Journal of Data Intelligence, Vol. 3, No.1 (2022) 116-130

PREDICTING PRE-TRIAL DETENTION OUTCOMES IN THE BRAZILIAN SUPREME COURT

THIAGO RAULINO DAL PONT

Federal University of Santa Catarina, UFSC, Department of Automation and Systems, Brazil

thiagordalpont@gmail.com

ISABELA CRISTINA SABO

Federal University of Santa Catarina, UFSC, Department of Law, Brazil

isabelasabo@gmail.com

PABLO ERNESTO VIGNEAUX WILTON

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

luzcka@gmail.com

VICTOR ARAÚJO DE MENEZES

Federal University of Santa Catarina, UFSC, Department of Law, Brazil

victormenezesx3@gmail.com

RAFAEL COPETTI

Federal University of Santa Catarina, UFSC, Department of Law, Brazil

rafaelcopetti1@gmail.com

LUCIANO ZAMBROTA

Federal University of Santa Catarina, UFSC, Department of Law, Brazil

lucianozambrota@uol.com.br

PABLO PROCOPIO MARTINS

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

pablopmartins@gmail.com

EDJANDIR CORRÊA COSTA

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

edjandir@gmail.com

EDIMEIA LILIANI SCHNITZLER

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

edililica@yahoo.com.br

PALOMA MARIA SANTOS

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

santos.paloma@gmail.com

RODRIGO RAFAEL CUNHA

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

cunhaegc@outlook.com

GERSON BOVI KASTER

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

gersonbk@gmail.com

AIRES JOSÉ ROVER

Federal University of Santa Catarina, UFSC, Department of Law, Brazil

Federal University of Santa Catarina, UFSC, Department of Knowledge Engineering and Management, Brazil

aires.rover@gmail.com

Brazil has a large prison population, which places it as the third country in the world with the most incarceration rate. In addition, the criminal caseload is increasing in Brazilian Judiciary, which is encouraging AI usage to advance in e-Justice. Within this context, the paper presents a case study with a dataset composed of 2,200 judgments from the Supreme Federal Court (STF) about pre-trial detention. These are cases in which a provisional prisoner requests for freedom through habeas corpus. We applied Machine Learning (ML) and Natural Language Processing (NLP) techniques to predict whether STF will release or not the provisional prisoner (text classification), and also to find a reliable association between the judgment outcome and the prisoners' crime and/or the judge responsible for the case (association rules). We obtained satisfactory results in both tasks. Classification results show that, among the models used, Convolutional Neural Network (CNN) is the best, with 95% accuracy and 0.91 F1-Score.

Association results indicate that, among the rules generated, there is a high probability of drug law crimes leading to a dismissed habeas corpus (which means the maintenance of pre-trial detention). We concluded that STF has not interfered in first degree decisions about pre-trial detention and that it is necessary to discuss drug criminalization in Brazil. The main contribution of the paper is to provide models that can support judges and pre-trial detainees.

Keywords: E-justice, Criminal Law, pre-trial detention, text classification, association rules, machine learning.

1 Introduction

Brazil has a large prison population. According to the National Penitentiary Department, there were more than 750 thousand people in detention in the country in 2019 [1]. This situation places Brazil as the third country in the world with the most incarceration rate [2].

There are two groups of prison population: permanent and provisional. The difference between them is that the provisional ones have not yet been definitively judged. Provisional prisoners should await the judgment with their freedom restricted because they represent imminent danger or may interfere with the process. This situation should be an exception and represent a smaller percentage of the prison population. However, the number of pre-trial detainees in Brazil is high. In 2019, there were 222,558 provisional prisoners, which corresponds to 29% of the total prisoners in the country [1].

Furthermore, 2.4 million new criminal cases entered the Brazilian Judiciary in 2019, of which 121.4 thousand (4.3%) were in the Superior Courts [3]. To minimize the problem of increasing caseload (which includes criminal cases about pre-trial detention), Brazilian Judiciary has been trying to advance in e-Justice with regulations encouraging the use of Artificial Intelligence (AI) in its domain [4,5].

For this purpose, this paper presents a case study with the Brazilian Supreme Federal Court (STF) judgments on pre-trial detention and the use of AI to generate models that can support judges and provisional prisoners. Our research questions are: 1. Can we accurately predict whether STF will release or not the prisoner in an habeas corpus judgment on pre-trial detention? 2. Can we find a reliable association between the judgment outcome and the prisoners' crime and/or the Judge-Rapporteur? To answer them, we resort to two Machine Learning (ML) tasks: classification and association rules.

