Bilge International Journal of Science and Technology Research

ISSN: 2651-401X
e-ISSN: 2651-4028
Research Article

2023, Volume: 7, Issue: 2, 128-141
Received: 20.06.2023; Accepted: 15.09.2023
DOI: 10.30516/bilgesci.1317525

Prediction of Turkish Constitutional Court Decisions with Explainable Artificial Intelligence

Tülay Turan [1]*, Ecir Uğur Küçüksille [2], Nazan Kemaloğlu Alagöz [3]

Abstract: Using artificial intelligence in law is a topic that has attracted attention in recent years. This study aims to classify the case decisions taken by the Constitutional Court of the Republic of Turkey. For this purpose, open-access data published by the Constitutional Court of the Republic of Turkey on the website of the Decisions Information Bank were used in this research. KNN (K-Nearest Neighbors Algorithm), SVM (Support Vector Machine), DT (Decision Tree), RF (Random Forest), and XGBoost (Extreme Gradient Boosting) machine learning (ML) algorithms are used. Precision, Recall, F1-Score, and Accuracy metrics were used to compare the results of these models. As a result of the evaluation showed that the XGBoost model gave the best results with 93.84% Accuracy, 93% Precision, 93% Recall, and 93% F1-Score. It is important that the model result is not only good but also transparent and interpretable. Therefore, in this article, using the SHAP (SHapley Additive exPlanations) method, one of the explainable artificial intelligence techniques, the features that affect the classification of case results are explained. The study is the first study carried out in our country to use explainable artificial intelligence techniques in predicting court decisions in the Republic of Turkey with artificial intelligence.

Keywords: Explainable Artificial Intelligence, Turkish Constitutional Court, Legal Jugment Prediction, SHAP, XGBoost.

[1]**Address:** Süleyman Demirel University, Instıtute of Scıence, Computer Engineering, Isparta, Türkiye
[2]**Address:** Süleyman Demirel University, Department of Computer Engineering, Isparta, Türkiye
[3]**Address:** Isparta University of Applied Science, Department of Computer Technologies, Isparta, Türkiye

***Corresponding author**: tulayturan@mehmetakif.edu.tr

Citation: Turan, T., Küçüksille, E. U., Kemaloğlu Alagöz, N. (2023). Prediction of Turkish Constitutional Court Decisions with Explainable Artificial Intelligence. Bilge International Journal of Science and Technology Research, 7(2): 128-141.

1. INTRODUCTION

Along with big data, artificial intelligence (AI) technology, which has developed rapidly in recent years, appears with various work examples in many sectors. In the literature review, it is seen that the application diversity is high in fields such as education (Roll and Wylie 2016; Knox 2020; Chen et al., 2020; Meço and Çoştu 2022), defense (Bistron and Piotrowski 2021), health (Yu et al., 2018; Jiang et al., 2017; Turan et al., 2022), trade (Di Vaio et al., 2020; Loureiro et al., 2021) and engineering (Goertzel and Pennachin 2007). It has been observed that the application diversity in the field could be higher. In this context, the development of AI solutions in the field of law is seen as an area that attracts the attention of scientists and needs a lot of new work.

When the studies in the field of law are examined, it is seen that it has developed in parallel with the development of natural language processing, a sub-branch of AI. It has been reported that AI solutions have been developed under the sub-titles of litigation decision prediction (Zhong et al., 2018; Long et al., 2019), document analysis (Zadgaonkar and Agrawal 2021), legal assistance (Socatiyanurak et al., 2021), contract creation, and review (Antos and Nadhamuni 2021; Labin and Segal 2021) on legal documents. Making models, analyzing, and producing results using case decision data from these titles is called Legal Judgment Prediction (LJP). LJP practices are an important field of application as it quickly concludes case decisions, works without the need for rest periods that people need, do not show emotionality in the decisions it takes, and offer recommendations. Niklaus et al., reported in their study that there are approximately 24 million cases per 17,000 judges in India and approximately 1.6 million cases on only five charges in Brazil (Niklaus et al., 2022). The duration of conclusion of these cases will inevitably be prolonged. LJP practices provide various benefits for law firms, clients, and legal professionals as a solution to this problem. These benefits are shown in Table 1.

Table 1: Benefits of LJP

Law Firm	Thanks to the LJP, they can predict the outcome of the cases they defend, improve their defense case and increase their winning rate (Velez and Kim 2017).
Client	Thanks to LJP, they can get an estimate of the outcome of the case they want to file before applying to any law firm. If the result is that they cannot win the case, they will not incur any costs because they will not apply to the law firm (Stevenson and Wagoner 2015).
Court	The workload of judges and courts will be reduced as faster solutions are achieved in less time with LJP (Ma et al., 2021).

The diversity of the data source is important in developing LJP applications. The Ministry of Justice of the Republic of Turkey has decided that in the "Human Rights Action Plan" published in April 2021, all court decisions in Turkey will be digitized and made accessible.[1] In the literature review, Chinese and English LJP practices were used by using open access court decisions shared by the European Court of Human Rights (Aletras et al., 2016; Kaur and Bozic 2019; Collenette et al., 2020), Supreme People's Court of China (Yan et al., 2019; Yang et al., 2020; Gan et al., 2022) and Supreme Court of the United States (Strickson and Iglesia, 2020). The development of Turkish legal datasets and LJP applications is still in its infancy, and it is seen that there are only two studies (Sert et al., 2021; Mumcuoğlu et al., 2021). In this context, it is aimed to create a new Turkish data set as a first step in the research and to contribute to Turkish LJP studies.

