



Predicting Court Judgment in Criminal Cases by Text Mining Techniques

Mohammad Farhadishad 

Mas., Department of Computer Engineering and Information technology, Razi University, Kermanshah, Iran, E-mail: m.farhadishad@stu.razi.ac.ir

Mohammad Kazemifard* 

*Corresponding author, Assistant Prof., Department of Computer Engineering and Information technology, Razi University, Kermanshah, Iran, E-mail: m.kazemifard@razi.ac.ir

Zahra Rezaei 

Assistant Prof., Department of Statistics and Information Technology, Institute of Judiciary, Tehran, Iran, E-mail: z.rezaei2010@gmail.com

Abstract

What is clear is that judges usually judge cases based on their knowledge, experience, personality, and sentiment. Due to high pressures and stress, it may be difficult for them to carefully examine documents and evidence, which leads to more subjective judgments. Legal judgment prediction with artificial intelligence algorithms can benefit judicial bodies, legal experts, and litigants as well as judges. In this research, we are looking at predicting legal sentences in drug cases involving the purchase, possession, concealment, or transportation of illicit drugs, using machine learning methods, and the effect of sentiment and emotions in case texts on predicting the severity of whipping, fines, and imprisonment. So, the text documents of 6000 Persian drug-related cases were pre-processed and then the translation of the NRC Glossary of Emotions and sentiment was used to give each item a score for positive or negative sentiment and a score for emotion. Then machine learning methods were used for modeling. BERT, TFIDF+Adaboost, and Skipgram+LSTM+CNN methods had the highest accuracy, respectively. Also, evaluation criteria were analyzed in situations where sentiment scores, emotional scores, or both were used in the prediction process along with judicial texts. Finally, it was found that the use of sentiment and emotion scores improves the accuracy of legal judgment predictions for all three types of sentences and that sentiments have a greater impact on the accuracy of legal judgment predictions than emotions.

Keywords: Legal Judgment Prediction, Text Mining, Sentiment Analysis, Emotions Analysis, Machine Learning.

Journal of Information Technology Management, 2023, Vol. 15, Issue 2, pp. 204-222.

Published by University of Tehran, Faculty of Management

doi: [https://doi.org/ 10.22059/jitm.2023.350464.3206](https://doi.org/10.22059/jitm.2023.350464.3206)

Article Type: Research Paper

© Authors

Received: January 22, 2023

Received in revised form: February 02, 2023

Accepted: March 25, 2023

Published online: May 23, 2023



Introduction

The last few decades have seen many positive socio-economic changes including remarkable developments in information dissemination. But still, all kinds of illegal activities occur daily, which make people pursue or defend their legal rights through judicial channels. Given the need for transparency in the judicial process, this is a field where almost every detail is meticulously recorded in text documents. This provides an opportunity for using intelligent automated data management and analysis tools, which could be especially useful in areas with demanding workloads (Yuriy, 2014). With the increasing use of digital documents in the legal domain, the sheer volume of textual content makes it increasingly difficult to manage legal information. However, it is not easy to meet the information needs of this field through ordinary document and record analyses, as this will require significant legal knowledge and a good understanding of legal terms. Therefore, to achieve cost and time-saving in this field, it is necessary to design a comprehensive framework or method for extracting and retrieving relevant legal data that meet specific conditions (Liu & Chen, 2017).

To issue a fair verdict, a judge must arbitrate impartially between the two parties. Because of the limited time and resources, it may become difficult for some judges to maintain an objective and impartial position when processing many cases. This poses two problems. First, judgments may become more subjective as it is difficult to accurately process the evidence of multiple cases (Wyner et al., 2010). Second, heavy workloads may put judges under severe pressure and stress, thereby reducing their ability to make quality judgments (Chou & Hsing, 2010). These two problems can be detected by predicting legal judgments and possible verdicts based on the details and outcomes of similar cases. Legal judgment prediction can also help litigants decide what actions to take before going to court and assist lawyers in advising their clients. A legal judgment prediction tool can also be turned into a publicly available online counseling service to inform people about the probable outcomes of litigations before filing a lawsuit, which may significantly reduce the number of cases and help poor people avoid the costs of hiring lawyers. While text mining is extensively used in a wide variety of fields, only a few studies have been performed on the text mining of legal documents (Şulea et al., 2017; Undavia et al., 2018).

Since judges decide cases based on not only their knowledge, experience, and personality, but also their sentiments (Conrad & Schilder, 2007), sentiment analysis can improve the accuracy of legal judgment predictions.

The subject of this paper is the prediction of judgments in Persian legal cases using text-mining techniques including Persian-specific text pre-processing, text vectorization, and sentiment and emotion analysis with the help of machine learning and deep learning methods. This is the first study to provide such a method for Persian cases as well as the first to use text mining techniques, emotion and sentiment analysis, and deep learning simultaneously in this field. The remainder of this article is as follows: In the second and third sections, Theoretical Foundations and Research Background are presented. In the fourth section, the proposed methods are discussed. Then, in the next section, the results of the implementation of the methods and data analysis are presented. and in the sixth section, the results are discussed.

Theoretical Foundations

Text Mining

Text mining is an automated or semi-automated text processing procedure for finding patterns and structures in texts and extracting relevant information from textual documents (Chen et al., 2013; Liu et al., 2015). Text mining is a rapidly developing field with several domains of application, including information retrieval and computational linguistics. In general, such natural language processing tools are created to extract knowledge from ontological data that are collected in various ways from different sources, including those written by humans. The importance of text mining techniques for the analysis of textual data stems from the fact that many textual data are poorly structured and cannot be analyzed without purpose-specific methods (Thamil Selvi & Laksmi, 2019). The text mining process consists of several steps, which tend to include the following:

- Text collection
- Text preprocessing
- Text vectorization (transformation of text into vectors)
- Feature selection
- Data mining/pattern selection
- Evaluation

Sentiment Analysis

Sentiment analysis is the process of assessing people's sentiments and opinions in their comments on social networking sites and other textual documents. Most studies in this field are focused on the English language, and much fewer studies have been conducted about

sentiment analysis in other languages including Persian. Persian is the official language of Iran, Tajikistan, and Afghanistan and is the first language of over 100 million people around the world. Persian is a challenging language for sentiment analysis and only a few tools have been developed for general Persian text processing. The processing of Persian documents and texts involves challenges such as misspellings, irregular word spacing, difficult stemming, and the presence of informal words (Basiri et al., 2014). In a legal case or record, the motive and thought process of the offender is described with several words or terms, some of which are sentimental or emotional. Examples of these sentimental terms include “intentional”, “cruel”, and “hateful” (Liu & Chen, 2017). In general, sentiments can affect all participants in the litigation process, including judges and juries (Eliot, 2020).