The paper is organized as follows: In section 2, we present some concepts on pre-trial detention and habeas corpus in Brazilian Law and overall ML tasks. In section 3, we expose some other AI works about judicial predictions in Superior Courts' judgments. In section 4, we describe the case study methodology, including the dataset construction and the techniques applied. In sections 5 and 6, we show and discuss the results. Finally, there are concluding remarks and new perspectives of study in section 7.

2 Background

2.1 Pre-trial detention and habeas corpus in Brazilian Law

According to the Brazilian Criminal Procedure Code [6], pre-trial detention is a kind of precautionary arrest and aims to provide security and effectiveness to criminal prosecution, preventing the accused and/or third parties from hampering the regular progress of the process. It can be decreed from the investigation stage until the end of the criminal proceedings.

The Criminal Procedure Code says that no person may be imprisoned except for flagrante delicto or from a decree with due justification written by the competent judicial authority [6]. The law does not provide for a maximum period of pre-trial detention (usually up to 120 days) and it is not allowed in cases of first time offenders accused of nonviolent crimes.
Brazilian Constitution [7] states that no one shall be arrested unless in “flagrante delicto” or by a written and justified order of a competent judicial authority, save in the cases of military transgression or specific military crime, as defined in law.

It is, therefore, a precautionary instrument, lasting as long as the reasons that gave rise to it. Pre-trial detention may be decreed as a guarantee of public order, economic order, for the convenience of criminal instruction or to ensure the application of criminal law, when there is evidence of the crime and sufficient evidence of authorship and danger generated by the state of freedom of the accused [6].
The habeas corpus is a constitutional action which aims to protect the individual against any restraining measure from the public power to his/her right of freedom. It is preventive (aims to cease imminent violence or coercion) or repressive (when a concrete prejudice occurs) [7].

It has a very summary procedure and the party can not use it for controversial issues of a fact. Anyone, on its own behalf or someone else’s, may fill habeas corpus without an attorney representation. In addition, it is possible for the Prosecutor to file the request. Overall, it is an instrument frequently used to prevent the maintenance of preventive detention in criminal actions in the STF.

The habeas corpus proceeding is the oldest and basic procedural institution for the protection of constitutional rights in Brazil and it is free of charge. Usually, habeas corpus is judged by STF when the constraining party is a Superior Court, or when the constraining party or the petitioner is an authority or employee whose acts are directly subject to the jurisdiction of the Federal Supreme Court or in the case of a crime, subject to the same jurisdiction in one sole instance [7]. Thus there are different ways to file a habeas corpus lawsuit in STF.

2.2 Machine Learning tasks: classification and association rules

AI aims to understand intelligence for the construction of intelligent entities. AI can be divided into six fields: Knowledge representation; Natural Language Processing (NLP); Automated Reasoning; ML; Computer vision; Robotics [8].

ML, one of the fields of AI, is concerned with the development of systems capable of learning from data. Géron [9] classifies the learning process in four major categories:

- supervised learning: the training dataset has the solution or the labels;
- unsupervised learning: the training dataset doesn't have the solution, that is, it is unlabeled;
- semi supervised learning: part of the training dataset is labeled;

- reinforcement learning: the system learns as it receives rewards.

Classification is the main ML task of the supervised learning category. It estimates a finite set of discrete labels, so the algorithms predict a class for unlabeled data. Metrics are used to assess the level of correctness of the result of the task [10]. The most used metrics to evaluate classification results are:

- accuracy: number of correct predictions divided by total of predictions made;
- precision: number of true positive divided by the sum of true positive and false positive;
- recall: true positive divided by the sum of true positive and false negative;
- F1-Score: harmonic mean of precision and recall: 2×precision×recall / (precision + recall).

Association rules is one of the ML tasks of the unsupervised learning category. It is used to find interesting associations in large sets of data items [11]. Algorithms can be used to implement association rules and metrics as support and confidence to measure the quality of the rules [12]:

- support: ratio of transactions containing the set of items of the association rule;
- confidence: the ratio of correct results of the association rule, considering the set of items of the support.

3 Related work

In this section, we discuss some previous predicting judicial decisions works, including researches based on text classification techniques. On behalf of the United States Supreme Court, experiments predicted the decisions of the Court regarding certain legal issues, using the votes of several judges who have integrated it over the years, in which a decision is confirmed or changed in a higher instance. For this experiment, variables such as the year of the case, the legal matter discussed, the location of the lower court and many others were taken into account. A Decision Trees and Random Forest (ensemble method) classifiers approach achieved 70.2% accuracy on the case outcome and 71.9% on the judge vote predictions [13, 14].

At the European Court of Human Rights, an experiment was able to foresee decisions based on text classification. The goal was to predict if there might be a violation or non-violation in the cases of Human Rights offenses carried out by a Member State, based on the premises of civil and political rights included in the European Convention on Human Rights. This Linear Support Vector Machines (SVM) classifier approach had 79% of accuracy [15].

