generating automatically an energy consumption model based on Principal Component Analysis (PCA) and an algorithm for node lifetime estimation based on Markov chain.

The authors in [2] also present a model for global energy consumption for WSN nodes using a simulator called Ns-2 and iMote2 hardware. The results show energy consumption simulated close to measured values on real devices under the same experimental conditions for radio-frequency circuits. A similar work is shown in [22], however, based on an event-trigger mechanism for each component (processor, radio-frequency circuit, sensors). Energy modeling in sensor networks is also discussed in [14].

3 Theoretical Introduction: Machine Learning Algorithms

In this section, the two algorithms used to estimate the current consumption will be introduced: Long-Short-Term Memory (LSTM) Networks and Decision Trees with AdaBoost.

3.1 Long-Short-Term Memory (LSTM) Networks

Long Short-Term Memory Network (LSTM) is a kind of deep learning algorithm called Recurrent Neural Networks (RNN) - dense neural networks connected to themselves: this feedback loop allows the information to be stored [11, 7]. LSTM networks have a high capability of predicting time series and it differs from other types of neural networks due to the introduction of a forget gate, that controls which states are remembered or forgotten.

A gate is a structure that adds, removes, or transmits data, being composed by a sigmoid neural network layer (here denoted by σ), that outputs a number within the interval $[0, 1]$ - where 1 denotes that all information is re-transmitted and 0 denotes to not allow the transmission of any information, and a Hadamard product (denoted by $\circ$). Let's denote by U and V the matrices connecting inputs and recurrent outputs, respectively, and $, x_t, h_{t-1}$, the input and the output of the previous cell, thus:

$$h_g = \sigma(Ux_t + Vh_{t-1}). \tag{1}$$

Each cell is composed of: (i) cell state (c_t), the internal memory of the cell that stores long and short-term memories; (ii) hidden state (c_h), which decides to retain short and/or long-term information on state cell to predict; (iii) input gate (i_t), which conditionally decides which input values will be updated on state memory; (iv) forget gate (f_t), which decides how much information of current input and from previous state cell is forwarded to current state cell; and, finally, (v) the output gate (o_t), which conditionally decides which will be the output according to input and memory block.

A piece of data is forgotten only when new data replaces it. LSTM starts deciding which data will be forgotten and then which new data will be stored.

The previous cell c_{t-1} is updated into a new cell c_t and then the old state is multiplied by a forgetting factor f_t. A new possible value is generated, so it is necessary to decide what will be transmitted to output based on the cell state applying a sigmoid layer. The tanh layer forces the value to be mapped in the interval $[-1, 1]$. Denoting U_g, V_g the weights for input and previous cell output, respectively, and b_g an input bias, we have:

$$g = \tanh(b_g + x_t U_g + h_{t-1} V_g). \tag{2}$$

The input gate is a hidden layer of sigmoid activation nodes with weighted inputs x_t, h_{t-1}. Then, the expression for the input gate can be written as:

$$i = \sigma(b_i + x_t U_i + h_{t-1} V_i). \tag{3}$$

Denoting by s_t the internal state of an LSTM cell, the forget gate is a set of sigmoid activation nodes multiplied by s_{t-1} to determine what must be remembered (output of forget gate close to 1) and what must be forgotten (output of forget gate close to 0), allowing LSTM to learn according to the scenario. So, the mathematical expression for the forget gate is given by:

$$f = \sigma(b_f + x_t U_f + h_{t-1} V_f). \tag{4}$$

So, the output of the forget gate works as a weight for the internal states. Thus, the output of this state s_t is expressed by:

$$s_t = s_{t-1} \circ f + g \circ i. \tag{5}$$

The last stage of the LSTM cell is the output gate, which is composed of a sigmoid layer and hyperbolic tangent (tanh) layer, responsible to create a vector with new possible values to be added to a $\tilde{c}_t$ state. The output is given by:

$$o = \sigma(b_o + x_t U_o + h_{t-1} V_o). \tag{6}$$

thus, cell output is expressed by:

$$h_t = \tanh(s_t) \circ o. \tag{7}$$

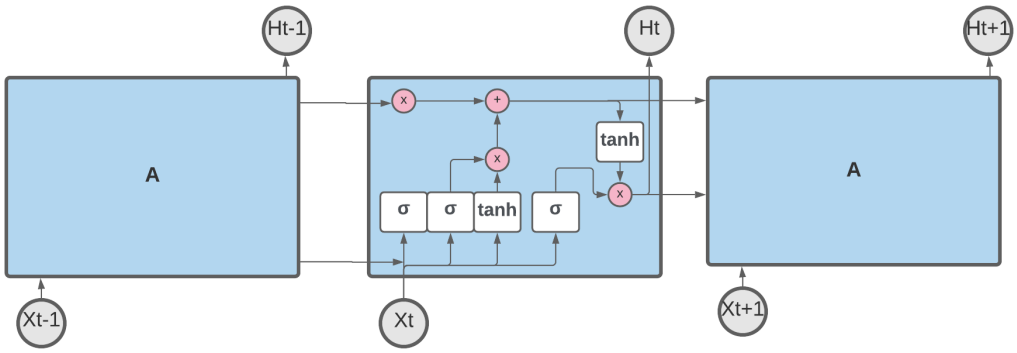


Figure 1: LSTM Architecture.

3.2 Decision Tree

A tree is a set T of positive integers with two functions $l(\cdot), r(\cdot) \geq 0$ denoting the left and right nodes respective with equality indicating a terminal node and from T to $T \cup \{0\}$ where each element of T denotes a node in the three. There is a unique parent of each node [18].

Decision trees are capable of modeling complex nonlinear decision boundaries. Associated with each internal node of the tree are a variable and a threshold and with each leaf (terminal node), is a class label. The decision tree divides the decision into simpler decisions at each node.

A sub-tree is a non-empty subset ($T_k \subset T$) associated to two functions l_k, r_r. Denoting $\{u(t), t \in \tilde{T}\}$ a partition of the data space $\mathbb{R}^p$ ($u(t)$ is a subspace of $\mathbb{R}^p$) and denoting $\omega_{j(t)} \in \{\omega_1, ..., \omega_C\}$ denote one of the class labels [18] then, a classification tree is a tree T associated with the class labels $\{\omega_{j(t), t \in \tilde{T}}\}$ and the partition $\{u(t), t \in \tilde{T}\}$ [18].

A classification tree is built using a labeled data set, $\mathcal{L} = (\mathbf{x}_i, y_i), i = 1, ..., n$ where $\mathbf{x}_i$ are the data samples and y_i the corresponding class labels [18]. Example: Properties {taste,color,shape,size}, pattern $\mathbf{x}$={sweet,yellow,long,medium} is classified as banana.

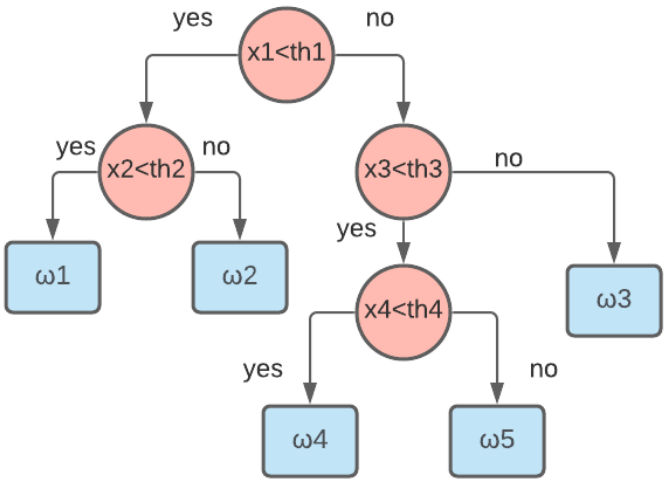


Figure 2: Generic example of a decision tree.

3.3 AdaBoost

Boosting is a technique for improving the accuracy of learning algorithms [3] and is a classifier in the form [20]:

$$F_T(x) = \sum_{t=1}^{T} f_t(x) \tag{8}$$

where each f_t is a weak learner that takes an object $\mathbf{x}$ and returns the class of the object. At each iteration of the training process, a weight $w_{i,t}$ is assigned to each sample in the training set equal to the current error on that sample.

AdaBoost (Adaptive Boosting) is a machine learning algorithm [20], an important ensemble method, and is a particular method of training a boosted classifier. It is very accurate, simple to implement, and presents good generalization [20]. AdaBoost often tends to be empirically resistant to overfitting [5], however, presents a sub-optimal solution. The AdaBoost algorithm was introduced by Freund and Schapire in 1995 [5, 20]. AdaBoost training process selects only features known to improve the prediction, reducing dimension and improving computation time once irrelevant features are not used [20]. The AdaBoost algorithm aims to build a strong classifier by doing a linear combination of weak learners (classifier) $h_t(x) : \xi \to \{-1, 1\}$ [5]:

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{9}$$

AdaBoost is capable of reducing bias and variance of weak classifiers, has a good generalization property, and its output converges to the algorithm of the likelihood ratio. AdaBoost can be seen as a feature selector and is close to sequential decision-making (it produces a sequence of gradually more complex classifiers).

Taking as input a training set $(x_i, y_i), i = 1, ..., m$ where x_i belongs to some domain and y_i is a label on some label set Y. For instance, assume $Y = \{-1, 1\}$. Let's consider $t = 1, ..., T$ the number of rounds. One of the ideas of the algorithm is to maintain a distribution or a set of weights over the training set. The weight of this distribution on training example i on round t is denoted $D_t(i)$. Initially, all weights are set equally, but on each round, the weights of incorrectly classified examples are increased to force the focus on the examples harder to learn. The objective of the learner is to find a weak hypothesis $h_t : X \to \{-1, 1\}$ appropriate for the distribution D_t. The goodness of a weak hypothesis is measured by its error [3]:

$$\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i] = \sum_{i:h_t(x_i) \neq y_i} D_t(i) \tag{10}$$

First, it assigns equal weights to all the training examples. Denoting D_t the distribution of the weights at the t-th learning round. From the training set and D_t, the algorithm generates a weak learner $h_t : \mathcal{X} \to \mathcal{Y}$ and then uses the training examples to test h_t. The weights of the incorrect classified examples will be increased. Such a process is repeated for T rounds. The final model is derived by weighted majority voting of the T weak learners [20]. The steps of the algorithm are shown below [3, 5, 17, 13, 20]:

Algorithm 1 AdaBoost

procedure AdaBoost($(x_i, y_i), i = 1, ..., N, \_i \in X, y_i \in Y = \{-1, 1\}$)
  Initialize $D_1(i) = 1/m$
  for $t = 1, ..., T$ **do**:
    Train the weak learner using distribution D_t
    Get weak hypothesis $h_t : X \to \{-1, 1\}$ with error $\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$
    Choose $\alpha_t = 0,5 \ln(\epsilon_t^{-1}(1 - \epsilon_t))$
    Update: $D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$, where Z_t is a normalization factor
  end for
Output the final hypothesis: $H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$
end procedure

4 Hardware

The system is composed of a microcontroller module, a WiFi Lora 32 (V2) module, a u-blox Gy-neo6mv2 GPS module, and an ADS 1115 16 analog-digital (A/D) converter. To activate and deactivate the GPS module, a BC 548 transistor was used to cut the ground (GND) of GPS module when a logic low level is applied to its base terminal that is connected to the microcontroller (See Fig. 3).

The current measurement is done through the voltage drop on a 1 Ohm resistor connected between the battery and the system. Two channels of A/D converter are connected between the terminals of the resistor. The microcontroller receives data from the converter and calculates the voltage between terminals. The obtained value is the current, once the resistance is equal to 1 ohm.

The system operates in cycles. The microcontroller activates the GPS and waits until the reception of a valid location. After receiving the valid location, GPS is deactivated and the coordinates are sent using the radio and the system waits five minutes to restart the cycle. The current reading is done each second.

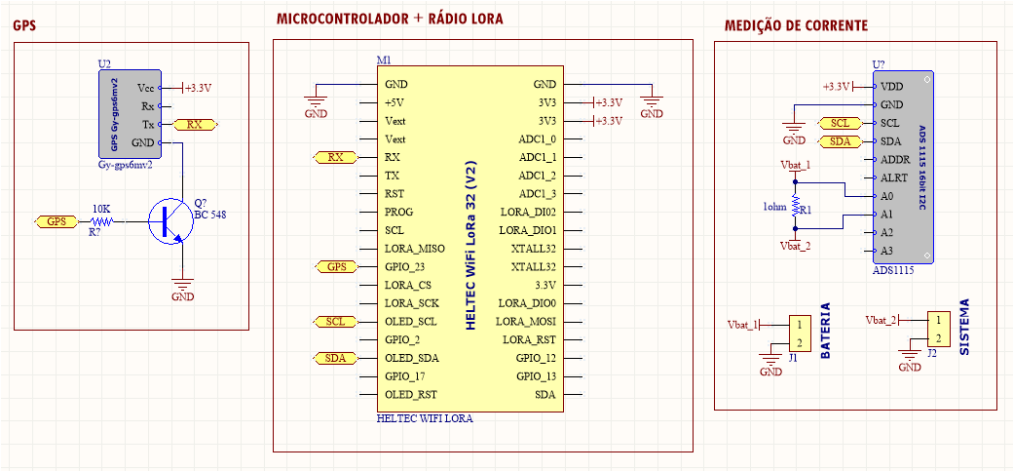


Figure 3: Circuit implemented.

5 Results and Discussion

The consumed current was estimated using LSTM networks (Figure 4) and a Decision Tree with AdaBoost (Figure 5). As is possible to see, 11,784 current samples were measured. 67% of the data were used for training and the 33% remaining were used for testing. The cycles of GPS activation/deactivation are quite evident on the graphics, showing the 5min period to restart the GPS data acquisition. When the GPS is activated, the circuit consumes 151.48 mA (average), while when deactivated, the circuit consumes 70.666 mA(average). It means that the activation of GPS module increases 114.36% the energy consumption.

In Figure 4, the blue curve represents the measured current, the orange one is the result of the estimation of LSTM during the training period and the green one is during the test period. The loss function used was Mean Squared Error and the Adam algorithm was used as the optimization function. Adam optimization is a stochastic gradient descent method adequate for large data and parameters problems based on adaptive estimation of first and second-order moments and which is computationally efficient, that consumes low memory, and which is invariant to diagonal re-scaling of gradients [9]. The calculated Root Mean Square (RMS) error for LSTM was 9.88. In figure 5, is possible to see, the blue curve represents the training samples (real data), the green one is the estimation with 1 estimator, and the red one with 300 estimators. The maximum depth of the tree was set to 10. Clearly, the accuracy when using only one estimator is much better, having an RMS error 0.0086.

As is possible to see, both algorithms can predict the behavior (profile of curve) of current consumption. So, the Decision Tree with AdaBoost with 1 estimator has better accuracy than LSTM and LSTM has better accuracy than the Decision Tree with AdaBoost with 300 estimators considering a depth equal to 10. However, if we increase the depth of the Decision Tree with AdaBoost to 20, the curve fits almost perfectly (see figures 6 and 7 for 1 estimator and 300 estimators, respectively) with a small advantage for the case using 300 estimators, presenting and RMS error equal to 0.00015.

While the architecture presented in [6, 16] is based on an AT-mega328p microcontroller, which contains an embedded 10-bit ADC, the present work is based on an ESP32 microcontroller with a 16-bit ADC. None of the papers mentioned in section 2 focuses specifically on GPS

consumption or on energy consumption forecasting. The models presented in [6] (PCA and Markov Chain) have errors between the experimental and the model varying between 1.09% and 15.19%.

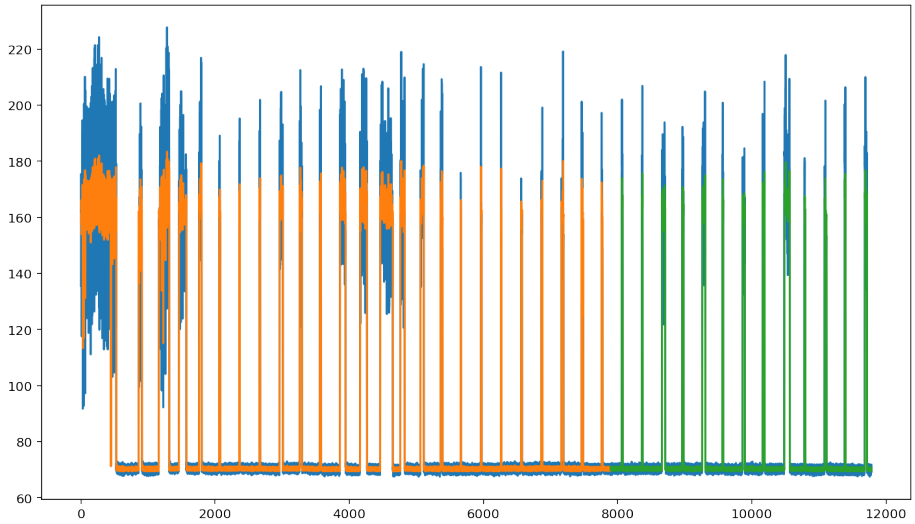


Figure 4: Current consumed estimated using LSTM.

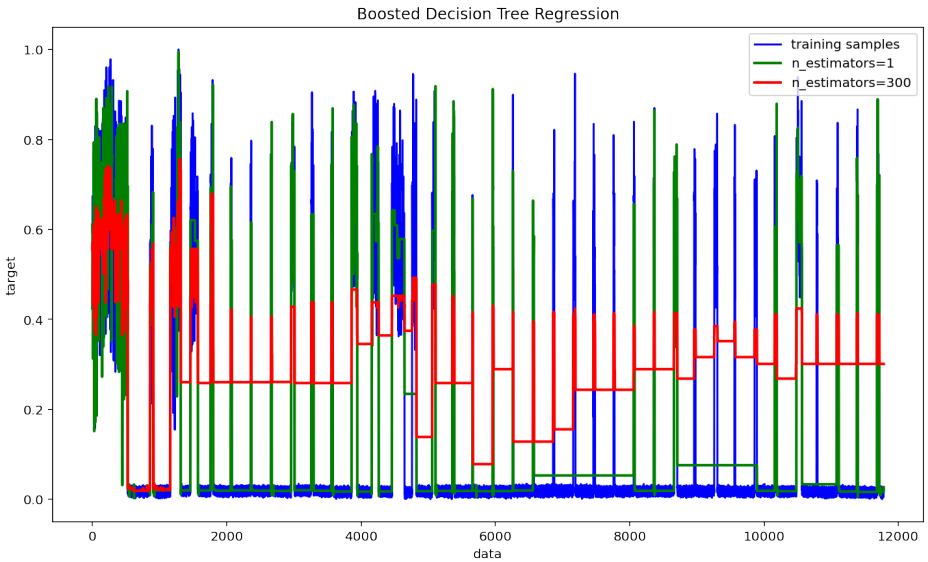


Figure 5: Current consumed estimated using Decision Tree with AdaBoost considering depth equal to 10.

8

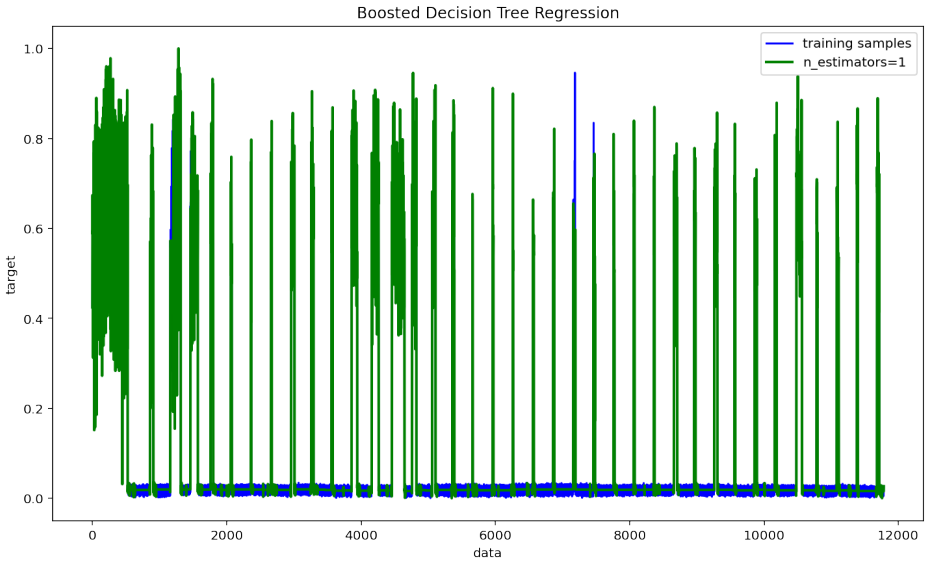


Figure 6: Current consumed estimated using Decision Tree with AdaBoost considering depth equal to 20 and 1 estimator only.

Figure 7: Current consumed estimated using Decision Tree with AdaBoost considering depth equal to 20 and 300 estimators.

6 Conclusion and Future Works

WSN has a wide number of applications. In our study, we used GPS to have a near real-time bovines location on large farms. The maintenance of the devices in distant and hard access areas may represent an important challenge for WSNs. In this context, is important to have an estimation of the energy consumption of WSN devices to predict their lifetime and have better maintenance schedules.

This paper showed the impact of the GPS module on current consumption and compared two approaches to predict the current consumption of WSN devices, LSTMs and Decision Tree with AdaBoost, with the last one presenting a better accuracy, mainly considering a depth equal to 20.

Possible interesting approaches to be explored in future works are RANdom SAmple Consensus (RANSAC) and Support Vector Machines (SVM).

References

[1] Jean Carle and David Simplot Ryl. Energy-efficient area monitoring sensor networks. *Computer*, pages 40–46, 2004.

[2] A. Coutray, A. Pegatoquet, M. Auguin, and C. Chabaane. Wireless sensor network node global energy consumption modeling. In *Conference on Design and Architectures for Signal and Image Processing (DASIP) 2010*, 10 2010.

[3] Yoav Freund and Robert E Schapire. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771 – 780, 9 1999.

[4] Nidia G. S. Campos, Atslands R. Rocha, Rubens Gondim, Ticiana L. Coelho da Silva, and Danielo G. Gomes. Smart and green: An internet-of-things framework for smart irrigation. *Sensors*, 20(1), 2020.

[5] Wei Gao and Zhi Hua Zhou. On the doubt about margin explanation of boosting. *Artificial Intelligence Journal*, 203:1–18, 7 2013.

[6] Román José Igual-Pérez. *Adaptive MAC layer for interference limited WSN*. PhD thesis, Université de Lille, 2015.