In the only study conducted so far on sentiment analysis in Chinese legal documents, the results have confirmed the positive effect of this analysis in improving the accuracy of legal judgment predictions (Liu & Chen, 2017). It should be noted that while this has used only one machine learning method, the present study makes use of several machine learning and deep learning methods for modeling.

Emotion Analysis

Emotion analysis is a challenging task (Scherer, 2021). This is because as Scherer puts it, the definition of emotion could be very complex (Mayer et al., 2008). Different definitions of emotion can be reached through different methodological and conceptual approaches. However, most theorists agree that emotion is a set of expressive, behavioral, physiological, and phenomenological characteristics (Lewis et al., 2016).

Emotion is an unconscious phenomenon that can become conscious during evocation. It is a social representation of sentiment and feeling that is influenced by culture and has an intensity and value as a form of emotional state (Plutchik, 1991).

Difference between Sentiment and Emotion

The concepts of emotion and sentiment are often used interchangeably. This is mostly because they both refer to experiences originating from the combined effects of biological, cognitive, and social causes. However, sentiments are distinguished from emotions by the length of time they are experienced (Kim & Klinger, 2018). Also, while sentiments tend to emerge towards a subject or situation, this is often not the case for emotions. For example, a person may wake up and be upset or happy for no reason.

Literature Review

This section provides a review of previously published articles on the application of text mining in legal and judicial documents

Table 1. List of published articles about legal judgment prediction

Author(s)	Method	Dataset language	Dataset description	Model accuracy
Aletras et al. (2016)	SVM	English	600 cases of the European Court of Human Rights (ECHR)	79%
Liu and Chen (2017)	SVM with sentiment analysis	Chinese	1,208 Chinese criminal records were collected manually over about 2 months	80.62%
Şulea et al. (2017)	SVM	French	131830 legal documents of the French Supreme Court	Case jurisdiction prediction: 96% Ruling prediction: 98% Ruling date prediction: 87%
Undavia et al. (2018)	W2V+CNN	English	8419 United States Supreme Court documents	72.4%
Alekseev et al. (2019)	Artificial neural network	Russian	167 cases of the Russian Economic Crimes Court	100%
Ait Yahia and Chakir (2019)	SVM	Arabic	1452 Arabic legal documents of the Moroccan courts in the field of real estate and traffic offenses	98%
Hammami et al. (2019)	CNN with max pooling	French	452 documents of the French Court of Appeals, extracted from data.gouv.fr	84.46%
Ryoma Kondo et al. (2022)	One-hot coding +XGB	Japanese	using data from approximately 110,000 court judgments from Japan spanning the period 1998–2018 from the district to the supreme court level	88.41
Intisar Almuslim and Diana Inkpen (2022)	Custom W2V + RCNN	English	4252 documents on Canadian Legal Information Institute (CanLII) website	93.46

Despite the extensive use of text mining in various fields, only a few studies have been performed about text mining in the legal domain, especially for court ruling prediction. It should be noted that this is the first study on the use of text mining and artificial intelligence in general in the field of legal judgment prediction for Persian cases.

Methodology

The goal of this study was to achieve the highest possible accuracy in legal judgment prediction. Therefore, in addition to the texts of court cases, two other characteristics, namely sentiments and emotions expressed in these texts were also incorporated into the prediction process. In this approach, to predict the outcome of a new case, the sentiment and emotion scores of previous cases are used in combination with the case texts in the legal judgment prediction model. The flowchart of the process of legal judgment prediction in this study is displayed in Figure 1.

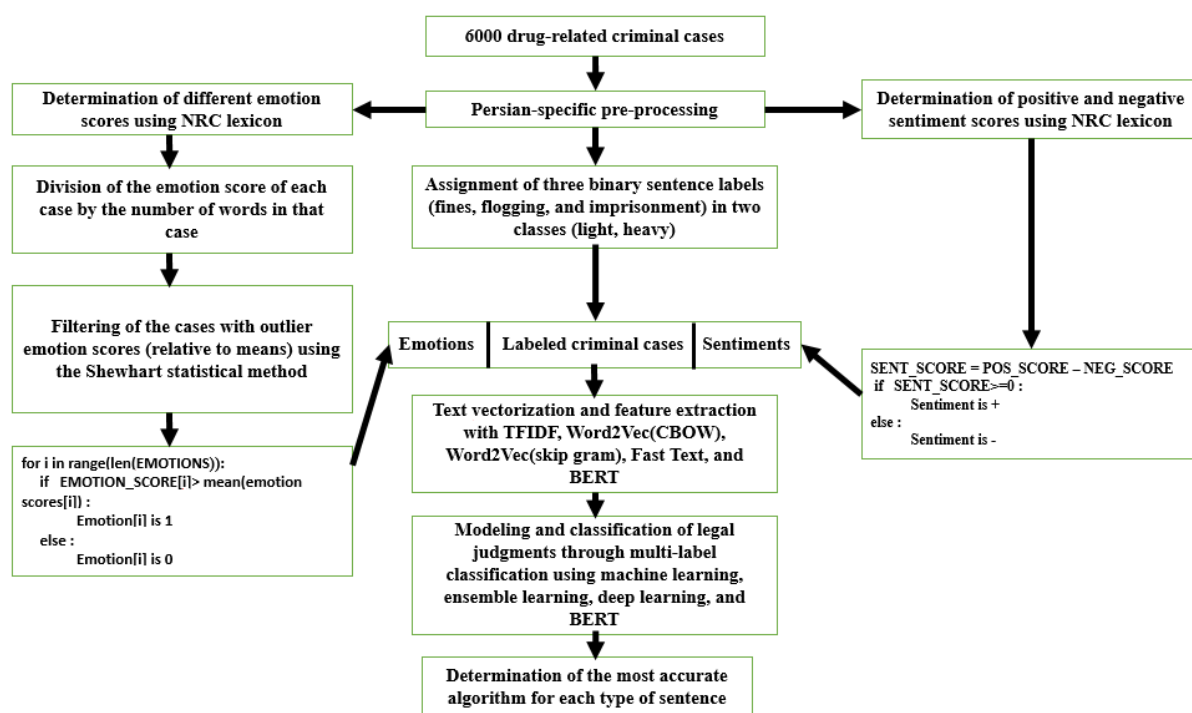


Figure 1. Flowchart of the legal judgment prediction process

Statistical Population and Dataset

The data of this study comprised 6,000 criminal cases involving the purchase, possession, concealment, or transportation of 500 grams to 5 kilograms of illicit substances, which, according to the Iranian criminal code, are punishable by fines of 15-60 million rials, 40-74 lashes, and 2-5 years of imprisonment. Legal court documents usually consist of multiple sections dedicated to the case number, identity information of litigants (plaintiffs, defendants, and their lawyers), case subject, main ruling, facts, and evidence, statements of litigants, cited legal laws and articles, ruling date, the name of the judge, and documents and evidence concerning the criminal motivation, process, and purpose.