In another experiment made at the Supreme Court of France, researchers carried out experiments that made it possible to predict case ruling with 96% of accuracy, 90% for predicting the law area of a case and 75.9% for estimating the decade to which the case belongs. For that, the researchers used cases and decisions from the 1880s to 2010 [16].

At the Supreme Court of the Philippines, in order to reduce court litigation and problems with pending cases, researchers carried out experiments to predict the outcome of the cases in the Court. The scope of such documents was limited to the criminal context. The authors made use of public processes as input and obtained 59% of accuracy on case outcomes using random forest classifiers [17].

In Brazil, the "VICTOR" project, financed by the STF, is currently underway and has the initial goal of automatically linking legal proceedings that constitutes general repercussion (GR). The GR is a procedural instrument that acts as a "recursal filter", allowing the Court to select the resources that it will analyze according to the criteria of legal, political, social or economic relevance. The solution based on deep learning had an accuracy of 90.35% in a preliminary evaluation with Convolution Neural Networks (CNN) [18].

Advancing the state of the art on this matter, our work brings not only classification but also association rules with attributes extracted from the texts. The legal matter that we choose for this experiment is a recurring theme in the legal discussion in Brazil, which is the pre-trial detention.

4 Methodology

The research is a case study since it investigates a contemporary problem with real data [19]. To achieve this, we constructed a dataset and applied ML and NLP techniques to generate models which embraced two tasks: classification and association. We used Orange 3 [20] in most of our classification and association experiments. As a complement, we applied Python Programming language with ML libraries in one of the classification models [21]. Fig. 1 presents the pipeline for this work, which is detailed in the following sections.

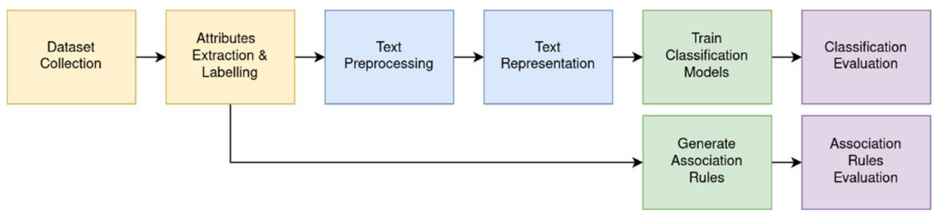

Figure 1. Pipeline for text classification and rule mining

4.1 Dataset: collection, attribute extraction and labelling

Our dataset is composed of around 2,200 judgments (collective judicial decision) from STF between June 2004 and February 2020. They are limited to the subject of pre-trial detention, where a provisional prisoner requests for freedom through habeas corpus. Once this type of data is public in Brazil, we collected the legal cases from STF's web platform [22] as Portable Document Format (PDF). However, we ignored documents composed only of scanned images, i.e., without any selectable text. Due to the structure of the documents, a more refined work would be necessary in order not to lose the formatting and consequently lose data. Thus, it was easier to extract the contents of the PDF files to raw text files. Using open-source tools such as the Python programming language [23] and open-source libraries for PDF and text processing [24], we extracted the text from the PDF documents to raw text files, which we used as part of our dataset in our experiments for classification (2,015) and association rules (1,776).
The dataset was analyzed by an interdisciplinary research group, which extracted 3 attributes from each judgment: a) final outcome; b) crime category; and c) Judge-Rapporteur. One group manually extracted attributes "a" and "b" while the other one automatically extracted "c" using NLP techniques like named-entity recognition and regular expressions which searches for patterns in the text since the rapporteur's name is usually in the same place in all documents.

a) Final outcome: Knowing that there are many ways to represent the same outcome in the decision text, for this attribute we came up with the assumption of a "released" or "not released" binary outcome. Thus, in the cases when house arrest was granted, the classification was "not released" because it did not imply the incarceration of such a person. For the study, "released" means that the punishment is not carried out in the penitentiary. The ratio of each final outcome in the dataset is: not released (75,73%) and released (24,27%).

b) Crime category: Considering the various forms of text that represents the crimes indicated by the judges, we created a framework for the classification of the dataset which resulted in 10 categories based on the Brazilian criminal laws. This categorization was done by a manual extraction of all the legal terms that could indicate the crimes and then by a computational processing to generate a new dataset with the terms categorized. For example: in our analysis, we found 93 terms related to drug-related law crimes, including different ways to express the same law and article; different articles for a similar legal concept and even orthographic errors for the writing of a crime, resulting in the category "drug law crime".
Thus, this category included individual drug traffic, criminal association for the drug traffic and international drug traffic, which are all crimes found in the same Act and that we choose to reduce to the same category. Fig. 2 shows the distribution of crimes across the documents considering the final outcome.