[7] Rafal Josefowic, Wojciech Zaremba, and Ilya Sutskever. An empirical exploration of recurrent neural networks. In *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

[8] Carlos Kamienski, Juha Pekka Soininen, Markus Taumberger, Ramide Dantas, Attilio Toscano, Tullio Salmon Cinotti, Rodrigo Filev Maia, and André Torre Neto. Smart water management platform: Iot based precision irrigation for agriculture. *Sensors*, (19), 2019.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[10] Abdelbasset Massouri. *Modélisation comportamentale SystemC-AMS d'interfaces RF et liaisons radio multipoint pour réseaux de capteurs*. PhD thesis, Université de Lille, 2012.

[11] L.R. Medsker and L.C. Jain. *Recurrent neural networks: design and applications*. CRC Press, 2001.

[12] Humberto Lauro Perotto Baldivieso, Susan Margaret Cooper, Andres Francisco Cibilis, Manuel Figueroa Pagan, Karen Udaeta, and Christina Michelle Black Rubio. Detecting autocorrelation problems from gps collar data in livestock studies. *Applied Animal Behavior Science*, (136):117–125, 2012.

[13] Raúl Rojas. Adaboost and the super bowl of classifiers: A tutorial introduction to adaptive boosting., 2009.

[14] D. Schmidt, M. Kramer, T. Kuhn, and N. Wehn. Energy modelling in sensor networks. *Advances in Radio Science*, (5):347–351, 2007.

[15] Viktor Toldov. *Adaptive MAC layer for interference limited WSN*. PhD thesis, Université de Lille, 2017.

[16] Viktor Toldov, Román Igual-Pérez, Rahul Vyas, Alexandre Boé, Laurent Clavier, and Nathalie Mitton. Experimental evaluation of interference impact on the energy consumption in wireless sensor networks. In *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–6, 2016.

[17] Liwei Wang, Massahi Sugiyama, Cheng Yang, ZhiHua Zhou, and Jufu Feng. On the margin explanation of boosting algorithms. In *21st Annual Conference on Learning Theory*, 7 2008.

[18] Andrew Webb. *Statistical Pattrern Recognition*. John Wiley and Sons, 2 edition, 2002.

[19] Gu Wei. *Robustness against interference in Internet of Things*. PhD thesis, Université de Lille, 2012.

[20] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Bing Ng, Angus amd Liu, Phillip S. Yu, ZhiHua Zhou, Michael Steinbach, David J. Hand, and Dan. Steinberg. Top 10 algorithms in data mining. *Knowledge Information Systems*, 14, 2008.

[21] Xin Yan. *Robustness against interference in sensor networks*. PhD thesis, Université de Lille, 2015.

[22] Hai Ying Zhou, Dan Yan Luo, Yan Gao, and De Cheng Zuo. Modeling of node energy consumption for wireless sensor networks. *Wireless Sensor Newtork*, (3):18–23, 2011.

[23] Ruogu Zhou and Guoliang Xing. Nemo: a high fidelity noninvasive power meter system for wirelss sensor networks. In *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 141–152, 2013.

A SmartNextGISSA for Monitoring and Predicting of Comorbidities

Daniel Fiuza[1], Francisco Junior[1], Fabio Gomes[2], Ivana Cristina[3], Odorico Monteiro[4], Mauro Oliveira[2]

[1] Federal Institute of Ceará - IFCE, Fortaleza, CE, Brazil,
[2] Federal Institute of Ceará - IFCE, Aracati, CE, Brazil,
[3] Oswaldo Cruz Foundation - Fiocruz, CE, Brazil,
[4] Federal University of Ceara - UFC, Fortaleza, CE, Brazil,
jwebsys@gmail.com, danielfiuzacm@gmail.com, prof.fabiojose@gmail.com, ivana.barreto@fiocruz.br, odorico@ufc.br, amauroboliveira@gmail.com

Abstract

A casual or late diagnosis given the lack of knowledge of the severity of a problem or asymptomatology of patients with comorbidity can cause serious consequences for the patient. Therefore, there is an urgent need for decision-making systems based on continuous monitoring in real time. This research presents SmartNextGISSA, an intelligent mechanism used in NextGISSA, an architecture developed with the objective of assisting decision-making in a situation room in monitoring patients with comorbidity, based on the crossing of data collected in real time from this patient and his health history. NextGISSA predictively and continuously monitors in real time patients with comorbidities (Pregnant Women, Hypertensives, Diabetics and Bedridden) of GISSA, an intelligent governance platform for decision-making in Digital Health, a product already operational in several municipalities by the startup "Avicenna Intelligent Governance". With the intention of validating the machine learning module of the architecture proposed in NextGISSA, the SmartNextGISSA module was implemented, as an emphasis of this article, which is characterized by a composition of the NextGISSA architecture responsible for the inference about the data of an individual classifying it with different comorbidity risk levels through machine learning algorithms. Thus, we present a development pipeline that consist of a comparative analysis between binary classification algorithms, such as Naive Bayes, Decision Tree and Random Forest, using the CRISP-DM methodology, a concept applied to data mining processes and, later, the model was implemented in a web application.

1 Introduction

Scientific and technological development has provided means of using digital information and communication technologies (TDIC), widely disseminated in the most diverse instances of society. Thus, many areas, such as health, have taken advantage of the benefits of these technologies, such as Machine Learning, applied to management, evaluation and analysis systems.

In the context of the Covid pandemic, the term comorbidity has become commonplace on a daily basis. The name is given to a set of causes that aggravate a disease and, in this way, can worsen a clinical condition. This study is justified, since a casual or late diagnosis given the lack of knowledge of the seriousness of the problem or even asymptomatology of patients with comorbidity urges the taking of measures with a fruitful and continuous follow-up in real time.

Machine learning can target different users with some comorbidity. With the intention of validating the machine learning module of the architecture proposed in NextGISSA, the SmartNextGISSA module was implemented, as an emphasis of this article, which is characterized by a composition of the NextGISSA architecture responsible for the inference about the data of an individual classifying it with different comorbidity risk levels through machine learning algorithms.

As for the choice of the research universe - population with comorbidity, it is known that this public is one of the priorities in care within the Primary Care policy of the public health network.

In view of this, the object of study is problematized through the following guiding question of this research: How to validate the machine learning module of the architecture proposed in NextGISSA from a scenario to monitor predictively and continuously in real time and in a way smart the patients with comorbidity found in the report of enrolled people (Pregnant Women, Hypertensive, Diabetic and Bedridden) of GISSA?

Thus, this article details the machine learning process applied to a dataset of individuals with hypertension to evaluate the application of supervised machine learning methods, specifically, binary classification algorithms, and implement them in the SmartNextGISSA module in order to to validate the architecture proposed in NextGISSA.

Therefore, the main contribution of this article includes validating the proposed implementation of the SmartNextGISSA module embedded in the NextGISSA architecture in order to provide a predictive model capable of inferring an individual's risk of comorbidity.

Thus, this article is structured as follows: Section two presents related works that use artificial intelligence are presented. In section three, the methodology used in the data mining process for the elaboration of the model. Section four presents an application scenario for SmartNextGISSA. In section five, the results obtained are presented and discussed and, finally, in section six, the conclusion and future aspirations of this article.

As for the choice of the research universe - population with comorbidity, it is known that this public is one of the priorities in care within the Primary Care policy of the public health network.

In view of this, the object of study is problematized through the following guiding question of this research: How to validate the machine learning module of the architecture proposed in NextGISSA from a scenario to monitor predictively and continuously in real time and in a way smart the patients with comorbidity found in the report of enrolled people (Pregnant Women, Hypertensive, Diabetic and Bedridden) of GISSA?

Therefore, the main contribution of this article includes validating the proposed implementation of the SmartNextGISSA module embedded in the NextGISSA architecture in order to provide a predictive model capable of inferring an individual's risk of comorbidity.

Thus, this article is structured as follows: Section two presents the methodology used in the data mining process for the elaboration of the model. In section three, related works that use artificial intelligence are presented. Section four presents an application scenario for SmartNextGISSA. In section five, the results obtained are presented and discussed and, finally, in section six, the conclusion and future aspirations of this article.

2 Related Works

BRAGA, Oton Crispim (2017) proposed an intelligent solution based on classifiers as an inference mechanism, capable of helping health professionals during the process of clinical management of diseases transmitted by the Aedes Aegypti mosquito, identifying the diagnosis based on symptoms and test results. The work was divided into two steps: one focused on pre-diagnosis, considering symptoms and clinical history, anamnesis; and another focused on the final diagnosis, also considering the results of specific tests, such as serology tests. The study uses a methodology based on Data Mining to guarantee knowledge based on examples. After several tests and adjustments in machine learning algorithms, it was possible to define two learning models capable of inferring the probability of a patient being infected with a certain disease, with an accuracy of up to 91.6%. From these models, they were able to build an intelligent API to support decision-making during the clinical management of dengue and chikungunya. The solution allows several applications to access the learning models. Among them, a popular consultation mobile application for the identification of dengue and chikungunya, MARCIA, an interoperable system for the clinical management of chikungunya.

Costa Filho (2021) proposes Smart GISSA, which is an architecture built from the Intelligent Governance System in Health Systems – GISSA, an innovative solution in the health area already highlighted in this work. This solution provides managers, servers and users with intelligent governance by promoting the systemic integration of information. Smart GISSA's architecture, in addition to extracting, transforming and loading data into dashboards, includes machine learning models to specialize decision-making in public health, in which two new Data Mining methodologies are proposed, focusing on risk analysis of death and in epidemiological surveillance to predict epidemics. It was possible to verify that this architecture already proposed to GISSA by Costa Filho (2021) does not use IOT and is not aimed at the hypertensive public.

In this way, the GISSA platform has undergone continuous evolution since its conception in 2009, as a result of research, development and innovation. Thus, research has been incorporated into GISSA, notably in the recent master's research (Francisco Junior, 2022).

From the study carried out on the related works and the proposed solutions, it was observed that almost all of the works directly related to the subject under study, Machine Learning, are aimed at different situations of case studies. Thus, this study arose from the need and possibility of converging Machine Learning and any comorbidity (hypertensive, diabetic, cardiac) on an existing platform, adding value to the architecture, in this case, of GISSA, in an expanded architecture.

3 Methodology

The research takes place from the performance of GISSA (Gissa, 2022), having as spatial delimitation the city of Fortaleza. The GISSA platform operates in a datacenter, in the cloud, provided by the company Avicena Software e Serviços Ltda. It can be accessed through the web, by computer terminals, smartphones and tablets and proactively communicating through the use of the system, accessing panels (dashboards), with access to summarized and analytical information, filterable, weighted, classified according to risks and compliance with standards quality for services provided, reporting and sending alerts and categorized messages.

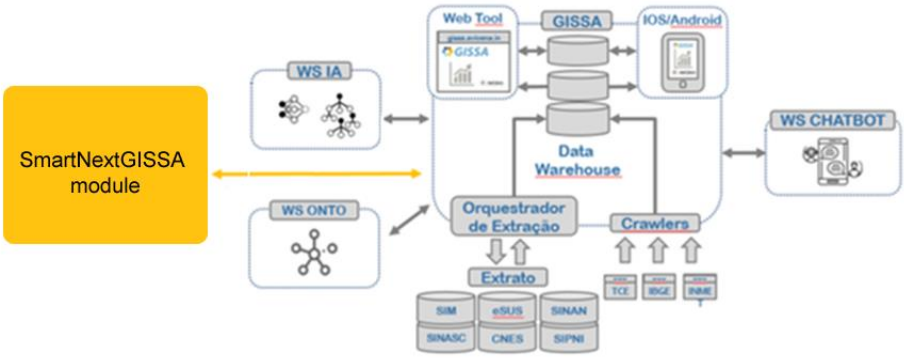


Figure 1: Proposal of the SmartNextGISSA System in the GISSA architecture

Considering that the pipeline was developed based on the CRISP-DM methodology (SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge M., 2021), the process was segmented into phases as proposed by the concept:

1. Understanding the Business: This step was characterized by understanding the problem based on the research carried out associated with the business domain of the NextGISSA platform, leading to the search for a dataset that corresponds to the scenario covered by NextGISSA.
2. Understanding the Data: Considering the definition of the problem and the compatibility of the data with the NextGISSA platform, the dataset was composed with health data from individuals characterized with 11 predictive variables (age, gender, height, weight, blood pressure, cholesterol, glucose level, smoker, alcoholic and active) and are classified into two groups of individuals (hypertensive and non-hypertensive). Data were obtained from resources available on the Kaggle platform and then exploratory data analysis was applied in order to describe and qualify them through statistical measurements. Among the analyses, the quality of the data was verified (presence of missing values, outliers, amount of data, among others) and the balance of data, indicating that they are approximately equally distributed between the two groups, with 34,979 cases of cardiac and 35,021 non-cardiac cases, totaling a sample of 70,000 cases.
3. Data preparation: This step consist of pre-processing the data, summarized by the steps of selecting columns and rows of interest (filtering), cleaning the data (removing missing values and outliers) and formatting the data (conversion of values categorical values into numerical values). After the pre-processing step, the data were normalized in order to transform the attributes into a similar scale.
4. Modeling: The process of defining the model is based on the pipeline prepared by Costa Filho (2021), since binary classification algorithms can be applied to the data set used in this article. Thus, some steps were performed to identify and evaluate the model to be defined, among them, the data loading and pre-processing, exploratory data analysis, hyperparameter optimization. The developed flow is presented in Figure 2. In the training stage, the binary classification algorithms, Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT) were applied using the Scikit-learn library. After training, the model was evaluated based on the main metrics, accuracy, precision, recall, f1 and the AUROC curve (Area Under ROC Curve). The features of the dataset are (F1) age in days of the individual, (F2) gender, (F3) height, (F4) weight, (F5) blood pressure, (F6) cholesterol level, (F7) glucose level, (F8) smoker, (F9) alcoholic and (F10) physically active.

5. Evaluation: In this stage, the performance of the supervised classifiers used in the training stage was evaluated. Then, the GridSearch strategy associated with the K-Fold Cross Validation (CV) method was applied, thus allowing the identification of the best parameters from combinations of hyperparameters.

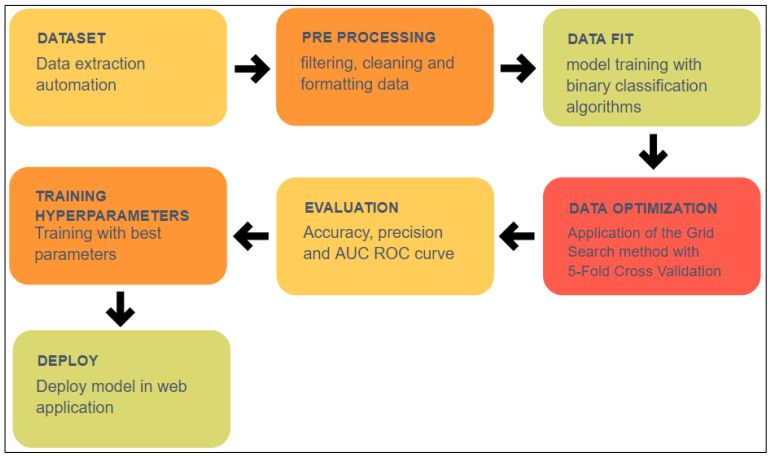


Figure 2: General flow for defining the model.

For the RF and DT classifiers, hyperparameters, number of estimators and maximum number of features were used, as shown in Table 1, while in LR penalty, C and solver from the Scikit-learn library were used, according to Table 2.

Parameters	Description	Tested Values
n_estimators	Number of trees in the forest	10, 60, 110, 160
max_features	Number of features to consider when looking for the best split	6, 11, 16, 21

Table 1 - Random Forest and Decision Tree classifier evaluation parameters

Parameters	Description	Tested Values
penalty	Specify the norm of the penalty	l1, l2
C	Inverse of regularization strength	0.0001, 0.01, 1, 100, 10000
solver	Algorithm to use in the optimization problem	liblinear

Table 2 – Logistic Regression classifier evaluation parameters

6. Implementation: After evaluating the model from the AUROC curves and performance metrics, its parameters were serialized and made available in a Rest API destined to a web application for validating the SmartNextGISSA module. In this context, this module's web service consumes the API that performs the inference based on the trained model. For this, an endpoint was made available that receives a POST request according to the format:

```
POST https://<host>/predict
```

```
{
"input": "[ 40, 140, 289, 172, 0, 0, 1, 0, 1, 0, 0,1, 0, 0, 1, 0,
1,0,0,0,1]"
}
```

The API returns, through the POST response, the value of the class (0 for non-hypertensive and 1 for hypertensive).

```
Response: [{"result": 1}]
```

Intelligent mechanisms are used to monitor vital data (weight, blood pressure, heart rate) and check these data against the patient's history through artificial intelligence algorithms.

4 SmartNextGISSA Architecture

This work concerns the use of Intelligent Systems in dealing with comorbidities, expanding the functionality of GISSA, a Public Health Governance product already in operation in several municipalities in Northeast Brazil (GISSA Manual, 2021). GISSA is a computing platform that automatically collects the databases of the Ministry of Health systems: E-SUS, CNES, SIM, SINASC, SI-PNI, SINAN through data extractor robots, analyzes this information using various intelligent and transforming them into integrated technologies information , useful for decision-making at different levels of health management.

Named SmartNextGISSA, the proposal presented here includes an intelligent module on the GISSA platform for monitoring and predicting cases of comorbidities, notably patients enrolled in the GISSA system (diabetic, hypertensive), in the city of Fortaleza-Ceará, as shown in figure 3.

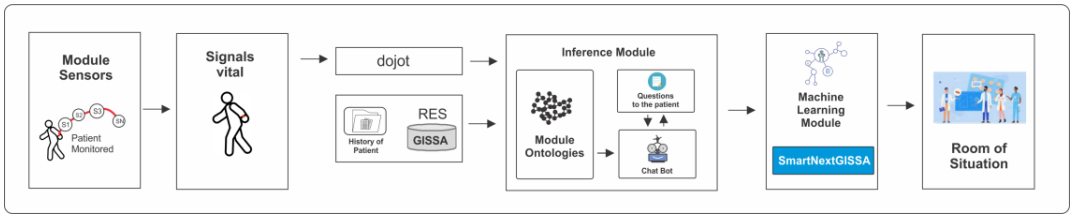


Figure 3: SmartNextGISSA Architecture

The SmartNextGISSA is proposed as an additional component to GISSA. In this context, the service will operate in the Machine Learning module of the NextGISSA scenario, receiving the history of patients from the databases monitored by the GISSA platform in order to filter individuals at potential risk through the inference of the trained model, thus allowing, that these cases move to the next stage of this scenario.

5 Results

Several experiments were carried out following the CRISP-DM methodology in order to optimize the model's performance. In the comparative performance analysis between the algorithms used, the

Logistic Regression (RL) classifier obtained better accuracy in relation to the others, as shown in Figure 4.

	Accuracy	Precision
Logistic Regression	0.719905	0.678717
Random Forest	0.716810	0.704431
Decision Trees	0.636381	0.643609

Figure 4: Accuracy and Precision

Accuracy represents the percentage of examples correctly categorized and is associated with the evaluation of the model's performance in all classes. This metric is accepted for the purpose of this work, since the dataset data are balanced and it is expected that the model presents assertive results.

In the evaluation, the AUROCC curve was also used, which represents the general performance of an estimator, since this metric considers all computed values of sensitivity and specificity. The greater the capacity of the estimator to discriminate individuals with and without hypertension, the more the curve approaches the upper left corner and the AUROCC will approach 1. In the application of the Grid Search optimization strategy associated with the Cross Validation (CV) technique, the Regression Logistic (RL) classifier was obtained as the best performance with the following parameters: C= 1.0, penalty = l1 e solver = liblinear.

After training with the mentioned values, improvements were obtained in the following metrics: accuracy = 0.720381, precision: 0.678812

The area under the ROC curve for this parameterization is shown in Figure 5 and presented as a value of 0.7220.

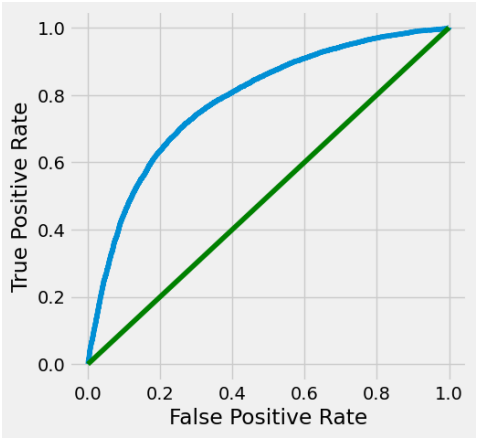


Figure 5: The area under the ROC curve

6 Conclusion

AI-based technologies using machine learning have the potential to transform healthcare. In practice, applications are used for more accurate diagnosis, identification of physiological patterns, decision support for processes in personalized medicine, etc. In the case of study, it is proposed to use it for patients with comorbidities, such as hypertension.

This article presented SmartNextGISSA, an intelligent system for the prediction of comorbidities, via monitoring of the patient's vital health signs, from their clinical history. SmartNextGISSA offers a monitoring environment by automating some services such as obtaining historical patient data, in addition to expanding the range of GISSA functionalities by including an intelligent module for monitoring and predicting comorbidities. To carry out the proof of concept of the proposed solution, a prototype will be built to carry out the tests in order to evaluate the effectiveness of SmartNextGISSA in carrying out online monitoring, integrating solutions that complement each other to signal alerts to the patient about their health status in case of any intercurrence.

As future work, we will assess monitoring for other comorbidities in addition to the report of people enrolled in GISSA, aiming to expand its scope and also identify improvements in the proposed solution.

References

AL-AZZAM, M. K.; ALAZZAM, M. B. Smart city and Smart-Health framework, challenges and opportunities. International Journal of Advanced Computer Science and Applications, v. 10, n. 2, p. 171–176, 2019.

BRAGA, Oton Crispim. Solução Inteligente Baseada em Aprendizado de Máquina para a Classificação de Doenças transmitidas pelo vetor Aedes Aegypti. Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Bacharelado em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE - Campus Aracati. 2017.

COSTA FILHO, Raimundo Valter. Smart-Gissa, Um Sistema Para Governança em Saúde Digital Baseado em Aprendizado de Máquina. Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Ciências e Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Doutor em Engenharia de Teleinformática. FORTALEZA.2021.

EL ATTAOUI, A., Largo, S., Jilbab, A. et al. Wireless medical sensor network for blood pressure monitoring based on machine learning for real-time data classification. J Ambient Intell Human Comput 12, 8777–8792 (2021). https://doi.org/10.1007/s12652-020-02660-1

FREITAS, Nicodemos. Predição de Casos de Arboviroses na Cidade de Fortaleza-CE. Utilizando Internet das Coisas e Aprendizado de Máquina/ Nicodemos Freitas. – 2021. Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2021.

STEPANIAN, Robert S.H. and LACAL, Jose C. Emerging Mobile Communication Technologies for Health: Some Imperative notes on m-health. Conference: Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. Octuber,2003.

JUNG, E.-Y. et al. Mobile healthcare application with EMR interoperability for diabetes patients. Cluster Computing, Sep. 2014. Volume 17, PP 871-880..