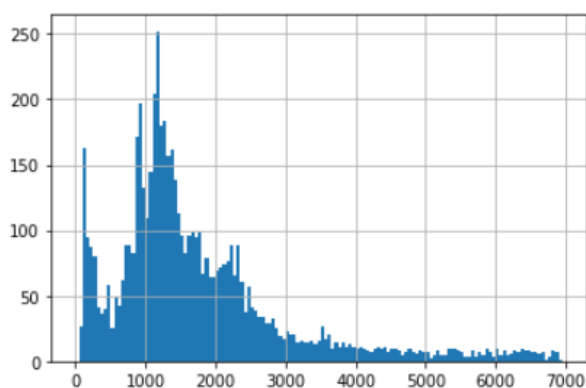


Figure 2. Histogram of the length of the cases in the dataset (in words).

As shown in Figure 2, the average length of cases considered in this study was about 2,000 words.

Pre-processing of Persian Legal Documents

The purpose of text pre-processing is to remove redundant and unimportant words from the text to prepare it for modeling. Because of the limited number of tools that support the Persian language, it is more difficult and complex to pre-process Persian texts than English documents. In this study, the pre-processing of legal documents was performed in ten steps. In the first step, tokenization is used which is the process of dividing a text into words or tokens. In the second step, all punctuation marks (e.g., question marks, commas, parentheses, exclamation marks) were removed. In the third step, all English letters and in the fourth step, all numbers in the documents were removed. In the fifth step, all Arabic characters were transformed into Persian characters, such as **ك** and **ي** which were converted to **ک** and **ی**, respectively. In the sixth step, stop words (e.g., conjunctions, pronouns, and propositions) were removed. After correcting the misspelled words in the seventh step, stemming and lemmatization methods were used to transform all words into their root form in the eighth and ninth steps. Considering that Lemmatization takes into account the morphological structure of words and uses a dictionary or a knowledge base to map each word to its root, it is naturally more accurate. However, due to the creation of new words in the Persian language, if a word is not in the knowledge base, stemming can be helpful in regular situations to convert words to their roots. Therefore, in this research, both techniques are used simultaneously. Finally, in the tenth step of pre-processing, the text was normalized. After these ten steps, the Persian judicial texts were ready for analysis and modeling.

Assignment of Sentence Labels to Court Cases

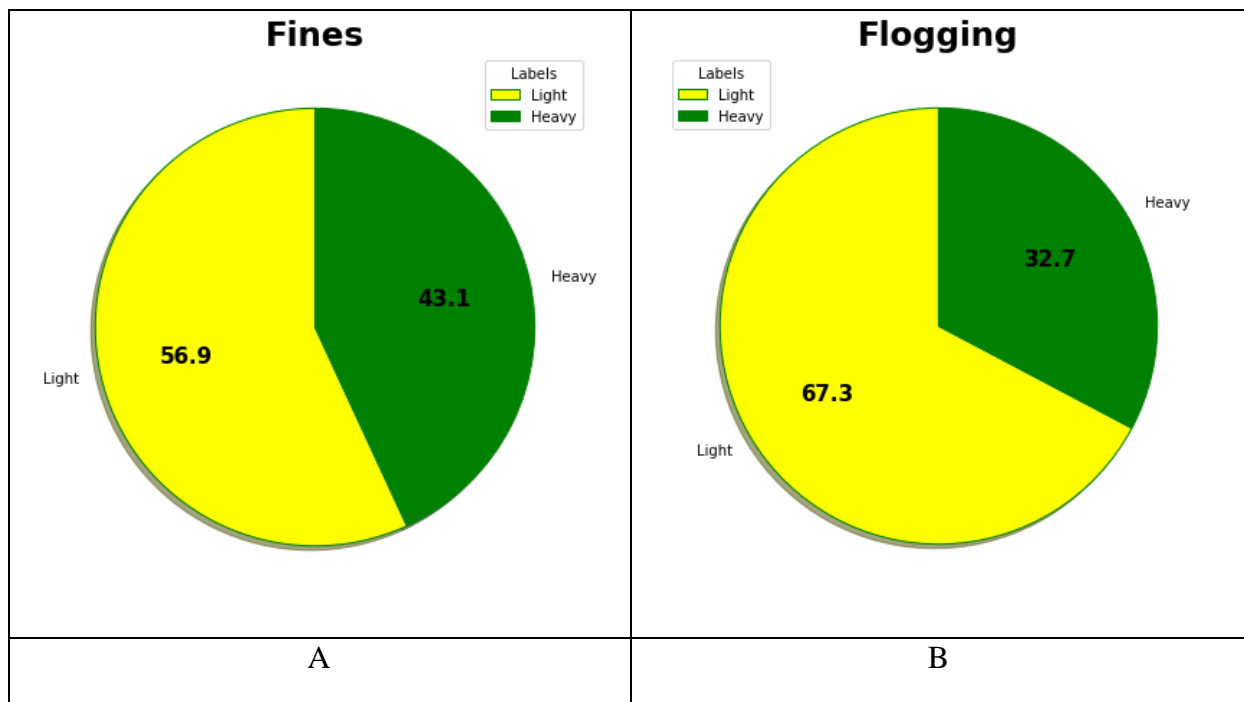
As mentioned, the cases considered in this study were 6,000 criminal cases involving the purchase, possession, concealment, or transportation of 500 grams to 5 kilograms of illicit substances, legal punishments include fines of 15-60 million rials, 40-74 lashes, and 2-5 years

of imprisonment. In this study, we defined three labels for sentences called “Fines”, “Flogging” and “Imprisonment”, each with two classes called “Light” and “Heavy”, which were assigned based on whether the sentence has been on the heavy side or light side of lawful punishment for the crime committed.

