2018 42nd IEEE International Conference on Computer Software & Applications

Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning

Michael Benedict L. Virtucio*, Jeffrey A. Aborot†, John Kevin C. Abonita, Roxanne S. Aviñante,
Rother Jay B. Copino, Michelle P. Neverida, Vanesa O. Osiana,
Elmer C. Peramo, Joanna G. Syjuco and Glenn Brian A. Tan
Advanced Science and Technology Institute, Department of Science and Technology
Quezon City, Philippines
Email: *michaelbenedict@asti.dost.gov.ph, †jep@asti.dost.gov.ph

***Abstract*—For the past decades, Philippine courts have been experiencing severe court congestion and case backlog problems. This study aims to provide a solution to alleviate these problems by predicting the outcome of court cases. As the Philippine Supreme Court case decisions are the only readily available data online, we use this as our dataset. We use Natural Language Processing, particularly the bag-of-words model to represent the case text into n-grams. Spectral clustering is also used to group these n-grams into topics. These n-gram and topic features are then input to the machine learning algorithms such as linear support vector machines and random forest classifiers. Linear support vector machine results reached 45% on the n-gram datasets and 55% on the topic datasets. The best result we obtained is 59% on the topic datasets using a random forest classifier. This is the first systematic study in predicting Philippine Supreme Court decisions based purely on textual content.**

I. Introduction

The collaboration between Artificial Intelligence (AI) and the legal domain has been going on for some decades already and goes by names such as AI and Law [1][2] and Jurimetrics [3]. We adopt the term AI and Law in this paper.

One current critical problem area in AI and its application to the legal domain is the case backlog problem common to courts in many countries [4][5]. One of these countries is the Philippines, which has an average of 1 million annual cases per court per year, and 4000 cases per court per day. This massive case backlog problem is highly evident in the lower courts [6]. In this paper, we assume that the semi-automation of the case classification task will aid the Philippine Judiciary in performing their duties.

We propose a Natural Language Processing (NLP) plus Machine Learning (ML)-based approach to predicting case outcome based on the framework used in [7]. We assume the following:

1) words or clusters of words appear in certain distinct contexts making them sufficient in defining case outcomes; and
2) the textual content of published case decisions can be used as proxy for the textual content of records of lodged cases in the court.

In this study, we use case decisions of the Philippine Supreme Court due to the inaccessibility of digital case decisions of the lower courts.

Specifically, we make the following contributions:

1) Once made publicly available, the processed dataset can be used by other researchers in doing related studies on the Philippine Judiciary.
2) We developed various machine learning models for predicting outcome of cases based on the textual content of the case decision (Section VII).
3) We showed that case outcome prediction using only case decisions is plausible (Section VIII).

We discuss related work in Section III and conclude in Section IX.

II. The Challenge

Deciding on cases is a complicated and time-consuming task. It requires the court staff to sift through various records to identify supporting statements for any possible case outcomes. The possible outcome with the strongest support based on the statements in a case decision will be defined as the predicted outcome of the case.

We limit the scope of the study to the set of cases wherein the Supreme Court decision is either affirmed or reversed. The case outcome prediction problem can then be framed as a binary case classification problem where the two classes, violation and non-violation of laws in criminal cases, correspond to the affirmed and reversed decisions of the Supreme Court respectively, as will be further discussed in Section V-B3b on Ground Truth Collection. The classes will hereon be referred to as affirmed and reversed for brevity.

III. Related Work

The work of N. Aletras, D. Tsarapatsanis, D. Preotiuc-Pietro and V. Lampos in [7] has been highly influential in our study. They used an NLP approach to predict the case decisions of the European Court of Human Rights (ECtHR) on cases relating to Articles 3, 6 and 8 of the European Convention on Human Rights. Unlike the theoretical frameworks in previous studies

978-1-5386-2666-5/18/$31.00 ©2018 IEEE
DOI 10.1109/COMPSAC.2018.10348

wherein the court decisions on cases were predicted mainly based on social, political, and economic factors encoded on logical data structures [8][9][10], the framework in [7] predicts the outcomes of cases based only on textual features extracted from case records. The effectiveness of textual content-based analysis in predicting or as a support to predicting case outcome in various contexts is evident in several previous works [11][12].

