Predicting Judicial Outcomes in the Brazilian Legal System Using Textual Features

Vithor Gomes Ferreira Bertalan[1][0000−0002−1585−7694] and
Evandro Eduardo Seron Ruiz[2][0000−0002−7434−897X]

[1] PPG-CA, DCM, FFCLRP, University of São Paulo (USP), vbertalan@usp.br
[2] Department of Computing and Mathematics, FFCLRP, University of São Paulo,
evandro@usp.br

Abstract. The combination of Natural Language Processing and Artificial Intelligence for the field of Law is a growing area, with the potential of radically changing the daily routine of legal professionals. The amount of text generated by those professionals is outstanding, and to this point, still unexplored by Computer Science. One of the most acclaimed research field covering both knowledge areas is Legal Prediction, in which intelligent systems try to predict specific judicial characteristics, such as the judicial outcome or the judicial class or a given case. This research intends to create a classifier to predict judicial outcomes in the Brazilian legal system. At first, we developed a text crawler to retrieve judicial outcomes from the official Brazilian electronic legal systems. Afterward, a few judicial subjects were selected, and some of their features were extracted. Later, a set of different classifiers was applied to predict the legal considering these textual features.

Keywords: Legal prediction · Digital humanities · Artificial Intelligence and Law

1 Introduction and Research Objectives

1.1 Introduction

The combination of Natural Language Processing and Artificial Computer Science has been revolutionizing many different fields of expertise. Subfields of Computer Science, like Natural Language Processing (also known as NLP), have steadily improved a myriad of professional and scientific activities. NLP helps researchers to understand how computers can process and analyze large amounts of natural language data and their meanings. Even simple NLP mechanisms, such as dictionaries and word counts, can offer interesting underlying facts that cannot always be noticeable without effective processing.

Law is one of the knowledge areas that are the most dependent on text data. Millions of legislation workpapers, court decisions, and appeals are produced daily, and many different job specializations, such as lawyers, judges, defendants, and plaintiffs, have various necessities that could be supplied by intelligent systems.

Over the last years, researchers have been dedicated to predicting judicial case outcomes using NLP application software and Machine Learning methods over those textual cases. See Section 1.3 below. However, no research with this intention has been done in Brazilian Portuguese for Brazilian courts, as of 2019.

Branting et al. [7] cites that automation of legal reasoning and problem-solving has been a goal of Computer Science research from its earliest days. However, according to the author, broad adoption of legal computer systems never occurred, and Computer Science and law remained a niche research area with little practical impact.

Hyman et al. [10] point out the Zubulake v. UBS Warburg case, a series of trials and decisions dealing with what data a litigant must preserve and under what circumstances the parties must pay for search and production costs, as a seminal case for AI in Law. According to the authors, this case became a landmark for the practical applications of the research. This case is mainly about the inability of the defendant to retrieve hundreds to thousands of emails that were claimed by the plaintiff to be relevant to the main issue in the lawsuit.

1.2 Research Objectives

As a primary objective, this research intends to develop a framework to predict judicial outcomes in the Brazilian state of São Paulo Justice Court. The São Paulo Justice Court is the most significant legal court on the planet, considering the number of cases per year. We believe that designing a predicting model that offers consistent results for this judicial court this model could be transferred, after fine-tuning, to any other court. Firstly, we will develop a text crawler to retrieve data from the legal outcomes. To pre-process these data, we combine NLP tools to extract characteristics from the text, selecting what is believed to be the primary information that can lead to valid predictions. After this step, we will insert these pre-processed data into machine learning frameworks. Finally, we evaluate all the methods and their respective results against the real judicial outcomes of the court.

1.3 Related Works

Recent advancements have significantly improved the state-of-the-art in the field of legal prediction. In the most influential work, Aletras et al. [3] have used a dataset of cases from the European Court of Human Rights, containing cases that violate Article 3 (*Prohibits torture and inhuman and degrading treatment*), Article 6 (*Protects the right to a fair trial*), and Article 8 (*Provides a right to respect for one's private and family life, his home and his correspondence*) of the

Convention. Katz et al. [12] have used random forests to predict the behavior of the Supreme Court of the United States.