For the “Fines” label, the sentence was considered “Light” if it was between 15 and 37.5 million rials and “Heavy” if it was between 37.5 and 60 million rials. Out of 6000 cases reviewed in this study, 3412 cases were labeled Light, and 2588 were labeled Heavy. Figure 3A shows the percentage of light and heavy fine sentences in the reviewed cases.

For the “Flogging” label, the sentence was considered “Light” if it was between 40 and 57 lashes and “Heavy” if it was between 57 and 74 lashes. Out of 6000 cases of this study, 4035 cases were labeled Light, and 1965 were labeled Heavy. The percentages of light and heavy flogging sentences in the reviewed cases are shown in Figure 3B.

For the “Imprisonment” label, the sentence was considered “Light” if it was between 2 and 3.5 years and “Heavy” if it was between 3.5 and 5 years. Out of 6000 cases considered in this study, 3720 were labeled Light, and 2280 were labeled Heavy. The percentages of light and heavy imprisonment sentences in the cases are shown in Figure 3C.



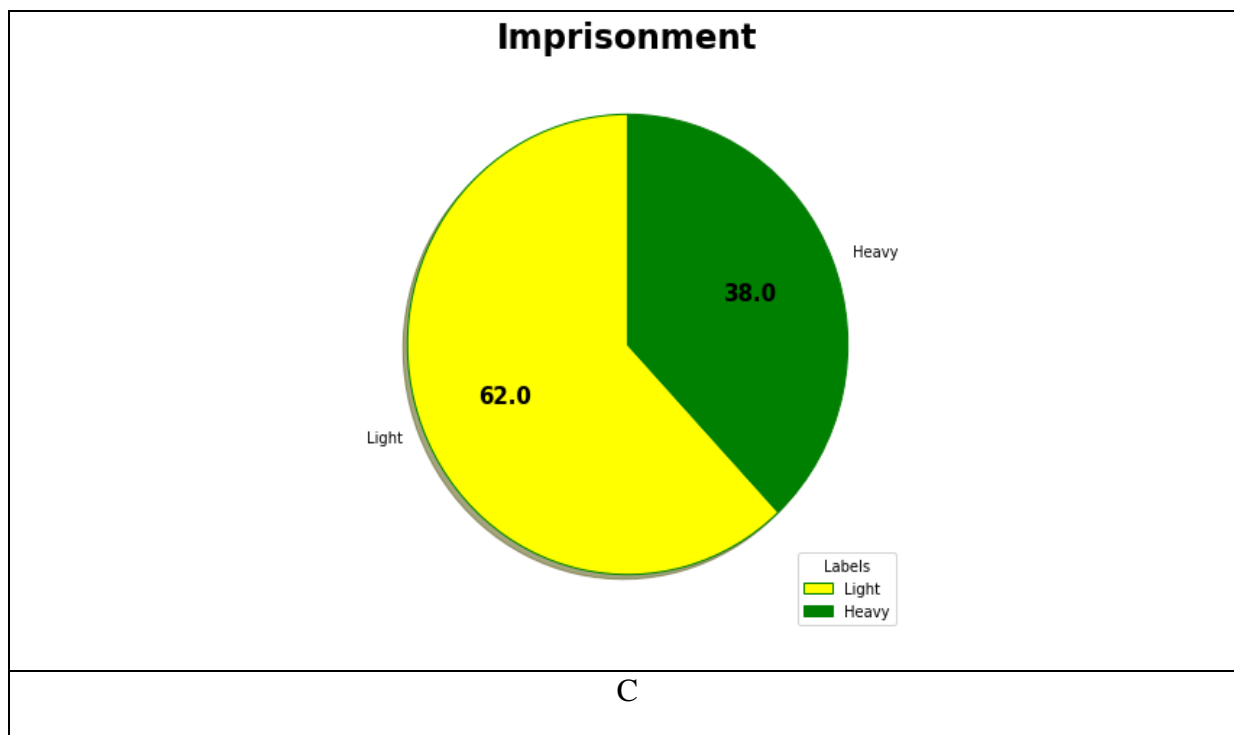


Figure 3. Percentage of light and heavy sentences in the reviewed cases: A) fines, B) flogging, and C) imprisonment.

Table 2 shows the number of cases by the type of sentences given and their intensity (whether they have been heavy or light).

Table 2. Number of cases by the assigned label (sentence type and intensity).

Intensity Sentence	Light	Heavy
Fines	3412	2588
Flogging	4035	1965
Imprisonment	3720	2280

Sentiment Scoring of Court Cases

In this study, sentiment scoring was performed using the NRC emotion lexicon. This lexicon consists of over 14,000 English words, but this study used its Persian translation. The lexicon covers eight types of emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The sentiment label interpretation for NRC was performed manually by crowdsourcing (Mohammad & Turney, 2010, 2013).

In the first step, the texts of the statements of the litigants in the court cases were tokenized. Then, all the words in the statements of each case were matched with those words in the NRC lexicon that convey positive or negative sentiments, and the total positive and negative sentiment scores of each case were calculated. The negative sentiment score was

then subtracted from the positive score. The sentiment variable was set to 1 if the answer was greater than zero (indicating positive sentiment) and was set to 0 otherwise (indicating negative sentiment). As shown in Figure 4, ultimately, out of 6,000 drug-related cases, 2,135 had positive sentiments and 3,865 had negative sentiments.

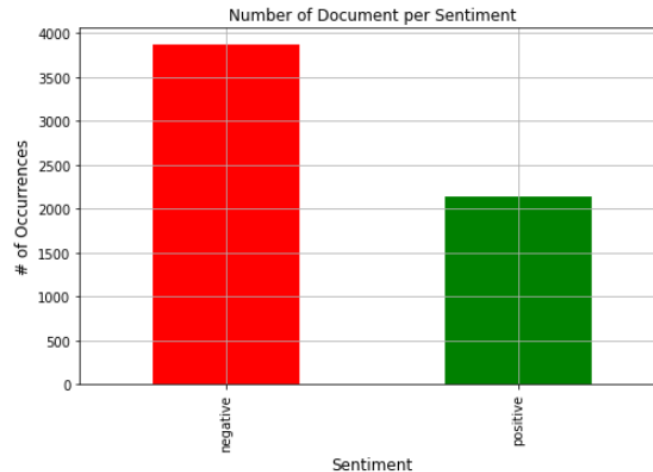


Figure 4. The number of drug-related criminal cases with positive and negative sentiments