The Constitutional Court of Turkey shares the decisions of norm review, individual application, supreme Court and political parties as open access on the "Decisions Information Bank" website. This study uses individual application court decisions shared by the Turkish Constitutional Court for a new Turkish legal dataset. Personal application decisions consist of approximately 10,000 lawsuits filed by individuals for alleged human rights violations since 2012. When the contents of the lawsuits are examined, it is seen that lawsuits have been filed for 24 fundamental rights and freedoms guaranteed by the Turkish Constitution. These lawsuits are based on "Violation, No Violation, Clearly Unfounded, Lack of Constitutional and Personal Importance, Non-Exhaustion of Recourses, Refusal of Application, Rejection, Objection to Administrative Rejection, No Place for Review, Incompetence in Terms of Person, Incompetence in Terms of Subject" for the relevant fundamental rights and freedoms. It is concluded with 15 different results, including "Incompetence of the Court, Time-Out of Time, Incompetence in terms of Place, Incompetence in terms of Time." The study aims to predict the result of "Violation" / "No Violation" from these results, and the data set was created accordingly.

Explainability is one of the important topics to be addressed while developing LJP applications. Classification models used in LJP applications can make decisions that are not easy to interpret with their non-linear and complex structures, and it is not known exactly how they reach the result. These results obtained by the models are defined as the black box (Mumford et al., 2021; Xu 2022). Explainable Artificial Intelligence (XAI) provides techniques to understand better and explain our model's results (Gunning and Aha 2019; Gunning et al., 2019). Thus, AI results defined as black box are converted into white box results. Disclosure of LJP model results is also important for legal professionals and clients awaiting explanations for why certain decisions were made. In this study, the results obtained by machine learning (ML) models are plotted using the SHApley Additive exPlanations (SHAP) method, one of the XAI techniques. Thus, the features that affect the model results are presented clearly and visually.

The contributions of this article to the literature are as follows.

1. Taking large volumes of legal data and automating the process of generating structured information and rules will reduce the time it takes for legal professionals to review hundreds of case documents and materials.
2. It will successfully expand the analysis of the case and help legal professionals reach the decision's conclusions faster.
3. It is the first study carried out in the country to explain the court decisions of the Republic of Turkey with XAI techniques.
4. Sharing a new publicly available Turkish legal dataset will significantly contribute to developing Turkish LJP practices.

2. RELATED WORK

Predicting the legal consequences of cases depends on many factors, such as evidence, witnesses, judges, opinions, and previous court decisions. Legal professionals try to predict a case's outcome using their experience, knowledge, professional judgment, and other cognitive skills, and for this, they have to review large amounts of data. Today, Lex Machina, Premonition Analytics, and Ravel Law have developed LJP software. These softwares draw attention with accelerating information access, optimizing time management, and successful forecasting results.

When the LJP studies in the world and our country are examined, it has been seen that the diversity of legal data

[1] https://insanhaklarieylemplani.adalet.gov.tr/

sets is not enough. Therefore, the scientists who work first start to develop applications by creating a legal data set. Xiao et al., revealed the first Chinese LJP dataset, CAIL2018. They created the data set using the decisions shared by the Supreme People's Court of China. Also, they pointed out that the transmitted data set was the largest LJP data set ever (Xiao et al., 2018). Chalkidis et al., (2019) stated that they created a data set in English using the decisions shared by the European Court of Human Rights. Their study stated whether the lawsuits' results are infringed or not by binary classification and multi-label classification. Niklaus et al., (2021) stated that they created a multilingual (German, French, and Italian) data set using the decisions shared by the Federal Supreme Court of Switzerland. They used BERT-based methods to develop models. In additionally, they stated that the Hierarchical BERT model with a Macro-F1-Score of 70% showed the best performance. Long et al., first created a Chinese dataset for automatic judgment estimation. In additionally, they propose AutoJudge, a new LRC model that captures complex semantics in their work. They reported that they obtained better estimation results with the AutoJudge model compared to many advanced models (Long et al., 2019). Guo et al., (2021) created the dataset they used in their studies from the Chinese referee document network page. In their work, they propose a new method called ModTen based on tensor models. They showed that their results were better than those obtained with the classification methods. Li et al., (2019) developed models with MANN, one of the deep neural network methods, using the Chinese dataset CAIL2018 in their study. Also, they compared the model they developed with the SVM, GRU, Bi-GRU, HAN, and TOPJUDGE models. As a result, they stated that the MANN model achieved the best results. Katz et al., (2017) created a data set using English decisions shared by the United States Supreme Court. The model they developed gave an accuracy rate of 70.2%. The studies of Mumcuoğlu et al., (2021) were available in the literature as the first comprehensive study of NLP for the Turkish legal system. In their studies, they classified deep learning methods using Decision Trees, Random Forests, Support Vector Machines and the decisions of the Constitutional Court and Supreme Court. Sert et al., (2021) have developed estimation models for "violation" or "non-violation" of Constitutional Court decisions. As a result of their studies, they achieved 90% success with MLP. The open-access court information used by the current LJP studies in the literature as the data source is shown in Table 2.

Table 2 Data Sources of LJP Articles

Year	Article	Dataset Source
2021	Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court (Sert et al., 2021)	Constitutional Court of Turkey
2021	Natural language processing in law: Prediction of outcomes in the higher courts of Turkey (Mumcuoğlu et al., 2021)	Constitutional Court of Turkey
2018	Cail2018: A large-scale legal dataset for judgment Prediction (Xiao et al., 2018)	Supreme People's Court of China
2019	Neural legal judgment prediction in English (Chalkidis et al., 2019)	European Court of Human Rights
2021	Swiss-judgment-prediction: a multilingual legal judgment prediction benchmark (Niklaus et al., 2021)	Federal Supreme Court of Switzerland
2019	Automatic judgment prediction via legal reading comprehension (Long et al., 2019)	Supreme People's Court of China
2021	TenLa: an approach based on controllable tensor decomposition and optimized lasso regression for judgment prediction of legal cases (Guo et al., 2021)	Supreme People's Court of China
2019	Mann: A multichannel attentive neural network for legal judgment Prediction (Li et al., 2019)	Supreme People's Court of China
2017	A general approach for predicting the behavior of the Supreme Court of the United States (Katz et al., 2017)	Supreme Court of the United States
2019	Convolutional Neural Network-based Automatic Prediction of Judgments of the European Court of Human Rights (Kaur and Bozic 2019)	European Court of Human Rights
2016	Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective (Aletras et al., 2016)	European Court of Human Rights
2020	An Explainable Approach to Deducing Outcomes in European Court of Human Rights Cases Using ADFs (Collenette et al., 2020)	European Court of Human Rights
2022	Exploiting Contrastive Learning and Numerical Evidence for Improving Confusing Legal Judgment Prediction (Gan et al., 2022)	Supreme People's Court of China
2019	Law article prediction based on deep learning (Yan et al., 2019)	Supreme People's Court of China
2020	Leniency to those who confess? Predicting the Legal Judgement via Multi-Modal Analysis (Yang et al., 2020)	Supreme People's Court of China
2020	Legal judgment prediction for UK courts (Strickson and Iglesia 2020)	Supreme Court of the USA
2022	ClassActionPrediction: A Challenging Benchmark for Legal Judgment Prediction of Class Action Cases in the US (Semo et al., 2022)	Supreme Court of the USA