In another influential research, Sulea et al. [18] have done a similar investigation, predicting the law area and decisions of the French Supreme Court using lexical features and support vector machine, SVM. The authors have used a diachronic collection of rulings from the French Supreme Court (*Court de Cassation*, in European French).

Recent researches have used machine learning successfully to improve Law decisions. Gokhale and Fasli [9] have developed a co-training algorithm to classify human rights abuses, using SVM and Logistic Regression. Branting et al. [7] have used hierar-chical attention networks, SVMs, and maximum entropy classifications for decision support in administrative adjudication, such and routine licensing, permitting, immigration, and benefits decisions. SVMs are also used by Fornaciari and Poesio [8] to automatically detect deception, such as defamation and false testimony, in Italian court cases. Remnits [16] has used Latent Dirichlet Allocation to discover the main topics of discussion in judicial outcomes of the United States Supreme Court. Mochales [15] has used argumentation mining to structure better legal arguments, capturing main issues and evidence of a given corpus.

Not directly related to the scope of legal prediction used in this paper, some recent research has also been conducted in Brazilian Portuguese. Aires et al. [1] have used deontic logic to identify norm conflicts in contracts. In contrast, Araujo, Rigo and Barbosa [5] have used ontology-based algorithms to classify legal documents in Brazilian judicial outcomes.

Liu and Chen [14] also write that natural language processing and machine learning have augmented possibilities based on the exploration of semantic of law and case texts. Also, according to Barraud [6], NLP and AI have expanded the limits of justice, transforming it into a predictive, quantitative, statistic, and simulative justice. As stated by Alarie, Niblett and Yoon [2], intelligent judicial systems can not only help lawyers with timely and objective assessments of their claims but also governments, by using legal classifiers to help evaluate claims and manage litigation risks.

2 Methodology

2.1 Domain Characterization

A corpus of judicial sentences, along with their outcomes, was collected from eSAJ, the electronic system of the São Paulo Justice Court (TJSP). To restrict the number of documents retrieved, a few previously defined judicial subjects were selected, which are: second-degree murder (in Brazilian Portuguese, *homicídio simples*), and active corruption (in Brazilian Portuguese, *corrupção ativa*).

Only subjects with very well defined outcomes will be selected. As well defined outcomes, we intend to choose those judicial outcomes with the condemnation or absolution of the defendant. Many different legal subjects, such

as divorce papers, do not have explicit terms for condemnation or absolution. Therefore, it is of utmost importance to find those judicial subjects with clear and well-established results. We also applied numerical labels to features as condemnations and to the absolutions. These binary labels are used to develop a mathematical/statistical model that one may predict the conclusion of the cases mentioned above based on their textual structure.

2.2   Data Collection

We have implemented a web text crawler to retrieve data from eSAJ, using Python. As the user can select from many different fields to exhibit the judicial opinions, such as classes, subjects, judges, and processes numbers, the crawler must be able to choose from those different query choices to compose a raw text file. The following fields were collected: judicial class, judicial subject, judge, county, release date, and full text of the judicial sentence.

In addition to the data collected from the text, additional fields were derived from the information collected: gender of the judge, and a boolean field indicating whether the county was the capital city of São Paulo, or a state city.

To classify each of the judicial outcomes, we have created the binary labels for condemnation (+1) and absolution (-1). We have also used the professional guidance of Brazilian lawyers, with the purpose of better understanding the texts. The language adopted in the field of Law worldwide might be notoriously obscure for a layman.

The data collection poses a compelling challenge, as the amount of data displayed on the website of the TJSP is indeed substantial. As of June 1st, 2018, a simple search for the Brazilian Portuguese correspondent of rape (*estupro*), for example, returns 5,658 hits. Another search for the Brazilian Portuguese term for drug (*droga*) returns 138,956 hits. Each hit is a judicial opinion of its own, containing many sentences and text topics. As each judicial class under the Brazilian law system contains different text topics, e.g. the topics in a text from the class *divorce papers* (in Brazilian Portuguese, *documentos de divórcio*) differ substantially from texts from the class release permits (in Brazilian Portuguese, *alvará de soltura*).