c) Judge-Rapporteur: During the judgement, one of the STF judges functions as the so-called Judge-Rapporteur, which is designated to analyze the case and send its vote to the other judges that must follow the vote or refrain to do so. This is a very relevant factor because the Judge-Rapporteur is the one who is going to expose the legal matters regarding the case in a report to the other judges. We noticed that some of them are more bound to have their votes followed-up by their counterparts. The ratio of reported cases for each Judges-Rapporteur in the dataset based on their corresponding name initials were: MA (48.42%), GM (10.25%), CL (7.10%), RL (6.98%), DT (6.76%), RW (4.96%), LF (4.62%), TZ (3.43%), CM (2.64%), AB (1.57%), RB (1.29%), EF (0.73%), JB (0.51%), EG (0.28%), CV (0.11%), CP (0.11%), AM (0.06%), HES (0.06%), IG (0.06%), OG (0.06%). Although we noticed that some judges were Rapporteurs more often than the others, this variance was not the objective of this study.

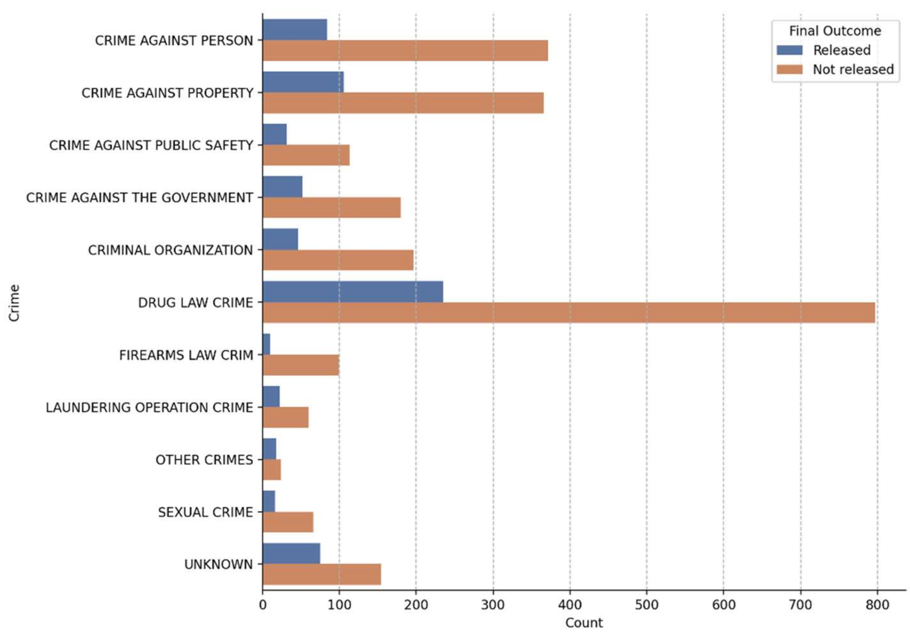

Figure 2 Distribution of crimes by final outcome

4.2 Text preprocessing

Text processing relates to the set of NLP techniques applied on raw texts so they can be used in a cleaner format as inputs for ML tasks [25]. We applied the text processing techniques as follows:

a) Noise Filtering: after extracting raw text from PDF documents, we need to remove the noise characters, where we keep only alphanumeric characters.

b) Normalization: To uniformly process the words, we convert the whole text to lowercase [26].

c) Tokenization: To interpret the structure of words we have to detect their starting and ending points to form tokens (pieces of text). We use regular expressions to detect sequences of letters and numbers [26].

d) Stemming: Words can be contracted to the root form or stem so we can reduce the variations of words [27]. In this work, we used the Porter Stemmer to the Portuguese language.

e) Filtering: In the text, there will be very common words, also called stopwords, which do not help to better capture the meanings of the text. These words include prepositions and articles. We removed those words from the corpus [28].

f) N-Grams: Words can have different meanings according to their surrounding words [29]. Thus, we consider, as a single unit, the sequences of two (bi-grams), three (tri-grams) or more, that consistently appear together [27] In this work, we used one-grams and bi-grams.

After the pre-processing steps, text representation techniques transform the corpus from a textual to a numerical format.

4.3 Text representation

ML techniques often require the data to be represented as vectors, that is, as sequences of numbers. Thus, we needed to transform each textual document to a sequence of numbers. A simple technique is the Bag of Words (BOW) model, where each number represents the Term Frequency (TF) of a word or N-gram in the document. This text representation technique achieves good results in Text Mining applications [30-32], but it loses the notion of sequences in the text [33]. In this work, we used the BOW model as input to classical ML techniques using Orange 3.