The comparative use of XAI in LJP applications is a new topic, and there are few study examples in the literature. Gorski et al., used the Grad-CAM image processing technique to make it explicable in legal texts. Gorski et al., (2020) of the DistilBERT expression with an accuracy of 85% of the best performance from the models they developed in their studies. Gorski and Ramakrishna (2021) stated in their studies that they explain legal texts with explainable methods such as Grad-CAM, LIME, and SHAP. They evaluated the results they obtained by contacting protective lawyers.

3. MATERIAL AND METHOD

3.1. Dataset

The data set was created from the individual application decisions shared by the Constitutional Court of the Republic of Turkey on the Decisions Information Bank website. When the transmitted cases are examined, it is seen that they were opened for 24 fundamental rights and freedoms guaranteed by the Turkish Constitution and were concluded with 15 different results. This study aims to predict the result of "Violation" and "No Violation" of fundamental rights and freedoms. For this purpose, all decisions with the result of "Violation" and "No Violation" have been obtained from the web page. When the cases taken were examined, it was seen that there were no exemplary decisions for some rights and freedoms. Table 3 shows the data set content of the study.

Table 3 Dataset Content

No	Right Freedom	Violation	No Violation
1	Right to Fair Trial (Penalty)	60	60
2	Right to Fair Trial (Law)	60	60
3	Right to Fair Trial (Administration)	60	60
4	Prohibition of Discrimination	3	3
5	Individual Application Right	0	0
6	Freedom of Religion and Conscience	1	1
7	Education right	3	3
8	Effective Right to Apply	2	2
9	Right to Request Review of the Provision	0	0
10	Freedom of expression	50	50
11	Excluded Rights	0	0
12	Right to Personal Freedom and Security	80	80
13	Abuse Ban	60	60
14	The right to protection of material and spiritual property	50	50
15	Freehold	80	80
16	Freedom of association	10	10
17	Protection of private and family life right	55	55
18	The right to vote, be elected, and engage in political activity	4	4
19	The legality of crimes and punishments principles	2	2
20	The right to organize meetings and demonstration marches	20	20
21	Right to life	50	50
22	Prohibition of forced labor and drudgery	0	0
23	Other Rights	0	0
24	Union Right	0	0
	Total	**650**	**650**
	Overall Total	**1300**	

When the decisions were examined, it was seen that they consisted of six main parts. These sections are the subject of the application, the application process, the events and facts, the relevant law, the examination, and the law and judgment section, respectively. To determine of these sections would be the independent variable and the dependent variable for the models, interviews were held with the lawyers. As a result, it has been decided to designate the "Examination and Justification" section as the "independent variable" and the "Provision" section as the "dependent variable." The dataset created for this study has been published on GitHub and can be accessed at https://github.com/tulayturan/KararListesi.

3.2. K-Nearest Neighbors (KNN)

The KNN algorithm is based on the logic of calculating the distance of the unknown data from other data and including it in the nearest class (Nikam, 2015). Euclidean, Manhattan, or Minkowski methods are used to calculate the distance measure (Shahid et al., 2009; Singh et al., 2013)

Euclidean distance calculation is calculated as given in Equation 1.

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

Manhattan distance calculation is calculated as given in Equation 2.

$$\sum_{i=1}^{n}|x_i - y_i| \tag{2}$$

Minkowski distance calculation is calculated as given in Equation 3.

$$\left(\sum_{i=1}^{n}(|x_i - y_i|)^q\right)^{1/q} \tag{3}$$

The n value in the equations can be defined as "no of dimensions," the x value as "datapoint from the dataset," and the y value as "new data point (to be predicted)."

3.3. Suport Vector Machine (SVM)

SVM is an algorithm that properly separates data from two or more classes (Ghosh et al., 2019). The separation of classes is determined by decision boundaries or hyperplanes (Somvanshi et al., 2016). In Figure 1, a red hyperplane and hyperplanes belonging to each class separate the two classes from each other. The region between +1 and -1 is called Margin. The wider the margin value, the better it classifies two or more classes (Brereton and Lloyd, 2010).

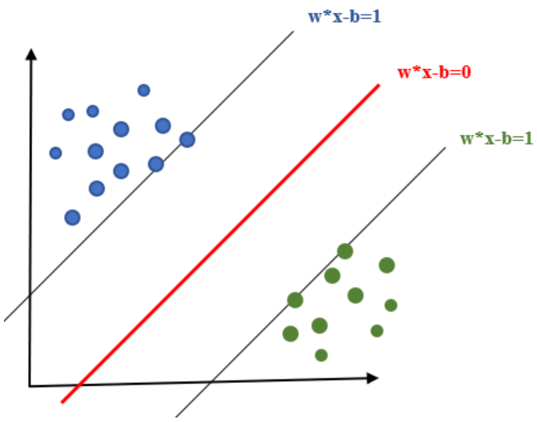

Figure 1. Class Separation with SWM

In Figure 1;
The w*x-b=1 hyperplane and each data represent the class members labeled 1.
The w*x-b=-1 hyperplane and any data below it represents class members labeled -1.