In [7], each section of the textual content of published case decisions of the ECtHR is modeled using the bag-of-words method in NLP. This produces n-grams, and after spectral clustering, clusters of n-grams called topics, which are used to train linear Support Vector Machine (SVM)-generated models for classifying cases. Each classification model trained for each section are then validated and tested to determine which set of features best predict the outcome of the case. The results of their model testing showed that the textual content of the *Circumstances of the Case* section best predict the outcome of the case. The average accuracy of 76% of the model for the *Circumstances of the Case* section using n-grams, and 78% accuracy of the model for the full/unsectioned text using topics, supports their hypothesis that the textual content and the parts of the case decision document can adequately model the decision-making process of the ECtHR. In this paper, we show that the same hypothesis and premise are supported to some level of accuracy in the context of the Philippine Supreme Court cases.

IV. Framework

We used the methodology of [7] as our model framework for the study with a few modifications to compensate for the differences between the Philippine Supreme Court dataset and the ECtHR dataset. Fig. 1a shows the steps done by [7] while Fig. 1b shows the methodology used in the study. There are some key differences to note.

First is the absence of Case Sectioning in our methodology. [7] used case sectioning to partition case decisions into *Facts of the Case*, *Circumstances of the Case*, *Laws*, and others. They were able to do this since the ECtHR follows a standard format in writing decisions. Since the Philippine Supreme Court does not strictly follow any standard format in penning decisions, it was not applied to the study.

Next is the presence of a Dataset Preparation step in our methodology. This is a necessary step to be added since available Supreme Court decisions online are just sorted according to the month and year they are released with no metadata on which cases tackle similar laws. This step is absent in [7] since the ECtHR has its own online repository of data that has all necessary metadata like the involved states, the articles of the convention that were discussed, etc., and may be filtered accordingly. With the Philippine Supreme Court having no such system, sorting the cases on our own was necessary.

Last is our addition of some processing steps to check for possible variations in results depending on the use or non-use of the said processing. These processing steps are word stemming and class balancing.

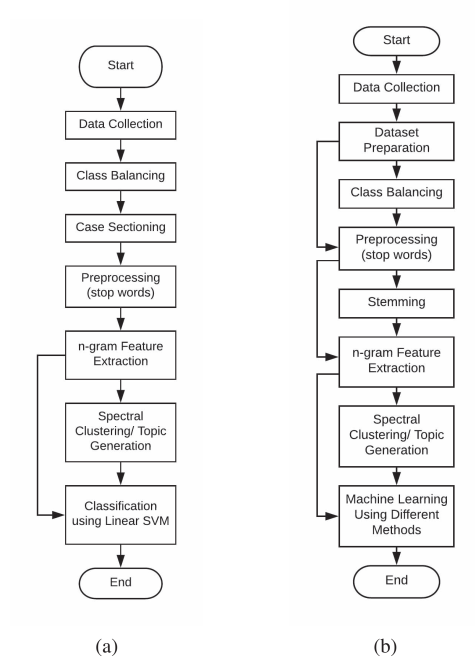

Fig. 1. (a) Framework used in [7], (b) Framework used in the study.

V. Data

A. Data Collection

Historical Philippine Supreme Court case decisions were scraped from the Chan Robles Virtual Law Library (chanrobles.com) and the Lawphil Project (lawphil.net). All cases from 1987 to 2017 were scraped and initially included in the dataset. Some preprocessing was also done including the removal of HTML tags from the downloaded raw documents and the removal of characters that do not conform to the ASCII encoding. Unnecessary headers and footers were also removed to facilitate text processing later on. We obtained a total of 27 492 cases.

B. Dataset Preparation

In [7], the dataset was divided into different articles, with each article pointing to a different aspect of human rights where there could have been a possible violation. To replicate this in the Philippine Supreme Court dataset, we filtered the dataset several times to come up with different subsets, with each subset involving a certain law or set of laws.

