Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 2393–2402

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Predicting Outcomes of Legal Cases based on Legal Factors using Classifiers

Rafe Athar Shaikh[a,*], Tirath Prasad Sahu[a], Veena Anand[b]

[a]*Department of Information Technology, NIT Raipur, Raipur-492010, India*
[b]*Department of Computer Science and Engineering, NIT Raipur, Raipur-492010, India*

Abstract

Predicting outcomes of legal cases may aid in the understanding of the judicial decision-making process. Outcomes can be predicted based on i) case-specific legal factors such as type of evidence ii) extra-legal factors such as the ideological direction of the court. The details of case-specific legal factors can be extracted from legal judgments. However, extracting these factors from legal texts is a tedious and time-consuming process. In this work, important factors affecting outcomes of murder related cases (taken from Delhi District Court) are identified and a database of 86 cases is prepared in order to use these factors as descriptors for outcome prediction. The outcome prediction is seen as a binary classification problem for classes 'Acquittal' and 'Conviction' of the accused person. Conventional machine learning classification algorithms are applied and Leave-one-out cross validation is used to produce the results. The performance of classifiers is evaluated and compared using metrics such as Accuracy, Precision, Recall, and F1 Score. The statistical distribution of features and the experimental results (Accuracy ranging from 85% to 92% and F1 Score from 86% to 92% for classifiers) show the success in identifying important factors of concerned cases and in turn predicting their outcomes.

Keywords: Outcome Prediction; Legal Cases; Legal Factors; Machine Learning; Classifiers; Classification Algorithms

* Corresponding author. Tel.: +91-8983444208
E-mail address: rafeatharmateen@gmail.com

1877-0509 © 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2019).
10.1016/j.procs.2020.03.292

1. Introduction

Human beings need to predict certain things in various aspects of their lives. The domain of law is no exception. Usually, the process of prediction starts with one or more questions. Whether to take a case in hand or not? Whether to settle the case outside or take it to the court? Will the settlement amount be worth it? What are the chances of winning the case? These are some of the questions that involve predicting the outcome of a case and the legal practitioners have to deal with, on a regular basis. These questions represent the importance of outcome prediction in case selection, making settlement decisions, and various aspects of legal processes [1]. Legal practitioners have been analyzing the cases retrospectively to identify and understand the elements or factors that play an important role in producing judgments. However, analyzing judgments after their declaration is not the only method to understand the decision-making process as noted in [2] that the explanatory theories ought to be tested against future outcomes as well. Hence, predicting outcomes of judgments and then analyzing successful and unsuccessful predictions of outcomes may gain us important insights into the decision making of legal cases [3].

The interpretation and implementation of the law are heavily dependent on legal texts. Hence, decision making in a legal case requires going through lots of legal documents. However, extracting relevant information from the legal text documents is a tedious and extremely time-consuming job due to their special characteristics like longer document size, a wide range of internal structure, extensive manual indexing, a complex pattern of relationships between documents and great reliance on citations [4]. It makes the task of outcome prediction difficult. Various legal factors affect the outcome of a legal case and these factors may differ greatly from one case type to another. Considering all the factors is infeasible for any predictive model. Hence, the factors that seem most relevant and useful for the problem are usually taken into account.

1.1. Approaches to outcome prediction of legal cases

Machine learning algorithms learn from past experience and have the ability to produce output when provided with new inputs. The input can vary greatly in its form, nature, and type and hence, can be used as a basis for grouping machine learning models. Accordingly, the scholarly works on legal case outcome prediction can be classified into the following categories based on the type and nature of descriptors (features) considered for predicting the outcome.

1) Political Science or Social Science Approach: Considering the fact that judges are human beings, this approach assumes that judges may be ideologically inclined towards some side in various issues, have their own perception and other biases, and various other social and political factors that affect their judgment such as mental resources of a judge [5] and decision of lower court [6]. Hence, descriptors considered for predicting outcomes are extra-judicial factors that may generate or represent human bias. Examples of these factors are votes of other justices [7], justice gender, case origin [8], petitioner type, respondent type, the ideological direction of the court [9], etc. The research works following this approach are [5-10].