When using deep learning techniques, we may need a more robust representation. In recent years, new techniques for representation based on neural networks have been proposed, which efficiently represent the text considering its syntactic and semantic structures [29, 34, 35]. In our experiments with deep learning techniques, we used the pre-trained representation for the GloVe [36] technique trained using Brazilian legal texts from Supreme, Superior and State Courts [37].

4.4 Classification models

In the classification task, we want to predict the result of the legal case having only the legal text as input and the result as output labels. We used two kinds of ML models for classification: classical and deep learning. Classical ML refers to the techniques that have few parameters in the training step while deep learning refers to the techniques represented as deep neural networks that have thousands, millions or billions of parameters. Deep learning techniques improve as they have more data to learn, while classical techniques improve until it reaches a plateau [38].

We employed available approaches of classical ML in Orange 3: linear, tree based, neural based, instance based. In terms of linear models, we used SVM, Naïve Bayes and Logistic Regression. As tree-based methods, we applied Decision Tree, Random Forest and AdaBoost. In terms of neural based, we used Multi-Layer Perceptron. And as instance based, we had K-Nearest Neighbors (KNN) [33]. In terms of deep learning, we applied the Convolutional Neural Network (CNN) [39], using word embeddings representation, as mentioned, and Keras framework with the Python programming language [21].

To evaluate our classification models, we divided our dataset into two groups, train and test, with cross-validation for classical ML techniques and random sampling for CNN. We applied the train set in the learning step of our models and used the test set in the evaluation step to make predictions. We checked the performance using accuracy and F1-Score metrics (explained in section 2.2).

4.5 Association rules model

In the association task, we want to find correlations between the extracted attributes from our dataset (Final outcome, Crime category and Judge-Rapporteur). To do so, we applied the FP-Growth algorithm through Orange 3 to create association rules with the attributes as input. FP-Growth generates a tree from frequent patterns by scanning the whole dataset following the support threshold. Then the rules are formed by constructing a conditional tree, which saves the costly dataset scans in the subsequent mining processes [40-41]. The rule is expressed in the form of an implication, e.g. A→B.

To evaluate our association rules, Orange 3 allows us to manipulate the support and confidence metrics (explained in section 2.2), as well as filter the rules with the desirable attributes in the consequent and antecedent.

5 Classification results and discussion

As a response to question 1, we set up a classification pipeline with 2.015 documents and the attribute/label “final outcome”, to predict whether the prisoner will be released or not. We followed the steps in the methodology, that is, pre-processing, representation, classification models training with train set and evaluation with test set. After evaluating our models with the test set, we calculated the accuracy and F1-Score as shown in Fig. 3 and Fig. 4, respectively.

In terms of accuracy, our models achieved good results, since all techniques had accuracy values higher than 70%. And the best accuracy comes from the deep learning technique, CNN, that achieved a high accuracy of 95%.

On the other hand, F1-Score also leveraged good results for most techniques except SVM. We also have to note that the best technique is no longer CNN but Logistic Regression. Furthermore, we note that we need both metrics to better perceive the performance of our models. When we have a label, such as “not released”, comprising almost 75% of the dataset, and we also have a model that assigns that label to all documents, we will have a higher accuracy. Thus, we need F1-Score which will balance the performance across the existing labels in the dataset.

Considering both metrics, we can point CNN as the best model, followed by Logistic Regression, AdaBoost, Multi-Layer Perceptron, Random Forest, KNN, Naïve Bayes and SVM.

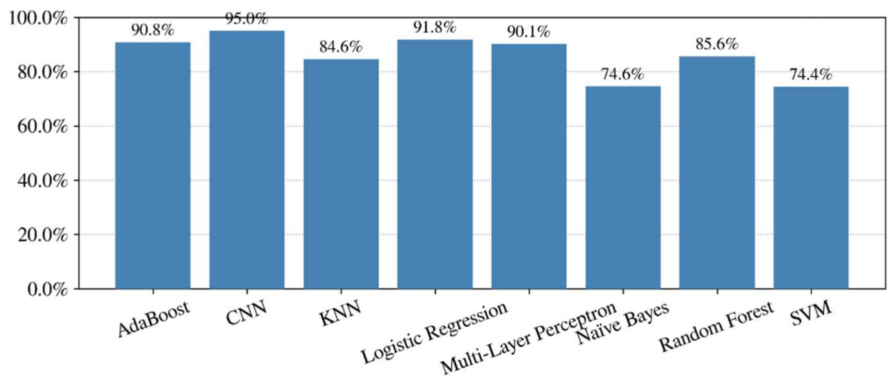

Figure 3 Accuracy for predictions in test set.