1) Filter 1: Criminal and Civil Cases: First, we classified the case decisions as either civil or criminal. This was done by searching for the pattern "People of the Phil" in the case titles/filenames since civil cases would have the name of the disputing parties as case title (e.g. Person A v. Person B), while criminal cases would have a different convention as the accused would be against the prosecutor representing the People of the Philippines (e.g. Person C v. People of the Philippines). This broke down the total cases into 19 360 civil

TABLE I
METADATA

Field	How it was obtained
1 - Case Title/ Name	Obtained from scraping
2 - Case Type	Obtained from Filter 1
3 - Year	Obtained from scraping
4 - Decision	Looking for keywords “so ordered”, “wherefore”, etc. and checking the succeeding words for the patterns “affirm” or “guilt” to mean that the lower court decision was affirmed, which also corresponds to the accused having found to violate certain laws, or “revers” or “acquit” to mean that the lower court decision was reversed hence a non-violation
5 - Classification	Checking the start of the decision for the pattern “appel”, which would correspond to a party being tagged as an appellant or an appellee, hence the case is an appeal; else the case is a petition
6 - Laws (Subfields: republic, act, presidential, batas, commonwealth, article, crime)	Locating patterns/keywords such as “article”/“art”, “republic act”/“ra”, “batas pambansa”/“bp”, “act”, “presidential decree”/“pd”, “commonwealth act”, “crime of”, “offense of”, “charge/charged with” and “violation of” and then extracting the succeeding 20 words
7 - Crime Category	Searching for offenses that are tagged to a certain category of crime according to the Philippines’ Revised Penal Code

cases and 8132 criminal cases. We chose to deal with criminal cases instead of the civil cases.

2) *Filter 2: Appeals and Petitions:* Criminal cases were further divided into appeals and petitions: appeals being cases previously decided by lower courts and escalated to the Supreme Court for reconsideration, while petitions may be for motions filed in the middle of the trial period. As decisions on appeals have finality and deal on the totality of the case, we chose to use these instead of the petitions. This split the criminal case dataset to 6931 appeals and 1201 petitions. Upon further cleaning of the dataset, some duplicates were found. These were removed and resulted to a dataset of 6483 cases.

3) *Filter 3: Laws/Crime Categories:*

a) *Database Generation/Metadata Tagging:* To easily sort the dataset into laws violated/crime categories, a database was created containing several fields on each case decision. Table I shows the metadata collected and stored in the database and how they were obtained.

b) *Ground Truth Collection:* Included in the generation of the database is the automated collection of ground truth as shown in the fourth field (*Decision*). To relate it to [7]’s methodology, they categorized cases as either violations or non-violations of a certain article in the European Convention on Human Rights. In the Philippine Supreme Court criminal case setting, we found this to correspond to affirmed and reversed decisions respectively. This is because it is the criminally accused that appeals a guilty verdict from the lower court to the Supreme Court hence an affirmation would mean he/she is guilty of violating criminal laws, and a reversal would mean he/she is innocent of the crimes charged.

TABLE II
FULL/SKEWED DATASET DISTRIBUTION

Dataset	Affirmed	Reversed	Total
Persons	3997	352	4349
Property	945	94	1039
Public Order	757	114	817
Drugs	755	139	894

c) *Sorting into Laws/Crime Categories:* The database was then queried to further classify the cases in the dataset into articles under the Philippines’ Revised Penal Code (RPC) which they fall under or into the laws possibly violated in a case. For example, the database was queried to return the list of cases that contains the phrase “Revised Penal Code” or “RPC”, and also the phrase “Article 114”; these cases will be tagged under Crimes Against Persons as Articles 114–120 correspond to Crimes Against Persons in the RPC. This was done for all crime categories identified in the RPC. Aside from crime categories listed in the RPC, we also queried the database for cases tackling the same laws. This was done for several laws such as the Dangerous Drugs Act (RA Nos. 6425 and 9165), RA No. 8353 (Anti-Rape Law), and others.