2) Linguistic based Approach: Another approach considers the linguistic features of legal judgments for predicting outcomes. Ngo [11] tried to predict outcomes on a database of 2019 Dutch legal cases according to their linguistic features. Some of the considered features were word count and frequency of different types of pronouns. In [12], the authors replaced case-specific names and instances by their role and used propositional patterns for predicting trade secret cases. Authors of [13] used unigrams and bigrams coupled with word type token to predict case ruling, law area, and timespan of French Supreme Court cases. Decisions of the European Court of Human Rights were predicted in [14] using N-grams and topics as features. Neural networks were used in [15] to predict charges of criminal cases of China. Documents were divided into three parts namely fact description, articles, and charges. Various combinations of fact descriptions and articles were tried as input. Liu and Chen used sentiment analysis to predict charges [16].

3) Legal Approach: A natural approach is predicting outcomes based on the case-specific legal factors of concerned legal cases. The descriptors of outcome in this approach depend on the nature of legal cases under consideration. In general, this approach works as follows. Identify important factors of the concerned cases, extract these feature values from each case text, prepare a database and apply machine learning algorithms to predict the outcome. Usually, the features are extracted manually by reading and analysis of the judgment. It is technologically

tough to extract features automatically due to the inherent complexity of legal texts. The values for each fact descriptor may be Boolean, numbers or values from a predefined set of features depending on the representation scheme. As the manual feature extraction is time consuming, data of 40 to 200 legal cases are generally collected. The same approach is adopted in this work. Research works following this approach are presented in the related work section.

The remaining part of the paper is organized as follows. The study on related work is presented in Section 2. The research methodology adopted is introduced in Section 3. Section 4 shows the impact analysis of the extracted features. Section 5 discusses the results obtained. Section 6 presents the conclusion and future scope.

2. Related work

Previous research works relying on case specific details of legal cases for outcome prediction are taken into account. One of the earliest outcome prediction models [17] used the nearest neighbor approach using 46 descriptors in 64 cases related to tax. The similarity metric used in this program was simply the no. of the features having the same values in the cases. A later work called SHYSTER [18] selected nearest neighbors by allocating weights to fact descriptors and using a more complex similarity metric.

For this seeming classification problem, Haar et al. [19] applied regression analysis with the dependent variable as the outcome for zoning amendment cases in Connecticut. Initially considered 167 features were reduced to 32 features and provided as input for the regression algorithm. Zeleznikow and Hunter applied the decision tree classifier to debt deferral cases using 5 case descriptors [20]. Originally created to make arguments, CATO [21] identified issues related to the factors applicable to the case, found the best cases according to relevance criterion and predicted the outcome using these cases. The database consisted of 184 trade secret misappropriation cases.

Brüninghaus and Ashley introduced an algorithm called IBP [22] and predicted 186 trade secret law cases. The model identified the issues raised in a case, analyzed which side the issue favored and predicted outcomes by combining analysis of these individual issues. Research work by Chorley and Bench-Capon [23] produced results comparable to IBP via assigning weights to factors and inducing rules from cases. Extending previous work of IBP, Ashley, and Brüninghaus were successful in developing a program called SMILE+IBP [24] that could reason about the legal case text automatically. The program classified the text according to facts and then predicted outcomes based on factors extracted from the classified texts.

Grabmair developed a system called VJAP [25] that predicted the outcomes of trade secret cases and could justify these predictions through legal arguments. The results were compared with the performance of IBP and incorrectly predicted cases were also analyzed. Christensen [26] predicted the behavior of the US Supreme Court in Indian Law cases. Logistic regression was applied and the statistical significance of 12 descriptors was identified on a dataset of 156 cases.

3. Research methodology

In brief, the research methodology adopted is as follows. First, the court and the type of legal cases to be predicted are selected. Second, important case factors that can best represent the problem of outcome prediction are identified. Third, these features are extracted by reading and analysis of judgments. Fourth, the extracted data is pre-processed to make it suitable for machine learning classification algorithms. Next, classification algorithms are applied to collected data and their performance is evaluated using relevant metrics. Finally, classifiers are compared based on their performance and results are analyzed. Fig. 1 and Algorithm 1 provide details about the research methodology.