3.4. Random Forest (RF)

The RF classifier is based on evaluating the predictions produced by multiple decision trees (Rokach, 2016). Observations for trees are determined by the Bagging method, and the variables are determined by the Random Subspace method. RF uses Gini as the feature selection measure, and Gini calculus is calculated as given in Eq. 4.

$$\sum\sum_{j\neq i}(f(C_i,T)/|T|)(f(C_j,T)/|T|) \tag{4}$$

An unexpected situation is selected for the T training set in the equation, and whether the selected value belongs to the Ci class is calculated.

3.5. Decision Tree (DT)

The first cells of the decision trees are called Root Nodes (Fletcher and Islam, 2019). There are nodes and leaves under the root. The root node is decided in decision trees by calculating Gini or Entropy values (Nanfack et al., 2022). Calculating the Gini value is shown in Equation 5, and the calculation of the Entropy value is shown in Equation 6.

$$Gini = 1 - \sum_j p_j^2 \tag{5}$$

$$\mathrm{Entropy} = -\sum_{j=1}^{c} \mathrm{p_j} \log_{2}(\mathrm{p_j}) \tag{6}$$

pj is the probability of occurrence of class j. It is calculated for each class, and the sum of the squares of the results is subtracted. The Gini value gets a result between 0 and 1, and the closer the result is to 0, the better the discrimination.

3.6. XGBoost

XGBoost is an ML algorithm proposed by Chen and Guestirn (2016). Boosting Tree algorithms are based on a decision tree known as a classification and regression tree (CART) (Dong et al., 2020).

The advantage of XGboost is that it supports linear classifiers and performs second-order Taylor expansion of the cost function to make the results more accurate. The loss function score used in the XGBoost algorithm and the solution of the weights are expressed as follows (Jiang et al., 2020).

$$w_j^* = -\frac{\sum g_i}{\sum h_i + \lambda} \tag{7}$$

$$obj^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{(\sum g_i)^2}{\sum h_i + \lambda} + \lambda.T \tag{8}$$

In the equation, obj∗ represents the score of the loss function. The smaller the score, the better the structure of the tree. w∗j represents the solution of the weights.

3.7. Model Evaluation Metrics

In this study, Confusion Matrix, Accuracy, Sensitivity, Precision, and Recall calculation methods were used to evaluate the performance success of the models. A confusion Matrix is an analysis tool that shows the extent to which a classifier can classify different class labels. In this study, the data set has 2 class labels and will be 2k*2 in size, as shown in Confusion Matrix Figure 2.

		ACTUAL	
		Negative	**Positive**
PREDICTION	**Negative**	True Negative (TN)	False Negative (FN)
	Positive	False Positive (FP)	True Positive (TP)

Figure 2. Confusion Matrix

Accuracy, Sensitivity, Precision, and Recall calculations of the models can be made using the Confusion Matrix values.

The Accuracy value is calculated as shown in Equation 9.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

The sensitivity value is calculated as shown in Equation 10.

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

The precision value is calculated as shown in Equation 11.

$$Precision \ = \frac{TP}{TP + FP} \quad (11)$$

The Recall value is calculated as shown in Equation 12.

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

3.8. Hyperparameter Tuning

Hyperparameters are parameters determined initially before the learning process starts in a machine learning model. The values of these parameters do not change when the learning process is over. Learning Rate, Epoch, etc., parameters can be given as examples.

Hyperparameter values given by default in machine learning models do not guarantee the best performance (Schratz et al., 2019). Therefore, the determination of hyperparameter values can greatly affect the model's performance (Mantodan et al., 2016). A large amount of hyperparameters in the models makes it almost impossible to adjust these values manually. For this reason, many hyperparameter tune methods are available and used in the literature. The most used methods are GridSearch and RandomSearch.

GridSearch creates a new model by trying all possible combinations from the given collection of values for each hyperparameter and returns the hyperparameter combination that provides the highest accuracy. The problem with this method is the process takes a long time when there are too many hyperparameters and values to try, and this causes the technique to run very slowly.

In the RandomSearch method, N combinations determined from each hyperparameter value collection are randomly selected and return the hyperparameter combination that provides the highest accuracy. With this method, a search can be made much faster and with an accuracy close to GridSearch.

3.9. Explainable Artificial Intelligence (XAI)

Today, where technological developments are advancing at a dizzying pace, AI; is used in many points that directly affect human life, such as health, autonomous vehicles, and military areas. Parallel to these developments, the fact that the mechanism used by AI in making decisions (black box) is not known exactly causes questions about reliability, transparency, bias, and fairness. In light of all this information, XAI can be defined as a field consisting of tools, techniques, and algorithms that can produce human-explainable explanations of AI decisions (Das and Rad, 2020). XAI is divided into different groups according to various approaches.

XAI scope is divided into local and global, depending on whether you understand the model from a local instance or as a whole. Local XAI focuses on disclosure on a single data basis and creates a description for each data. In Global XAI, the model is tried to be explained as a whole (Anders et al., 2021; Spinner et al., 2019). Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016) and SHApley Additive exPlanations (SHAP) (Lundberg and Lee 2017) methods can be given as examples for Local XAI. For Global XAI, Class Model Visualization [80] and Spectral Relevance Analysis (SpRAy) (Lapuschkin et al., 2019) methods can be given as examples. At this point, LIME and SHAP can also be used for global XAI.

BackProb and Pertubration, depending on the algorithmic approach used, whether it focuses on the input sample or the model parameters. BackProb XAI's description depends on the gradients propagating back from the prediction layer to the input layer. In Perturbation XAI, however, the explanation depends on random or carefully chosen changes in properties in the input data (Nie et al., 2018; Lin et al., 2020; Ivanovs et al., 2021). For BackProb XAI, DeConvolutional Nets (Zeiler and Fergus, 2014) and SpRAy methods can be given as examples. For Perturbation, XAI, LIME, and SHAP methods can be given as examples.