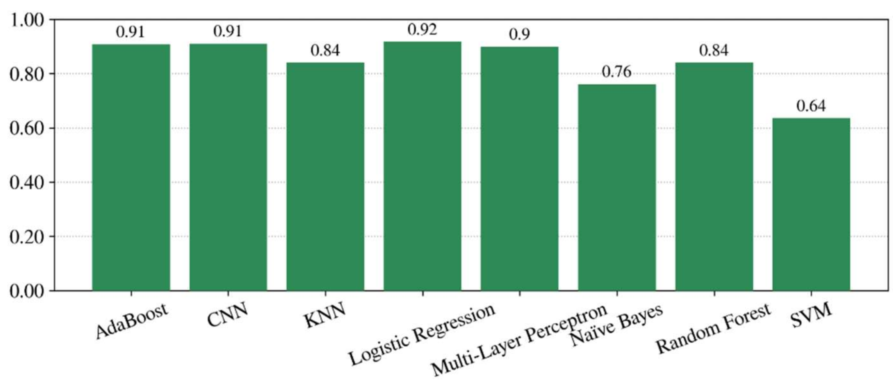

Figure 4 F1-Score for predictions in test set.

6 Association results and discussion

As a response to question 2, we generated association rules with transactions from 1.776 judgments and the three attributes extracted. We organized the results into three groups: a) rules with minimum support 5%, minimum confidence 70% and "not released" as the consequent (Table 1); b) rules with minimum support 1%, minimum confidence 90% and "not released" as the consequent (Table 2); and c) rules with minimum support 1%, minimum confidence 30% and "released" as the consequent (Table 3). According to Orange 3 tutorials [42], for large datasets it is normal to set a lower minimal support (e.g. between 2%-0.01%), so one must increase confidence.

We did not find rules with "released" as the consequent within reasonable confidence. It means that there is no strong correlation between this outcome and any crimes and/or any Judge-Rapporteur in the dataset.

Table 1. Group of association rules “a”.

Supp	Conf	Antecedent		Consequent
0.055	0.860	Criminal organization	→	Not released
0.058	0.831	Judge-Rapporteur RL	→	Not released
0.398	0.822	Judge-Rapporteur MA	→	Not released
0.073	0.822	Judge-Rapporteur MA, Crime against property	→	Not released
0.076	0.813	Judge-Rapporteur MA, Crime against person	→	Not released
0.057	0.810	Judge-Rapporteur CL	→	Not released
0.178	0.806	Crime against person	→	Not released
0.164	0.804	Judge-Rapporteur MA, Drug law crime	→	Not released
0.141	0.749	Crime against property	→	Not released
0.292	0.729	Drug law crime	→	Not released

This first group (Table 1) represents the rules that have a balance between support and confidence. We can infer from this group that the habeas corpus on pre-trial detention in drug law crime and criminal organization, also in crimes against person and property, will probably have the judgment as “not released”. In other words, the prisoner who commits these crimes will be kept in pre-trial detention.

Another inference is that certain judges as rapporteurs (MA, CL and RL) suggest the probability of the outcome being “not released”. Also, the presence of the Judge-Rapporteur MA minister with crimes against property and person, and drug law crime, indicates the probability of the outcome “not released”. However, we emphasize that the judgment is composed of the votes of all the Judges, and the vote of Judge-Rapporteur MA may not be “not released”.

Considering both metrics, we can point that the strongest association rules are “Drug law crime → Not released” and “Judge-Rapporteur MA → Not released”.

Table 2. Group of association rules “b”.

Supp	Conf	Antecedent		Consequent
0.016	1.000	Judge-Rapporteur MA, Firearms law crime	→	Not released
0.011	1.000	Judge-Rapporteur MA, Drug law crime, Firearms law crime	→	Not released
0.017	0.968	Drug law crime, Firearms law crime	→	Not released
0.013	0.958	Judge-Rapporteur RL, Crime against person	→	Not released
0.012	0.955	Drug law crime	→	Not released
0.019	0.944	Judge-Rapporteur MA, Crime against property	→	Not released
0.025	0.936	Firearms law crime	→	Not released
0.033	0.921	Crime against property	→	Not released
0.042	0.914	Judge-Rapporteur MA, Criminal organization	→	Not released
0.016	0.906	Judge-Rapporteur RL, Crime against property	→	Not released

This second group (Table 2) represents the rules with the highest confidence, but not necessarily with high support. We realize that it contains the same or similar rules to the previous group. The main difference between Table 1 and Table 2 is a new crime that appears: firearms law crime. This crime appears associated with the drug law crime, so that both have a high probability of resulting in "not released".