Table 1 illustrates the number of documents retrieved from the eSAJ system.

Table 1. Number of texts collected by judicial subject

Judicial Subject	Number of Cases	Absolutions	Convictions
Second-degree Murder	591	255	336
Active Corruption	191	31	158

2.3   Data Preprocessing

The data retrieved was pre-processed to remove unnecessary information. We began by tokenizing the text and eliminating stopwords. For this task, we used

the Natural Language Processing Toolkit in Python, called NTLK. NLTK can deal with different languages other than English, such as Brazilian Portuguese, which makes this framework a strong candidate for this research. Unusual characters, such as hyphens or parentheses, were also removed. After those steps, we have proceeded with the stemming of the resulting text. For the testing rounds, we used a 10-fold cross validation.

2.4 Data Transformation

In this step, we transformed the data into a mathematical sequence that can be passed through machine learning algorithms. We used TFIDF (term frequency–inverse document frequency) to transform each of the sentences into numbers that are processed through various machine learning (ML) algorithms.

3 Preliminary Results

3.1 Homicides Data Set

We have already tested the two different datasets with various ML methods. The methods selected were: Logistic Regression (LR), Linear Discriminant Analysis (LDA), Gaussian Naïve Bayes (GNB), K-Neighbor (KN), Support Vector Machines (SVM) and, Regression Trees (RT). All the tests were initially run by splitting the training set with 75% of the whole data set, and the test set with the remaing 25% .

For the first data set, the homicides legal texts, we found the calibration plot in Fig. 1. In calibration plots using Platt Scaling, the closest to the perfect calibrated curve, the better. Therefore, from this figure, we can infer that Regression Trees show better results. Support Vector Machines presented the lowest performance, predicting values very differently to those that would be expected based on the reviewed literature.

At last, we run a k-fold cross validation, with 10 epochs, for each of the algorithms. The results are shown in Table 2 and Fig. 2. LDA and Linear Regression, in these tests, were the best choices among all the algorithms.

Table 2. Results (by mean values) on the homicides database, after 10 epochs k-fold