Gissa - Avicena - Governança Inteligente. Avicena, 2022. Disponível em: https://avicena.in/gissa/ . Acesso em: 20, Novembro, 2022.

SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge M. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science. v. 181, p. 526-534, 2021.

CidadeSocial: A social sensing model for urban dynamics *

Eliel R. Silva[1], Jonice Oliveira[1], and Tiago Cruz de França[2]

[1] PPGI, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
elielsilva@ufrj.br, jonice@dcc.ufrj.br
[2] DECOMP, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, Brazil
tcruzfranca@ufrrj.br

Abstract

In everyday urban life, people see and face different problems and challenges. Citizens may face criminal activity, poor public transportation, and other difficulties. If their claims were gathered and used to highlight urban problems, they could draw attention from authorities to such issues, saving time to solve the problems and thus improving their quality of life. People would behave as (very sophisticated) sensors of the urban environment. However, allowing anyone to share and interact with others about topics of interest to them is a necessary step toward such big-picture thinking. As a proposal to reach this goal, we describe CidadeSocial, a platform for participatory and passive sensing in a dynamic urban setting. Next, we describe the main features of a developed mobile client. Afterwards, we exemplify how to use the platform in two different application scenarios. The proposed platform contributes to the development of smart and participatory communities in which everyone plays an important role in the information economy.

1 Introduction

A city is defined as an "urban agglomeration" of continuous urban areas (or built-up areas) [2]. In this area, it is possible to observe activities that are performed with frequency between the geographical space and its inhabitants, like commuting to work, going to school, and getting public transport. Because of the large number of people who live in cities, these activities, known as "urban dynamics," have become even more important. Nowadays, most of the population lives in cities. For example, in Europe, more than 70 percent of people already live in urban areas [1].

In urban dynamics, residents and public organizations collaborate to address issues related to the environment that they live in. For example, people who go to similar places (such as classrooms, bus stops, or hospitals) face similar challenges in these settings. They can interact with each other (actively or involuntarily) to solve problems or simply to share information that interests them. This collaboration can take place in person or via social media (e.g., Facebook, LinkedIn, Instagram, or Twitter).

Although social media platforms facilitate remote interactions, we would like to draw attention to the fact that, on such platforms, the information generated by user interactions just flows through their pre-existing networks (a group of pre-connected users) [4]. This means that the information spreads in a connection-oriented way. So, even if someone is outside of a network where information is propagated, he will not receive it. To exemplify, we can consider a scenario where a certain public service uses a Twitter profile to share useful public information (such as schedules, alerts, and problems with a mode of public transportation). Such public

*This work was supported in part by Oracle Cloud credits and related resources provided by the Oracle for Research program (award number CPQ-2160239). Also, we would thank CNPq and CAPES for all support.

information will only spread up to the boundary of the group of a service's profile followers. As a result, other users of the same platform won't even get information relevant to their interests.

Aside from the restrictions on information dissemination, it is possible to notice that opportunistic interactions—spontaneous interactions that do not make use of social connectivity—are not taken into account on social media platforms such as Twitter, Facebook, and so on. Their communication models are insufficient to support the day-to-day, opportunistic interactions [3, 13].

Opportunistic interaction are spontaneous interaction that occurs between strangers. In addition, the citizens' information available on the digital world could improve services in cities. When a person shares information about a problem in a city, public departments may be able to use such information. In other words, humans could provide useful information about the city.

When someone shares information related to urban dynamics, he behaves like an electronic sensor. To put it another way, everyone could utilize their five senses to record experiences and share them as data in the digital world. Citizens are indeed the most adaptable (and much smarter) sensors. They understand context well, and unlike electronic sensors, humans can go wherever they want, observe, and share information about what they deem relevant [14].

However, there are some needs to be addressed. What parts of information and communication technology can help people act as "sensors" of urban spaces (sources of information about how urban dynamics work) and help them work together in a way that allows them to share and use this information?

As a proposal, this paper presents an adaptation of a platform named CidadeSocial, which means "Social city" in free translation to English [3, 13, 10, 9]. Our proposal, which is a work in progress, makes use of mobile devices, sensors (e.g., GPS), and a recommendation system (oriented to geolocation and interests) in order to allow citizens to monitor the city at the same time that they consume information and interact with other users in a connectionless way. Then, citizens can act as "sensors" of urban dynamics and perform opportunistic collaboration by sharing and consuming as much information as they want.

Our goal is to define a system that will aid in the formation of opportunistically dispersed networks of humans acting as urban dynamics sensors. Humans can sense a lot of physical variables and deal with different contexts. The sensed information could flow beyond their already-established social connections into a different network composed by human sensors.

This paper is organized as follows: the methodology is presented in Section 2. In Section 3, we present the literature. In Section 4, we describe the proposal. In Section 5, we present possible application scenarios. Finally, we present the final considerations in Section 6.

2 Methodology

The motivation for CidadeSocial development is a lack of appropriate models that use ICT—Information and Communication Technologies—to support humans in sharing information about urban dynamics without the need to be part of a network of people. To fulfill this motivation, it will be necessary to implement an artifact that will give humans this sensing capability.

First, the research question was defined and analyzed from different perspectives in order to enhance our understanding of the problem statement. This was followed by the development of prototypes, designs, and models based on the functional requirements collected from the problem statement. The requirements highlighted the need to make solutions that focus on

people, since people are the most important part of this architecture and work as an intelligent network of human sensors.

Considering the requirements, in the second stage of the methodology, we proposed a model that combines different ICT such as mobile devices, GPS, a gamification approach, a content recommendation engine, and other minor components to address the elucidated requirements. This work divided the solution into two parts—client and server—to aid comprehension. The first is the back-end, with its services like the recommendation service. Following the description of the mobile client and its components.

3 Related Work

A city is a concept with numerous definitions that vary depending on the context and domain in which it is used. Dijkstra et al. [4] say a city may be defined based on the number of people who live in a specific area. Other countries consider factors such as the presence of public essential services—public health and education—or the number of people engaged in non-agrarian activities. In the context of this work, the main factor is the level of urban dynamics performed in these places. Urban dynamics include, among other things, commuting, migration, the formation of dynamic events like protests, and changes in population density caused by new housing or industrial developments in the city.

In this context, each city is unique in terms of characteristics, dynamics, and issues that affect its inhabitants. These differences have motivated different researchers in the last few years to understand and document these phenomena[5, 7]. In that work, their intention was to document and understand aspects of mobility, emergency events, urban violence, energy use, communications, economics, and so on. [5] referrers to these categories of dynamics as the "pulse of the city". In general, the researchers aimed to identify anomalies and identify new trends while analyzing the crowd's behaviors.

Santos et al. have evaluated the use of data in the macrourban region of São Paulo. They discovered some interoperability issues after analyzing governmental open data platforms, which may make the use of city data for understanding city dynamics difficult. The main contribution of this work was the proposal of a structure responsible for ingesting and processing this data from different sources and making it available to the public. In [6], the authors made use of passive sensing to collect data about the levels of rain in the city of Fortaleza, Brazil, in order to predict the emergence of dengue fever cases. These studies try to understand, document, and model urban dynamics in the domains of governmental data ingestion and public health, respectively. They did not propose an artifact to enhance the inhabitants' communication and collaboration in their day-to-day lives. CidadeSocial is concerned with the human exchange of information in urban areas.

Another important related concept is "passive sensoring" which refers to the capture of information that does not require users' active input while they go about their daily lives [8]. In this field, Kubicek et al. (2019) provided a case study of cellular phone information for situational awareness. Cellular phones, perhaps the most ubiquitous personal technology available to society, provide a unique platform from which people can communicate not only verbally, but also through social media. Although cellular phone data is not easy to access because of privacy concerns, this data contains a wealth of useful information for understanding the site and situation of users. In this context, CidadeSocial proposal differs from the other study because it not only focus to gather information passively from a part of the city, however, it also gives the possibility to users posts actively information.

Trivedi et al. investigated how mobile phones functioned as contact tracing sensors. In

their work, they have modeled a network-centric contact tracing method that uses enterprise Wi-Fi networks for contact tracing in office and campus settings. These models were capable of determining when a human in an office had contact with another person who was infected with another disease, using passive sensing as a tool for controlling diseases' spreading. The present work also aims to use involuntary sensing techniques, but it is not limited to the use of Wi-Fi networks to perform the data gathering.

The Stack Exchange[1] is a tool that supports the web forums creation. It proposes is that each forum had its own subject. Users can collaborate by sharing both doubts (as topics) and answers (as topic's comments). The effective management of topics to avoid repeating questions and the gamification approaches has been leading the Stack Exchange to get success. Different instances of Stack Exchange are running nowadays, each one focuses on specific subjects. The most well known by information technology professionals are the Stack Overflow[2] and the Ask Ubuntu[3]. This software application does not explore functionalities as user localization as CidadeSocial do. Besides, in CidadeSocial, the communications are based on interests related to urban day-to-day subjects.

An Instagram[4] and Foursquare[5] evaluation was made in [12]. The authors wanted to investigate these media as being participatory sensing networks, where data from sensors (e.g., GPS and barometers, etc.) captured by the mobile devices are attached to the user-provided data [12]. Then, they described Instagram as being a media focused on cultural issues, and Foursquare as being a participatory sensing network for the definition of way finding. Thus, these authors pointed out the usefulness of these media as tools to understand both the dynamics of displacement (by tourists, for example) and the acquisition of information concerning cultural events. The aim of CidadeSocial is in communication centered on user interests (which are related to city dynamics) and not only on cultural or issues of location.

4 Proposal

This proposal intends to take an important step for society to collaborate on understanding the main social demands and problems of urban centers in collaboration with CidadeSocial. We believe that it is necessary to have a means to obtain and recover information that can be useful to improve urban spaces and people's quality of life. Therefore, we would like to have and provide data in order to support (but not address) future analyses (e.g., data fusion or new scales) and get new knowledge on different urban domains (such as public transport and air pollution, among others).

Using CidadeSocial, people will be able to send data about their urban spaces. Such data could be about physical infrastructure, services, or events that happened nearby. Since humans use their senses (and even perception) to obtain information from the real world, we made an analogy using the expression human acting as a sensor. When combined with computer technology, sensors are capable of taking information from the real world and sending it (in some representation of measurements) in the digital world.

Humans are much smarter, more adaptable, and more sophisticated than electronic sensors. They can create, feel, and do not have to follow repetitive procedures [14]. They send what they want, when they want. People open up new possibilities, but they are also far more complex

[1] https://stackexchange.com/
[2] https://stackoverflow.com/
[3] https://askubuntu.com/
[4] https://www.instagram.com/
[5] https://foursquare.com/

than an electronic sensor, which brings new challenges.

CidadeSocial is a platform for exchanging information about citizens' daily lives and their surroundings. Like in a sensor network platform, people can also send messages in a distributed way. However, human nature and ability enable them to create new connections, creating new and dynamic networks according to their interests.

4.1 The CidadeSocial Description

The communication on CidadeSocial is based on users' common interests and locations. The interests are defined as the strategy to link users, their interests, and message content. All information is associated with an interest. Therefore, users receive information according to their interests. The user's interests are defined upon account creation and can be modified at any time. The set of predefined interests is shown in Table 1. In CidadeSocial, communication (the content of messages) comes before a person-to-person relationship, unlike more popular social media platforms (such as Facebook). In other words, the primary goal of CidadeSocial is to facilitate information dissemination in order to reach everyone and anybody interested in a subject (or interest), even if they are strangers to one another.

Topics can be linked to a location or an organizational unit. Places can be found in identifiable areas or regions. Places include, but are not limited to, streets, squares, bus stops, subway or train stations, beaches, tourist attractions, and so on. an organizational unit in which a unit represents an organization with several units (or departments) The City Hall of Rio de Janeiro (an organizational unit), for example, is located in Floriano Square (a place), Cinelândia, at coordinates -22°54'36.175" N -43°10'36.108" E. Administrators record places and organizational units in the database, each with a unique name and location (geography coordinates are also used, for example, to pin places and organizations on a map). The primary goals are to connect places and organizations with subjects.

The user timeline is made up of topics categorized by interests and published near the user's current geo-coordinates. The timeline appears as the user's initial screen after logging in. It is a map with icons indicating the user's timeline's subjects. Users can also view a list of recommended topics. The topic list is a custom-built object list of the data generated on the server, and there is a recommendation engine that takes care of the user's interests. Another screen is the discover screen (topic visualization), which displays all of the user's topics (topics created by the user).

Table 1: List of interests adapted from [3]

Interests			
Art	Comedy	Crime	Elections
Employment/Work	Entertainment	Event	Exposition
Fashion	Food	Government	Health
Infrastructure	Judiciary	Legislation	Mass Gathering
Mobility	Municipal Council	Music	Politics
Protest	Recreation	Security	Service
Social Network	Sport	Trade	Transport
Violence	Other		

Table 2: Topic Types adapted from [3]

Topic Types			
Complaint	Denouncement	Disclosure	Lament
Joke	News	Praise	Other

Due to external factors, users' access to the Internet might be lost during the day; however, there is an offline mode that enables usage of some features without the Internet, and as a result, there is a client sync component. A user can create topics, view their timelines (limited to the content currently there), rate topics, and comment on subjects that have become accessible since the last synchronization while in offline mode. When an Internet connection is or becomes accessible, synchronization invariably takes place. To ensure that the user will have access to updated information, the mobile device automatically initiates an update with the server (for instance, every 30 seconds). When the connection to the Internet is restored, the data that were generated while it was disconnected and are delivered to the server.

Every time the user utilizes the system, the system may record his geocoordinates. When it becomes available, the GPS will be used to determine the coordinates. When a user conducts a search for information, the coordinates are used. The data will be arranged in accordance with how near the user is to the location where the messages retrieved by the search were published. There are other search options available, including searches by interests, locations, or organizational units. A system for gamification was created. It is thoroughly explained in [12]. Its goal is to motivate users to collaborate. The user's actions within the application result in points and level advancement. There is a ranking that is dependent on user progress.

4.2 CidadeSocial Architecture

The distributed system CidadeSocial has a client-server architecture. To construct the application over HTTP, we use the Representational State Transfer (REST) architectural style. Every communication is conducted over HTTPS in secure channels. User authorization was controlled using the OAuth2 protocol. The distributed model of the CidadeSocial is shown in Fig. 1. Both client and server are essentially made up of components, and each has a unique interface that enables communication between them. The client-side components offer user interfaces for tasks like syncing, logging in, and signing up. The server side, on the other hand, will set up and offer end destinations for web requests.

4.2.1 Client-side components

The client-side of CidadeSocial is currently a mobile application that runs on top of the Android operating system. The client, as previously stated, is an offline-first application. To accomplish this, it includes data management components that allow the app to keep all data on the mobile in case of an internet connection failure. Furthermore, the mobile software includes components aimed at determining the client's geolocation. These components update the app based on the geo-location, allowing it to construct a feed of topics based on interests and proximity.

The components of participatory sensing are in charge of topic posting, timeline processing, and evaluation registration. They work together to let users share information like text to express thoughts, pictures, or videos from the real world and to rate how useful a topic is.

There is also a component specialized in implementing gamification to help engage users

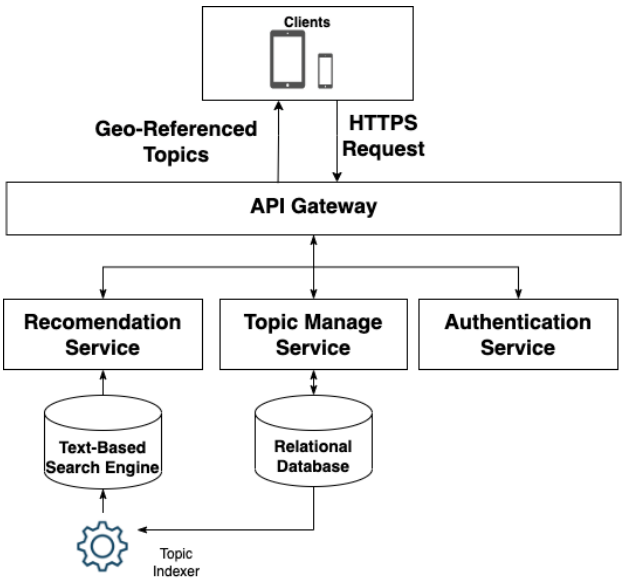


Figure 1: High-Level Overview of CidadeSocial architecture

in the task of actively contributing to the platform by posting or commenting on topics. The gamification component awards points to users based on their actions. This score is used to rank CidadeSocial users. For example, if a user posts a topic and it is evaluated 100 times, the user should be rewarded with points and a badge. [11] contains more information on CidadeSocial Gamification. Also, a more detailed explanation about the client side can be found in [3].

4.2.2 Server-side components

In this application architecture, the server is responsible for acting as a single source of truth. As this work is describing a distributed system that aims to gather and process data coming from the physical world, due to consistency purposes, all messages (or evaluations) published by clients are sent to the server, which registers the activities and stores the data in a database.

The API Gateway is the component that acts as a gateway to all other components via the web. It is called CidadeSocialAPI and is built on REST principles. All HTTP requests for each service (resource) offered by the server are received by the CidadeSocialAPI. Based on the REST principles, all resources have at least one URL. Table 4 depicts some URL examples.

All requests rely on the authorization component to obtain the client application's credentials. This component is in charge of handling client credentials. To authorize access, the OAuth2 protocol is used. The authorization is defined by the component using the client's credentials (which are associated with their level of permission for each server resource). The client must send the credentials associated with each request. Because the application has no state, all requests must be self-contained (they must have all the information necessary for its processing).

The Topic Manage Service contains all of the logic for performing operations such as timiline building operations, data CRUD (create, read, update, and delete), user ranking based on gamification techniques, and information retrieval. This service is also in charge of processing

Table 3: Example of URLs of CidadeSocial API. Adapted from [3]

Request Method	Path	Description
POST	/users	Send the new user JSON representation to create a new user
PUT	/users/{user_id}	Send the JSON user representation with new data to update the logged user
GET	/topics/timeline/{user_id}	Topics recommended to a user
GET	/topics/{user_id}	A list of 20 first topics of a user

and responding to API Gateway requests. It is built on a relational database that stores all information about topics, places, users, and interests. Correa et al. [3] previously explained specific details about this service.

Finally, the recommendation service is used to generate personalized topic recommendations based on the users' ratings, interests, favorite locations, and current location. This service uses the user's latitude and longitude to perform geo-queries against a text-indexed repository of topics and interests. A process called ETL (extract, transform, load) is used by this base to index data from the Topic Manage Service.

4.3 The CidadeSocial Implementation

A solution for a real-world problem must be developed and validated, as described in the methodology session. To accomplish this, the proposal described in the architecture section was implemented to test its viability. To accomplish this, both the Topic Manage Service and the Recommendation Service were built on the server side with Java and the SpringBoot[6] framework. The Hibernate Object-Relational Mapping framework was used to connect Topic Manage Service and PostgresSQL[7].

Elastic Search[8], a search engine based on the Lucene library, was also used to implement the recommendation service. It provides a multitenant, distributed full-text search engine. The need to perform geo-queries based on user location was the motivation. In addition, Logstash[9] was used as an ETL tool to index data from the relational database.

The client was built with React Native with Platform Expo in order to generate the app in the future for IOS. This paper presents the main flows of use of CidadeSocial as an app in order to better understand the app. The topics-geolocation features are connected to the use cases. Along with this description, the client's user interfaces will be presented as usage flows.

The first use case is the sign in, where the user could insert his personal information and select his interests from a list of interests. Fig. 2 shows the sign in flow. After completing the sign in, a user could access a map with topics published by other users in the same region (defined by a radius size of 1 km).

[6]https://spring.io/projects/spring-boot
[7]https://www.postgresql.org/
[8]https://www.elastic.co/elasticsearch/
[9]https://www.elastic.co/logstash/

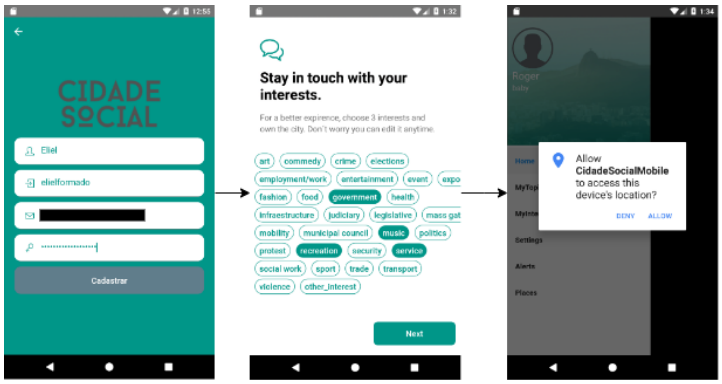


Figure 2: The Flow of a sign in CidadeSocial

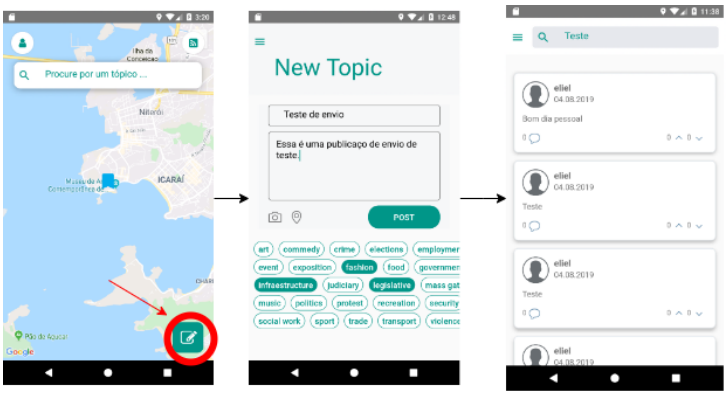


Figure 3: Publish Topic Flow

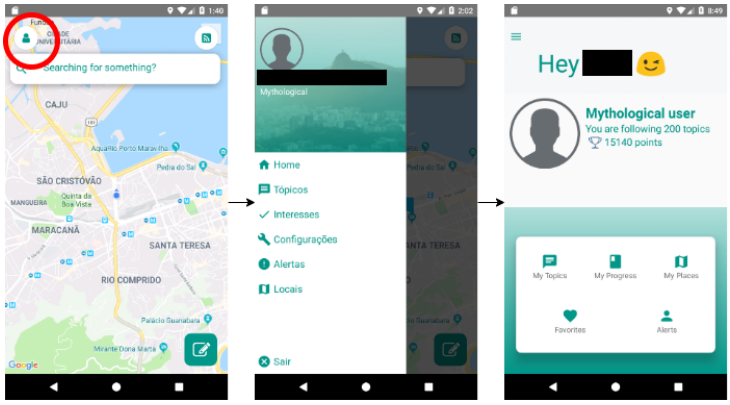


Figure 4: CidadeSocial Gamification Flow

As shown in Fig. 3, the main application's screen shows a map in which the user can see the nearby topics according to his or her interests. Other important screens related to the user's timeline enable users to just list the topics. The topics that will compound the user timeline are returned as pageable responses that enable users to get more and more topics up to the end of available data.