Emotion Scoring of Court Cases

Emotion scores of the cases were also determined using the NRC emotion lexicon, which as mentioned in Section 2.4, has eight types of emotion (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). To analyze emotions, we first tokenized the texts of the statements of the litigants and then matched their words with the words in the NRC lexicon that convey these eight types of emotions. Then, we calculated the total score of each type of emotion for each case and divided this score by the total word count of that case (this was because cases vary in length and longer cases have a higher probability of obtaining higher emotion scores). Considering the presence of outlier (extremely high) scores for some emotions in some cases, which could be problematic for giving emotion variable binary values based on the average of scores in all cases (i.e. setting the emotion variable of a case to 1 if its score is higher than the total average and 0 if its score is lower than the total average), we had to remove the skewness to a great extent. Thus, we used the Shewhart statistical method to find and remove outliers. This method involves finding an upper control limit (UCL) and a lower control limit (LCL) and removing all values that are above UCL or below LCL. For further explanation, first, the score of each emotion in the texts was calculated based on a match with the emotion dictionary. Then, the emotion score was divided by the length of the words in the text, taking into account that longer texts with more emotional words should not have higher emotional scores than shorter texts. Additionally, the Shewhart method was used to remove scores of texts that were above or below the threshold for averaging. Finally, averaging was performed, and after the aforementioned normalization, texts with emotional scores higher than the

average were categorized according to their respective emotions. The Shewhart charts for the scores of each emotion are shown separately in Figure 5.

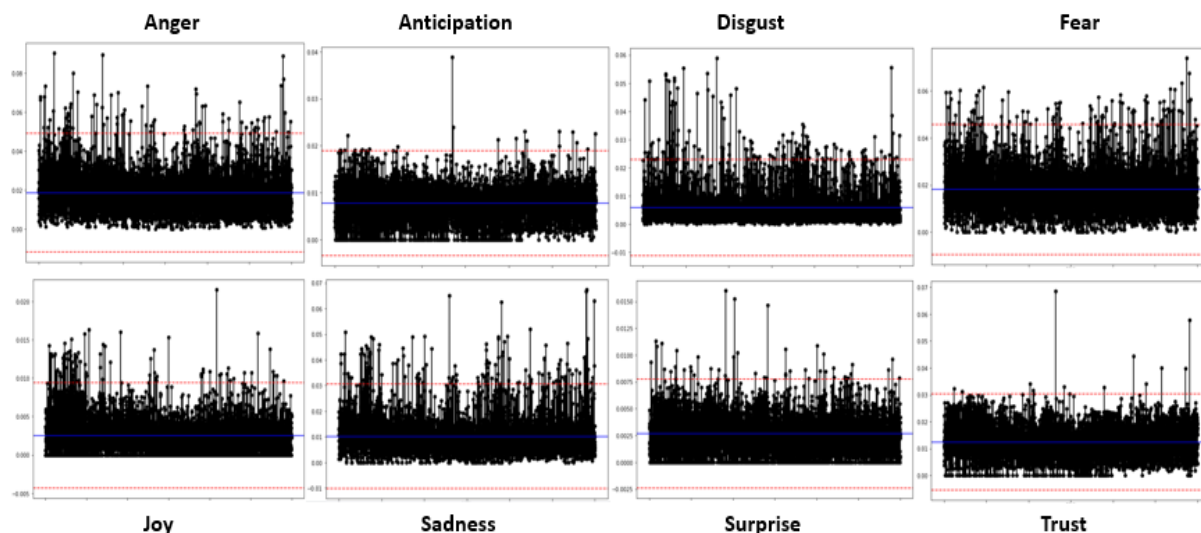


Figure 5. Shewhart chart for the scores of each emotion

It should be noted we did not completely discard the scores that were higher than UCL or lower than LCL. Rather, we just did not include these scores in averaging. For example, if the surprise emotion score of a case was higher than UCL or lower than LCL, this score was not included in the averaging of surprise emotion scores of all cases, but the surprise emotion variable of the case was nevertheless set to 1 if it was higher than the total average or 0 if it was lower than the total average.

Figure 6 shows the number of cases containing each emotion. Each case file can contain multiple emotions or no emotions at all.

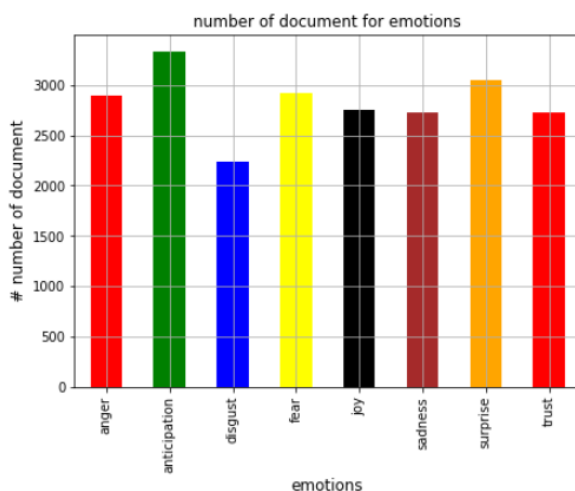


Figure 6. The number of cases with each type of emotion.

Text Vectorization and Feature Extraction

In tabular data, each column of the table is a feature of the data. But in textual data, features should be identified and extracted by text vectorization, which involves transforming the text into numbers. This can be done by various methods. In this study, TF-IDF and Word2Vec methods including Continuous Bag of Words, Skip-gram, Fast Text, and BERT were used for this purpose.

Modeling and Classification of Legal Judgments

As in Section 4.1. It was mentioned that in this research, judicial cases were investigated which include the purchase, possession, concealment, or transportation of 500 grams to 5 kilograms of drugs, which, according to the Iranian criminal code, are punishable by fines of 15-60 million rials, 40 - 74 lashes, and 2-5 years of imprisonment. Therefore, the punishment determined for this research should have a value within the specified range in the law for all three types of punishment, and in this research, models were created to predict the severity of punishment using machine learning algorithms. For this purpose, three binary classifications were created for each punishment using different machine learning algorithms.

Thus, after text vectorization, classification, and modeling were performed with combinations of eleven machine learning methods, namely Support Vector Machine(SVM), Artificial Neural Network(ANN), decision tree, Naive Bayes(NB), Logistic Regression and K-Nearest-Neighbor (KNN), Random Forest, Adaboost, Long Short-Term Memory (LSTM), Convolutional Neural Network(CNN), and Bidirectional Encoder Representations from Transformers(BERT).

Results

In this section, we evaluate the accuracy of different algorithms in predicting legal judgments to determine which offers the highest accuracy for each sentence label (i.e. fines, flogging, and imprisonment). To examine the effect of sentiment and emotion scores on legal judgment prediction, in the next section, sentiment and emotion scores of the considered cases are also included in the prediction process.