As criminal cases may often involve multiple crimes, it cannot be avoided that a case may be tagged under multiple crime categories as well (e.g. Crimes Against Persons and Crimes Against Property). This results to duplicate cases. To avoid too much duplicity, we hard-coded some restrictions in the queries we used. For example, we restricted cases that fall under certain laws such as the Dangerous Drugs Act from mixing with the RPC cases.

Upon inspection of the distribution of cases thus far, we observed that some crime categories/laws had very few cases tagged under them. Since this is defined as a binary classification problem, we ensured that the crime categories/laws we will be selecting for inclusion in the dataset have at least around a hundred case decisions for either class. To meet this criterion, some sets were dropped, while the others were merged together. This resulted in the use of four laws/crime categories namely: (1) Crimes Against Persons, Crimes Against Chastity and Republic Act 8353 (Anti-Rape Law), (2) Crimes Against Property, (3) Crimes Against Public Order, and (4) the Dangerous Drugs Act. These subsets will hereon be referred to as Persons, Property, Public Order, and Drugs respectively for brevity. The distribution of the skewed/full dataset is shown in Table II.

4) *Class Balancing:* Upon further inspection of the full dataset, we noticed that there is class imbalance since the number of affirmed cases is much greater than the number of reversed cases. To combat this, we made a separate dataset where we decreased the count of affirmed cases to equal that of the reversed cases. The selection of cases was done randomly. The distribution of the balanced dataset is shown in Table III. Note however that both datasets were still used to verify any difference in results.

TABLE III
BALANCED DATASET DISTRIBUTION

Dataset	Affirmed	Reversed	Total
Persons	352	352	704
Property	94	94	188
Public Order	114	114	228
Drugs	139	139	278

VI. NATURAL LANGUAGE PROCESSING

Regular expressions were used to remove punctuation marks, numbers, and any other characters that were not letters. Part-of-speech (POS) tagging was also used to tag the part of speech of the remaining tokens.

We also removed proper nouns and stop words from the case decisions. Stop words are frequent words that do not add any significant meaning to the text. The NLTK library has a corpus of proper nouns and English stop words [13], and added to that, we also included a list of Tagalog stop words.

A. Feature Extraction

1) n-grams: We used the bag-of-words model to represent the case decision dataset. For each subset, we obtained the top-2000 most frequent n-grams for $n \in \{1\}, n \in \{1,2\}, n \in \{1,2,3\}$, and $n \in \{1,2,3,4\}$. This resulted to several matrices with size $m \times 2000$, where m is the number of cases in the subset. Each element in the matrix corresponds to the normalized frequency of that particular n-gram (column) for that particular case (row). This, combined with stemmed and unstemmed versions of the dataset resulted to eight different datasets so far, with a combined total of 32 subsets. The distinction between the full/skewed and balanced datasets, raised this total to 16 datasets with a total of 64 subsets.

2) Topics: We also created topics for each article by clustering together n-grams that appear in similar contexts. We did this by computing the cosine metric of the datasets obtained thus far and from there, created n-gram $\times$ n-gram similarity matrices. We then applied spectral clustering on the similarity matrices to obtain 30 clusters/topics for each subset. Note here that since this is hard clustering, an n-gram can only be part of a single topic. 16 topic datasets were generated with a total of 64 subsets.

B. Consolidation of Datasets

After feature extraction, a total of 32 datasets were generated with a combined total of 128 different subsets.

VII. MACHINE LEARNING

A. Support Vector Machines

As [7] showed good results when SVM was used for text classification, we also used it for our study. We used four different kernels to build the models: linear, radial basis function (RBF), polynomial, and sigmoid. Grid search was used to find the optimal parameters for each kernel. The dataset was split into the training and test sets with 80% being used for training and 20% used for testing. We also used 10-fold cross-validation.

Upon implementation of the different SVM kernels, we found that the level of accuracy across kernels did not deviate much and that the same trends in results were found. We will be discussing these trends in Section VIII but only for linear SVM to be able to directly compare it to the results of [7]. Also, as expected, the results on the full datasets across all kernels had the machine classifying all cases as affirmed. Again, this is caused by the huge class imbalance. We thus opt not to include these results anymore in Section VIII.