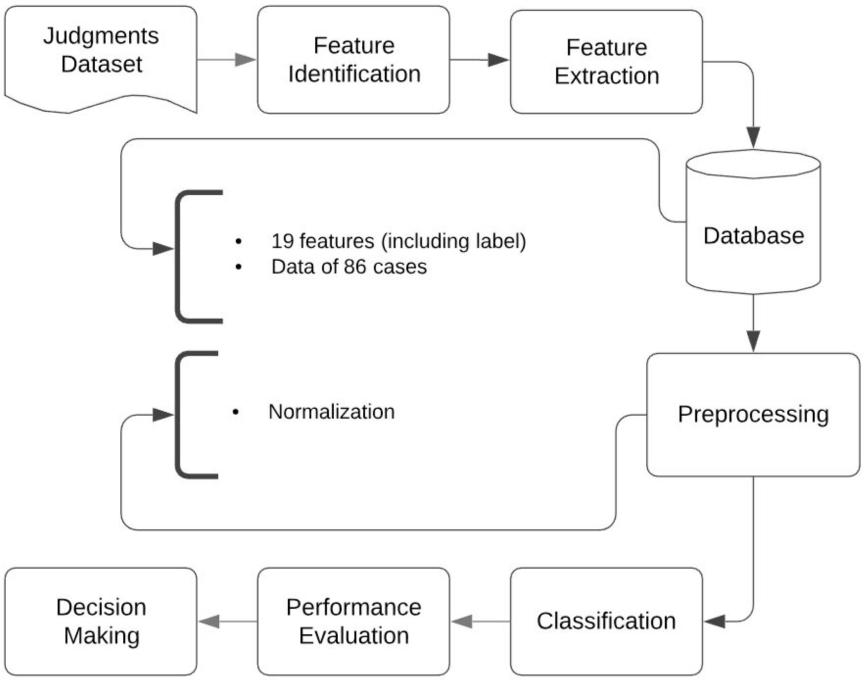

Fig. 1. Research methodology.

Algorithm 1. Research Methodology

1. **Case Selection:** Fetch judgments of concerned cases.
2. **Feature Identification:** Read judgments and identify important factors affecting the outcome.
3. **Feature Extraction: For** each judgment **do**
   i. Read and extract features identified in step 2.
   ii. Represent features as per the representation scheme.
4. **Preprocessing:** Perform preprocessing on the data.
5. **Classification:** Apply classification algorithms and build models.
6. **Performance Evaluation:** Calculate relevant metrics for each model to measure performance at prediction.
7. **Comparison and Selection:** Compare results and select the best model.

END.

3.1. Description of dataset

The dataset consists of cases from the Delhi District Court. The cases selected are criminal cases related to the murder. The judgments considered are of the year 2017 and 2018 made available by legal publisher Indiankanoon [27]. Data of the total 86 cases are collected.

3.2. Feature identification

After reading and analyzing lots of cases and taking inputs from legal professionals, the factors that play an important role in deciding the outcome in concerned cases are identified. List of these factors (features) and their explanation are provided in Table 1.

Table 1. Description of features.

S. No.	Features	Description
1	alive	Injured alive or dead?
2	identified	Victim/injured identified accused?
3	evidence	Type of evidence (ocular/circumstantial)
4	motive	Motive absent/presented/established by the prosecution?
5	npw	No. of public witnesses
6	nimpw	No. of important witnesses (non-police, informal witnesses)
7	new	No. of eye witnesses
8	newsupp	No. of eye witnesses supporting prosecution
9	ewcontra	Contradiction in statements of eye witnesses?
10	notherw	No. of important witnesses that are not eye witnesses (other)
11	nothsupp	No. of other witnesses supporting the prosecution
12	othercontra	Contradiction in statements of other witnesses?
13	ndw	No. of defense witnesses
14	ndwsupp	No. of defense witnesses supporting accused
15	dwcontra	Contradiction in statements of defense witnesses?
16	newsuppbynew	Supporting eye witnesses among all eye witnesses
17	nothsuppbynotherw	Supporting other witnesses among all other witnesses
18	ndwsuppbyndw	Supporting defense witnesses among all defense witnesses
19	outcome	Accused convicted or acquitted?