Table 3 shows the accuracy of different machine learning methods in predicting each type of sentence (fines, flogging, and imprisonment).

Table 3. Accuracy of different methods in predicting sentence labels (fines, flogging, and imprisonment).

Classification method and data type	Method of Persian text vectorization	Algorithm	Accuracy in predicting flogging	Accuracy in predicting fines	Accuracy in predicting imprisonment
Three binary Classification for the prediction of legal judgments with three labels of fines, flogging, and imprisonment in two classes of heavy and light for 6000 court cases related to the purchase, possession, concealment, or transportation of illicit substances	TF-IDF	SVM	87.80	85.68	87.13
	TF-IDF	ANN	85.52	85.60	86.11
	TF-IDF	Decision tree	85.5	83.33	89.01
	TF-IDF	Naive Bayes	81	85.61	83.87
	TF-IDF	Logistic regression	84.02	85.69	86.86
	TF-IDF	KNN	77.20	78.78	82.36
	TF-IDF	Random forest	86.53	86.36	86.89
	TF-IDF	Adaboost	88.90	89.39	90.91
	Word2Vec(CBOW)	LSTM	83.07	86.02	86.91
	Word2Vec(CBOW)	LSTM and Dropout	82.05	86.39	89.5
	Word2Vec(CBOW)	LSTM and CNN	85.75	87.11	89.47
	Word2Vec(Skip-gram)	LSTM	84.17	86.32	88.17
	Word2Vec(Skip-gram)	LSTM + Dropout	82.50	86.70	90.67
	Word2Vec(Skip-gram)	LSTM + CNN	86.65	88.33	91.12
	Fast Text	LSTM	82.23	84.17	87.33
	Fast Text	LSTM + Dropout	85.57	85.80	86.5
	Fast Text	LSTM + CNN	84.73	85	85.60
	BERT	BERT	85.41	90.06	90.64

Results for the Prediction of Flogging Sentences

Figure 7 shows the accuracy of different machine learning methods in predicting flogging sentences in the considered cases (the methods are arranged in order of accuracy).

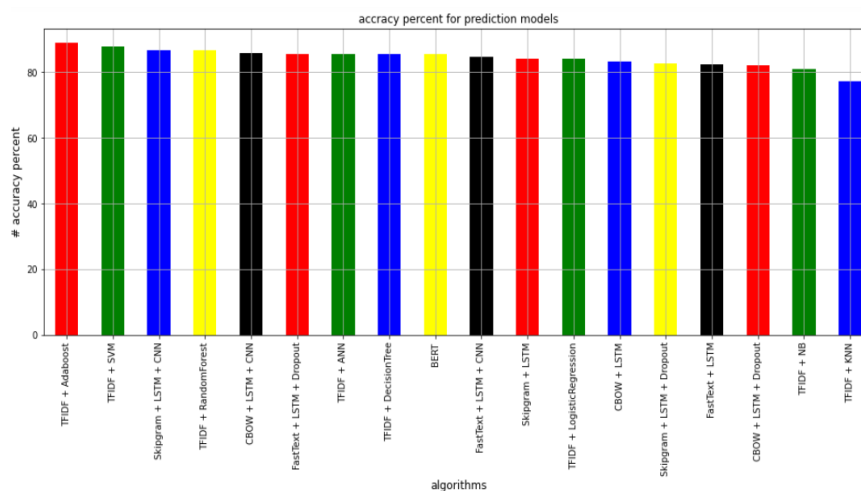


Figure 7. Accuracy of different methods in predicting flogging sentences in the studied legal cases

As shown in Figure 7, the top three most accurate methods in predicting flogging were TFIDF+Adaboost with 88.90% accuracy, TFIDF+SVM with 87.80% accuracy, and Skipgram+ LSTM+CNN with 86.65% accuracy, and the three least accurate methods for this sentence were TFIDF+KNN with 77.20% accuracy, TFIDF+Naive Bayes with 81% accuracy, and CBOW+LSTM+Dropout with 82.05% accuracy. Because there are three different algorithms with the highest accuracy for three types of punishments, to predict the sentence of each case, one must use these three algorithms. However, for each type of punishment, one should rely on the algorithm that has the highest accuracy for that specific punishment

Results for the Prediction of Fine Sentences

The accuracy of the tested machine learning methods in predicting fine sentences in the considered cases is illustrated in Figure 8 (the methods are arranged in order of accuracy).

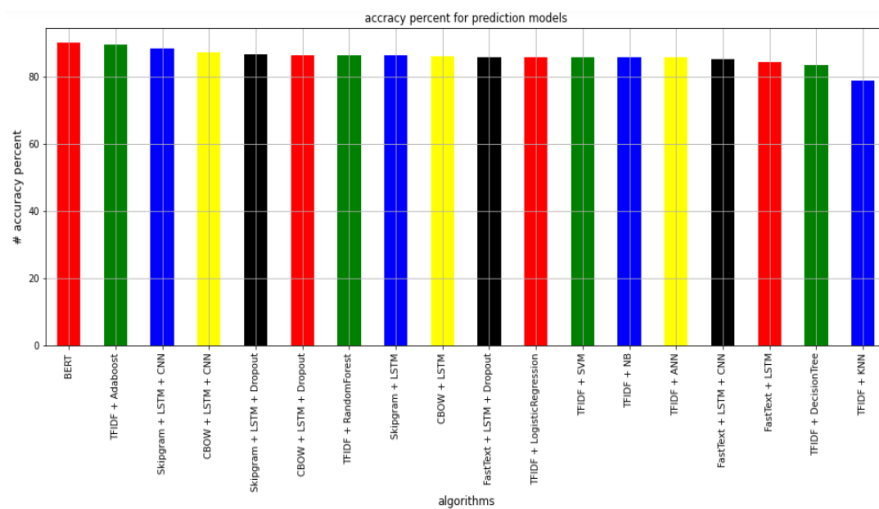


Figure 8. Accuracy of different methods in predicting fine sentences in the studied legal cases.

As Figure 8 demonstrates, the top three most accurate methods in predicting fine sentences were BERT with 90.06% accuracy, TFIDF+Adaboost with 89.39% accuracy, and Skipgram+LSTM+CNN with 88.33% accuracy, and the three least accurate methods for this sentence were

TFIDF+KNN with 78.78% accuracy, TFIDF+DecisionTree with 83.33% accuracy, and FastText+LSTM with 84.17% accuracy.