B. Other Methods

We also tried using logistic regression and Naïve Bayes for our datasets but results will not anymore be shown as we found them to be inconclusive.

TABLE IV
LINEAR SVM RESULTS ON N-GRAM DATASETS

Method*		Persons	Property	Public Order	Drugs	Average
NS	n=1	**0.49**	0.37	0.43	0.38	0.42
NS	n=2	0.48	0.39	0.46	0.45	**0.45**
NS	n=3	0.48	0.37	**0.48**	0.45	**0.45**
NS	n=4	0.47	0.37	**0.48**	**0.48**	**0.45**
S	n=1	0.46	**0.42**	0.46	0.41	0.44
S	n=2	**0.49**	0.37	**0.48**	0.43	0.44
S	n=3	0.48	0.39	0.46	0.45	**0.45**
S	n=4	**0.49**	0.39	0.46	0.45	**0.45**

*NS - no stemming; S - with stemming
*n - maximum size of n-gram obtained

TABLE V
LINEAR SVM RESULTS ON TOPIC DATASETS

Method*		Persons	Property	Public Order	Drugs	Average
NS	n=1	0.52	0.45	0.52	0.50	0.50
NS	n=2	0.46	**0.58**	**0.59**	0.57	**0.55**
NS	n=3	0.43	0.50	0.54	**0.66**	0.53
NS	n=4	0.39	0.50	0.50	0.48	0.47
S	n=1	0.43	0.53	0.46	0.48	0.48
S	n=2	0.48	0.50	0.52	0.61	0.53
S	n=3	**0.53**	0.45	0.52	0.45	0.49
S	n=4	0.48	0.45	0.50	0.59	0.51

VIII. RESULTS AND DISCUSSION

Tables IV and V show the results from the runs of linear SVM on both the n-grams and topic datasets. Bold font denotes best accuracy in a particular law/crime category or on average across laws/crime categories.

With [7] having an average accuracy of 75% on their full/unsectioned n-gram dataset, we can say that our results are not on the same level as theirs. Table IV in particular shows rather poor results with a test set accuracy not even reaching 50% or random chance in any subset and only reaching 45%

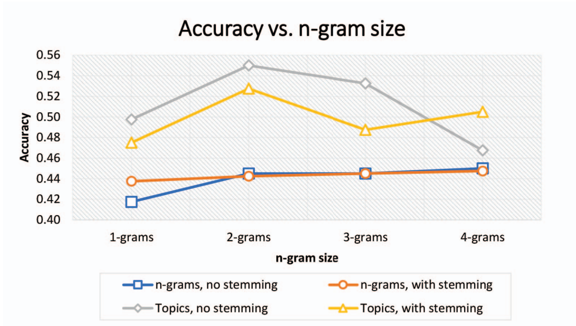

Fig. 2. Accuracy vs. n-gram size

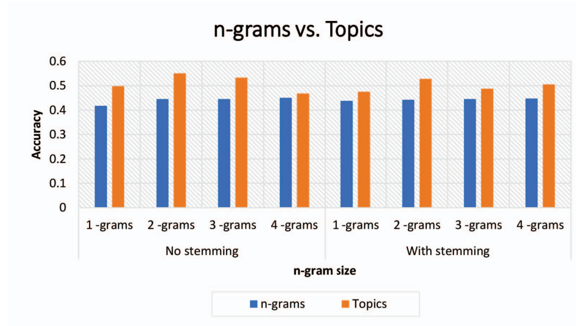

Fig. 3. n-grams vs. Topics

on average. This discrepancy in the level of test set accuracy can be attributed to several reasons:

a) Highly structural case decision of ECtHR: ECtHR case decisions have a standard format, and this was used by [7] to discover that the section pertaining to the circumstances of the case gives the best results. Any insight as to whether the same can be said for Philippine Supreme Court cases, or whether proper sectioning would have made an improvement to the prediction of the machine cannot be made since Philippine Supreme Court cases do not follow any strict format on the penning of decisions.