Algorithm	Acccuracy	Accuracy σ	Recall	Precision	F1 Score
Logistic Regression	**0.893390**	0.040169	0.912864	0.896453	**0.903230**
LDA	**0.908588**	0.035791	0.909055	0.922748	**0.914758**
K-Neighbors	0.815621	0.045490	0.873097	0.813255	0.839270
Regression Trees	0.881610	0.019635	0.887465	0.907709	0.895757
Gaussian Naïve Bayes	0.700678	0.085623	0.968779	0.661592	0.783059
Support Vector Machines	0.568644	0.080258	1.000000	0.568644	0.900421
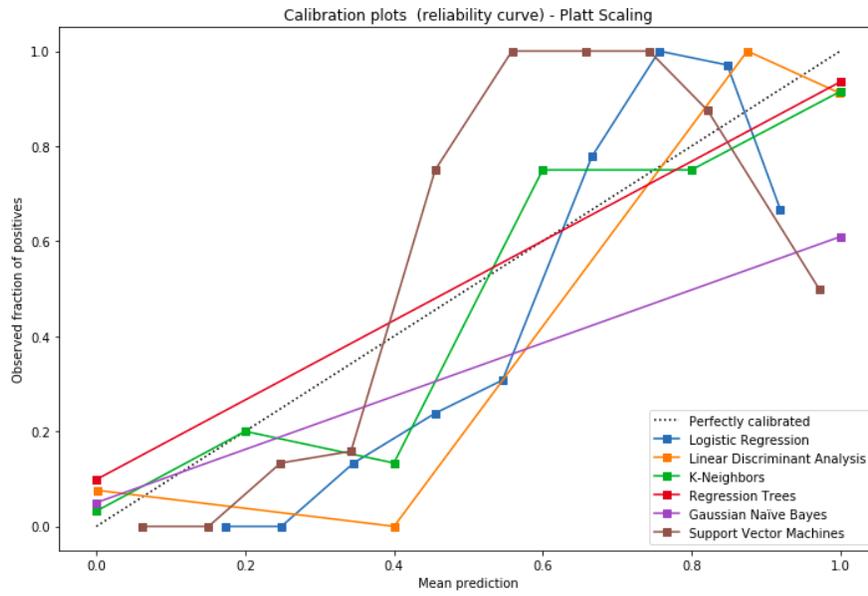
Fig. 1. Calibration plot for the homicides database.
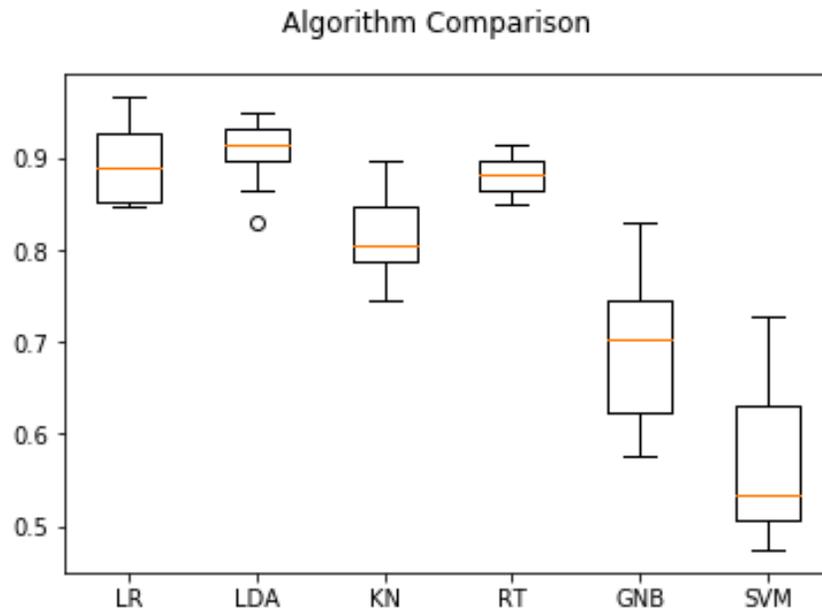
Fig. 2. Candlesticks for the algorithms with 10 k-fold

3.2 Corruption Data Set

For the second data set, with the corruption legal texts, we found the calibration plot shown in Fig. 3. In Platt Scaling, the closest to the perfect calibrated curve, the better. As the previous results with the Homicides data set, we can infer that Regression Trees are the best choice. K-Neighbors showed the lowest performance, predicting values very differently to those that would be expected. All the tests were initially run by splitting the training set considering 75% for the training set and 25% for the test set.
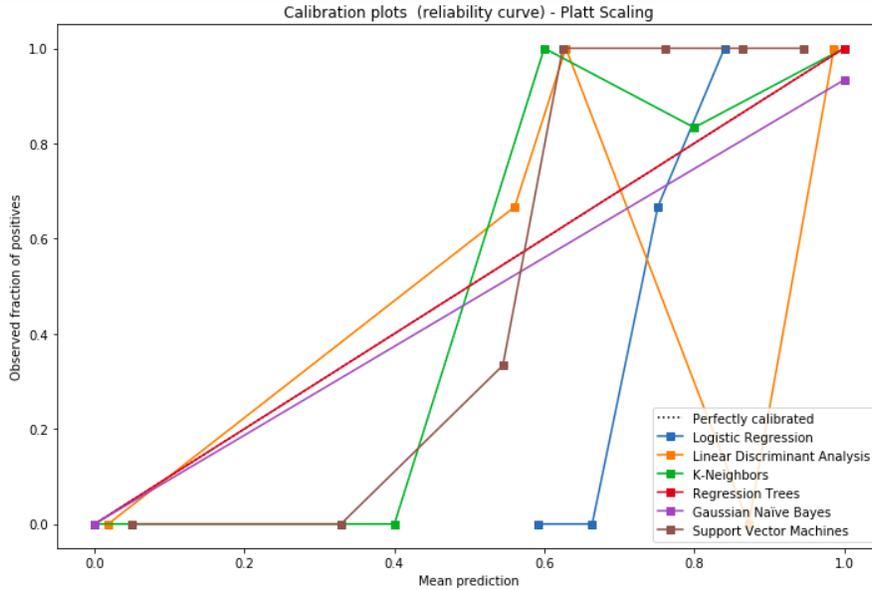
Fig. 3. Calibration plot for the corruption database.