3.3. Feature extraction

Features are extracted through manual reading and analysis of judgments. The example below gives a brief idea about the process of feature extraction. Words in the bold, shed light on respective features (contained in braces). Not all the features are included in the example as some of them require analysis of the whole judgment.

Example “Vinod Kumar Gupta father of **deceased** (alive, identified) met them on the spot …The prosecution examined **27 witnesses** (npw) to prove its case…Smt. Jyoti Gupta was examined as **DW-1** (ndw)…Ld. Defense counsel submitted that it is not the case based upon the ocular evidence. There is **no eye witness (**new, newsupp, ewcontra, evidence**)**. The case is based upon **circumstantial evidence** (evidence)…Vinod Kumar and Sushma were examined as PW-9 and PW-10. They are the parents of the deceased but they have also **not supported** (nothsupp, nothsuppbynotherw) the prosecution case. Ld. Counsel submitted that under the circumstances **motive itself is not proved and established** (motive)…It is established that it was the accused who murdered his wife by strangulating her. I, therefore, hold him guilty and **convict** (outcome) him for the offense punishable u/s 302 IPC."

3.4. Preprocessing

As the features are spread over different ranges compared to one another, min-max normalization is applied to these features to make them suitable for classification algorithms. A value x of a feature can be normalized as follows.

$$xnormalized = \frac{x - x\min}{x\max - x\min} \tag{1}$$

Where xmin is minimum and xmax is the maximum value in the observed values of the feature under consideration.

4. Feature impact analysis

The characteristics of features for classes ‘Acquittal’ and ‘Conviction’ are investigated through various measures. Table 2 and Table 3 show class wise statistical details of the actual and normalized data. The central tendencies (mean, median) and standard deviation of the extracted features represent the characteristics of classes 'Acquittal' and 'Conviction'.

At first glance, it seems that greater value for features of S. No. 9, 12-14 and 18 should support the accused, while the greater value for remaining features should lead to Conviction. It is evident from the statistical distribution that the features are following this intuition. There are some exceptions though, as seen in features of S. No. 6 and 13-14 i.e. features nimpw, ndw, and ndwsupp are behaving otherwise. However, a closer look explains this behavior. It may happen that a good no. of witnesses are executed but there are some contradictions in their statements that can lead to the advantage of the opposite party. Features of S. No. 9, 12 and 15 fill this gap.

Table 2. Characteristics of the actual dataset.

S. No.	**Features**	**Acquittal**			**Conviction**		
		Median	Mean	SD	Median	Mean	SD
1	alive	1	0.512	0.506	1	0.558	0.502
2	identified	0	0.163	0.374	1	0.581	0.499
3	evidence	0	0.233	0.427	0	0.349	0.482
4	motive	0	0.07	0.258	0	0.721	0.908
5	npw	10	12.302	7.933	15	17.86	9.086
6	nimpw	3	2.791	1.521	3	3.535	1.882
7	new	1	1.744	1.347	2	1.93	1.595
8	newsupp	0	0.442	0.854	1	1.837	1.632
9	ewcontra	1	0.791	0.412	0	0.07	0.258
10	notherw	0	1.047	1.43	1	1.605	2.227
11	nothsupp	0	0.535	1.099	1	1.419	2.003
12	othercontra	0	0.256	0.441	0	0.163	0.374
13	ndw	0	0.419	0.879	0	0.674	0.919
14	ndwsupp	0	0.163	0.485	0	0.326	0.644
15	dwcontra	0	0.047	0.213	0	0.116	0.324
16	newsuppbynew	0	0.217	0.391	1	0.721	0.454
17	nothsuppbynotherw	0	0.202	0.384	0.5	0.474	0.483
18	ndwsuppbyndw	0	0.109	0.306	0	0.233	0.427

Table 3. Characteristics of the normalized dataset.