b) Highly specific contextual content of ECtHR case decisions: The ECtHR is also highly specialized in the sense that only cases on human rights reach it. This leads to highly specific contextual content of the case decisions and this may have had a significant effect on the n-grams obtained and their frequencies in the different texts.

c) Room for improvement in dataset preparation compared to properly indexed cases of ECtHR online: Dataset preparation and ground truth collection were done on our own, in a semi-automated way. As we only picked patterns to search and based the sorting of cases on the presence or absence of these patterns in the text, we were not able to inspect the context in which these patterns appear in the case decisions. For example, a case text may have mentioned a certain crime/law in passing, but is more heavily focused on a different crime/law in actuality; our current sorting method will sort the case under both laws/crime categories. With the absence of a domain expert on our team, this is the assumption we made: that a single mention of a certain law or crime in a case decision would be enough reason to sort it under that law or crime category. With the results we obtained, this may not have been a strong assumption.

Additional insight can also be gleaned from the results in Tables IV and V: (1) there doesn't appear to be any correlation between test set accuracy and stemming, (2) there also doesn't appear to be any strong correlation between test set accuracy and increasing or decreasing n-gram size but the results on 2-grams datasets display better accuracy than others. Correlation between test set accuracy and n-gram size is more clearly shown in Fig. 2.

TABLE VI
RANDOM FOREST RESULTS ON TOPIC DATASETS

Method*		Persons	Property	Public Order	Drugs	Average
NS	n=1	0.48	0.58	0.57	0.54	0.54
NS	n=2	**0.52**	**0.68**	**0.61**	0.54	**0.59**
NS	n=3	0.48	0.47	0.50	0.50	0.49
NS	n=4	0.51	0.47	0.43	0.39	0.45
S	n=1	0.47	0.53	0.50	0.48	0.50
S	n=2	0.45	0.66	0.59	0.54	0.56
S	n=3	0.49	0.58	0.59	0.50	0.54
S	n=4	0.50	0.53	0.52	**0.62**	0.54

[7] displayed better results on their topic dataset. They attributed this to topics forming a more abstract and better representation of the information contained in each case. The same trend can also be observed in our results listed in Table V and as more clearly shown in Fig. 3.

With the topic dataset showing much potential, we tried using this dataset as an input to a random forest classifier to gain insight on whether an ensemble of decision trees that chooses between the frequency or lack of these topics in a case to predict its decision can do better than linear SVM. Deciding based on clusters of n-grams that appear in similar contexts also seem more sensible compared to deciding simply on the frequency of a certain individual n-gram. Similar to the SVM methodology, the dataset was split 80-20 for the training and test sets. The results are shown in Table VI.

Table VI shows a big improvement from Table IV and even from Table V results. Here, a subset test set accuracy can reach as high as 68% and even 59% on average.

Future Direction

Even with the aforementioned problems and issues on dataset preparation that may have hurt the test set accuracies, the improvement of results on the topic datasets as shown in Table V, and even further with the use of a random forest classifier as shown in Table VI, shows that the study has enough promise to be continued further. The next application would be to the lower courts in the Philippines as this is where case backlog is most severe according to a consultation we had with a legal researcher from the lower court.

As we learned from a law professor from the University of the Philippines in one of our consultations, cases in the lower courts are very different from cases dealt with by the Supreme Court. Lower court cases deal with questions of facts (e.g. "Where were you at this time?", "How bright was the light at this time?") while the Supreme Court deals with questions of law (e.g. "Is this evidence admissible?", "Does this law apply to this case?"). The Supreme Court does not deal anymore with questions of facts as they tend to agree already with the lower court judges' determination on these matters. As the Supreme Court deals with questions of law, which are more abstract than questions of facts, the current bag-of-words model may not have been capable enough to sufficiently represent these matters. Dealing with lower court data and questions of facts may be more easily represented and may be less abstract, thus possibly leading to better results using the current framework.