The user could also navigate using other functionalities. As shown in Fig. 3, a user posted a topic from the main screen. The right button on the first screen will give them access to the form to publish a topic. The user enters the title and text and then proceeds to select interests related to the topic. At the end of this flow, the user would see his topic on the screen of topics. That screen also shows some elements of the gamification approach.

Gamification is the strategy used to encourage users to use CidadeSocial, by given then the feeling of progress while using the plataform. People tend to feel motivated by actions that bring a direct reward, like when they interact with game like elements (gamefulness) [11].Users receive insignias according to the comments they receive and publish topics. The insignias are bronze, silver, gold, and platinum. Other activities contribute to the user experience, and they receive titles on the app. The user starts as a baby and grows in experience until he reaches a mythological level (see the last screen in Fig. 4). Gamification provides the user with a sense of evolution and significance. The evaluation of topics and messages allows other users to identify the best answers (comments) to a question and the top topic evaluated by people.

4.4 Application Scenarios

The major motive for this proposal and its implementation is to give people the ability to act as sensors and report problems and ideas connected to their reality and community. In this context, we discovered that citizens may desire to communicate and exchange information in cities on a daily basis in a variety of settings. As a result, we've identified two scenarios in which the CidadeSocial may be useful.

4.4.1 Sensory Participation and Passive Sensing in Urban Mobility

Consider the following scenario: the government invests heavily in a new bus model (bus rapid transit) and heavily in new autos. The bus service has been adjusted in the hopes that the new transportation service will improve people's journeys. However, in a large city, it may be difficult to track which routes have buses in bad condition or if drivers are driving dangerously. CidadeSocial encourages people's engagement in this scenario by making it simple to speak with others, even strangers. They can upload articles with pictures showing that a certain car is not well taken care of, and CidadeSocial can passively collect GPS data to see if the bus is taking the right route.

4.4.2 Engage pople by making easy get and share information

CidadeSocial helps people get information from others or find the information they need because others have had the same problems. In a previous experience, CidadeSocial was instantiated as "CampusSocial" [13]. In this use case, the academic community of the Federal University of Rio de Janeiro was able to share information based on their interests and get answers to daily questions like "Where are the libraries located on the campus?" "Is the university open 24 hours a day?" or "When is the university restaurant less crowded?" At this scene, the academic. The community was a human network sensor, sharing knowledge that machine sensors wouldn't be able to do.

5 Final Considerations and Future Works

This paper presented a platform of opportunistic collaboration that results in networks of information produced by humans about urban dynamics. This platform is named CidadeSocial. CidadeSocial is a model that employs participatory and passive sensing. Humans are capable of detecting and measuring events in urban dynamics. CidadeSocial supports the sharing of human-produced data and combines it with others produced by the electronic sensors of mobile devices. Furthermore, the application may offer content depending on the user's interests and location, as well as provide gamification approaches to engage and motivate users. The goal is to support the broad spread of information in urban areas, as the city could be sensored and the new information could be useful for improving the lives of citizens.

The primary purpose of this research was to create and implement a model that would aid in the establishment of opportunistically distributed networks of individuals acting as urban dynamics sensors. This goal was accomplished through the design of the "CidadeSocial" architecture and the deployment of both the backend service and the mobile client, which have the ability to help gather, process, and generate information derived from urban dynamics. CidadeSocial's methodology differs in that by using humans as active and passive sensors, the platform may transform itself into a powerful network of sensors.

5.1 Challenges to the Validity of the Proposal

The effectiveness of CidadeSocial is dependent on widespread adoption and use because all information and its validity are dependent on human usage. Even the proposed gamification approach may not be sufficient for this purpose. Besides, important issues such as fake news or rumors were outside the scope of this paper.

5.2 Future Works

There is the interest of evaluating and aligning CidadeSocial with the content-based publish-subscribe communication paradigm because of its characteristic of proactive distribution of interest-based content, as we intended for CidadeSocial. Furthermore, in the future, it will be necessary to address the adoption and engagement of CidadeSocial's users (experience, acceptance, and intent to use). The evaluation may be made through the application of well-known models of adoption of new technology. We hope to do that in alignment with the real use on Brazilian university campuses, since some of them are managed as cities. The real and free use and adoption of a social sensor platform could increase our knowledge of it.

References

[1] Vito Albino, Umberto Berardi, and Rosa Dangelico. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology*, 22:2015, February 2015.

[2] Jeff Cody and Francesco Siravo. *Historic Cities: Issues in Urban Conservation*. Getty Publications, July 2019. Google-Books-ID: bzGRDwAAQBAJ.

[3] Ana Correa, Eliel Silva, Tiago França, José Gomes, and Jonice Oliveira. CidadeSocial: An Application Software for Opportunistic and Collaborative Engagement of Urban Populations: First Workshop, BiDU 2018, Rio de Janeiro, Brazil, August 31, 2018, Revised Selected Papers. pages 141–155. January 2019.

[4] Lewis Dijkstra, Aneta Florczyk, Sergio Freire, Thomas Kemper, and Martino Pesaresi. Applying The Degree Of Urbanisation To The Globe: A New Harmonised Definition Reveals A Different Picture Of Global Urbanisation.

[5] Gregory Dobler, Masoud Ghandehari, Steven E. Koonin, Rouzbeh Nazari, Aristides Patrinos, Mohit S. Sharma, Arya Tafvizi, Huy T. Vo, and Jonathan S. Wurtele. Dynamics of the urban lightscape. *Information Systems*, 54:115–126, December 2015.

[6] Nicodemos Freitas and Emanuel Bezerra Rodrigues. Predição de Casos de Dengue na Cidade de Fortaleza-CE Utilizando Internet das Coisas e Aprendizado de Máquina. In *Anais do VI Workshop de Computação Urbana (CoUrb 2022)*, pages 196–209, Brasil, May 2022. Sociedade Brasileira de Computação - SBC.

[7] Akira Kinoshita, Atsuhiro Takasu, and Jun Adachi. Real-time traffic incident detection using a probabilistic topic model. *Information Systems*, 54, July 2015.

[8] Sujen Man Maharjan, Anubhuti Poudyal, Alastair van Heerden, Prabin Byanjankar, Ada Thapa, Celia Islam, Brandon A. Kohrt, and Ashley Hagaman. Passive sensing on mobile devices to improve mental health services with adolescent and young mothers in low-resource settings: the role of families in feasibility and acceptability. *BMC Medical Informatics and Decision Making*, 21(1):117, April 2021.

[9] Jonice Oliveira, Andressa Silva, and Raphael Franckini. *UFRJ Social - Propagação Colaborativa e Recomendação De Informações Utilizando Computação Móvel e Dados Georreferenciados*. October 2012.

[10] Jonice Oliveira, Andressa Silva, and Raphael Franckini. Collaborative information gathering and recommendation using mobile computing. In *Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 67–71, June 2013.

[11] Eliel Silva, Tiago França, and Jonice Oliveira. *Aumento da Adesão e do Engajamento de Usuários do Campus Social com Uso de Mecanismos de Gamificação*. October 2016.

[12] Thiago H. Silva, Pedro O. S. Vaz De Melo, Jussara M. Almeida, and Antonio A. F. Loureiro. Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21(1):42–51, February 2014. Conference Name: IEEE Wireless Communications.

[13] Pamela Tabak, Eric Figueiredo, Tiago França, Fabricio Faria, and Jonice Oliveira. *Campus Social: uma ferramenta para trocas oportunísticas de informações em campi universitários*. July 2015.

[14] Dong Wang, Tarek Abdelzaher, and Lance Kaplan. Social sensing trends and applications. pages 13–20. December 2015.

Technical Session 4 (Full Papers)

Application in Autocad to automate the calculation of mechanical efforts in fiber optic network projects

Douglas de Paiva Avelino, Jeann de Souza Teixeira and Antônio Alan Rodrigues de Araújo

[2] Federal University of Ceara, Sobral, Brasil.
eng.douglaspaiva@gmail.com, jeann.jst@gmail.com e alanaraujo.eng@gmail.com

Abstract

The topic addressed in this academic work describes a tool for capturing and extracting data applied to AutoCAD software using the Autolisp dialect. The plugin has the function of automating the calculation of mechanical efforts in fiber optic network projects. With the growing demand after the growth of Fiber the home (FTTH) networks, there was a need to install new cables on electricity distribution poles. Consequently, the development of FTTH networks has caused overcrowding on poles in large urban centers with a limited number of fixing points. The authorization of shared use of aerial infrastructure is regulated by NBR 15214/2005. Municipal FTTH networks occupy a significant percentage of a city's total poles. As a result, the regulatory process becomes very repetitive, thus leading to the possibility of human error. This work proposes the creation of a plugin that automates a large part of the technical process known as the sharing of poles project to use aerial cables in energy distribution poles. Therefore, a spreadsheet is used as a kind of database. From this worksheet, all the attributes necessary for the elaboration of the project are extracted, such as the vector orientation of the effort applied by the cable. The plugin developed in this work performs a georeferenced analysis of the entire plant that involves the network, extracts specific data from each pole and plots all points on the map, automates the process of orienting double T-type poles and creating blocks with the resulting stress values, in addition to calculating all the mechanical stresses in the network. The tool obtained convincing results, certifying the excellent functioning of the application to the projects, both in getting and in the implementation of data, thus legitimizing the proposal presented in the work.

1 Introduction

The incentive for the creation of the first telecommunication operators in Brazil emerged with the idea of making the state monopoly of telecommunications systems, conceived by Telebras and Embratel, more flexible. That occurred due to the approval of the General Law of Telecommunications (LGT) in 1995. The law nº9.472 determined the rules that the National Agency of Telecommunications (ANATEL) should follow, like the supervision and regularization of the telecommunications companies (CARVALHO et al.,).

The incentive for creating the first telecommunication operators in Brazil emerged with the idea of making the state monopoly of telecommunications systems, conceived by Telebras and Embratel, more flexible. That occurred due to the approval of the General Law of Telecommunications (LGT) in 1995. The law nº9.472 determined the rules that the National Agency of Telecommunications (ANATEL) should follow, like the supervision and regularization of the telecommunications companies (CARVALHO et al.).

Internet providers, also known as Internet Service Providers (ISP), emerged in the '90s, more precisely on June 22, 1995. But the difficulties encountered were enormous compared to the big operators, who needed the infrastructure to operate (LIGEIRO,2019).

In the past, the rules determined by ANATEL were very harsh to the providers because there was an excessive bureaucracy, the contribution of the charges was high, the difficulties of investments in infrastructure (like equipment, qualified professionals, etc.), and that ended up creating clandestine providers. According to the LGT (art. 183 from Law 9.472/97), the following crime is foreseen: "To clandestinely perform telecommunication activities," with the penalty is 2 to 4 years of imprisonment, in addition to a fine (ANATEL,2020).

In the second decade of 2000, it was initialized, by the providers, a process of internet services offers that came out from the countryside to the big urban centers (BONILHA,2019). That movement had as its idea the opposite displacement usually made by the big operator's cause; naturally, they went from the big urban centers to the countryside. The regional providers saw that it had a great market to invest in those far away regions, like city districts and suburbs.

The growth of FTTH networks coincided with the development of internet providers. Consequently, years later, around 2016, the demand for optical cables increased significantly, causing the cost of equipment related to the infrastructure of wired internet networks to present a sharp price drop.

Since then, in 2016, the FTTH networks became popular, and regional providers started to use this new market trend. That's how the providers began the process of space occupation on the posts of the electric dealership, which is ENEL in the state of Ceará, once the big telecommunications operators only occupied those posts.

2 Problem Statement

In 2005, the ABNT NBR 15,214 technical standard was created to establish the minimum technical conditions necessary for sharing the infrastructure of aerial and underground electrical energy networks with telecommunications networks. The norm is intended for the occupation of electrical power network infrastructure up to 34.5 kV, in urban and rural regions, by Telecommunications networks.

At the time, the ABNT NBR 15,214 technical standard was intended only for large operators. The ABNT NBR 15,214 technical standard wasn't popular, the providers needed to be made aware of it, and even energy concessionaires needed to have well-defined customer clarification sectors. In a short time, the demand for space on power poles increased due to the presence of regional providers that expanded quickly, reaching dozens in large cities.

In a short time, it was noticed that the excess (crowding) of cables on the poles was causing a risk to people's safety, as they are present on sidewalks, close to avenues, streets, buildings, etc. The poles were not designed to support the load (effort) generated by fiber optic cables, so there were several changes in the types of poles, now being developed to cover a certain number of occupants in addition to the wires of the electrical network.

To have control over the integrity of the poles and public safety, currently, energy concessionaires require that infrastructure-sharing projects calculate the effort exerted by Telecommunications cables.

However, this work of elaborating the calculation carried out by each cable present on the power pole can be considered somewhat exhaustive due to the large amount of necessary information that must be extracted from the technical drawing, in addition to the fact that the flow of work is repetitive, that is, it is subject to calculation errors.

Therefore, this work presents a practical, dynamic, and efficient solution for the described problem. This solution, developed mainly in AutoCAD software, intends to quickly and automatically calculate the effort resulting from each fiber optic network project pole.

3 Solutions to the problem

The main goal of this work is to automate the effort calculation of fiber optic cables, aiming at the minimum error in the elaboration of projects and consequently the saving of time. In addition, prove from the results of all the mathematician's development that the current standard (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2005) is outdated since it provides a lower number of fixation points (occupants) of fiber optic cables than the pole can support it (currently there are five positions), that is, energy concessionaires could assign more points to internet companies, thus solving the problem of overcrowding on poles. Some specific steps will be taken to automate the effort calculation process and are listed below:

- To use graphic software where the project will be made;
- The use of a tool to collect and export information from a graphic design;
- To collect the exported data into a table to find the resulting calculation;
- To import the resulting data into the graphic software showing the values resulting from the effort calculation.

4 Related works

The theme of this work has also been explored in other bibliographies, such as (MABJO, 2021), which presents an editable worksheet to insert values such as the height of the pole, type of pole, the geographic orientation of the fiber, type of cable, among others, and later returns the effort value and vector orientation of the effort. Something similar can also be seen in (SILVA, 2019). Another bibliography that is also related to stress calculation is (NEXANS,), which details the mathematical concept used to calculate the mechanical stress on poles. The works cited above are quite effective but do not present a solution that interacts with CAD (Computer Aided Design) software, allowing greater flexibility in the project. However, works that use Lisp language to automate projects and perform calculations are discussed (MORENO; BAZÁN, 2017), (AMIRASLANOVA, 2012), (MUHAMMAD; HASSAN, 2015), and (Kumar; Singh, 2008). Such works are more focused on the area of Civil Engineering due to the vast field of application.

5 Methodology

In this section, the method used to automate the effort calculation process will be addressed, including AutoCAD's programming to the Excel spreadsheet.

5.1 Application of Lisp to calculate mechanical stress in fiber optic network projects

As shown, Autolisp contains very effective tools to guarantee an excellent personalized functioning of a particular process executed in AutoCAD.

To better explain how broad this data extraction process is, it will be presented which elements and situations are mandatory to perform the effort calculation on a pole of the electric utility.

- Situation 1: On a pole where the fiber optic cable is at the end of the FTTH branch;
- Situation 2: On a pole that has an angular orientation between the anchorages of the fiber optic cable less than or equal to 160° (always considering the smallest angle);
- Situation 3: On a pole that has 3 fiber optic cable anchorages;
- Situation 4: On a pole that has 4 fiber optic cable anchorages.

For each situation described above it is necessary to seek the following information:

- Pole identification: In the sharing project, identifiers are placed for each pole to ensure organization and thus allow its location;
- Height and Effort of the pole: During field collection, this information is recorded to know what type of pole will be used for the passage of the optical fiber, that is, it is known which electric power network is passing through the pole, low or /and medium voltage, the height of the pole and the total effort it supports;
- Span: You need the value in meters of the span of each fiber optic cable that is anchored on the pole;
- Regulator Span: It is necessary to know the value of the regulatory span in meters for each of the fiber optic cables that are anchored on the pole until the next anchorage;
- Geographic orientation of the fiber: For each fiber optic cable anchored on the pole, it is necessary to know the angle it forms with the reference (zero degrees, hollow face of the pole) which is oriented towards geographic east;
- Location on the plank: For each calculation performed, it is necessary to identify which plank the pole is inserted in the drawing.

Given this, a task routine was created so that it is possible to capture all the information described and analyze each situation. Figure 1 shows a summary of the process, which starts with calling the function in AutoCAD until creating an "Excel.csv" file.

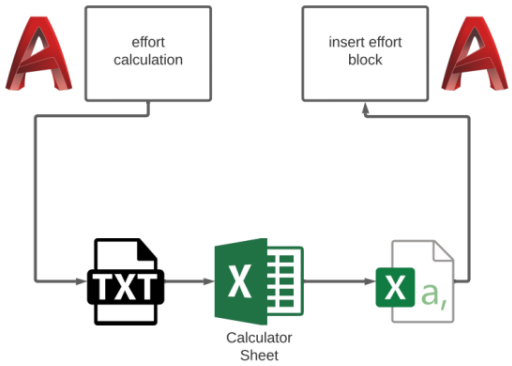


Figure 1 - Summary of the effort calculation process.

As it is necessary to summarize the effort calculation in a table, a Lisp was used to extract the AutoCAD data in text format and then insert it into an Excel spreadsheet. The tables present in the spreadsheet's calculation model automatically calculate, based on equation 1, the effort and angle resulting from each pole, respectively. The second Lisp was also created to save time, as it automates this procedure of finding the post, inserting the block with effort and angle attributes, and writing the values manually; the only action the user needs to perform is to click on the desired location to insert the block.

6 Case study

A small pole-sharing project was made for the case study to observe its results, following the order of figure 1. The example project is shown below in figure 2. It has a fiber optic network route passing through 18 electricity utility poles.

The points were chosen randomly, but with accurate data from a conventional project, this choice was made due to data protection from partner companies. It is possible to observe in figure 2 that the project has two layouts and a small number of posts, 18. This choice aims at the easy execution of the codes and presentation of results since, generally, the projects have an average of 500 to 1000 poles depending on the City.

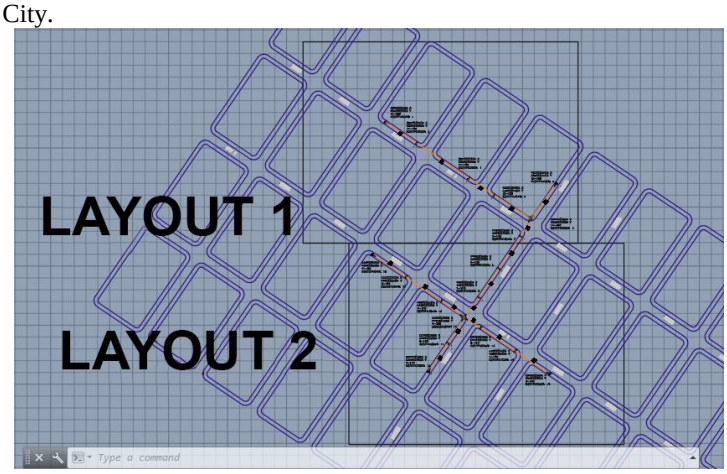


Figure 2 - Sample Project Overview

The purpose of the case study is also to verify the total load exerted by the telecommunication cables on the poles and to verify if it is possible to increase the number of points available on the pole because in this way the problem of capacity on the poles would be solved with its greater availability of points. The total effort exerted by the cable will be compared to the nominal effort of the pole.

7 Results and discussions

The first step is to use Lisp to extract the necessary data to calculate the mechanical stress on the posts. Below in figure 3 shows part of the information collected at the end of the code execution.

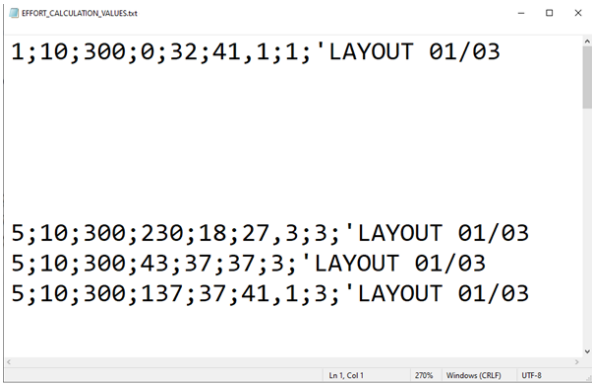


Figure 3 - txt file with extracted data

Each data referring to the pole is separated by ";", so the order of organization of the information follows the following configuration: id, effort, height, orientation in degrees, span, regulator span, amount of forces, and sheet number. Once inserted in the worksheet, the resulting values are instantly found on the pole. Below is the calculation table for posts 1 and 5.

Table 1 - Calculation of Pole 1

Pole:	1	Height (m):	10,0			pole capacity (daN):	300		model:	DT	layout:	01/03	Wire rope	NOT		splice box	NOT
Company:		**Force**	**Angle**	**average distance**	**distance**	**cable type**	**N° ways**	**Effort at the tie point (daN)**	**wind load (daN)**	**total effort (daN)**	**Multiplication factor (20 cm from the top of the pole)**	**X-Axis Reflected Effort (daN)**	**Y-Axis Reflected Effort (daN)**				
Company X		F1	0	41,1	32	CFOA-SM-AS80-12F	1	27,7	3,4	31,1	0,69	21,5	0,0				
Resulting effort (daN):	21,5		**Resultant force angle (°):**	0		**Load used on the pole (daN):**	150		**Swap pole?**	NOT	**Total Effort on Face (daN):**	21,5	0,0				

Table 2 - Calculation of Pole 2

Pole:	5	Height (m):	10,0			pole capacity (daN):	300		model:	DT	layout:	01/03	Wire rope	NOT		splice box	NOT
Company:		**Force**	**Angle**	**average distance**	**distance**	**cable type**	**N° ways**	**Effort at the tie point (daN)**	**wind load (daN)**	**total effort (daN)**	**Multiplication factor (20 cm from the top of the pole)**	**X-Axis Reflected Effort (daN)**	**Y-Axis Reflected Effort (daN)**				
Company X		F1	230	27,3	18	CFOA-SM-AS80-12F	1	21,7	2,3	24,0	0,69	-10,6	-12,7				
Company X		F2	43	37	37	CFOA-SM-AS80-12F	1	19,4	3,1	22,5	0,69	11,3	10,6				
Company X		F3	137	41,1	37	CFOA-SM-AS80-12F	1	24,0	3,4	27,4	0,69	-13,8	12,9				
Resulting effort (daN):	17,0		**Resultant force angle (°):**	141		**Load used on the pole (daN):**	216		**Swap pole?**	NOT	**Total Effort on Face (daN):**	-13,1	10,8				

With the worksheet finalized, the values resulting from post 1 (21.5 daN and 0°), post 5 (17.0 daN and 141°), and the others are inserted in the AutoCAD Software; below are shown in figure 4 with the results of the values.