S. No.	**Features**	**Acquittal**			**Conviction**		
		Median	Mean	SD	Median	Mean	SD
1	alive	1	0.512	0.506	1	0.558	0.502
2	identified	0	0.163	0.374	1	0.581	0.499
3	evidence	0	0.233	0.427	0	0.349	0.482
4	motive	0	0.035	0.129	0	0.36	0.454
5	npw	0.157	0.202	0.156	0.255	0.311	0.178
6	nimpw	0.3	0.279	0.152	0.3	0.353	0.188
7	new	0.2	0.349	0.269	0.4	0.386	0.319
8	newsupp	0	0.088	0.171	0.2	0.367	0.326
9	ewcontra	1	0.791	0.412	0	0.07	0.258
10	notherw	0	0.116	0.159	0.111	0.178	0.247
11	nothsupp	0	0.076	0.157	0.143	0.203	0.286
12	othercontra	0	0.256	0.441	0	0.163	0.374
13	ndw	0	0.105	0.22	0	0.169	0.23
14	ndwsupp	0	0.081	0.242	0	0.163	0.322
15	dwcontra	0	0.047	0.213	0	0.116	0.324
16	newsuppbynew	0	0.217	0.391	1	0.721	0.454
17	nothsuppbynotherw	0	0.202	0.384	0.5	0.474	0.483
18	ndwsuppbyndw	0	0.109	0.306	0	0.233	0.427

5. Experimental results

In this section, evaluation parameters for performance of classifiers are introduced and results are discussed. The dataset of 86 cases is balanced with 43 entries each for classes 'Acquittal' and 'Conviction'.

5.1. Evaluation parameters

The prediction of outcome is seen as a binary classification problem. Hence, performance metrics such as Accuracy, Precision, Recall, and F1 Score are considered for evaluating the correctness of prediction results. These measures are calculated by formulae given below.

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \tag{5}$$

Where TP, FP, TN, and FN in the context of this work are as mentioned in Table 4. Hence, Precision is the proportion of actual conviction cases among all the cases in which conviction is predicted. Recall is the proportion of cases in which conviction is predicted among all the actual conviction cases. F1 Score conveys the balance between Precision and Recall and hence, it is a good measure to evaluate the performance of classifiers.

5.2. Discussion of results

Conventional classification algorithms belonging to different categories are applied with Leave-one-out cross validation for generating predictions. The predictions are analyzed from the following perspectives: 1) Predictions on all cases by each classifier, and 2) Predictions on each case by all classifiers.

Table 4. Confusion matrix.

	Predicted Acquittal	**Predicted Conviction**
Actual Acquittal	True Negative (TN)	False Positive (FP)
Actual Conviction	False Negative (FN)	True Positive (TP)

1) Algorithm Wise Analysis: Table 5 shows the Accuracy, Precision, Recall and F1 Score for applied classifiers. Classification and Regression Trees (CART) is performing best in terms of accuracy while Bagging and Random Forest (RF) share the second rank. Bagging is the most precise in predicting the conviction cases with CART and RF being second most precise. In terms of identifying most of the conviction cases, Support Vector Machines (SVM) is at the forefront followed by k-Nearest Neighbors (kNN), CART and Naïve Bayes (NB) with equal recall rate for conviction cases. Logistic Regression (LR) is consistent for all metrics. In terms of being able to identify most of the conviction cases and being precise at the same time, CART is again most successful as F1 Score suggests.

The predictions of classification algorithms are analyzed in detail as well. There are 2 cases in which only 1 classifier was correct. CART and SVM each predicted one case correctly in which all other classifiers failed. There are 5 cases in which only 1 classifier predicted incorrectly. Among these 5 cases, SVM failed in 2 cases, NB in 2 cases, and kNN in the remaining 1 case. Similarly, 5 cases are there in which only 2 classifiers were incorrect. Among these 5 cases, SVM was unsuccessful in 3 cases, while NB and LR each in 2 cases. It can be seen from these numbers that SVM is behaving quite differently than other algorithms, whether in the positive or negative sense. NB may be ranked after SVM for predicting incorrectly when most of the algorithms were correct. CART is the only classifier that does not come in these 2 lists (classifiers having incorrect predictions when other classifiers were correct). Hence, CART should be most accurate as it is predicting correctly when others failed, and is also not among those classifiers that failed when most of the other classifiers were correct. These numbers are in alignment with the results presented in Table 5.

Table 5. Performance of classifiers.