XAI is divided into two depending on integration into a particular model and can be applied to the desired model in general. In Intrinsic XAI, explainability resides in the synargy architecture and cannot be transferred to other architectures. In Post-Hoc XAI, however, the algorithm does not depend on the model architecture and can be pre-trained neural networks (Weber and Wermter 2020; Tritscher et al., 2020; Kenny et al., 2021, Colaner 2021). Neural Additive Models (Agarwal et al., 2020) and Bayes Rule Lists (Letham et al., 2015) are examples of Intrinsic XAI. For Post-Hoc XAI, LIME and SHAP can be given as examples. Besides, Automatic Concept-based Explanations (Ghorbani et al., 2019) are one of the Post-Hoc methods.

3.10. SHapley Additive exPlanations (SHAP)

SHAP is a game theory-based method used to describe the performance of a machine learning model. In SHAP, an output model is defined as a linear sum of input variables (Mangalathu et al., 2020). SHAP is defined as a model-independent interpretability method because it can derive post-hoc explanations for the predictions of any classification model by associating inputs with output (Tideman et al., 2021).

Equation 13, which explains the importance of the i feature as a Shapley value, is given below (Rodríguez-Pérez and Bajorath 2020).

$$\emptyset_i = \frac{1}{|N|!} \sum_{S \subseteq N\{i\}} |S|!\,(|N| - |S| - 1)!\,[f(S \cup \{\dot{I}\}) - f(S)] \quad (13)$$

Here f(S) is the model's output, and N is the set of all features. In the equation, the Shapley value of the feature i expresses the mean of the contributions in all possible permutations of a feature set. In this method, features are added to the set separately. The model output change shows that variable's relevance.

4. RESULTS

The developed system consists of 5 parts. In the first stage, the content of the Turkish legal data set was passed through the data preprocessing steps. In the second stage, the vector space model of each word was created with the tf-idf (Term Frequency - Inverse Document Frequency) method. The supervised learning model classification techniques KNN, SVM, DT, RF, XGBoost, and LJP models were developed in the third stage. In the fourth stage, the classification performances of the models were compared with precision, recall, F1-score, and accuracy evaluation metrics. In the last section, the XGBoost model, which achieved the best result with an accuracy rate of 93.07%, is explained with SHAP plot techniques. The developed system diagram is given in Figure 3.

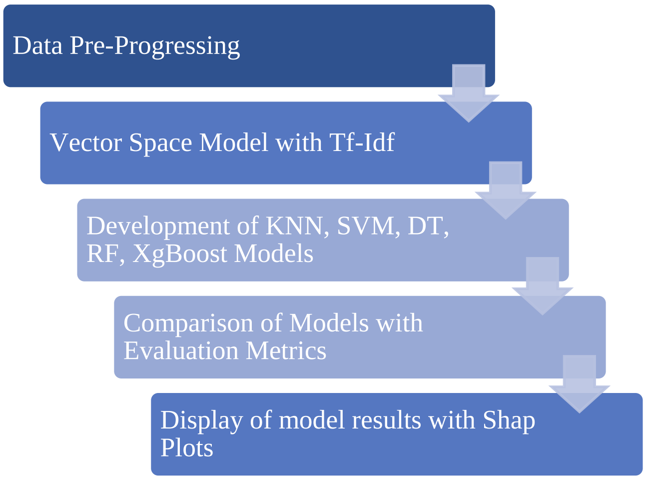

Figure 3 Developed System Diagram

Data preprocessing can be defined as the operations performed on the data set before AI models are developed. It is an important step in applications developed with natural language processing. Data preprocessing, transformation, deletion of duplicate data, and editing of noisy, incomplete, or contradictory data are performed on the data. As a result of this process, the model performance increases. The data preprocessing steps performed on the data in the study are shown in Figure 4.

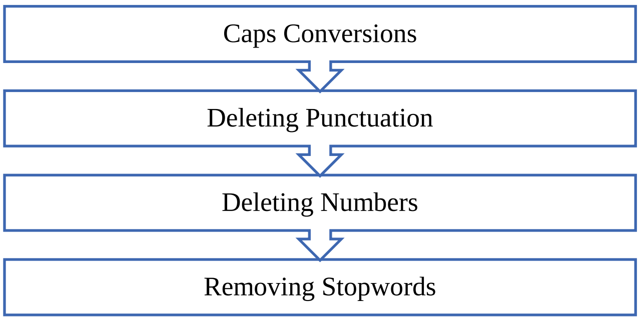

Figure 4. Data Preprocessing Steps

After the data preprocessing step, the vector space model of each word was created with the tf-idf method. The tf-idf method is used in the digitization of the text. Each text is an MxN vector of available words. The document term matrix is formed by superimposing the vectors. This matrix consists of M texts and N terms. If terms are mentioned in the text, the weight value of that term will be different from "0". Tf-idf looks at the frequency of the related term in the text "Tf" and its importance in the text "idf."

While calculating the Tf text frequency value, the ratio of the number of terms in the sentence to the total number of words is considered. Eq.14 shows the calculation method.

$$tf(t,d) = \frac{f_{t,d}}{\sum t' \epsilon \, d \; ft',d} \tag{14}$$

d : Document
t : Relative Frequency of Term

When calculating the Idf importance value, it is the logarithm of the total number of sentences, the ratio of the selected term to the total number of sentences in all sentences. Eq. 15 shows the calculation method.

$$idf = (t,D) = \log \frac{N}{|\{d \; \epsilon \; D : t \; \epsilon \; d\}|} \tag{15}$$

After calculating the tf and idf values, the tf-idf value of each word is obtained by multiplying the two values found. Eq. 16 shows the calculation method.

$$tfidf = (t,d,D) = tf(t,d).idf(t,D) \tag{16}$$

The data for which the vector space model was created was divided into an 80% training set and a 20% test set by the Holdout method in model validation methods. As a result, of the 1300 case texts in the data set, 1040 were used for training and 260 for evaluation.