Those results show that, at least for the two databases being evaluated, Regression Trees are the best choice for predicting the outcomes of legal texts. In both data sets, it showed to the the best choice, or being among the best choices available.

4   Conclusions and Future Steps

After the tests done, we can infer that Regression Trees are the best method to predict results in both data sets being analyzed. Even though the other algorithms showed varying results. SVM, as an example, showed a good performance in the corruption database, but the lowest value in the homicides database. Regression Trees have always kept good predicting outcomes. Those results match

Table 3. Results (by mean values) on the corruption database, after 10 epochs k-fold.

Algorithm	Acccuracy	Accuracy (σ)	Recall	Precision	F1 Score
Logistic Regression	0.824269	0.100378	1.000000	0.824269	0.900421
LDA	0.840351	0.108212	0.963918	0.864229	0.907444
K-Neighbors	0.824854	0.096681	0.963918	0.846981	0.897678
Regression Trees	**0.973684**	0.042433	0.988070	0.974167	**0.980524**
Gaussian Naïve Bayes	0.866374	0.101778	0.989474	0.876901	0.925169
Support Vector Machines	0.824269	0.100378	1.000000	0.824269	0.900421
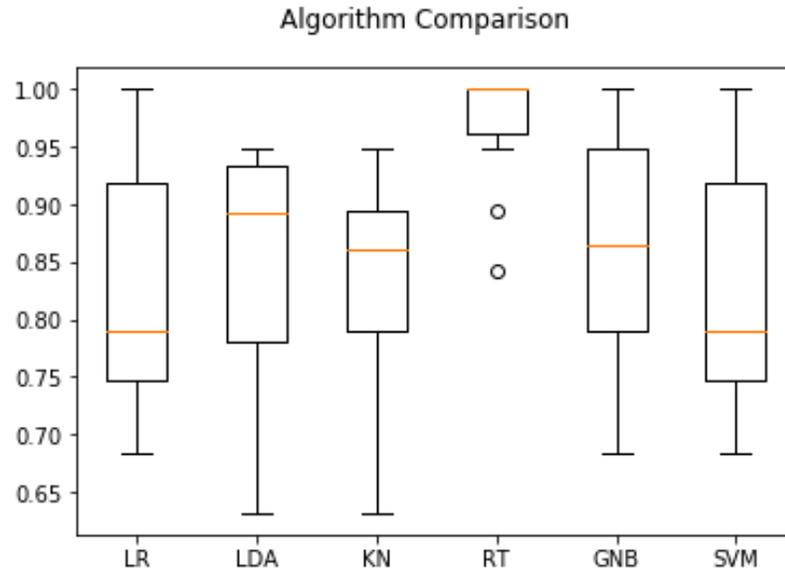
Fig. 4. Candlesticks for the algorithms with 10 k-fold

other results found by other researches in the legal area, in many different countries, such as the one conducted by Kastellec [11], who obtained good outcomes by using Regression Trees in the American legal system. The author mentions that Regression Trees have the capability of studying legal conceptions of Law, revealing patterns that other methods cannot emulate as effectively.

Other researchers also used the same method, such as Rios-Figueroa [17], who used Regression Trees to analyze the concept of judicial independence and corruption among Supreme Courts in Latin America, Antonucci et.al. [4], who adopted Regression Trees to measure the efficiency of Italian courts, and Kufandirimbwa and Kuranga [13], who used the same method to predict outcomes in Zimbabwe.

Those researches show that, even though legal systems are extremely different around the world and throughout different languages and countries, such as Brazil, USA, Italy and Zimbabwe, they do have similar characteristics that can be effectively measured by the correct algorithms.