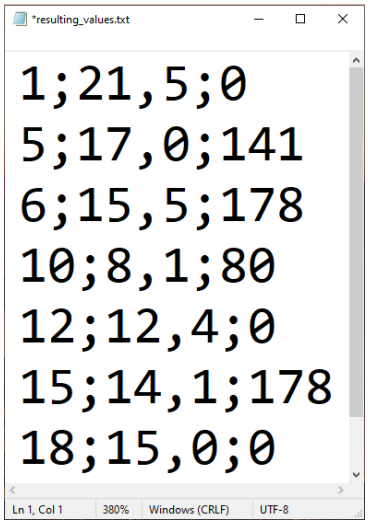


Figure 4 - txt file with resulting values

The file containing the resulting values is used as input for the second AutoLISP code, whose function is to insert the resulting blocks containing the values already filled in, automating the final process of effort calculation. Figure 5 below shows the block resulting from pole 1 (21.5 daN and 0°).

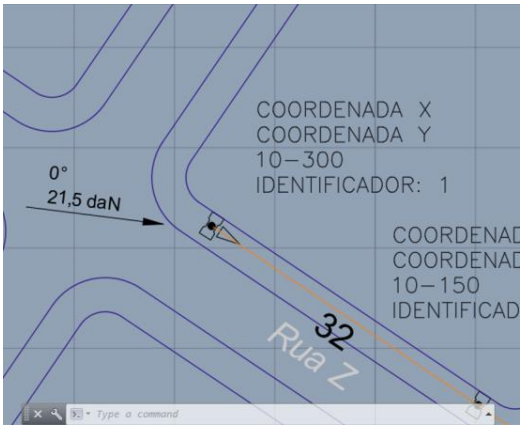


Figure 5 - Resulting Block on post 1 (Project-example).

Another significant result found is the relationship between the resulting effort exerted by the fiber optic cables and the nominal effort of the pole. It is possible to notice that the difference is significant; this can be seen below.

The resulting effort on pole 1 is 21.5 daN and represents 7.16% of the nominal effort of the pole, 300 daN. On pole 5, the resulting effort is 17.0 daN, representing 5.6% of the nominal value of the pole, 300 daN. Soon it is seen that it is possible to make more occupants available on the power poles.

8 Conclusion

In this work, a study was proposed regarding calculating mechanical efforts in fiber optic network projects using an application in Autolisp in AutoCAD software to automate its elaboration process.

Difficulties were encountered in the AutoCAD platform regarding how the data needed to perform the effort calculation would be automatically extracted without needing a designer to search for each piece of data throughout the project. An application of the Lisp language was found on AutoCAD, which adapted to AutoCAD is called Autolisp. It uses the programming language to perform various tasks in AutoCAD, and all the information necessary for the effort calculation is exported through it.

The first contact with the Autolisp language was complicated, as its syntax is different from other more common languages in Engineering, such as C or C++. In addition, it is peculiar to the AutoCAD environment, which requires knowledge of AutoCAD itself. To get around this, teaching materials such as videos and handouts were used to study as much content as possible involving AutoCAD and Autolisp. Despite the challenges, it was easy to deal with the language, as it is intuitive and straightforward in its applications.

Aiming the correct use of the tool, the entire composition of the graphic elements in the AutoCAD software was studied, in which they present everything from geometric data such as points, angles, and radius, to information on the classification of an object: lines, circle, polyline, rectangle, arc, insert block.

To validate the tool's effectiveness, it was used to do projects at the company Ágil Engenharia de Telecomunicações EIRELI. During the execution of the works, it was validated that this automation improved the tasks performed, reducing the project development time. Before, a project was done in approximately five days; now, with the in AutoLisp tool, the time has been reduced to approximately one day. Another improvement was to avoid design errors associated with calculations that were once common due to human error. The projects needed to be reviewed after they were finalized, and now, with the use of the tool, it is no longer necessary to review the calculations.

The goals indicated in the work were successfully achieved, obtaining satisfactory results consistent with the idea proposed at the beginning of this project.

Finally, it is essential to point out that through the results found for effort and orientation resulting from the example project, it is possible to conclude that the load exerted by the fiber optic cables is minimal about the total nominal effort of the pole. With this, this work serves as a justification for the increase in the number of occupants destined to the telecommunications cables on the poles of the electric power concessionaires, showing that the ABNT NBR 15.214 standard is outdated once the cables used in the fiber optic networks today are technologically more advanced compared to the time when the standard was written. Therefore, it is necessary to reformulate the norm so that it is possible to minimize the current major problem, which is the considerable crowding of cables on power poles.

References

Moreno, R., and A. M. Bazán. "Design automation using script languages. High-level CAD templates in non-parametric programs." IOP conference series: materials science and engineering. Vol. 245. No. 6. IOP Publishing, 2017.

Amiraslanova, Zarbaf. "Automation of working drawings and calculation of reinforcement weight in concrete structures." 2012 IV International Conference" Problems of Cybernetics and Informatics"(PCI). IEEE, 2012.

Muhammad, Akilu, and Hassan Ibrahim Ogiri. "Autocad Programming For Reinforced Concrete Structural Detailing." (2015).

Kumar, S., and R. Singh. "Automation of strip-layout design for sheet metal work on progressive die." Journal of materials processing technology 195.1-3 (2008): 94-100.

ANATEL, A. N. de T. Lei nº 9.472, de 16 de julho de 1997. 2020. Disponível em: https://www.anatel.gov.br/legislacao/leis/2-lei-9472. Acesso em: 17 jan. 2021.

CARVALHO, J.; ARITA, C.; NUNES, A. A política de implantação da Internet no Brasil. Disponível em: http://www.portcom.intercom.org.br/pdfs/5be0d57f5fde664d948d9c2cbc80b619.PDF. Acesso em: 15 nov. 2019.

CEMIG - COMPANHIA ENERGÉTICA DE MINAS GERAIS. Projetos de Redes de Distribuição Aéreas Urbanas. Belo Horizonte – Minas Gerais - Brasil, 2014.

COELCE. UTILIZAÇÃO DE MATERIAIS EM LINHAS DE DISTRIBUIÇÃO AÉREAS DE MT E BT: Dt-042/2010. [S. l.], 2010.

DECORA., V. Saiba o que é AutoCAD e porque arquitetos, designers e engenheiros usam o programa! 2019. Disponível em: https://www.vivadecora.com.br/pro/tecnologia/o-que-e-autocad/. Acesso em: 14 jan. 2021.

FONSECA, J. M. R. S. T. e J. O. AutoLISP - IV Ambiente de programação Visual Lisp. 2011. Disponível em: https://paginas.fe.up.pt/~tavares/ensino/CFAC/Downloads/Apontamentos/AutoLISP\%20IV_p.pdf. Acesso em: 26 jan. 2021.

FRED, A. LISP - Introdução. 2002. Disponível em: http://users.isr.ist.utl.pt/~lmmc/iasd/Lisp_intro.pdf. Acesso em: 17 jan. 2021.

FURUKAWA, E. CABO OPTICO CFOA-AS (ABNT), AUTOSUSTENTADO PARA VÃO ATÉ 200m: Especificação técnica. [S. l.], 2020. 6 p.

LABEGALINI, P. R.; LABEGALINI, J. A.; FUCHS, R. D.; ALMEIDA, M. T. PROJETOS MECÂNICOS DAS LINHAS AÉREAS DE TRANSMISSÃO. SÃO PAULO: EDGARD BLÜCHER LTDA, 1992.

LIGEIRO, G. ISP e seu apetite histórico pela inovação. 2019. Disponível em: https://www.khomp.com/pt/isp-apetite-historico-pela-inovacao/. Acesso em: 11 nov. 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 5422: Projeto de linhas aéreas de transmissão de energia elétrica. Rio de Janeiro, 1985. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 15214: Rede de distribuição de energia elétrica - compartilhamento de infra-estrutura com redes de telecomunicações. Rio de Janeiro, 2005.

AUTODESK. AutoLISP Developer's Guide. 2013. Disponível em: http://docs.autodesk.com/ACDMAC/2013/ENU/PDFs/acdmac_2013_autolisp_developers_guide.pdf. Acesso em: 23 jan. 2021.

BONILHA, C. A Revolução Silenciosa no Brasil: O Crescimento dos Provedores Regionais de Banda Larga. 2019. Disponível em: https://digital.futurecom.com.br/especialistas/revolu-o-silenciosa-no-brasil-o-crescimento-dos-provedores-regionais-de-banda-larga. Acesso em: 15 jan. 2020.

MABJO. Passo a Passo Uso do Aplicativo Customizado para Fibra Óptica. 2021. Disponível em: https://www.mabjo.com.br/downloadcalculomecanicoesforcoposte.aspx. Acesso em: 20 fev. 2021.

NEXANS. Cálculos mecânicos. Disponível em: https://www.nexans.com.br/SouthAmerica/group/doc/pt_BR/CalcMec_pt.pdf. Acesso em: 14 fev. 2021.

SILVA, B. V. D. ANÁLISE DOS ESFORÇOS ATUANTES EM POSTES DE CONCRETO ARMADO DEVIDO A CABOS. CEARÁ: [S. n.], 2019.

On Modeling Network Slicing Communication Resources with SARSA Optimization

Eduardo S. Xavier[1], Nazim Agoulmine[2], and Joberto S. B. Martins[1]

Abstract

Network slicing is a crucial enabler to support the composition and deployment of virtual network infrastructures required by the dynamic behavior of networks like 5G/6G mobile networks, IoT-aware networks, e-health systems, and industry verticals like the internet of vehicles (IoV) and industry 4.0. The communication slices and their allocated communication resources are essential in slicing architectures for resource orchestration and allocation, virtual network function (VNF) deployment, and slice operation functionalities. The communication slices provide the communications capabilities required to support slice operation, SLA guarantees, and QoS/ QoE application requirements. Therefore, this contribution proposes a networking slicing conceptual model to formulate the optimization problem related to the sharing of communication resources among communication slices. First, we present a conceptual model of network slicing, we then formulate analytically some aspects of the model and the optimization problem to address. Next, we propose to use a SARSA agent to solve the optimization problem and implement a proof of concept prototype highlighting its results.

1 Introduction

Network slicing is a crucial enabler to support the composition and deployment of virtual network infrastructures required by the dynamic behavior of networks like 5G/6G mobile networks, IoT-aware networks, e-health systems, and industry verticals like the internet of vehicles (IoV) and industry 4.0 [22] [4] [24]. In general, the slicing process results from the need to share resources among existing infrastructures to improve performance, provide cost-efficient solutions, and optimize operation [20].

This technology is already used in the context of 5G networks [22] and provided as a service (slice-as-a-Service: SlaaS) by network operators. This allows customs to create their private virtual networks (slices) tailored to their specific application domains and to develop their own business models. Network slicing is expanding its use in other scenarios of telecommunication networks, content provider networks (ISPs), experimental networks, and IoT systems, among others [15].

Network slice instance life cycle process such as commissioning, operating, and decommissioning [22] requires appropriate network communication resources. A communication slice [1] eventually represents a set of communication resources that can be used in the slicing process. It holds resources like links, optical slots, virtual private networks (VPNs), and other communication facilities necessary to provide the exchange of information among logical slices, and architectural slicing entities and for supporting the slicing process functionalities.

The communication slice resources significantly impact the performance of the resulting sliced virtual network (SVN) or virtual network operator (VNO). Among the most common network characteristics that impact the network slicing process, we can mention delay-aware network slicing like in 5G deployments [18], quality of service (QoS) aware network slicing [24], energy-aware network slicing [23], and, in general, application-dependent and multi-domain network slicing [19].

[1] A specialized slice that provides communication services among network slicing entities

The objective of this paper is therefore to propose a conceptual model of slice communication and formulate analytically some of its aspects. The model should be able to capture the set of communication resources to support the optimization of the allocation of communication resources to the different slices on top of various underlying technologies (e.g. Elastic Optical Networks - EON [7], MultiProtocol Label Switching - MPLS, others).

This paper is organized as follows. Section 2 presents the related work and Section 3 introduces the concept of multidomain sliced virtual networks. Section 4 presents a conceptual and analytical model for a communication slice used in the network slicing process. Section 5 presents a proof of concept of using the models with a SARSA agent optimizing the allocation of bandwidth resources for a communication slice. Finally, Section 6 presents the final considerations.

2 Related Work

There have been a very significant number of state-of-art research projects launched in the area during the last decade such as SFI2 (Slicing Future Internet Infrastructures) [6] [16], NECOS (Novel Enablers for Cloud Slicing) [5], SELFNET [17] and MATILDA [11], standardization initiatives launched by the IETF (Internet Engineering Task Force) [12], 3GPP (3rd Generation Partnership Project) [2], ITU (ITU-T - Telecommunication Standardization) [13], ETSI (European Telecommunications Standards Institute) [8] and ONF (Open Networking Foundation) [10] and published surveys [4] [25] [9] [14] [3]. These different initiatives have focused on different technical aspects, architectures, and slicing strategies, and all require communication slices to operate and manage the provided functionalities.

However, these slicing architectures, projects, and initiatives did only address the conceptual and analytical modeling of the basic structures and functionalities that compose the slicing process in a preliminary way or did only indicate them as future challenges to solve. To the best of our knowledge, the conceptual and analytical modeling of communication slices is a new contribution to the network slicing domain.

3 Resources, Slice and Sliced Virtual Network (SVN)

A multi-domain Sliced Virtual Network (SVN) as illustrated in Figure 1 is a multi-domain or a multi-tenant[2] infrastructure that is dynamically configured and deployed by requesting and orchestrating resources from a pool of providers on domains.

3.1 The Slice

For the scope of this paper aiming at the slicing model and deployment understanding, it is essential to conceptualize the vision of a *slice* as a component of the sliced virtual network.

We define a slice as a specific resource, service, function, or set of resources, services, and functions virtualized, shared, and grouped using any software or hardware facility. The slice with its resources, services, and functions physically resides in nodes or another physical or virtual deployment in domains.

As such, slice resource examples are virtual machines, virtual switches with hosts deployed with OpenFlow, chunks of bandwidth belonging to a physical link, slots of a fiber EON de-

[2]For the scope of this paper, a tenant can be a network domain, a service provider, a business unit, or a specific multi-tier or single-application tier providing resources for network slicing.

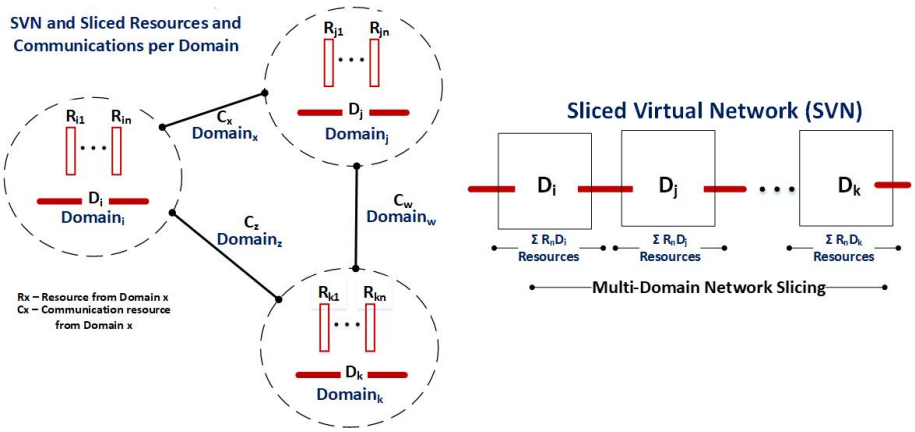


Figure 1: A Multi-Domain Sliced Virtual Network (SVN) and its Resources

ployment, LSP MPLS connections, shared spectrum in 5G radio access networks (RAN), and others. Slice function and service examples are virtual network functions (VNFs) deployed over a network providing specific services or facilities to the user.

Considering this slice basic concept, an SVN encompasses resources, services, and functions with the necessary communication resources to interconnect them inside domains and between domains as illustrated in Figure 1. In general, resources belonging to the same SVN reside in different domains and are physically or virtually attached to nodes in their respective domains.

The network slicing architecture functionalities (resource marketplace, resource broker, resource orchestrator, slice instantiation, slice monitoring, and others) are distributed in terms of the domains participating in the SVN deployment and certainly, depend on the proposed architecture and the deployed functional blocks of the network slicing architecture (SELFNET, NECOS, SFI2, MATILDA, other).

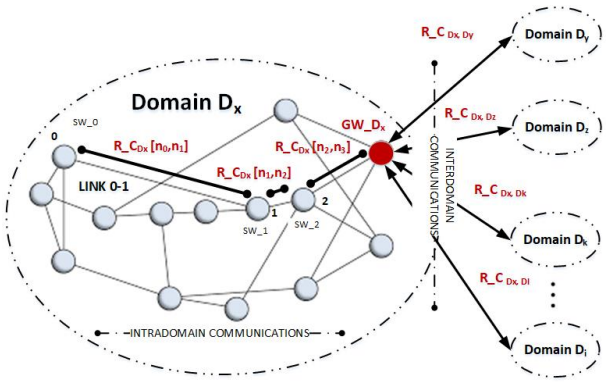


Figure 2: Intradomain and Interdomain Communication Slices

3.2 Communication Resources and Communication Slice

In order to allow the execution of the network slicing process and functionalities in any deployed slicing architecture, it is necessary to allocate communication resources allowing communication among the entities involved in the slicing process. Furthermore, once the SVN is deployed, communication resources are also necessary to support the communication requirements of the applications running (slice operation).

The generic view of communication resources used by a network slicing infrastructure to enable resource orchestration, deployment, and slice operation is illustrated in Figure 2.

We assume that the slicing process to create a sliced virtual network (SVN) involves single or multiple domains $(D_x, ..., D_z)$. Each domain is generically configured by a single or a set of nodes $(n_i, ...n_j)$ hosting resources and domains that are interconnected by communication resources.

A *communication slice* is then defined as a set of communication resources orchestrated and allocated between slices, nodes, network-slicing entities, and domains. As such, the domain nodes $(n_i, ...n_j)$ hosting resources and domains are interconnected by communication slices $(C_x, ...C_y)$.

We identify two types of communication slices that are orchestrated and deployed with distinct configurations and characteristics:

- Intradomain communication slices; and
- Interdomain communication slices.

In infrastructures composed of network domains, the modeling assumes that a gateway concentrates all communications between different domains.

We focus in this paper specifically on interdomain communications and how to model it in terms of communication slices.

4 Network Slicing Interdomain Communications

The objective of a network slicing interdomain communication model is to formally structure and capture the needs in terms of communications for the slicing process. It also allows the identification of parameters leading to the optimization of the resource allocation process.

4.1 Network Slicing Assumptions

We first introduce the following assumptions in the context of network-slicing interdomain communications that are necessary for our modeling and problem formulation:

- Each network domain is SDN-compatible;
- Each network domain gateway GW_D_i (Figure 2) is an SDN-enabled switch whose programmed behavior is to route packets between domains;
- Each network domain implements monitoring mechanisms to collect performance monitoring parameters;
- All intradomain and interdomain links are configurable in terms of allocated resources; and
- All network domains support network resource identification and have capabilities for resource allocation.

Notation	Description
$D_i^{l_i}$	The domain i located in physical location l_i
$RD_i^{l_i}$	Domain's set of shareable resources at a physical location
$R_i^{D_i^{l_i}}$	A shareable resource at domain $D_i^{l_i}$
$R\_IS_{D_i}^{l_i}$	The infrastructure and service resources
$R\_C_{D_i}^{l_i}$	The network communication resources
B_{D_i,D_j}	Bandwidth between domains
L_{D_i,D_j}	Packet loss between domains
Dl_{D_i,D_j}	Delay between domains
B_{n_i,n_j}	Bandwidth between nodes
L_{n_i,n_j}	Packet loss between nodes
Dl_{n_i,n_j}	Delay between nodes
$P\_RC_{D_k,D_j}^{l_i}$	Set of communication's link parameters between domains

Table 1: Notation and variables

4.2 Network Slicing Model

Based on these assumptions, we can now specify an analytical model of multi-domain SVN considering a set of network domains federating together their resources and infrastructures to the slicing process:

$$\aleph =< D_i^{l_i}, D_j^{l_j}, D_k^{l_k}, ..., D_z^{l_z} > \tag{1}$$

Where: $D_i^{l_i}$ is a network infrastructure domain located at site l_i.
Each network infrastructure domain $D_i^{l_i}$ has a set of shareable resources such as:

$$RD_i^{l_i} =< R_i^{D_i^{l_i}}, R_j^{D_i^{l_i}}, R_k^{D_i^{l_i}}, ..., R_z^{D_i^{l_i}} > \tag{2}$$

Where:

- $RD_i^{l_i}$ is the set of shareable resources provided by D_i and located at site l_i; and
- $R_i^{D_i^{l_i}}$ is one particular shareable resource.

There are different types of resources at each network infrastructure domain location $D_i^{l_i}$:

- Infrastructure appliances like virtual machines, access points, and IoT devices;
- Computing services like virtual network functions (VNF), storage and computing services; and
- Communications services like physical links, LSPs (MPLS Link Switched Paths), fiber lambdas, and 5G connections.

For the purpose of the SVN model, we distinguish between two types of resources:

- Infrastructure and service resources - $R\_IS_{D_i}^{l_i}$; and

- Communications resources - $R\_C_{D_i}^{l_i}$.

Users (clients) request infrastructure, service, and communication resources that are orchestrated by a network slicing software (NECOS, MATILDA, other) to create their sliced virtual network (SVN) as illustrated in Figure 1.

The communication resources $R\_C_{D_i}^{l_i}$ provide the interconnection of infrastructure and service resources $R\_IS_{D_i}^{l_i}$ for intradomain and inter-domain connections. As such, for the SVN modeling, there are two distinct communication resources or communication slices (Figure 2):

- Intradomain communication slices used between internal nodes of the domain: $R\_C_{D_i[n_j,n_k]}^{l_i}$; and
- Interdomain communication slices used between domains: $R\_C_{D_i,D_k}^{l_i}$

The communication slices are characterized by a set of parameters related to interdomain (Equation 5) and intradomain (Equation 4) communications:

$$P\_RC_{D_i,D_j} =< B_{D_i,D_j}, L_{D_i,D_j}, Dl_{D_i,D_j} > \tag{3}$$

$$P\_RC_{n_i,n_j} =< B_{n_i,n_j}, L_{n_i,n_j}, Dl_{n_i,n_j} > \tag{4}$$

Where:

- B_{D_i,D_j} is the available bandwidth between domains D_i and D_j;
- L_{D_i,D_j} is the packet loss between domains D_i and D_j;
- Dl_{D_i,D_j} is the delay between domains D_i and D_j;
- B_{n_i,n_j} is the available bandwidth between nodes n_i and n_j in a domain;
- L_{n_i,n_j} is the packet loss between nodes n_i and n_j at a domain; and
- Dl_{n_i,n_j} is the packet delay between nodes n_i and n_j at a domain.