In the third stage of the study, models were developed with KNN, SVM, DT, RF, and XGBoost. Hyperparameter optimization was performed on the models to obtain the best accuracy result. The GridSearchCv object in the scikit–learn library is used for this. With GridSearchCv, separate models were established for the model's hyperparameter values desired to be tested, and the hyperparameter values that gave the most successful results were determined. The hyperparameter values that provide the best accuracy result for the models are shown in Table 4.

Table 4 Hyperparameter results of models

Models	HyperParameters	Value
K-Neaest Neighbors Model	n_neighbors	8
	p(metric)	Manhattan
Support Vector Machines Model	c	3
	kernel	rbf
Random Forest Model	max_features	8
	min_samples_split	5
	n_estimators	100
Decision Tree Model	criterion	Gini
	max_depth	3
	min_samples_split	2
XGBoost Model	learning_rate	0.1

	max_depth	3
	n_estimators	500
	subsample	0.6

After the hyperparameter optimization, the final models were established, and the classification performances of the models were compared with the precision, recall, F1-score, and accuracy evaluation criteria. As a result of the assessment showed that the XGBoost model gave the best results with 93.84% Average Accuracy, 93% Precision, 93% Recall, and 93% F1-Score. The performance values of the KNN, SVM, DT, RF, and XGBoost models are shown in Table 5.

Table 5 Model Performance Values

Model	**Class**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
KNN Model	No violation (0)	0.80	0.92	0.86	
	Violation (1)	0.93	0.83	0.88	90.76
	Mean/ Total	0.87	0.88	0.87	
SVM Model	No violation (0)	0.84	0.94	0.88	
	Violation (1)	0.95	0.87	0.91	89.61
	Mean/ Total	0.89	0.90	0.90	
DT Model	No violation (0)	0.83	0.89	0.86	
	Violation (1)	0.92	0.87	0.89	88.07
	Mean/ Total	0.87	0.88	0.88	
RF Model	No violation (0)	0.81	0.95	0.87	
	Violation (1)	0.95	0.84	0.89	90
	Mean/ Total	0.88	0.89	0.88	
XGBoost Model	No violation (0)	0.90	0.94	0.92	
	Violation (1)	0.95	0.93	0.94	93.84
	Mean/ Total	0.93	0.93	0.93	

Classification estimation successes of the models are also shown in Figure 5 with ROC (Receiver Operating Characteristic) curves.

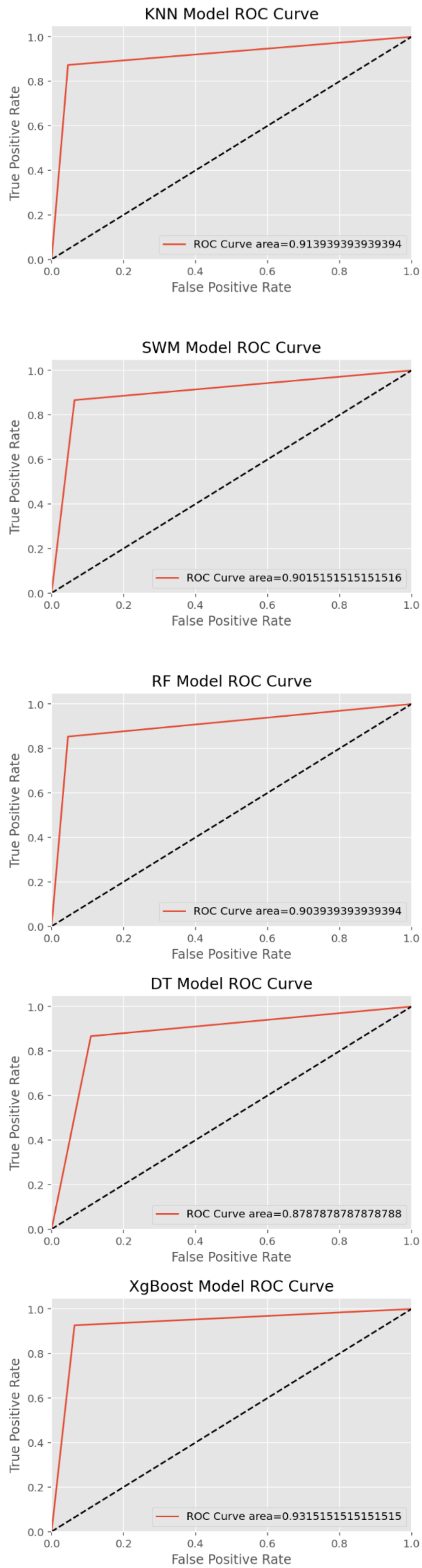

Figure 5. ROC Curve Area

XGBoost, where the best prediction result is obtained in our LJP application, is a tree-based complex machine learning algorithm. Although it provides very high success, the results cannot be interpreted directly. In this study, using the SHAP method, one of the XAI techniques, the variables that affect the classification of the case decisions of the XGBoost model are explained with graphics. The waterfall plot shows the variables that push the model output from the baseline to the predicted model output. In the graph, variables that make the forecast higher are shown with a positive (red) contribution value, and variables that push the forecast down are shown with a negative (blue) contribution value. Figure 6 shows how much each variable affects the estimation result for the first case decision. This is also the result of the local interpretability of the initial decision. In the figure, it is seen that the variable "kararında" is the variable that most affects the f(x) value with the positive value of "+1.64". It is seen that the bars in the graph are mostly red and have positive values. Accordingly, the variables increase the SHAP value for the first case decision. As a result, since the f(x) value moves away from E[f(x)] value positively, it is seen that the case may result in "Violation" for the first case decision.