Classifier	Accuracy	Precision	Recall	F1 Score
LR	88.37	88.37	88.37	88.37
kNN	88.37	86.67	90.70	88.64
CART	**91.86**	92.86	90.70	**91.76**
NB	84.88	81.25	90.70	85.71
SVM	86.05	81.63	**93.02**	86.96
Bagging	90.70	**94.87**	86.05	90.24
RF	90.70	92.68	86.37	90.48
Boosting	88.37	90.24	86.05	88.10

In brief, CART is performing best in terms of both Accuracy and F1 Score. However, it should be noted that all the classification algorithms are performing well as evident from accuracy ranging between 85% and 92% and F1 Score ranging between 86% and 92%.

2) Case Wise Analysis: Predictions by the classifiers for each case are also analyzed. Table 6 presents the count of cases versus the no. of classifiers that correctly predicted those cases. It is found that 64 cases are correctly predicted by all 8 classifiers. Hence, 22 cases are having at least 1 incorrect prediction.

There are 2 cases that were incorrectly predicted by all the classifiers. These incorrectly predicted cases are actually the same two other cases already present in the database, but the details of two different accused persons were collected in these two entries (as the data for different accused is collected, we considered them as different cases instead of the same case). Most of the facts in a case are generally the same for all accused persons in that case. Hence, a small no. of factors differ for these accused persons. These wrongly predicted entries had a different (opposite) outcome from their respective co-case entries present in the database for different accused.

Table 6. Count of cases according to the number of classifiers correctly predicting those cases.

Correct Predictions	0	1	2	3	4	5	6	7	8
Number of Cases	2	2	1	2	2	3	5	5	64

There are 5 entries like these (entries for the same 5 cases already present in the database but for different accused). As mentioned above, 2 entries were incorrectly predicted by all classifiers but the remaining 3 cases are correctly predicted by all classifiers. In these entries, the outcome was the same for the accused as in respective entries for other accused. From the prediction results of these 5 cases, it is clear that this model is weak in predicting the cases where more than one accused is present and outcomes differ for the accused persons.

Cases with only 1 and 2 correct predictions are also analyzed. The first category cases were based on circumstantial evidence and the later on ocular evidence. All the cases had some contradictions in statements of witnesses. These results seem justifiable as the nature and intensity of contradictions could not be taken into account. The contradictions are somewhat subjective in nature and human beings might differ at times in their interpretation. Some more factors need to be included that can inculcate more details about contradictions or perform as proxies for the same.

The inclusion of more factors like delay in registering FIR, filing case or recording statement of injured, recovery of weapon, etc. will provide greater details about cases by providing other aspects of the facts and make the prediction more authentic and convincing. As the performance of machine learning algorithms depends greatly on the quality and quantity of data, the inclusion of more cases may improve the results.

6. Conclusion and future scope

In this paper, a model to predict outcomes of murder related cases is build using judgments of Delhi District Court. Machine learning classification algorithms were applied to predict 'Acquittal' or 'Conviction' of accused after identifying and collecting important legal factors for 86 murder related cases. The results are promising, demonstrating the relevance of considered factors and success in predicting outcomes. There seems a weakness though i.e. poor performance in predicting outcomes for cases that have more than one accused and outcome is different for different accused persons.

As per our knowledge, this is the first research work on an Indian legal dataset for outcome prediction of legal cases. Unlike most of the works on outcome prediction of legal cases, the methodology adopted is very relevant from the legal viewpoint, the cases selected for prediction belong to the category of criminal cases as opposed to civil cases. The important factors that might affect the outcome of murder related cases are identified and a tedious job of extracting these factors and preparation of a useful database to accomplish the concerned task is carried out in this work. Though the work is limited in terms of number of factors considered for prediction and the size of the database (due to manual extraction of features), it will provide a base about how to approach the problem of outcome prediction in legal cases which is relevant for legal practitioners and how to extract the features through legal texts manually as well as automatically. With modifications and advancements, the work can be helpful for legal professionals and useful insights about the decision-making process can be gained through this work.

Numerous other possibilities are available for future research. The inclusion of factors that could not be considered in this work will result in more authenticity and acceptability of the research. Though difficult, the task of extracting features from the text of judgment could be automated, leading to a reduction in time for data collection and enabling the collection of a large quantity of structured data. A similar approach of using case-specific details for predicting outcomes could also be applied in other types of legal cases as well. The study could also be extended to predict charges in case of conviction instead of just predicting outcomes of legal cases.