As future steps, we plan to adopt different methods of converting word embeddings into whole texts, so that we can also utilize methods, such as neural networks, and, eventually, compare these with the ones mentioned in this work.

References

1. Aires, J.P., Pinheiro, D., De Lima, V.S., Meneguzzi, F.: Norm conflict identification in contracts. Artificial Intelligence and Law **25**(4), 397–428 (2017)
2. Alarie, B., Niblett, A., Yoon, A.H.: How artificial intelligence will affect the practice of law. University of Toronto Law Journal **68**(supplement 1), 106–124 (2018)
3. 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**, e93 (2016)
4. Antonucci, L., Crocetta, C., D'Ovidio, F.D.: Evaluation of Italian Judicial System. Procedia Economics and Finance **17**(September 2015), 121–130 (2014). https://doi.org/10.1016/s2212-5671(14)00886-7, http://dx.doi.org/10.1016/S2212-5671(14)00886-7
5. de Araujo, D.A., Rigo, S.J., Barbosa, J.L.V.: Ontology-based information extraction for juridical events with case studies in brazilian legal realm. Artificial Intelligence and Law **25**(4), 379–396 (2017)
6. Barraud, B.: An algorithm that can predict judges' decisions: Toward a robotization of justice? Les Cahiers de la Justice (1), 121–139 (2017)
7. Branting, L., Yeh, A., Weiss, B., Merkhofer, E., Brown, B.: Inducing predictive models for decision support in administrative adjudication. In: AI Approaches to the Complexity of Legal Systems: AICOL International Workshops 2015-2017: AICOL-VI@ JURIX 2015, AICOL-VII@ EKAW 2016, AICOL-VIII@ JURIX 2016, AICOL-IX@ ICAIL 2017, and AICOL-X@ JURIX 2017, Revised Selected Papers. vol. 10791, p. 465. Springer (2018)
8. Fornaciari, T., Poesio, M.: Automatic deception detection in italian court cases. Artificial intelligence and law **21**(3), 303–340 (2013)
9. Gokhale, R., Fasli, M.: Deploying a co-training algorithm to classify human-rights abuses. In: 2017 International Conference on the Frontiers and Advances in Data Science (FADS). pp. 108–113. IEEE (2017)
10. Hyman, H., Sincich, T., Will, R., Agrawal, M., Padmanabhan, B., Fridy, W.: A process model for information retrieval context learning and knowledge discovery. Artificial Intelligence and Law **23**(2), 103–132 (2015)
11. Kastellec, J.P.: The Statistical Analysis of Judicial Decisions and Legal Rules with Classification Trees. Journal of Empirical Legal Studies **7**(2), 202–230 (2010). https://doi.org/10.1111/j.1740-1461.2010.01176.x
12. Katz, D.M., Bommarito, M.J., II, J.B.: A general approach for predicting the behavior of the supreme court of the united states. PloS one **12**(4) (2017)
13. Kufandirimbwa, O., Kuranga, C.: Towards Judicial Data Mining : Arguing for Adoption in the Judicial System **1**(2), 15–21 (2012)
14. Liu, Z., Chen, H.: A predictive performance comparison of machine learning models for judicial cases. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI). pp. 1–6. IEEE (2017)
15. Mochales, R., Moens, M.F.: Argumentation mining. Artificial Intelligence and Law **19**(1), 1–22 (2011)
16. Remmits, Y.L.J.A.: Finding the topics of case law: Latent dirichlet allocation on supreme court decisions. B.Sc., Faculteit der Sociale Wetenschappen (7 2017), supervisors: Kachergis, G.E. ; Kuppevelt, D. van ; Dijck, G. van
17. Rios-Figueroa, J.: Judicial Independence and Corruption: An Analysis of Latin America. SSRN Electronic Journal (212) (2011). https://doi.org/10.2139/ssrn.912924

18. Şulea, O.M., Zampieri, M., Vela, M., van Genabith, J.: Predicting the law area and decisions of french supreme court cases. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017. pp. 716–722 (2017)