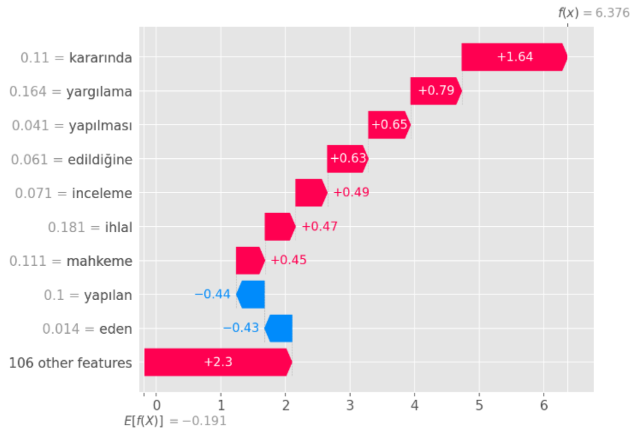

Figure 6. Waterfall Plot for First Case Decision

The force plot in Figure 7 is one of the local interpretability plots that show how variables contribute to the model's prediction. The graph shows the variables that contributed the most to the initial lawsuit decision and their marginal contributions, similar to the Waterfall graph. It is seen that the bars in the graph are mostly red and have positive values. As a result, the case may result in a "Violation" for the first case decision.

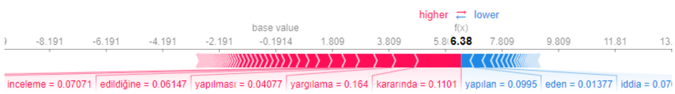

Figure 7. Force Plot for First Case Decision

The bar graph clearly shows the effects of the variables on the model output, thanks to its simple appearance. Figure 8 shows how much each variable affects the estimation result for the 100th example of the case decisions. In the graph, it is seen that the variable "edilmediğine" is the variable that affects the f(x) value the most, with the negative value of "-1.5". It is seen that the bars in the graph are mostly blue and have negative values. Accordingly, for the 100th case decision sample, the variables decrease the SHAP value. As a result, since the f(x) value deviates from the E[f(x)] value in a negative direction, it is seen that the case for the 100th case decision may result in "No Violation."

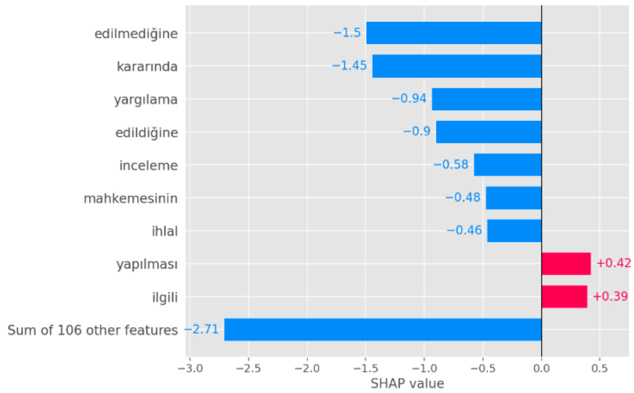

Figure 8 Bar Chart for Centenary Verdict

Beeswarm and Summary plot charts are used to explain the importance of the variables and their contribution to the model on the whole data set. The global interpretability of the model trained with XGBoost in this study is shown in Figure 9 and Figure 10. Each graph dot represents a case decision, while the X-axis shows the SHAP values. When the examine the results obtained in both graphs, it is seen that the variable "kararında" makes the most marginal contribution to the estimates. It is also seen that with the increase of this variable value, the SHAP value also increases. As a result of this, it is seen that the probability that the case will result in "Violation" increases.