Although the low support, we highlight that the rules "Judge-Rapporteur MA, Firearms law crime → Not released" and "Judge-Rapporteur MA, Drug law crime, Firearms law crime → Not released" obtained 100% confidence.

Table 3. Group of association rules "c".

Supp	Conf	Antecedent		Consequent
0.014	0.532	Judge-Rapporteur CM	→	Released
0.018	0.432	Judge-Rapporteur GM, Drug law crime	→	Released
0.039	0.379	Judge-Rapporteur GM	→	Released
0.011	0.328	Judge-Rapporteur TZ	→	Released
0.016	0.318	Judge-Rapporteur RW	→	Released
0.021	0.308	Judge-Rapporteur DT	→	Released

This third group (Table 3) represents the rules with the lower confidence, specifically rules with "released" as a consequent. It means that there is no strong correlation between this outcome and any crimes and/or any Judge-Rapporteur in the dataset. However, we observe that the Judges-Rapporteur that appear here (CM, GM, TZ, RW and DT) are different from the Judges-Rapporteur that appear in the rules of previous groups (Table 1 and Table 2).

7 Conclusion and future work

The two research questions outlined for this research were answered since, according to the indicated metrics, we obtained satisfactory results both in terms of classification and association rules. We concluded from these experiments that STF has not interfered in first degree decisions about pre-trial detention. Notably, no reasonable metrics were found in association rules with the outcome "released", dispelling the impression of impunity or that certain crimes, such as those against the government, could revoke pre-trial detention.

In terms of application, while the classification results can speed up the judgment, for example, when the period of pre-trial detention has already expired, the association results can identify patterns in judgments and thus reduce biases. Or even point out legal issues for debate, such as the drug criminalization in Brazil, since "drug law crime" is strongly correlated with the outcome "not released". We emphasize that these experiments had the character of formulating a model based on past judgments and verifying the main variables involved. It was not the scope of this work to discuss automating judicial decisions or hindering access to justice, since the judgment must be the result of a human evaluation. We understand that both results from classification and association rules can be useful as an assist tool and not a replacement for the magistrate, mainly when the object of the case is someone's freedom.

Future prospective studies may address the possible correlations between other attributes, such as the prisoners' location, the other STF judges present at the judgment and their votes.

References

1. Prisoners in Brazil Homepage, https://bityli.com/Bk8Ir, last accessed 2020/09/10.
2. CONECTAS Homepage, "Brazil ranks as the third country with the largest prison population in the world", https://bityli.com/lN18R, last accessed 2020/09/10.
3. CNJ.: Justiça em Números 2020: ano-base 2019. CNJ, Brasília (2020). https://bityli.com/ORZNy, last accessed 2021/04/01.
4. CNJ Homepage, Resolution no. 332 of August 21, 2020, https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf, last accessed 2021/04/01.
5. CNJ Homepage, Ordinance no. 271 of December 4, 2020, https://atos.cnj.jus.br/files/original234208202012155fd949d04d990.pdf, last accessed 2021/04/01.
6. Brazilian Criminal Procedure Code Homepage, http://www.planalto.gov.br/ccivil_03/decreto-lei/del3689.htm, last accessed 2021/03/25.
7. Brazilian Federal Constitution Homepage, http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm, last accessed 2021/03/25.
8. Russell, S. J., Norvig, P., Davis, E.: Artificial intelligence: A modern approach. 3rd edn. Prentice Hall, New Jersey (2013).
9. Géron, A.: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. 2nd edn. O'Reilly Media, Inc (2019).
10. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. Knowledge-Based Systems 46, 109–132 (2013), https://doi.org/10.1016/j.knosys.2013.03.012, last accessed 2021/04/02.
11. Cios, K. J., Pedrycz, W., Swiniarski, R. W., Kurgan, L. A. (orgs.).: Data mining: A knowledge discovery approach. Springer, New York (2007).
12. Brin, S., Motwani, R., & Silverstein, C.: Beyond market baskets: Generalizing association rules to correlations. ACM SIGMOD Record 26(2), 265–276 (1997), https://doi.org/10.1145/253262.253327, last accessed 2021/04/04.
13. Katz, D., Bommarito M., Blackman J.:. Predicting the behavior of the Supreme Court of the United States: a general approach. ArXiv, 1407.6333 v1, pp. 1-17. (2014).
14. Katz, D., Bommarito M., Blackman J.: A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE (2017), https://doi.org/10.1371/journal.pone.0174698, last accessed 2021/04/06.
15. Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. PeerJ Computer Science 2(93), pp. 1-19. PeerJ (2016).
16. Șulea, O., Zampieri, M., Vela, M., Genabith, J.: Predicting the Law area and decisions of french supreme court cases. In: RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING, 11., 2017, Varna, Bulgaria. Proceedings of a meeting held 2-8 September 2017, pp. 716-722. INCOMA Ltd Shoumen (2017).
17. Virtucio, M., Aborot, J., Abonita, J., Aviñante, R., Copino, R., Neverida, M., Osiana, V., Peramo, E., Syjuco, J., Tab, G..: Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER SOFTWARE AND APPLICATIONS, 42., Tokyo (2018), https://doi.org/10.1109/COMPSAC.2018.10348, last accessed 2021/04/06.
18. Silva, N., Braz, F., Campos, T., Guedes, A., Mendes, D., Bezerra, D., Gusmão, D., Chaves, F., Ziegler, G., Horinouchi, L., Ferreira, M., Inazawa, P., Coelho, V., Fernandes, R., Peixoto, F., Maia Filho, M., Sukiennik, B., Rosa, L., Silva, R., Junquilho, T., Carvalho, G.: Document type