Results for the Prediction of Imprisonment Sentences

The chart plotted in Figure 9 shows the accuracy of different machine learning methods in predicting imprisonment sentences in the considered cases (The methods are arranged in order of accuracy).

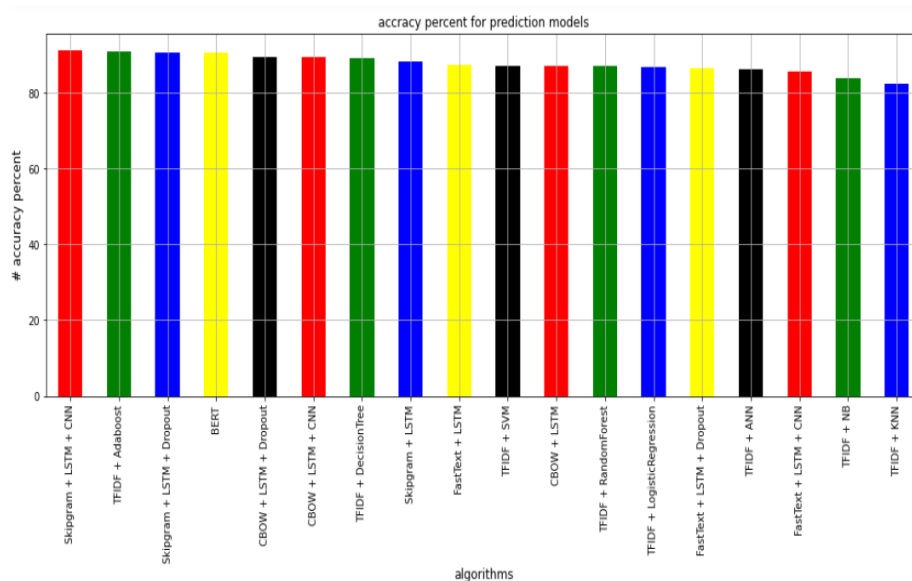


Figure 9. Accuracy of different methods in predicting imprisonment sentences in the studied legal cases

As shown in Figure 9, the top three most accurate methods in predicting imprisonment were Skipgram+LSTM+CNN with 91.12% accuracy, TFIDF+Adaboost with 90.91% accuracy, and Skipgram+LSTM+Dropout with 90.67% accuracy, and the three least accurate methods for this sentence were TFIDF+KNN with 82.36% accuracy, TFIDF+Naive Bayes with 83.87% accuracy, and FastText+LSTM+CNN with 85.60% accuracy.

Investigation of the Effect of Sentiment and Emotion Scores on the Accuracy of Legal Judgment Prediction

After determining the accuracy of the methods (Table 3), we identified the algorithm that offers the highest accuracy for each sentence label (fines, flogging, and imprisonment) and determined its accuracy in legal judgment prediction under the following conditions.

- 1) When the input data are the judicial texts plus the sentiment scores of cases.
- 2) When the input data are the judicial texts plus the emotion scores of cases.
- 3) When the input data are the judicial texts plus both sentiment and emotion scores of cases

Table 4 shows the accuracy of the algorithms that were most accurate in predicting each sentence label when sentiment and emotion scores of the considered cases were also incorporated into the prediction process.

Table 4. Effect of sentiment scores and emotion scores on the accuracy of legal judgment prediction

The algorithm with the highest accuracy in predicting the sentence based on the texts of the case file	Data			Accuracy
	Case text	Sentiment score	Emotion score	
TFIDF+Adaboost (the algorithm with the highest accuracy in predicting flogging sentences)	✓			88.90
TFIDF+Adaboost (the algorithm with the highest accuracy in predicting flogging sentences)	✓	✓		90.12
TFIDF+Adaboost (the algorithm with the highest accuracy in predicting flogging sentences)	✓		✓	89.36
TFIDF+Adaboost (the algorithm with the highest accuracy in predicting flogging sentences)	✓	✓	✓	91.69
BERT (the algorithm with the highest accuracy in predicting fine sentences)	✓			90.06
BERT (the algorithm with the highest accuracy in predicting fine sentences)	✓	✓		93.13
BERT (the algorithm with the highest accuracy in predicting fine sentences)	✓		✓	92.02
BERT (the algorithm with the highest accuracy in predicting fine sentences)	✓	✓	✓	93.49
Skipgram+LSTM+CNN (the algorithm with the highest accuracy in predicting imprisonment sentences)	✓			91.12
Skipgram+LSTM+CNN (the algorithm with the highest accuracy in predicting imprisonment sentences)	✓	✓		91.20
Skipgram+LSTM+CNN (the algorithm with the highest accuracy in predicting imprisonment sentences)	✓		✓	91.49
Skipgram+LSTM+CNN (the algorithm with the highest accuracy in predicting imprisonment sentences)	✓	✓	✓	91.52

As shown in Table 4, the most accurate algorithm in predicting flogging sentences based exclusively on case texts was TFIDF+Adaboost with 88.90% accuracy, which increased to 90.12% (1.22% improvement) when sentiment scores were added to the dataset and to 89.36% (0.46% improvement) when emotion scores were incorporated into the prediction process. Adding both sentiment and emotion scores to the input data increased the accuracy of TFIDF+Adaboost in predicting flogging sentences to 91.69%, which is a 2.79% improvement.

The most accurate algorithm in predicting fine sentences based on case texts was BERT with 90.06% accuracy. The accuracy of this method increased to 93.13% (3.07% improvement) when sentiment scores were introduced to the dataset and to 92.02% (1.96% improvement) when emotion scores were added to the dataset. Once sentiment and emotion scores were both incorporated into the prediction process, the accuracy of this method in predicting fine sentences increased to 93.49%, which is a 3.43% improvement.

Finally, the most accurate algorithm in predicting imprisonment sentences based only on case texts was Skipgram+LSTM+CNN with 91.12% accuracy, which increased to 91.20% (0.08% improvement) when sentiment scores were also used in the prediction process and to 91.49% (0.37% improvement) when emotion scores were added to the dataset. Introducing

both sentiment and emotion scores to the dataset raised the accuracy of Skipgram+LSTM+CNN in predicting imprisonment sentences to 91.52%, which is a 0.4% improvement.