Figures 1 and 2 illustrate a generic view of the slicing process and related interdomain communications. The network slicing infrastructure setup from the point of view of communication resources is as follows:

- A set of domains (D_i);
- A single communication slice (configurable link or another communication resource) between domains;
- An SDN OpenFlow-capable switch (gateway) handling the interdomain packet routing among domains; and
- A SDN switch (gateway) programmed to handle the interdomain packet routing among domains.

The interdomain slice communication parameters $P\_RC^{l_i}_{D_k,D_j}$ are configured during the slicing commissioning phase, as proposed in the 3GPP network slicing reference architecture and model [1].

An SVN will require resources of distinct domains to be allocated end-to-end:

$$SL_k^{D_i} =< R_i^{D_i^{l_i}}, R_j^{D_i^{l_i}}, R_k^{D_y^{l_i}}, ..., R_z^{D_y^{l_i}} > \quad (5)$$

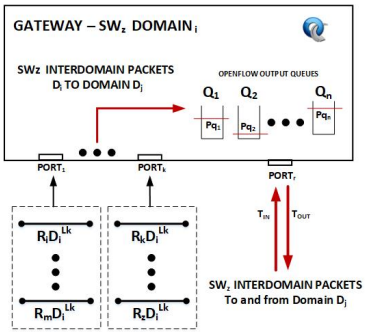


Figure 3: Openflow Switch Handling Operation and Management Slicing Generated Packets

The communication slice modeling assumes that each domain contributes to a set of different resources that are located in various physical sites (domains).

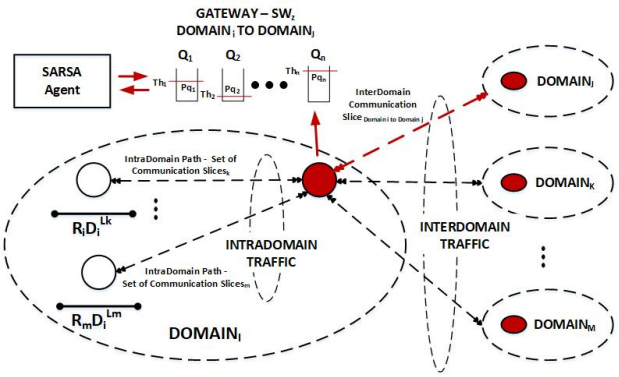


Figure 4: Interdomain Communication Slice and Gateway at Domain i

The model is agnostic to the issue of traffic distinction between packets generated with the slices already instantiated (slice operation) and packets generated by the network slicing management software installed (orchestrator, resource marketplace, monitoring, others).

The slicing-related interdomain traffic between domains is handled by an SDN switch as illustrated in Figure 3.

In summary, the interdomain traffic at the gateway is composed of the packets generated (operation and management) by all resources belonging to the domain D_i having as destination the domain D_j.

The slicing communication model assumes that domains have only one network connection among them. In other words, the domains do not act as intermediate domains switching packets in the path to a destination domain.

For the interdomain packets at the gateway, the following definitions hold (Figure 3):

- All packets belonging to a set of resources $R_i^{D^{l_i}}$ at domain D_i with the same performance parameters constraint use a specific queue Q_n;
- N switch queues handle the packet generated by the shareable resources at domain D_i;
- The switch queues have SDN resources control capabilities controlled by SDN Controllers [21] for resource control;
- A priority is assigned to each output queue; and
- Each queue has a threshold level control parameter P_{Q_n}.

The priority and threshold level assigned to the queues are used to support optimization (e.g. optimization controller as shown in the following section).

In summary, the model assumes that packets generated from any sliced resource with similar performance constraints are grouped in the same controlled queue in the gateway.

The following hypotheses are considered for the control of the intradomain packets and the gateway queues as highlighted in Figures 2 and 4):

- Intradomain communications will be based on existing underlying communication technologies (MPLS LSPs connections, EON fiber slots, other);
- A gateway handles all the inbound and outbound interdomain traffics;
- In a domain, each node hosting sharing resources for the slicing creates a path to the gateway; and
- Each path associated with a resource provided by a node is associated with a particular queue in the gateway.

The intradomain slice communication analytical model is not the focus of this paper, and these premises make clear its interrelation with the interdomain modeling and allows the independent modeling of it.

The optimization problem to solve here is the sharing of the communication resources between the different slices taking into account the performance requirement (e.g.: QoS) of each slice. This means scheduling the packet originating from the different slices towards the different available queues in the gateway. This a complex engineering problem that is difficult to solve in an analytical way considering all the parameters that need to be taken into account. For that, we propose to investigate the use of a Reinforcement Learning SARSA agent which is explained in the following section.

5 SARSA Agent to Optimize Resources Sharing

The interdomain communication slice model is now applied to the network slicing deployment setup illustrated in Figure 4 in which we have:

- A multidomain slicing infrastructure with n domains;
- A single communication slice between domains; and
- A SDN-capable switch (gateway) handling bidirectional interdomain packets between the domains.

In terms of the proof of concept, each interdomain communication slice has a reinforcement learning SARSA agent aiming to optimize the allocation of communication resources. The RL-SARSA agent acts during slice operation to dynamically keep performance parameters accordingly to management-defined objectives.

The interdomain slice communication parameters ($P\_RC^{l_i}_{D_k,D_j}$) are configured during the slicing commissioning phase and are dynamically adjusted by the SARSA agent during the slice operation phase.

5.1 SARSA Agent Model and Configuration

The objective of the SARSA agent is to control the queue flushing transmission rates to preserve the performance parameters defined by the manager while sharing unused resources.

The slice communication queues (Q_i) are configured as follows: i) Three queues corresponding to three performance parameters controlled by the agent; ii) Each configured queue threshold (Th_i) corresponds to the performance parameter assigned to the queue and served to packets generated by sliced resources with this requirement; and iii) Each queue Q_i has two states: below threshold (BT) and above threshold (AT).

The actions defined for the queues in the AT state are to increase the transmission rate, reduce the transmission rate, and do nothing. Each executed state/action has a defined reward.

The SARSA agent and communication slice parameters and initial conditions for running are as follows:

- Agent configuration parameters: i) Epsilon-greedy policy $\epsilon = 8\%$; ii) Learning rate $\alpha = 20\%$; and iii) Discount factor $\gamma = 80\%$
- Other parameters are: i) Threshold limit (triggers agent action) = 50%; ii) Agent actions: bandwidth increased or reduced by 10%; iii) the Maximum number of attempts = 500; and iv) Queue priorities are: $p1$, $p2$ and $p3$ with $p1 > p2 > p3$.

5.2 Implementation and tests

The simulation environment was configured on a Linux (Ubuntu 22.04.1 LTS) Intel(R) Core(TM) i5-3470 CPU @ 3.20GHz desktop. Visual Studio Code v.1.73.0 and Python v3.10.6 are used to execute the tests and the statistical analysis.

Each test run scenario has a minimum process cycle of 10^4 packet production for each queue with a Poisson distribution.

The SARSA agent is called each time any queue reaches its configured threshold. The SARSA agent processes up to 500 episodes in search of a new configuration of the flushing bandwidth distribution among queues to keep buffer occupation in the configured threshold limit.

5.3 The Slice Communication Evaluation Results

A series of tests have been undertaken. It aims to overload the queues to evaluate the behavior of the agent. The three defined scenarios are the following: i) Scenario 1 - One of the queues is overloaded; ii) Scenario 2 - Two queues are overloaded; and iii) Scenario 3 - All queues are overloaded.

The dynamics of the overloaded queues are configured as follows: i) First set traffic 30% above the queue's defined limit for 10 minutes; ii) Increase to 50% above its defined limit for additional 10 minutes; iii) Increase to 80% above its defined limit for additional 10 minutes, and iv) Increase to 100% above its defined limit for additional 10 minutes.

Figures 5a and 5b illustrate the SARSA agent's behavior for scenario one. Figure 5a plots the state of the queues while they are being saturated with overload traffic of packets. The queue transmission rate (flushing rate) configured by the SARSA agent is illustrated in Figure 5b. We observe that the total available bandwidth for the link is distributed and reconfigured among the queues according to the dynamic need to flush packets from a specific queue and keep queue occupation below the defined threshold.

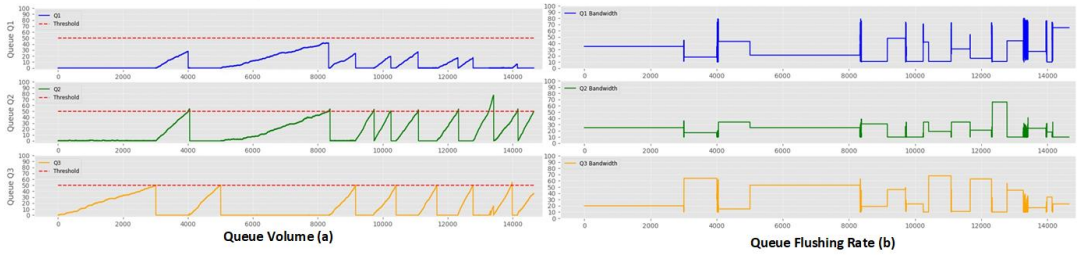


Figure 5: Test Scenario 1 - Queue Size and Transmission Flushing Rate

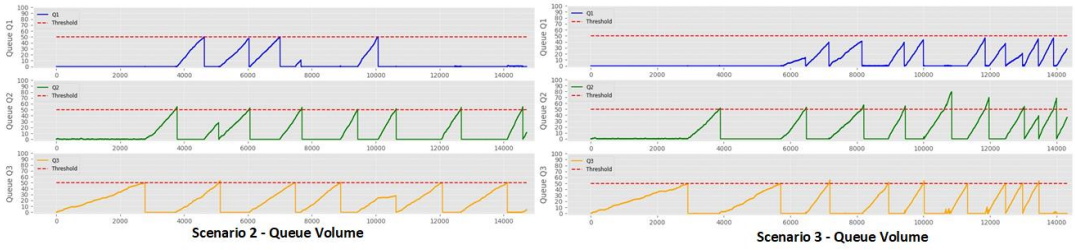


Figure 6: Test Scenarios 2 and 3 - Queue Size

For scenarios two and three, the behavior of the SARSA agent is illustrated in Figure 6. In scenario two, two queues may overload, and, as observed in scenario one, the SARSA agent reconfigures the queue's transmission rate to keep buffer occupation below the defined threshold. The agent can deal with simultaneous overload for the simulation-defined parameters by keeping queue occupation as required. The behavior of the SARSA agent for scenario 3 is equivalent to its behavior in scenario two.

6 Final Considerations

This paper presents a conceptual model of network slicing and present an analytical model to allocate communication resources between slide processes. The conceptual model is along with a SARSA agent that optimizes the allocation of communication resources among slices. The SARSA agent uses the conceptual model to formulate the required communication resources of each slice. A proof of concept implementation of the SARSA agent aims to demonstrate that the SARSA agent contributes to dynamically adjusting and controlling the slice communication parameters between domains. The proposed conceptual model demonstrates the feasibility and ease of handling different types of communication resources for optimizing the communication slice. Future work includes the leverage of the conceptual model with the integration of intradomain and interdomain models and the new formulation of the distributed optimization problem to solve by a federation of SARSA agents.

7 Acknowledgments

The authors would like to thank the University of Paris Saclay, University of Evry (IBISC Lab), the Instituto ANIMA, and the FAPESP MCTIC/CGI cooperation agreement under the thematic research project 2018/23097-3 for supporting this work.

References

[1] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and Orchestration; Concepts, Use Cases and Requirements. Technical Specification 3GPP TS 28.530 V15.0, 3GPP, 2019.

[2] 3GPP. 5G-Evolution-3GPP. Technical Report Release 16-17, 3GPP, 2020.

[3] I Afolabi, T Taleb, K Samdanis, A Ksentini, and H Flinck. Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions. *IEEE Communications Surveys Tutorials*, 20(3):2429–2453, 2018.

[4] A. A. Barakabitze, A Ahmad, R Mijumbi, and A Hines. 5G Network Slicing Using SDN and NFV: A Survey of Taxonomy, Architectures and Future Challenges. *Computer Networks*, 167:106984, February 2020.

[5] S Clayman, A Neto, F Verdi, S Correa, S Sampaio, I Sakelariou, L Mamatas, R Pasquini, K Cardoso, F Tusa, C Rothenberg, and J Serrat. The NECOS Approach to End-to-End Cloud-Network Slicing as a Service. *IEEE Communications Magazine*, 59(3), March 2021.

[6] G. N. Dias, J. F. Rezende, L N. Ciuffo, I. Machado, Flavio Silva, Tereza C. de Brito, F. Redigolo, Joberto S. B. Martins, Leobino Sampaio, and Antonio Abelem. SFI2 - Slicing Future Internet Infrastructures project. In *Proceedings of the The Global Experimentation for Future Internet*, pages 1–3, Coimbra, Portugal, November 2019.

[7] Gilvan Durães, Rafael Reale, Romildo Bezerra, Alexandre Fontinele, André Soares, and Joberto S. B. Martins. Evaluating the Applicability of Bandwidth Allocation Models for EON Slot Allocation. In *Proceedings of the IEEE International Conference on Advanced Networks and Telecommunications Systems (IEEEANTS)*, pages 1–6, India, December 2017.

[8] ETSI. Mobile Edge Computing A key technology towards 5G. Technical Report WP No. 11, European Telecommunications Standards Institute, September 2015.

[9] X Foukas, G Patounas, A Elmokashfi, and M K Marina. Network Slicing in 5G: Survey and Challenges. *IEEE Communications Magazine*, 55(5):94–100, May 2017.

[10] ONF Open Networking Foundations. Applying SDN Architecture to 5G Slicing. Technical Report TR-526, ONF - Open Networking Foundations, 2016.

[11] P Gouvas, A Zafeiropoulos, C Vassilakis, E Fotopoulou, G Tsiolis, R Bruschi, R Bolla, and F Davoli. Design, Development and Orchestration of 5G-Ready Applications over Sliced Programmable Infrastructure. In *29th International Teletraffic Congress (ITC)*, volume 2, pages 13–18, September 2017.

[12] IETF. Framework for IETF Network Slices. RFC- Request for Comments draft-ietf-teas-ietf-network-slice-framework-00, Internet Engineering Task Force, March 2021.

[13] Telecommunication Standardization ITU-T. Framework of Network Virtualization for Future Networks. Technical Report ITU-T Y.3011, ITU-T, January 2012.

[14] Alexandros Kaloxylos. A Survey and an Analysis of Network Slicing in 5G Networks. *IEEE Communications Standards Magazine*, 2(1):60–65, March 2018.

[15] I Kovacevic, A Shafigh, S Glisic, B Lorenzo, and E Hossain. Multi-Domain Network Slicing With Latency Equalization. *IEEE Transactions on Network and Service Management*, 17(4):2182–2196, December 2020.

[16] Joberto S B Martins, Tereza C. Carvalho, Flavio Silva, and Rodrigo Moreira. SFI2 Network Slicing Reference Architecture. Technical Report TR03/2022, Universidade de São Paulo - USP, 2022.

[17] J Nightingale, Qi Wang, J M. Alcaraz Calero, E Chirivella-Perez, M Ulbricht, J A. Alonso-López, R Preto, T Batista, T Teixeira, M J Barros, and C Reinsch. Qoe-driven, energy-aware video adaptation in 5g networks: The SELFNET self-optimisation use case. *Int. J. Distributed Sens. Networks*, 12(1), 2016.

[18] Jonathan Prados-Garzon and Tarik Taleb. Asynchronous Time-Sensitive Networking for 5G Backhauling. *IEEE Network*, 35(2):144–151, March 2021.

[19] K. Samdanis, X. Costa-Perez, and V. Sciancalepore. From Network Sharing to Multi-Tenancy: The 5G Network Slice Broker. *IEEE Communications Magazine*, 54(7):32–39, July 2016.

[20] P K Thiruvasagam, A Chakraborty, and C. Murthy. Resilient and Latency-Aware Orchestration of Network Slices Using Multi-Connectivity in MEC-Enabled 5G Networks. *IEEE Transactions on Network and Service Management*, 18(3):2502–2514, September 2021.

[21] Eliseu Torres, Rafael Reale, Leobino Sampaio, and Joberto S. B. Martins. A SDN/OpenFlow Framework for Dynamic Resource Allocation based on Bandwidth Allocation Model. *IEEE Latin America Transactions*, 18(5):853–860, April 2020.

[22] Shalitha Wijethilaka and Madhusanka Liyanage. Survey on Network Slicing for Internet of Things Realization in 5G Networks. *IEEE Communications Surveys Tutorials*, 23(2):957–994, 2021.

[23] Y. Xiao and M. Krunz. Dynamic Network Slicing for Scalable Fog Computing Systems With Energy Harvesting. *IEEE Journal on Selected Areas in Communications*, 36(12):2640–2654, December 2018.

[24] F. Yousaf, M. Gramaglia, V. Friderikos, B. Gajic, D. von Hugo, B. Sayadi, V. Sciancalepore, and M. Crippa. Network Slicing with Flexible Mobility and Qos/Qoe Support for 5g Networks. In *IEEE International Conference on Communications*, pages 1195–1201, May 2017.

[25] Shunliang Zhang. An Overview of Network Slicing for 5G. *IEEE Wireless Communications*, 26(3):111–117, June 2019.

Technical Session 5 (Short Papers)

Tooled Methods for the Simulation of Deployment of Safe Services in the Internet of Vehicles

Audrey Ahoukeng Donwoung[1], Nazim Agoulmine[2], and Bachir Djafri[3]

Paris Saclay University - Université of Évry Val d'Essonne - IBISC Laboratory
nazim.agoulmine@univ-evry.fr, dreyswoung@gmail.com, bachir.djafri@univ-evry.fr

Abstract

Cars are evolving into full mobile computers capable of executing advanced applications and communicating directly with each others using V2V (Vehicle to Vehicle) communications and with far end applications running in data centers or Cloud Computing using V2I (Vehicle to Infrastructure) communications. In this emerging context, it is now possible to envision innovative services such as information dissemination, safe driving, road accidents mitigation and so on that can be dynamically deployed in this infrastructure. The objective of this work is to investigate tooled methods for the simulation of such a new connected car environment in an exemplar services. The selected scenario is the waste collecting management service deployment in connected trucks operating in a city.

Keywords: Internet of Vehicles | Edge Computing | Cloud Computing | Transportation Services | Formal Modelling | SysML | SUMO Simulator | Distributed Systems.

1 Introduction

Nowadays vehicles have become essential items in daily life of people and in the management of big cities. The growing number of these vehicles on the roads increases the need to significantly reduce the traffics, pollution emission, risk of personal injury, material damage and so on. According to Transport and Environment organisation,[1], the almost a third of trucks tested emit dangerous levels of air pollution. Similarly, road safety situation report carried out by the World Health Organization (WHO), road accidents are responsible for 1.35 million[2] deaths every year worldwide. Many researchers have looked at the Internet of Vehicles (IoV) as one potential solution to help improving the situation. This concept originated to connect multiple users with vehicles to ensure safe and efficient road conditions through one communication from vehicle to any entity (V2X) which must be reliable and in real time.

Since vehicles are systems equipped with computing and communication capabilities and more and more edge computing infrastructure is being deployed (Edge Computing), vehicles can communicate directly together if close enough (V2V) and/or with the infrastructure (V2I) at any time. With their more and more increased computing capacity, they are also able to collect and process a large amount of information in real time (weather, road conditions, traffic conditions, etc.) coming from their embedded sensors or from the infrastructure [2]. However, since vehicles are mobile and move at variable speeds, the conditions in which they evolve are not fully predictable including their connectivity [5]. Hence some services can not be executed only onboard and therefore the vehicles need to interact with far-end applications, such as intelligent driving assistance, and high resolution maps creation services. Cloud Computing and Edge Computing are those technologies that will help to execute these applications and reduce

[1] https://www.transportenvironment.org

[2] 1.35 million: number from https://apps.who.int/iris/bitstream/handle/10665/277372/WHO-NMH-NVI-18.20-fre.pdf

the response times distributing the processing between the vehicles, the edges and the cloud computing. Thus, this system formed by connected vehicles, the edges and the cloud computing can be seen as a distributed system of fixed and mobile components, connected through an underlying communication infrastructure such as 5G [12].

In this paper, the objective is to model, analyze and validate a representative since that is a route optimization service for the collection, transportation and deposit of waste in a city. The aim is to formally design the waste management planning service and verify its operation and behaviour by simulation to increase confidence in the services before its real deployment. The used simulator is called SUMO (Simulation of Urban MObility) [10]. It is an open-source and highly portable simulator tool that allows modelling of intermodal traffic systems including road vehicles, public transport and pedestrians.

This paper is structured in 5 sections. Section 1 is a general introduction to the context. Section 2 presents the objectives of the paper. Section 3 presents the problem statement where we describe the system model and system properties formal modelling. Section 4 presents an implementation and an evaluation in SUMO. Section 5 is a performance evaluation after doing simulation tests in SUMO.

2 Related Works

Internet of Vehicles has been the topic of numerous contributions in the literature. Several surveys have been published [13] highlighting the challenges in the area in terms of infrastructures and mechanisms to support efficient services deployments and connectivity among vehicles and infrastructures and between vehicles [7]. More recently, more focused papers have been published addressing the interactions between the vehicles, the Edge and the Cloud Computing infrastructures (that corresponds to our considered context). For instance, in [9] authors present a framework to build advanced computing systems for autonomous vehicles. The paper addresses three categories of problems: cutting-edge computer system designs (real-time data processing as autonomous vehicles are mobile and often have power restrictions in term of energy), V2X applications (how vehicles cooperate with each other and with the infrastructure) and autonomous vehicle security (protecting autonomous driving on-board computer systems from attack). In [8] authors focus on the design of a web-based car monitoring solution that allows network vehicles to exchange their mobility information to get quick road awareness using the specification of the service vehicle information and vehicle signal specification. In [3], authors propose a solution to optimize waste collection and optimize vehicles service routes in real time based on signals sent by intelligent waste bins. In [11], authors investigate using artificial intelligence algorithms for solving the vehicle routing problem with time windows for a heterogeneous fleet of waste collection vehicles. Authors of [4] provide a contribution that aim to evaluate the performance and mobility of vehicular networks using the SUMO simulator used in this work. In [1], authors aim to reduce the delay and guarantee the delivery of messages in city environment where the density of vehicles is high. The authors introduce the concept of Dynamic Fog (DFCV) that consists on using the Fog Computing and Cloud Computing for the dissemination of messages.