classification for Brazil's supreme court using a Convolutional Neural Network. In: International conference on forensic computer science and cyber law, pp. 7-11. HTCIA (2018).
19. Yin, R. K.: Estudo de Caso: Planejamento e métodos. Bookman (2015).
20. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar ,J., Žitnik, M., Zupan, B.: Orange: Data mining toolbox in python. J. Mach. Learn.Res.14(1), pp. 2349–2353 (2013).
21. Keras GitHub Homepage, https://github.com/fchollet/keras, last accessed 2021/03/10.
22. Brazilian Supreme Federal Court Homepage, http://portal.stf.jus.br, last accessed 2021/04/01.
23. Van Rossum, G., Drake, F.L.: Python 3 Reference Manual. CreateSpace, Scotts Valley, CA (2009).
24. Bird, S., Loper, E.: NLTK: The Natural Language Toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions. pp. 214–217. Association for Computational Linguistics, Barcelona, Spain (2004).
25. García S., Luengo J., Herrera F.: Data preprocessing in data mining. Springer International Publishing, Switzerland (2015).
26. Jurafsky D., Martin JH.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 3rd edn. Draft, Stanford University (2019).
27. Aggarwal, C.C.: Machine Learning for Text. In: Machine Learning for Text, pp. 1–16. Springer International Publishing, Cham (2018).
28. Kotu V., Deshpande B.: Data science: concepts and practice, 2nd edn. Morgan Kaufmann (Elsevier), Cambridge (2019).
29. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings pp. 1–12 (2013).
30. Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. International Journal of Computer Vision 104(2), pp. 154–171 (2013).
31. Meriño, L.M., Meng, J., Gordon, S., Lance, B.J., Johnson, T., Paul, V., Tank, U.S.A.: A bag-of-words model for task-load prediction from EEG in complex environments. Engineering pp. 1227–1231 (2013).
32. Milosevic, N., Dehghantanha, A., Choo, K.K.R.: Machine learning aided Android malware classification. Computers and Electrical Engineering 61, pp. 266–274 (2017).
33. Kowsari, K., Meimandi, K.J., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D.: Text classification algorithms: A survey. Information (Switzerland) 10(4) (2019).
34. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of DeepBidirectional Transformers for Language Understanding pp. 4171–4186 (2018).
35. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan,A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G.,Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C.,Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish,S., Radford, A., Sutskever, I., Amodei, D.: Language Models are Few-Shot Learners (2020).
36. Pennington, J., Socher, R., Manning, C.: Glove: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). vol. 19, pp. 1532–1543. Association for Computational Linguistics, Stroudsburg, PA, USA (2014).
37. Dal Pont, T.R., Sabo, I.C., Hübner, J.F., Rover, A.J.: Impact of Text Specificity and Size on Word Embeddings Performance: An Empirical Evaluation in Brazilian Legal Domain. Lecture

Notes in Computer Science, vol. 12319, pp. 521–535. Springer International Publishing, Cham (2020).

38. Sewak, M., Sahay, S.K., Rathore, H.: Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection. In: 201819th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel / Distributed Computing (SNPD). pp.293–296. IEEE (2018).
39. Kim, Y.: Convolutional Neural Networks for Sentence Classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). vol. 2017, pp. 1746–1751. Association for Computational Linguistics, Stroudsburg, PA, USA (2014).
40. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. ACM SIGMOD Record 29(2), pp. 1–12 (2000).
41. Agarwal, R.C., Aggarwal, C.C., Prasad, V.V.V.: Depth first generation of long patterns. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’00. vol. 2, pp. 108–118. ACM Press, New York, New York, USA (2000).
42. Orange Data Mining Homepage, https://orangedatamining.com, last accessed 2021/04/09.