The main challenge in continuing this study using lower court data, is the lack of lower court data in digital format. During an interview we had with the Office of the Court Administrator of the Supreme Court, we found out that efforts to automate court processes, and along with it digitize court data, is still ongoing. This means that currently, lower court case decisions and case files in general, still exist purely in hard copy form. It is of utmost importance therefore, that automation of court processes, and digitization of court data, be the first step accomplished before studies involving lower courts can be made.

Another possible route is looking at other junctures of a case that can also be fitted with an AI/ML solution instead of such a high-level task such as deciding on a case. Currently, the US Judiciary, which the Philippine Judicial System is highly influenced by, has AI solutions in areas like pretrial and bail, criminal sentencing, and parole [14]. An assessment on whether these solutions can also be applied to the Philippine setting, and whether these solutions: (1) will have high impact especially on case backlog, (2) is a current need of the judiciary, and (3) if the judiciary is ready to adopt these solutions, should be made.

IX. Conclusion and Recommendations

We conducted an experiment on Philippine Supreme Court case decisions and applied NLP and ML approaches to predict the case outcome. Results from linear SVM yielded an accuracy of 45% at best on n-grams datasets and 55% on topic datasets. Succeeding runs using a random forest classifier yielded the best accuracy of 59%. Future direction of this work is targeted toward lower court applications. However, there is a problem with the lack of lower court data in digital format, as efforts on court automation and case file digitization are still ongoing.

Future work can also improve on dataset size and preparation; domain experts can also be tapped to manually sort case decisions into laws/crime categories with more credibility. Other AI methods (e.g. deep learning) can also be used.

References

[1] E. L. Rissland, K. D. Ashley, and R. P. Loui. "AI and Law: A fruitful synergy". In: *Artificial Intelligence* 150.1-2 (2003), pp. 1–15.

[2] T. Bench-Capon et al. "A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law". In: *Artificial Intelligence and Law* 20.3 (2012), pp. 215–319.

[3] L. Loevinger. "Jurimetrics: Science and prediction in the field of law". In: *Minn. L. Rev.* 46 (1961), p. 255.

[4] *Supreme Judiciary And The Backlog Of Cases*. 2013. URL: https://www.lawteacher.net/free-law-essays/constitutional-law/supreme-judiciary-and-the-backlog-of-cases-constitutional-law-essay.php?vref=1.

[5] L. Garcia-Navarro and P. Moura. *Brazil: The Land Of Many Lawyers And Very Slow Justice*. 2014. URL: https://www.npr.org/sections/parallels/2014/11/05/359830235/brazil-the-land-of-many-lawyers-and-very-slow-justice (visited on 04/11/2018).

[6] *Philippine judiciary and criminal justice system under pressure: An inside look*. 2017. URL: https://www.aseantoday.com/2017/02/philippine-judiciary-and-criminal-justice-system-under-pressure-an-inside-look/.

[7] N. Aletras et al. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective". In: *PeerJ Computer Science* 2 (2016), e93.

[8] S. Bruninghaus and K. D. Ashley. "Predicting outcomes of case based legal arguments". In: *Proceedings of the 9th international conference on Artificial intelligence and law*. ACM. 2003, pp. 233–242.

[9] A. D. Martin et al. "Competing approaches to predicting supreme court decision making". In: *Perspectives on Politics* 2.4 (2004), pp. 761–767.

[10] D. M. Katz, M. J. Bommarito II, and J. Blackman. "A general approach for predicting the behavior of the Supreme Court of the United States". In: *PloS one* 12.4 (2017), e0174698.

[11] Y. Sim, B. Routledge, and N. A. Smith. "Friends with Motives: Using Text to Infer Influence on SCOTUS". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016, pp. 1724–1733.

[12] S. Agrawal et al. "Affirm or Reverse? Using Machine Learning To Help Judges Write Opinions". In: (2017).

[13] S. Bird, E. Klein, and E. Looper. *Natural Language Processing with Python*. O'Reilly Media, 2009.

[14] A. Christin, A. Rosenblat, and D. Boyd. "Courts and Predictive Algorithms". In: Data & Civil Rights: A New Era of Policing and Justice. (Washington, DC, Oct. 27, 2015). Oct. 2015. URL: http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf (visited on 01/26/2018).