Conclusion

The subject of this article was the prediction of legal judgments in criminal cases related to the purchase, possession, concealment, or transportation of illicit drugs using text mining, machine learning, and deep learning methods and the impact of sentiments and emotions in the texts of case files and documents on the prediction of the type (flogging, fines, and imprisonment) and intensity (heavy, light) of sentences. For this purpose, we first pre-processed the textual documents of 6000 Persian drug-related court cases and then used a translation of the NRC emotion and sentiment lexicon to give each case a positive or negative sentiment score and an emotion score based on scores given for eight types of emotions. Next, we classified the sentences into two classes of light and heavy, and modeled them with a variety of machine-learning methods.

The methods with the highest accuracy in predicting flogging, fine, and imprisonment sentences based exclusively on case texts were found to be BERT, TFIDF+Adaboost, and Skipgram+LSTM+CNN, respectively. Finally, the accuracy of these algorithms (algorithms that were most accurate for each type of sentence based only on case texts) under the conditions where sentiment scores, emotion scores, or both were incorporated into the prediction process was investigated. The results of this investigation showed that the use of sentiment and emotion scores improved the accuracy of legal judgment prediction for all three-sentence types (flogging, fines, and imprisonment). Among the three sentence types, predictions for a flogging were most greatly affected by sentiment and emotion scores, and predictions for imprisonment were least affected. Overall, sentiment had a greater effect on the accuracy of legal judgment predictions than emotion, although the opposite was true for imprisonment sentences.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author (s) received no financial support for the research, authorship, and/or publication of this article.

References

- Ait Yahia, I., & Chakir, L. (2019). Arabic Text Classification in the Legal Domain. In: *IEEE*, Marrakech, Morocco.
- Alekseev, A., Zuev, D., Katasev, A., Kirillov, A., & Khassianov, A. (2019). Prototype of classifier for the decision support system of legal documents. In: *21st Scientific Conference, Scientific Services & Internet*, pp. 40-51.
- Aletras, N., Tsarapatsanis, D., Preotiuc-Pietro, D., & Lampos, V. (2016). Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ in Computer Science*, 2, e93.
- Almuslim, Intisar & Inkpen, Diana. (2022). Legal Judgment Prediction for Canadian Appeal Cases. 163-168. 10.1109/CDMA54072.2022.00032.
- Basiri, E., Naghsh-Nilchi, A., & Ghassem-Aghaee, N. (2014). A Framework for Sentiment Analysis in Persian. *Open Transactions on Information Processing*, 1(3), 1-14.
- Chen, Y.-L., Liu, Y., & Ho, W.-L. (2013). A Text Mining Approach to Assist the General Public in the Retrieval of Legal Documents. *Journal of the American Society for Information Science and Technology*, 64, 280-290.
- Chou, Sh., & Hsing, T.-P. (2010). Text Mining Technique for Chinese Written Judgment of Criminal Case. In: H. Chen, M. Chau, Sh. Li, S. Urs, S. Srinivasa, & G. A. Wang (Eds.), *Intelligence and Security Informatics*. PAISI 2010. Lecture Notes in Computer Science, vol 6122. Springer, Berlin, Heidelberg.
- Conrad, J., & Schilder, F. (2007). Opinion mining in legal blogs. In: *Proceedings of the International Conference on Artificial Intelligence and Law*, pp. 231-236.
- Eliot, L. (2020). Legal Sentiment Analysis and Opinion Mining (LSAOM): Assimilating Advances in Autonomous AI Legal Reasoning. Working paper.
- Hammami, E., Akermi, I., Faiz, R., & Boughanem, M. (2019). Deep Learning for French Legal Data Categorization. In: K. D. Schewe & N. Singh (Eds.), *Model and Data Engineering*. MEDI 2019. Lecture Notes in Computer Science, vol. 11815. Springer.
- Kim, E., & Klinger, R. (2018). A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. In: *Zeitschrift für digitale Geisteswissenschaften*. Erstveröffentlichung vom 16.12.2019. Version 2.0 vom 23.07.2021.
- Kondo, Ryoma & Yoshida, Takahiro & Hisano, Ryohei. (2022). Masked prediction and interdependence network of the law using data from large-scale Japanese court judgments. *Artificial Intelligence and Law*. 1-33. 10.1007/s10506-022-09336-5.
- Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (2016). *Handbook of emotions*. The Guilford Press.
- Liu, Y., & Chen, Y.-L. (2017). A two-phase sentiment analysis approach for judgement prediction. *Journal of Information Science*, 44, 016555151772274.
- Liu, Y., Chen, Y.-L., & Ho, W.-L. (2015). Predicting associated statutes for legal problems. *Information Processing & Management*, 51, 194-211.
- Mayer, J., Roberts, R., & Barsade, S. (2008). Human Abilities: Emotional Intelligence. *Annual Review of Psychology*, 59(1), 507-536.
- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.

- Mohammad, S., & Turney, P. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29, 436-465.
- Plutchik, R. (1991). *The Emotions Facts, Theories, and a New Model*. Lanham, Md. University Press of America.
- Scherer, K. R. (2021). What are emotions? And how can they be measured? *Social Science Information*, 44, 693-727.
- Şulea, O.-M., Zampieri, M., Malmasi, Sh., Vela, M., Dinu, L., & Genabith, J. (2017). Exploring the Use of Text Classification in the Legal Domain. In: *Proceedings of 2nd Workshop on Automated Semantic Analysis of Information in Legal Texts (ASAIL)*. London, United Kingdom.
- Thamil Selvi, C. P., & Laksmi, P. (2019). A Survey on Text Analytics and Text Mining. *International Journal of Recent Technology and Engineering (IJRTE)*, 7.
- Undavia, S., Meyers, A., & Ortega, J. (2018). A Comparative Study of Classifying Legal Documents with Neural Networks. In: *IEEE*, Poznan, Poland.
- Wyner, A., Mochales, R., Moens, M.-F., & Milward, D. (2010). Approaches to Text Mining Arguments from Legal Cases. In: E. Francesconi, S. Montemagni, W. Peters, & D. Tiscornia (Eds.), *Semantic Processing of Legal Texts*. Lecture Notes in Computer Science, 6036. Springer, Berlin, Heidelberg.
- Yuriy, Y. Ch. (2014). Privetstvennoe slovo General'nogo prokurora Rossiyskoy Federatsii. *Russian Journal of Legal Studies*, 1(4), 7.

Bibliographic information of this paper for citing:

Farhadishad, Mohammad; Kazemifard, Mohammad & Rezaei, Zahra (2023). Predicting Court Judgment in Criminal Cases by Text Mining Techniques. *Journal of Information Technology Management*, 15 (2), 204-222. [https://doi.org/ 10.22059/jitm.2023.350464.3206](https://doi.org/10.22059/jitm.2023.350464.3206)