All these works have been inspiring this contribution. However our focus in this contribution is different as it aims to address the specification of a vehicular service evaluate its behavior in term of adaptation in a dynamic context that is the collection of waste in a city.

3 Problem Statement

The research questions we aim to address in this paper are the following: How to model a vehicular service using a formal modelling language? How to model constraints in terms of mobility starting points and ending points (i.e. trucks parks)? And how to model constraints in terms of collecting point of waste containers?

The used method consists on specifying the different representative scenarios of the waste containers collection and identify the functional and non-functional goals (i.e. requirements) of such a service in the frame of these scenario. In addition, the method aims also to identify the constraints on the connected vehicles in the context of each scenario i.e. the context in which the waste collection planning service will be deployed (i.e. planning of paths of connected trucks in a city to collect waste containers) and verify whether the designed service is compliant with the identified requirements at run-time.

4 Proposed Solution

4.1 System Model

The used high level defined scenario is the following : A company is responsible of waste collection in a city. It has deployed a set of waste collection points in the city where a set of connected containers are installed. Citizens use these containers to deposit their waste over time. Containers could only be unloaded at specific time intervals by the company defined by the city regulation. For that, the company has a set of electrical and connected trucks that are used for collecting the waste. Trucks are parked in different sites in the city. Containers are also installed in different sites in the city. When they are filled above a certain predefined level (threshold), a message is sent by the containers embedded sensors to the back-end waste collection management service deployed in the cloud (via appropriate wireless network and protocol). At periodic time, this cloud service processes the messages and identifies the set of containers that need to be unloaded and their respective GPS positions. Next, this cloud service selects the set of trucks to use for the collection of waste and plan accordingly all their trips.

The proposed system model introduces three layers namely the connected vehicles layer (i.e trucks), the Edge layer and the Cloud layer:

- The three layers are connected together using network infrastructures. The Cloud Computing and the Edge layers are connected via fixed backbone network.
- The Edge layer consists on a set of road side computing and networking infrastructure to which vehicles (in the vehicular layer) can connect using wireless communications technologies (e.g. 802.11p).
- The terminal layer is composed of all the connected vehicles, connected waste containers, traffic lights, etc. These objects are responsible for the collection of data environment and transmit it to the edge layer so that data can be processed and stored.

The Cloud layer contains all services that are accessible by the connected vehicles. It is responsible for processing and storing the information received from each Edge node and sometimes returns a response following a request.

4.2 System Properties Model

The services offered by the distributed system (based on mobile and fixed components) in connection possibly with the safety of vehicles and the safety of passengers and pedestrians, may have requirements and constraints in terms of reliability and response time (so-called functional requirements) and so-called non functional requirements that result from the cooperation between distributed components of the system, fixed (Edge and Cloud) and mobile (vehicles).

Non functional requirements are not directly related to the services offered by the system, but can be nevertheless critical for the proper functioning of the system. These requirements may be related with time and/or space: V2I and V2V communication constraints (e.g. delay, speed), availability of processing resources (e.g. CPU, memory), in relation to arrivals and departures of vehicles (e.g. mobility model, breakdowns). These requirements may also be neither temporal nor spatial, but simply resulting from a failure or unavailability of resources in the infrastructure. Therefore, these non-functional requirements may no longer be satisfied (temporarily, partially or totally) over time.

In order to model the system properties, we will base on specific use-cases to identify them. These properties include the functional and the non functional requirements. Once these requirements are identified, we will derive sub-goals from these goals which will identify sub properties of the system. The process will stop when we'll know that these properties are either intrinsic properties of the vehicles, network, edge computing or cloud computing infrastructure or these are properties that should be provided by the cloud services or the connected vehicles controllers.

In the following, we present the different considered scenario to identify these system properties and model them formally using SysML modeling language [6]. SysML supports all the cycle of software engineering from the specification, analysis, design, to the verification and validation of systems and systems-of-systems design. This language is chosen in this work for its high capacity to model any engineered system and also because it is a full ISO standard since 03/2017 [6].

The requirements of the system have been captured in Papyrus SysML and specified as shown in figure 1. We mention in the diagram to verify that certain requirements are indeed satisfied by blocks from the block diagram. To also make a verification using the activity diagram.

The block diagram describes the architecture of a system as a block which can be decomposed recursively into other blocks to the basic elements of the structure. The structure of system which resulting is therefore a hierarchy of blocks. The blocks are linked together by relation which can mainly be either a composition or an aggregation relation. Our system is essential made up of vehicles, wastes containers, edges nodes and Cloud, hence the use of composition link to connect our different blocks. The wastes containers are equipped with an electronic card and sensor for their operation. However we will not enter into the composition of the vehicle. In each block we define the different variables, operations to be performed and if necessary the constraints.

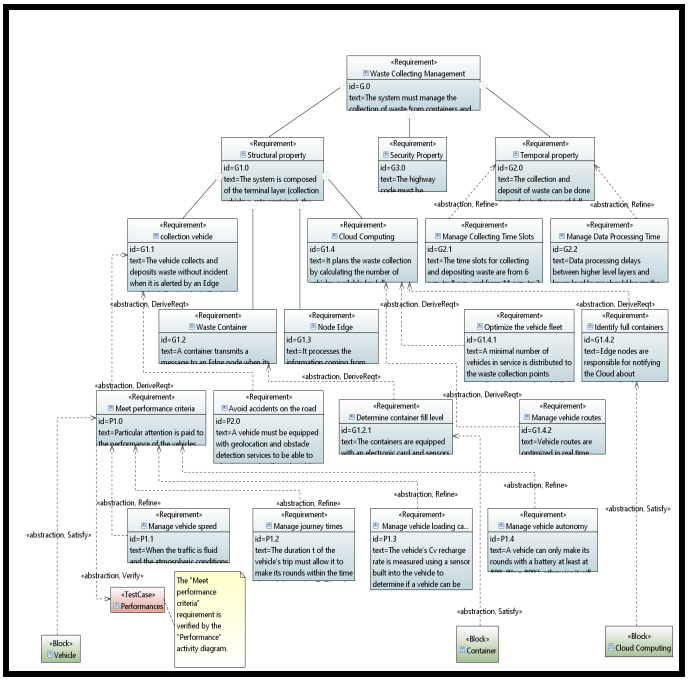


Figure 1: SysML Requirements Diagram Specification

5 Preliminary Implementation and Evaluation

We have implemented these scenarios and corresponding use cases in the SUMO simulator (Simulation of Urban MObility).[1] In this tool, each vehicle is individually simulated and has its own route. We have then used SUMO version 1.14.1 for Windows and OpenStreetMap[2] to create a map of the City of Evry (in France), convert it in a format understood by the simulator using **netconvert tool** and finally imported it in SUMO. Next, We have added additional objects (polygons) in the map representing mobility constraints (e.g building, rivers, etc.)[3].

6 Conclusion and future work

In this paper, we presented a work which objective was to formally model a innovative service in the context of connected vehicles. The chosen services is the waste collecting management service in a city and the chosen modeling language is SysML.
We have conducted all the phase of the proposed tooled method from the formal specification of the service towards it simulation in a realistic urban mobility simulator that is SUMO. The learned lesson is that this global process is very important for the specification, development and test of innovative services in the context of Internet of Vehicles to understand their functioning and make them more optimized and safe.

[1] https://www.eclipse.org/sumo/
[2] https://www.openstreetmap.org
[3] https://sumo.dlr.de/docs/Networks/Import/OpenStreetMap.html

Our future work is to advance the modelling part by introducing more artefact in the formal models related to this ecosystem and also enhance the capaV2V communications that have not been considered in this work. These V2V communication can be of paramount benefit to this future ecosystem to leverage the provided services.

References

[1] Mohammed Atiquzzaman. Dfcv: A novel approach for message dissemination in connected vehicles using dynamic fog. In *Wired/Wireless Internet Communications: 16th IFIP WG 6.2 International Conference, WWIC 2018, Boston, MA, USA, June 18–20, 2018, Proceedings*, volume 10866, page 311. Springer, 2018.

[2] Antonia Bertolino, Antonello Calabro, Felicita Di Giandomenico, Giuseppe Lami, Francesca Lonetti, Eda Marchetti, Fabio Martinelli, Ilaria Matteucci, and Paolo Mori. A tour of secure software engineering solutions for connected vehicles. *Software Quality Journal*, 26(4):1223–1256, 2018.

[3] Xinghan Chen, Zhihan Lv, Ruichang Li, and Shanshan Fan. Optimization of classified municipal waste collection based on the internet of connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*, PP:1–10, 04 2020.

[4] Francisco Edigleison da Silva Barbosa, Francisco Ferreira de Mendonça Júnior, and Kelvin Lopes Dias. A platform for cloudification of network and applications in the internet of vehicles. *Transactions on Emerging Telecommunications Technologies*, 31(5):e3961, 2020.

[5] Rejab Hajlaoui. *Résolution à base d'heuristiques du problème de routage dans les réseaux ad hoc de vehicules*. PhD thesis, Bourgogne Franche-Comté, 2018.

[6] Information technology — Object management group systems modeling language (OMG SysML). Standard, International Organization for Standardization, Geneva, CH, March 20@articlearticle, author = Chen, Xinghan and Lv, Zhihan and Li, Ruichang and Fan, Shanshan, year = 2020, month = 04, pages = 1-10, title = Optimization of Classified Municipal Waste Collection Based on the Internet of Connected Vehicles, volume = PP, journal = IEEE Transactions on Intelligent Transportation Systems, doi = 10.1109/TITS.2020.2981564 17.

[7] Moez Jerbi. *Protocoles pour les communications dans les réseaux de véhicules en environnement urbain : routage et geocast basés sur les intersections*. PhD thesis, 2008. Thèse de doctorat dirigée par Agoulmine, Nazim Informatique Evry-Val d'Essonne 2008.

[8] Junhee Kwon, Bien Aime Mugabarigira, and Jaehoon Paul Jeong. Web-based car control and monitoring for safe driving of autonomous vehicles. In *2022 International Conference on Information Networking (ICOIN)*, pages 435–440. IEEE, 2022.

[9] Shaoshan Liu, Liangkai Liu, Jie Tang, Bo Yu, Yifan Wang, and Weisong Shi. Edge computing for autonomous driving: Opportunities and challenges. *Proceedings of the IEEE*, 107(8):1697–1716, 2019.

[10] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using sumo. In *The 21st IEEE International Conference on Intelligent Transportation Systems*, pages 2575–2582. IEEE, November 2018.

[11] Piotr Nowakowski, Krzysztof Szwarc, and Urszula Boryczka. Vehicle route planning in e-waste mobile collection on demand supported by artificial intelligence algorithms. *Transportation Research Part D: Transport and Environment*, 63:1–22, 2018.

[12] Abdennour Rachedi, Toufik Ahmed, and Mohamed Mosbah. End-to-end network slicing for its-g5 vehicular communications. pages 266–271, 06 2021.

[13] Xu Xiaolong, Li Haoyuan, Xu Weijie, Liu Zhongjian, Yao Liang, and Dai Fei. Artificial intelligence for edge service optimization in internet of vehicles: A survey. *Tsinghua Science and Technology*, 27(2):18, 2022.

Intelligent Orchestration of Containerized Applications in a Cloud Infrastructure (Short Paper)

Julie Farkouh[1], Audrey Ahoukeng Donwoung[1], and Nazim Agoulmine[1]

Paris Saclay University - Université of Évry Val d'Essonne - IBISC Laboratory
julie.farkouh@etd.univ-evry.fr, audrey.ahoukeng-donwoung@etd.univ-evry.fr, nazim.agoulmine@univ-evry.fr

1 Introduction

Resource contention and performance interference between co-located applications can be a serious problem in a multi-tenant infrastructure. This problem is difficult to address by existing orchestrators and may cause applications performance downgrade, extra maintenance costs, as well as service-level agreement (SLA) violations. Therefore, a refined and robust container orchestration system is the key factor in controlling overall resource utilization, energy efficiency, and application performance. However the highly dynamic nature of modern applications, diverse features of cloud workloads and environments to take into account considerably raise the complexity of orchestration mechanisms[3]. One particular orchestration problem is the deployment of containers in appropriate hosts with the right required resources as resources play a major role in the containerized applications performance. This problem is not fully taken into account by modern orchestrator since the type of application and the deployment mode are not known to the orchestrators and therefore it is not able to determine the appropriate required resources.

Deploying Applications or Websites that are in production mode is an important concern for many companies willing to externalize their Applications and services in the Cloud Computing. Since allocated resources play a major role in their performance and the end-users satisfactions, determining the most appropriate level of resources is of paramount importance. Today mainstream deployment is using containers and orchestrator such as Kubernetes. The concerns is the difficulty to identify the required level of resources to allocate and maintain to these applications during their life cycle.

To overcome this problem, we propose to artificial intelligence and more precisely Decision Tree to predict the applications types and modes to take into account this information during orchestration. We adopt in this work a use case where the deployed applications are containerized Web Applications (CWA) and containerized DataBase Applications (CDA) which could be in test mode or production mode. The orchestrator that receives this information can deploy therefore deploy them in the most appropriate nodes with the most appropriate resources (CPU/Memory/Network) and achieve better resource allocation decision and performances.

Some related works on this issues are presented here. In [1], Ishak et al. proposed an intelligent scheduler in Kubernetes environment using ML based on the execution prediction of the applications and the nature of the processed data in order to better select the appropriate device, CPU or GPU. In [2] , Jose et al. proposed Network-Aware scheduler for container-based applications in Smart City deployments that added the parameters of network quality to the calculation score of the Kubernetes scheduler in order to take it into account when selecting the appropriate deployement node.

2 Problem Statement

Kubernetes which is one of the most popular orchestrating tools for containers. It is an open source scheduling platform that automates Linux container operations: Distribution, Replication, Load-balancing, Availability, Higher-level interfaces to composition features and Docker which is an open source containerization platform - with standardized executable components that combines application source code with the operating system(OS), libraries and dependencies that are required to run that code in any environment.

Kubernetes is doing a great job in scheduling, deploying and provisioning containers based on the user-requested RAM and CPU for their containers in proportion to the free RAM and CPU available in the infrastructure. Unfortunately, those are not only the resources that should be taken into account when deploying containers as other resources may play an important role in the application performance based on what it really needs such as Network resources and Disk Input and Output.

Eventually, these addition requirements should be taken into account during orchestration and appropriate related resources should be computed based on the type of application container service and the context in which they are deployed. If applications are deployed in production mode, more attention to all resources needs to be addressed to cope with the potentially large amount of service requests. Whereas deploying an application in a development mode for a single user developer is not that important as there is no online demands and as a result, high performance requirements are not necessary. Unfortunately, Kubernetes Scheduler that it is responsible for scheduling and deploying the containers into the nodes does not take into account these parameters (e.g. Network Quality , I/O ..) and the mode in which applications are deployed (development, stage, production). As a result, the deployment solution could be inappropriate resulting in bad performances.

3 Proposed Solution

As a solution, we introduce a model which uses ML to automatically predict the type of containerized applications to execute and their corresponding mode for deployment. Once predicted, the solution triggers the appropriate deployment of the applications in the appropriate nodes of the containerized infrastructure using an orchestrator (in our case Kubernetes).

In our model, we have addressed two types of applications: Containerised Web-based Applications (CWA) and Containerised Database Applications (CDA) in production mode. The rational behind the selection of these two applications is that there are the most commonly deployed industry grade applications.

Therefore, the ML Model will predict the Container Type and Mode and based on these predicted values, the model will assign the container to the appropriate computing node. This means that if the ML Model predicts that the container is for a Web application Service in a production mode, it will automatically assign and deploy it into one of the nodes with the highest network speed interface and if the predicted value is database container in production mode, then it will assign it to one of the nodes with the highest available RAM capacity.

During the prediction phase, the model reads the desired containers configurations and

extracts the features that correspond to their type and deployment mode. These features are then introduced in a previously trained ML Model that uses classification learning to classify the containers based on their type and mode.

Regarding the deployment phase, the proposed solution first searches for the nodes with the highest network interface speed from all the available nodes in its cluster to host the Web-application containers (since in production mode, this feature is one of the most important). Similarly, it searches for the node with the highest available RAM capacity to host to the Database Server containers.

- **Container Image** and **Port** that the container will use – these two features are the most important indicators to predict the application type and the ones commonly used. The used images are the ones available in the docker hub - official website of the dockers images - and widely known for web services like Apache, Nginx or IIS and the relevant ports would be also the generally the most used one: 80, 443, 8080 . For Data Base Service, the name of the images that are the most used are Mysql, Oracle or Postgress Sql and corresponding ports are usually 3306, 1521.
- **Name** of the Container, as in IT, we should always use as a best practices in naming process, names that indicate clearly the service to help maintenance such as, logging, debugging, auditing, etc. For the web application, it could vary between the following names "Websites", "Front ends", "Webs". Similarly, for the Database application, names could be "Database", "MyDataBase", "DB" in repeated way like for web services.

  RAM we assume in our work the RAM would start from 256/512 M bytes of RAM for the development mode to up to 4 G bytes of RAM for applications in the production mode..

These are the features we took into account to train our ML Model to classify them into two types of outputs (Web or database) and two predicted environment mode outputs (Production or Development).

4 Implementation and Experimentation

For our Deployment, we first trained our model in Matlab Classification App – Analysis Trees on simple Dataset of the Web and Database services properties previously introduced based on containers' image, Port, Name indicating reserved RAM. We then exported our Model into Matlab Production Server in order to use it for production. The provided interface is a Rest API.
Next, we used Kubernetes configured into one master node and two Worker nodes with different RAM Capacities and Network Interface Speed.

Worker Node1 = 8 Gg RAM , 10,000 Mb/s Traffic Speed
Worker Node2 = 10 Gg RAM, 1000 Mb/s Traffic Speed

After training the ML Mode, the sequence of performed tasks are the following:

1. Our Model first finds the node wirh the highest speed interface which is worker-node 1 and label it "prodweb" as for Webservices in Production.

2. It also finds the node with the highest RAM Capacity Node which is worker-node 2 and labels it “proddb” for DataBase Production
3. Then it reads the configuration file of the pod configuration containing the container specifications (YAML File) and extracts the features of the containers and send them via a HTTP Request API to our trained ML Model, The predicted values of the ML Model would be one of three cases:
   - **Containerised Web-based Applications (CWA)** in **Production Mode**.
   - **Containerised Database Applications (CDA)** in **Production Mode**.
   - **CWA** or **CDA** in **Development Mode** or Other application types in **Production / Development** Mode but we didn't follow-up thi czse in the deployment case we did not handle the deployment, just the first two first.
4. The pod configuration file is then updated to launch the the deployment in the corresponding node.

We tested the accuracy of our ML Model. Since our data was simple and the output is only composed of three possible values, the model achieved 93% accuracy.

We tested also the robustness of the prediction with general values in the dataset. For example, we changed the names to some unrelated strings (that do not contain the pattern) to the application while keeping the other values as it in the datase (e.g. the Ram to a lower/higher values than from the ones in our dataset , Changing two features like the Name and Image. Despite these changes the ML was still able to predict correctly the type and mode, thanks to the Decision Tree (DT) approach.

5 Limitations and Future Works

In our work we presented a simple Model which is able to predict the containerized application type and mode based on features contained in the configuration file. Based on this prediction, the scheduler aims to deploy these applications in the most appropriate nodes in the cluster. The prediction worked well with the type and mode, but we aim to extend it to address other type of services like Domain Name Service (DNS), Mail Services, Virtual Private Networks services (VPN), etc. However, this work has several limitations we plan to overcome. Our scheduler state is considered in an initial status with no consideration for other potentially running containers competing for the same computing and networking resources. Indeed, the solution identifies the nodes with the highest network interface and RAM capacity without considering the current bandwidth or RAM utilization. A possible approach would be to combining our model with the native Kubernetes scheduler score tool that it used in node selection i.e. using the result as factor to computing the node selection in a pod, similarly to the work presented in Jose et al.[2]. we envision also to expand our dataset to predict several more containerized applications types and have more real accuracy results for better orchestration mechanisms based on the container application type.

6 Conclusion

In this work, we presented a ML based scheduler which relays on ”Type” and ”Mode” Awareness of containerized applications for better resources allocation and node selections decisions. With

this solution, it is possible to detect the nature of applications: Web Containerized Applications and Data Base Containerized Applications and their production mode (test, production). This information is used to improve the allocation of resources (Network/RAM). We used two popular open-source projects that are Docker and Kubernetes, to validate our model which helps to identify additional needed resources beyond RAM and CPU (that are used in the original Kubernetes scheduler). We implemented a poc of our model a separate scheduler. Our future works is to integrated our model with Kubernetes scheduler and allow is to be aware of more information to perform a better scheduling and deployment .

References

[1] Ishak Harichane, Sid Ahmed Makhlouf, and Ghalem Belalem. A proposal of kubernetes scheduler using machine-learning on cpu/gpu cluster. In *Computer Science On-line Conference*, pages 567–580. Springer, 2020.

[2] Jose Santos, Tim Wauters, Bruno Volckaert, and Filip De Turck. Towards network-aware resource provisioning in kubernetes for fog computing applications. In *2019 IEEE Conference on Network Softwarization (NetSoft)*, pages 351–359. IEEE, 2019.

[3] Zhiheng Zhong, Minxian Xu, Maria Alejandra Rodriguez, Chengzhong Xu, and Rajkumar Buyya. Machine learning-based orchestration of containers: A taxonomy and future directions. *arXiv preprint arXiv:2106.12739*, 2021.

Comparative Analysis of Service Mesh Platforms in Microservices-Based Benchmark Applications

Francisco A. A. Gomes[1], Paulo A. L. Rego[2], Fernando A. M. Trinta[2], and José N. de Souza[2]

[1] Federal University of Ceará (UFC), Crateús, CE, Brazil
almada@crateus.ufc.br
[2] Group of Computer Networks, Software Engineering and Systems (GREat)
Federal University of Ceará (UFC), Fortaleza, CE, Brazil
pauloalr@ufc.br, fernando.trinta@dc.ufc.br, and neuman@ufc.br

Abstract

The development of monolithic systems brings several challenges related to system maintenance and scalability. To mitigate this problem, there is currently a trend in developing systems to use the composition of microservices. The application is divided into independently deployable services, which can quickly implement on any infrastructure resource, and each service runs in an isolated process. Furthermore, the development of technology for microservices also increases the operational complexity associated with modern applications. The service mesh is a promising approach to mitigate this situation, introducing a dedicated infrastructure layer over microservices without imposing modifications to the service implementations. The proposed work aims to present a survey about service mesh platforms and application benchmarks based on microservices and performance tests performed. As a result, tests show that despite the benefits of the service mesh, it impacts application latency.