References

[1] Osbeck, M. K. (2018). Lawyer as Soothsayer: Exploring the Important Role of Outcome Prediction in the Practice of Law.

[2] Martin, A. D., Quinn, K. M., Ruger, T. W., & Kim, P. T. (2004). Competing approaches to predicting supreme court decision making. *Perspectives on Politics, 2*(4), 761-767.

[3] Ashley, K. D., & Brüninghaus, S. (2006). Computer models for legal prediction. *Jurimetrics*, 309-352.

[4] Turtle, H. (1995). Text retrieval in the legal world. *Artificial Intelligence and Law, 3*(1-2), 5-54.

[5] Brüninghaus, S., & Ashley, K. D. (2006). *Progress in textual case-based reasoning: predicting the outcome of legal cases from text.* Paper presented at the AAAI.

[6] Corley, P. C., Collins Jr, P. M., & Calvin, B. (2011). Lower court influence on US Supreme Court opinion content. *The Journal of Politics, 73*(1), 31-44.

[7] Guimerà, R., & Sales-Pardo, M. (2011). Justice blocks and predictability of us supreme court votes. *PLoS ONE, 6*(11), e27188.

[8] Katz, D. M., Bommarito, I., Michael, J., & Blackman, J. (2014). Predicting the behavior of the supreme court of the united states: A general approach. *arXiv preprint arXiv:1407.6333*.

[9] Ruger, T. W., Kim, P. T., Martin, A. D., & Quinn, K. M. (2004). The Supreme Court forecasting project: Legal and political science approaches to predicting Supreme Court decisionmaking. *Columbia Law Review*, 1150-1210.

[10] Katz, D. M., Bommarito II, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PloS one, 12*(4), e0174698.

[11] Ngo, M. (2017). Classification of Legal Verdicts Using a Corpus Linguistic-Based Approach.

[12] Brüninghaus, S., & Ashley, K. D. (2003). *Combining case-based and model-based reasoning for predicting the outcome of legal cases.* Paper presented at the International Conference on Case-Based Reasoning.

[13] Sulea, O.-M., Zampieri, M., Vela, M., & van Genabith, J. (2017). Predicting the Law Area and Decisions of French Supreme Court Cases. *arXiv preprint arXiv:1708.01681*.

[14] Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science, 2*, e93.

[15] Luo, B., Feng, Y., Xu, J., Zhang, X., & Zhao, D. (2017). Learning to Predict Charges for Criminal Cases with Legal Basis. *arXiv preprint arXiv:1707.09168*.

[16] Liu, Y.-H., & Chen, Y.-L. (2018). A two-phase sentiment analysis approach for judgement prediction. *Journal of Information Science, 44*(5), 594-607.

[17] Mackaay, E., & Robillard, P. (1974). *Predicting judicial decisions: The nearest neighbour rule and visual representation of case patterns*.

[18] Popple, J. (1996). *A pragmatic legal expert system*: Dartmouth (Ashgate).

[19] Haar, C. M., Sawyer Jr, J. P., & Cummings, S. J. (1977). Computer power and legal reasoning: A case study of judicial decision prediction in zoning amendment cases. *Law & Social Inquiry, 2*(3), 651-768.

[20] Zeleznikow, J., & Hunter, D. (1994). *Building intelligent legal information systems: Representation and reasoning in law*: Kluwer Law and Taxation Publishers.

[21] Aleven, V. (2003). Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence, 150*(1-2), 183-237.

[22] Brüninghaus, S., & Ashley, K. D. (2003). *Combining case-based and model-based reasoning for predicting the outcome of legal cases.* Paper presented at the International Conference on Case-Based Reasoning.

[23] Chorley, A., & Bench-Capon, T. (2005). An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law, 13*(3-4), 323-371.

[24] Ashley, K. D., & Brüninghaus, S. (2009). Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law, 17*(2), 125-165.

[25] Grabmair, M. (2017). *Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs.* Paper presented at the Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law.

[26] Christensen, G. (2018). Predicting Supreme Court Behavior in Indian Law Cases.

[27] https://indiankanoon.org/.