1 Introduction

The development of monolithic systems brings several challenges related to the maintenance and lack of autonomy of the development teams to adopt new technologies that differ from the adopted architecture [1]. There is currently a trend in systems development to use the composition of microservices, that is, small pieces of cohesive and autonomous software. With the microservice architecture, the application is divided into independently deployable services that can be rapidly developed on any infrastructure resource, and each service runs in an isolated process (container) [3]. Additionally, these services communicate through an HTTP API[1]. Thus, this architectural style allows developers to decompose software into small units, facilitating the scalability of only the most demanded services at that moment, which strongly contrasts with the architectures of corporate systems that are usually a single implementable component.

Several studies point to challenges related to microservices, for example, methodologies and tools for multilayer monitoring, functional adaptation algorithms at runtime, advanced adaptation features at runtime, and high availability support [2]. Furthermore, developing technology for microservices also increases the operational complexity associated with modern applications. This led to the emergence of service mesh, a promising approach to mitigate this situation, which introduces a dedicated infrastructure layer on top of microservices without imposing modifications on service implementations.

[1] Application Programming Interface

The proposed work aims to present a survey about service mesh platforms and its application in a benchmark of applications based on microservices. The benchmark works as a reference to an application that has several microservices related to each other and reflects the idea of a corporate application. Furthermore, the work carried out a performance experiment to analyze the impact of these platforms.

The rest of this article is organized as follows: Section 2 presents the background: microservices, service mesh and platforms. Section 3 presents the results of experiments carried out. Finally, Section 4 concludes the article and exposes possible future works.

2 Background

The microservices emerged empirically from architectural patterns used in the real world, where systems are composed of services that collaborate to achieve their goals, communicating through lightweight mechanisms (e.g., Web APIs) [3]. The idea of microservices is to build small applications, developed independently, which tend to present efficient processing and interoperability aspects, allowing the continuous deployment/delivery of large and complex applications [5]. Given these characteristics, it is possible to implement each microservice with a different technology *stack*. According to Zimnermann (2017) , microservices are a way of implementing and deploying services in SOA using state-of-the-art software engineering practices (i.e., development and deployment paradigms and technologies).

A microservice that communicates with other services incorporates business logic and network communication logic. Each microservice contains a significant part of its network communication-related code, independent of the service's business logic. Implementing functionality related to service-to-service communication from scratch is costly [4]. Instead of focusing on business logic, the developer needs to spend a lot of time creating service-to-service communication functionality. Service mesh is a communication infrastructure between services. With a service mesh, a given microservice will not communicate directly with the other microservices. Instead, all service-to-service communication occurs in a software component called a Proxy sidecar. The sidecar is a software component co-located with the service in the same Virtual Machine or pod (Kubernetes). The Proxy sidecar layer is known as the Data Plane. All these Proxies sidecars are controlled via a control plane. This is where all settings related to inter-service communications apply.

Service mesh provides built-in support for networking functions like resiliency, service discovery, etc. Therefore, service developers can focus more on business logic, while most of the work related to network communication is offloaded to service mesh. The microservice for sidecar communication always takes place via standard protocols such as HTTP1.x/2.x, gRPC, etc. Thus, service mesh is a technology and language-independent. Regarding the control plane, all service mesh sidecars are centrally managed by a control plane. This is useful when supporting service mesh features such as access control, observability, service discovery, etc. All changes made to the control plane are sent to the sidecars.

Several platforms offer *service mesh* for microservice applications. This section will introduce two of them: Istio and Linkerd. Istio[2] is an *open-source* platform for connecting, managing, and securing microservices. It provides an infrastructure for communication between microservices, with resiliency, routing, load balancing, service-to-service authentication, observability, and more, without requiring any changes to the service code. By deploying Sidecar, the developer can add the service to Istio's *service mesh* . Deploying Istio is deeply tied to Kubernetes, but

[2] https://istio.io/

deploying it to other systems is also possible. Linkerd[3] is a *service mesh* for Kubernetes. It makes running services easier and safer by offering run-time debugging, observability, reliability, and security, all without requiring changes to source code. Linkerd is fully open source, licensed under Apache v2, and is a graduate project of the Cloud Native Computing Foundation.

3 Experiments and Results

To verify the use of service mesh platforms and to analyze the impact of these platforms on benchmarks applications in microservices, the work carried out a performance experiment taking into account the latency of a request.

For the experiment, a software called Apache JMeter[4] was used, a tool that allows the performance of load and stress tests in Web applications. The tests were performed on a machine with the following configuration: a node with a 4th generation 1.8 GHz i3 processor, with 2 cores, 8 GB of RAM, 128 GB of SSD, running on the Mint 20 Operating System. In addition, Kubernetes version 1.21 was installed.

For the present research, the test consists of analyzing the Teastore application[5] in three scenarios: with Kubernetes only, Kubernetes with Istio, and Kubernetes with Linkerd. With JMeter, the Teastore application was submit tests considering the return of stored data and calculating the average latency. Each scenario was subjected to different request rates: 600, 1200, and 2400 requests per minute. 10 users were used to running each scenario in concurrency at each request rate.

Figure 1 presents the results of the tests carried out. Baseline is the Kubernetes-only scenario. The obtained results show a better performance of Linkerd compared to Istio in all request rates. When verified at the rate of 2400 requests per minute, the increase in latency was 12.9% comparing the two platforms (Linkerd - 482,2 and Istio - 426,8). Istio even showed better results in some tests than Linkerd, but its overall average was impaired due to the high incidence of errors during the execution of the tests. This can be explained by the fact that Istio has a higher CPU and memory consumption than Linkerd. Another result found is about the impact of service mesh platforms. When comparing Istio with Baseline, the average latency value is up to 6.15 times higher when subjected to a rate of 2400 requests per minute (Baseline - 67,35 and Istio - 482,2).

4 Conclusion and Future Works

The article presented a survey about relevant topics, such as microservices, service mesh and tools, benchmark application, and a performance test performed. Important concepts and definitions for understanding the work were presented, as well as tools that can be used to carry out analyzes in microservices applications. Service mesh brings advantages to the developer because instead of focusing on business logic, it was necessary to spend time creating service-to-service communication functionalities, and with the service mesh is no longer necessary. The results showed that despite the benefits of service mesh, it impacts the latency of applications. The conclusion is that, in terms of performance, Linkerd is better than Istio, which is explained by its lighter system, that consumes fewer resources. However, there are still reasons why a developer might prefer to use Istio, more popularity and greater community support.

[3] https://linkerd.io/
[4] https://jmeter.apache.org/
[5] https://github.com/DescartesResearch/TeaStore

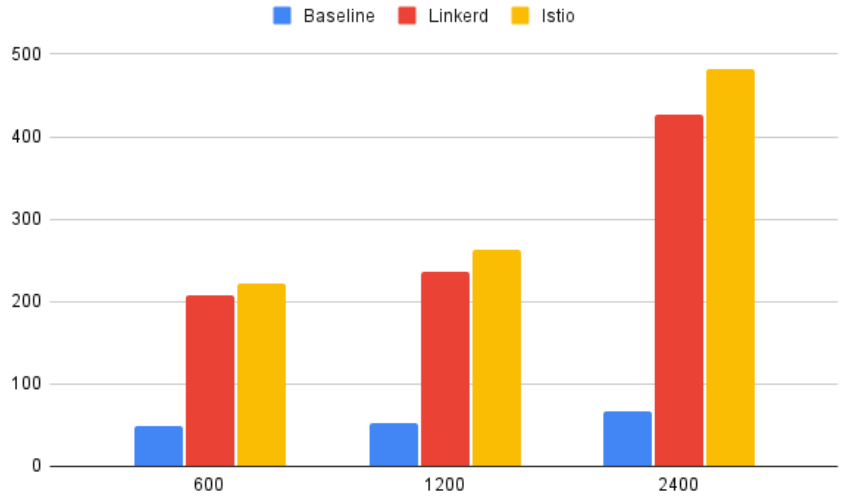


Figure 1: Comparative Analysis of Tests

As future work to improve this research, it is intended to carry out:

- Tests in other benchmarks in order to verify if the same behavior remains in relation Kubernetes, Linkerd, and Istio;
- Use another Service Mesh platform for analysis in performance tests; and
- Perform more performance tests and analyze other metrics (CPU, RAM, Disk)

References

[1] Keith H. Bennett and Václav T. Rajlich. Software maintenance and evolution: A roadmap. In *Proceedings of the Conference on The Future of Software Engineering*, ICSE '00, page 73–87, New York, NY, USA, 2000. Association for Computing Machinery.

[2] Emiliano Casalicchio and Stefano Iannucci. The state-of-the-art in container technologies: Application, orchestration and security. *Concurrency and Computation: Practice and Experience*, 32(17):e5668, 2020. e5668 cpe.5668.

[3] Martin Fowler and James Lewis. Microservices. 2014.

[4] Kasun Indrasiri and Prabath Siriwardena. Service mesh. In *Microservices for the Enterprise*, pages 263–292. Springer, 2018.

[5] Sam Newman. *Building microservices: designing fine-grained systems.* " O'Reilly Media, Inc.", 2015.

[6] Olaf Zimmermann. Microservices tenets. *Computer Science-Research and Development*, 32(3-4):301–310, 2017.

Blockchain Security in the Internet of Things: Literature Review

Joyce Quintino[1], Carina T. de Oliveira[2], and Rossana M. C. Andrade[1]*

[1] Federal University of Ceara - UFC - Fortaleza, Ceara, Brazil
joycequintinoalves@alu.ufc.br, rossana@ufc.br
[2] Federal Institute of Ceara - IFCE - Fortaleza, Ceara, Brazil
carina@lar.ifce.edu.br

Abstract

The Internet of Things (IoT) allows different devices in our daily routines, such as refrigerators, light bulbs and autonomous cars, to be part of the Internet. Accordingly, the amount of data circulating in the network increases. Thus, a secure and reliable IoT environment becomes essential to avoid security vulnerabilities. In IoT, resources are limited and existing security solutions used in the traditional Internet can be ineffective, making data security still a challenge. In this scenario, Blockchain emerges as a promising technology to improve security in IoT, as it enables the recording of data in a decentralized, encrypted and immutable way with the consensus of the network participants in a packet data structure called a block. In Blockchain, the consensus is a protocol responsible for maintaining the state of the network and allowing the addition of new blocks with data. However, many consensus protocols require high computational power and power consumption to solve the mathematical problems that make possible data records in the blocks. Since many devices have low processing power in the IoT environment, developing new consensus approaches is needed to use Blockchain in IoT effectively. Therefore, this paper presents the results of a literature review about the use of Blockchain technology, specifically consensus protocol approaches, in IoT security.

Keywords: Internet of Things; Security; Blockchain; Consensus Protocols.

1 Introduction

The Internet of Things (IoT) is a network of objects and embedded devices connected to the Internet [6]. IoT connects devices of our everyday life, such as refrigerators, light bulbs, wearables devices and autonomous cars, through the Internet. According to [3], the number of these devices will reach about 25 billion by 2025. Data that comes from IoT devices can be useful, for example, to connect patients to doctors, enable more efficient farming, aid decision-making in smart environments, optimize the industrial sector, and contribute to economic growth. So, the IoT has significant benefits that contribute to society.

On the other hand, as the amount of connected devices increases, so does the amount of data traversing the network and, thus, a secure and reliable IoT environment becomes essential to avoid security vulnerabilities [1]. However, the limitations of IoT devices in computing power, storage, and power consumption hinder the development of a secure IoT. Existing security solutions used in the traditional Internet can be ineffective [5]. In this way, IoT devices are subject to several security threats, considering generic vulnerabilities for Internet-connected devices and specific threats inherent to the technology of these devices [9].

In this scenario, Blockchain emerges as a promising technology to improve security in IoT, as it enables the recording of data in a decentralized, encrypted and immutable way with the consensus of network participants in a packet data structure called block [2]. Considering that many devices have low processing power in the IoT environment, there is a need to develop new approaches to consensus protocols, as the current versions are not suitable for IoT due to the high processing power required [7].

This paper presents the results of a literature review for using Blockchain technology, specifically consensus protocol approaches, to improve the Internet of Things security. We hope to bring to the community the state of the art about consensus protocols, the low-cost consensus approaches, applications and mechanisms, which have been developed to improve security in IoT environment, as well as the most frequent attacks related to Blockchain consensus, where consensus solutions are applied, and what tools are used for Blockchain simulations, development, and testing.

*CNPq's Productivity Scholarship in Technological Development and Innovative Extension - DT - Level 1D

2 Literature Review

Our literature review follows the precepts of a systematic mapping [4] and the PICOC (Population, Intervention, Comparison, Outcome, and Context) structure, which is a conceptual model to support the research questions [10]. Furthermore, the Parsif.al[1] tool was used to document the review process.

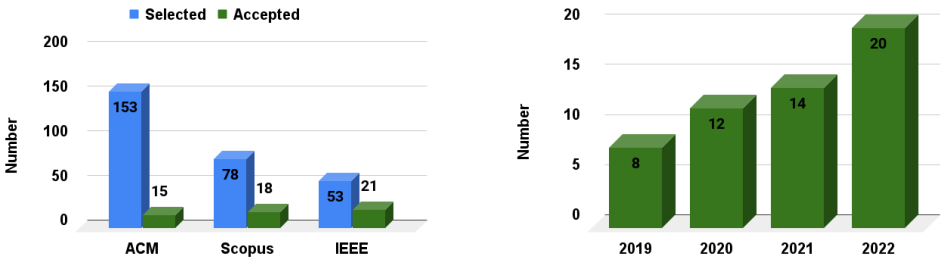
We selected three well-known digital libraries to find the relevant papers for this research: ACM Digital Library[2], Scopus[3] and IEEE Xplore[4]. In the research for the most relevant papers published in recent years, this review considered papers published from 2019 to October 2022. To find the papers that are part of this review, a search string was formulated and can be viewed on the github repository[5]. We defined five research questions (RQ) to address in the literature review that are presented in Section 3.

The selection process of the papers was initiated considering the defined inclusion and exclusion criteria. The inclusion criterias are papers about low-cost consensus used to improve security in blockchain-based IoT and english papers. The exclusion criterias are extended abstracts, unrelated papers about low-cost consensus used to improve security in blockchain-based IoT, secondary studies (surveys and systematic mappings) and papers published before 2019. Also, others two filters were considered to improve the selection of the papers. The first filter involves reading the papers' titles, abstracts, and keywords. The second filter is the complete reading of the papers to fill in the form with the research questions.

3 Results

Figure 1(a) shows the total number of papers selected from each base and the number of papers accepted at the end of the selection process. It is noticeable that most of the accepted papers are from IEEE Xplore. With the support of the Parsif.al tool, 35 duplicates papers were identified. As a way to validate the 35 duplicates, the titles of each paper were checked. After the duplicates were removed, 249 papers remained, which went through the remaining stages of the review. After the selection phases, 195 papers were rejected, and 54 were accepted.

Figure 1(b) presents the distribution of the accepted papers per year. Among the 54 accepted papers, 8 papers were published in 2019, 12 papers in 2020, 14 papers in 2021, and 20 papers in 2022. This result shows a growth in the number of publications in recent years on low-cost consensus protocols for improving the security of the Blockchain-based Internet of Things.



(a) By base

(b) By year

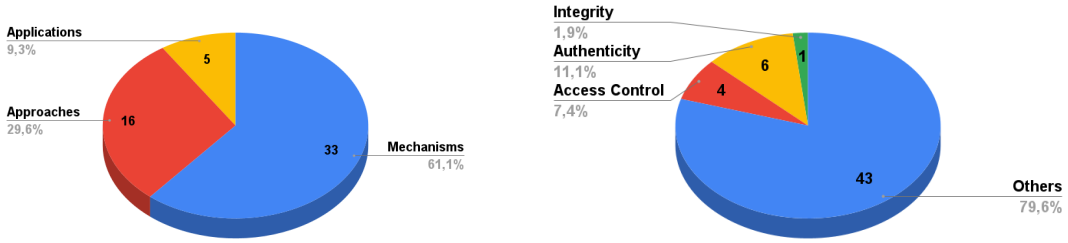
Figure 1: Number of accepted papers.

RQ1 - What are the approaches/applications or low-cost consensus algorithm/protocol mechanisms for improving security in the Blockchain-based Internet of Things? We categorized the research according to the focus of the paper's proposal. The categories defined are application, mechanism, or approach. The application category in this work means using a consensus protocol already known in the literature to test its performance or compare it with others in the IoT context. Mechanism represents new proposed consensus

[1] https://parsif.al/
[2] https://dl.acm.org/
[3] https://www.scopus.com/
[4] https://ieeexplore.ieee.org/Xplore/home.jsp
[5] https://github.com/JoyceQuintino/masterResearch/blob/literaturereview/searchstring

protocols. Finally, the approach represents a solution based on some consensus protocol modified for IoT. Among the 54 selected, 33 papers were classified as mechanisms, 16 were classified as approaches, and 5 were classified as applications. Figure 2(a) shows the number of approaches, applications, and mechanisms.

RQ2 - What is the main focus of the low-cost consensus approach/mechanism/application? Some consensus solutions found are aimed at integrity, authenticity, and access control. According to [8], integrity is defined as the property in which information has not been altered unauthorizedly. Authenticity is defined as the property of being genuine and capable of being verified and trusted. Access control is limiting and controlling access to systems and applications through links. According to our research, it was observed that 6 of the papers focus on Authenticity, 4 papers on Access Control, and 1 on Integrity. Also, 43 papers have this diversified response and were classified as Others. Figure 2(b) represents the division of the categories by solution focus. This part of the study shows a trend in research toward authenticity focused solutions.



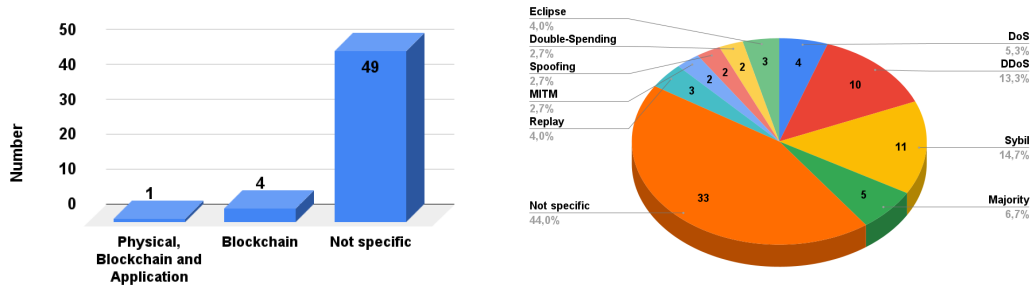
(a) Applications, approaches, and mechanisms

(b) Main focus of the solutions

Figure 2: Research Questions RQ1 and RQ2.

RQ3 - What level/layer is the low-cost consensus solution applied? This question identifies at which level/layer the consensus solution is applied. The results show that only 4 specified the layer. More specifically, 1 of the solutions applied at 3 different layers (Physical, Blockchain, and Application), and 4 papers focused on the Blockchain layer. The remaining 49 papers did not specify which layer they suit. Figure 3(a) shows how many papers have specified the layer.

RQ4 - What are the attacks associated with blockchain consensus? The results show that 21 papers cited one or more types of attacks, while 33 cited no types of attacks at all. The most frequent attacks are Denial of Service (DoS), Distributed Denial of Service (DDoS), Sybil, Majority Attack (*Majority*, also known as 51% Attack), Eclipse, Double-spending, Spoofing, Replay, and Man in the Middle (MITM). Twenty-one papers have mostly cited DDoS and *Sybil* attacks. In the future, a study based on these two types of attacks can be conducted to determine how they affect Blockchain-based IoT security. Figure 3(b) shows the frequency of the mentioned attacks.



(a) Level/layer of the low-cost consensus solution

(b) Frequency and types of attacks

Figure 3: Research Questions RQ3 and RQ4.

RQ5 - What are the tools used? It was observed that some of the papers used tools to test, implement, and validate the solution. Among the 54 accepted papers, 25 papers used one or more tools. To facilitate the presentation of the tools, we have grouped them according to their context of use. The tools are grouped into the following categories: blockchain platforms, simulators, Integrated Development Environment (IDE), software testing, network testing, and browser extension. Table 1 shows the classification of the tools according to these categories.

Category	**Tools**
Blockchain Platforms	Ethereum, Hyperledger Fabri, Hyperledger Sawtooth and Ganache
Simulators	NS-3, NS-2 and TeraSim
IDE	Remix IDE online
Software Testing	HPE LoadRunner and Caliper
Network Testing	Fabric test blockchain
Browser extension	Metamask

Table 1: Tools categories.

4 Final Remarks

The results of this literature review suggest that new consensus solutions should be developed to run on low-processing IoT devices. Another recommendation is to study how new consensus solutions can prevent Sybil and DDoS attacks to avoid future vulnerabilities. For future work, we intend to study how the consensus solutions proposed by the selected papers improve the IoT security, and to identify which methodologies are used for developing consensus protocols.

References

[1] Muhammad Ahmad, Qaiser Riaz, Muhammad Zeeshan, Hasan Tahir, Syed Ali Haider, and Muhammad Safeer Khan. Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using unsw-nb15 data-set. *EURASIP Journal on Wireless Communications and Networking*, 2021.

[2] Muhammad Nasir Mumtaz Bhutta, Amir A. Khwaja, Adnan Nadeem, Hafiz Farooq Ahmad, Muhammad Khurram Khan, Moataz A. Hanif, Houbing Song, Majed Alshamari, and Yue Cao. A survey on blockchain technology: Evolution, architecture and security. *IEEE Access*, 9, 2021.

[3] Yulin Fan, Yang Li, Mengqi Zhan, Huajun Cui, and Yan Zhang. Iotdefender: A federated transfer learning intrusion detection framework for 5g iot. In *IEEE International Conference on Big Data Science and Engineering (BigDataSE)*, 2020.

[4] Barbara Kitchenham and Pearl Brereton. A systematic review of systematic review process research in software engineering. *Information and Software Technology*, 55(12), 2013.

[5] Imran Makhdoom, Mehran Abolhasan, Justin Lipman, Ren Ping Liu, and Wei Ni. Anatomy of threats to the internet of things. *IEEE Communications Surveys & Tutorials*, 21(2), 2019.

[6] Roberto Minerva, Abyi Biru, and Domenico Rotondi. Towards a definition of the internet of things (iot). *IEEE Internet Initiative*, 1(1), 2015.

[7] Arshdeep Singh, Gulshan Kumar, Rahul Saha, Mauro Conti, Mamoun Alazab, and Reji Thomas. A survey and taxonomy of consensus protocols for blockchains. *Journal of Systems Architecture*, 127, 2022.

[8] William Stallings. *Cryptography and Network Security: Principles and Practice*. Pearson, 2015.

[9] Nazar Waheed, Xiangjian He, Muhammad Ikram, Muhammad Usman, Saad Sajid Hashmi, and Muhammad Usman. Security and privacy in iot using machine learning and blockchain: Threats and countermeasures. *ACM Computing Surveys*, 53(6), 2020.

[10] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in Software Engineering*. Springer, 2012.

Author Index