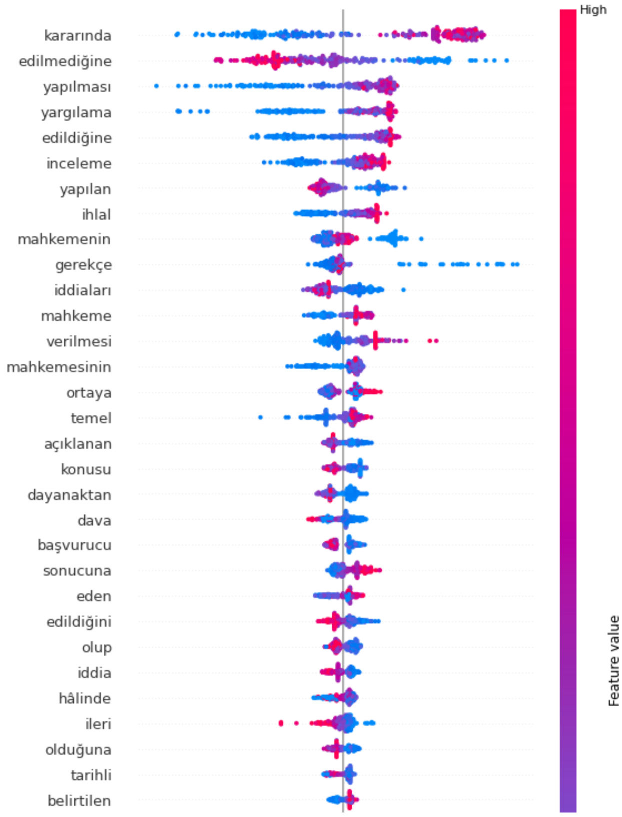

Figure 9. Besswarm Plot

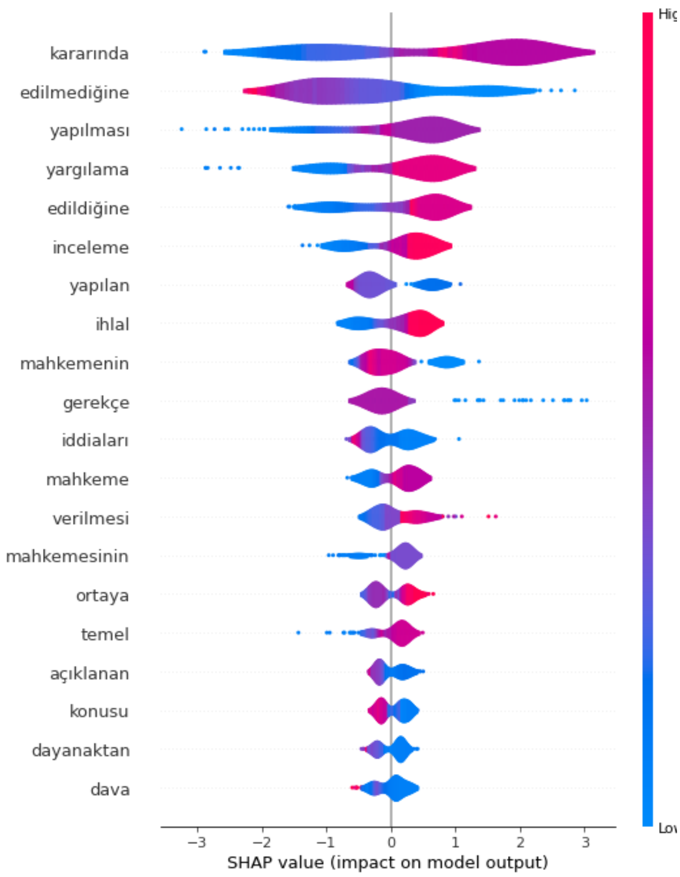

Figure 10. Summary Plot

The last SHAP chart used in the study is the Heatmap plot shown in Figure 11. This graph shows the global interpretability of the trained model. In the graph, the X-axis represents the decision examples, and the y-axis represents the variables. The f(x) curve at the top of the graph is the model estimates of the samples. To the right of the graph are the SHAP values encoded in the color scale. According to the graph, the word "kararında" is the most important variable, and each decision's impact value is shown.

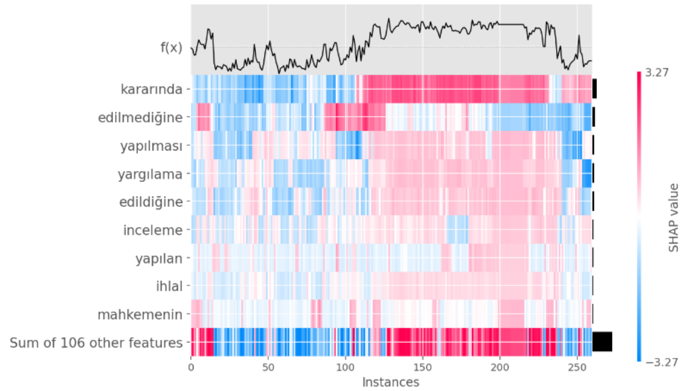

Figure 11. Heatmap Plot

5. DISCUSSION AND CONCLUSIONS

With the LJP applications developed due to the combination of artificial intelligence and legal data, some studies have started a new process in the legal sector. Litigation decisions that take a long time can be concluded quickly, the workload of the legal staff will be alleviated, and they can produce logical, clear results because they do not show emotionality in the decisions. With this study, the Turkish LJP application, which has very few working examples in Turkey, has been developed. The results of the Turkish Constitutional Court case decisions were estimated using the KNN, SVM, DT RF, and XGBoost algorithms. Among the developed models, the XGBoost model achieved the best prediction result with an accuracy rate of 93.07%. The results of the XGBoost model were made more transparent and interpretable by using the SHAP plots techniques such as Waterfall plot, Force plot, Beeswarm plot, Summary plot, Bar plot, and Heatmap plot. According to the results obtained, it was seen that SHAP values produced consistent results. As a result, end-users trust the models and that the decisions are fair and reliable. The AI models, prediction success rates, and the methods used by the Turkish LJP applications in the developed study and the literature are compared in Table 6. Accordingly, it is seen that the study stands out with its success rate and the XAI methods it uses.

Table 6 Comparison of Turkish LJP Studies

Year	Article	Court	Machine Learning Models			XAI Plots
2021	Predicting Turkish Constitutional Court Decisions using Artificial Intelligence (Sert et al., 2021)	Constitutional Court of Turkey	MLP	% 90.0	None	None
2021	Using natural language processing to predict decisions in Turkey's higher courts (Mumcuoğlu et al., 2021)	Constitutional Court of Turkey	DT	% 85.1	None	None
			RF	% 87.6		
			SWM	% 85.5		
			DL	% 91.8		
2022	This study	Constitutional Court of Turkey	KNN	[illegible]	Shap	Force Plot
			SWM	[illegible]		Bar Plot
			DT	% 87.6		Summary Plot
			RF	% 88.4		Beeswarm Plot
			XGBoost	% 93.0		Heatmap Plot

The study is the first in Turkey to explain the model results using SHAP graphics, one of the XAI techniques. The new Turkish law data set has been brought to the literature with the study. In future studies, models with different AI algorithms can be built on the data set, applications can be developed, and Turkish LJP studies can be contributed.

Ethics Committee Approval
N/A

Peer-review
Externally peer-reviewed.

Author Contributions
All authors have read and agreed to the published version of
manuscript.

Conflict of Interest
The authors have no conflicts of interest to declare.

Funding
The authors declared that this study has received no financial support.

REFERENCES

Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., & Hinton, G. E. (2021). Neural additive models: Interpretable machine learning with neural nets. Advances in Neural Information Processing Systems, 34, 4699-4711.

Altreas, 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 Comput Sci.

Anders, C. J., Neumann, D., Samek, W., Müller, K. R., & Lapuschkin, S. (2021). Software for dataset-wide XAI: from local explanations to global insights with Zennit, CoRelAy, and ViRelAy. arXiv preprint arXiv:2106.13200.

Antos, A., Nadhamuni, N. (2021). Practical guide to artificial intelligence and contract review. In: Research Handbook on Big Data Law, ed. Vogl, R., 467-481, Edward Elgar Publishing.

Bistron, M., Piotrowski, Z. (2021). Artificial intelligence applications in military systems and their influence on sense of security of citizens. Electronics, 10(7), 871.

Brereton, RG., Lloyd, GR. (2010). Support vector machines for classification and regression. Analyst, 135(2), 230-267.

Chalkidis, I., Androutsopoulos, I., Aletras, N. (2019). Neural legal judgment prediction in English. arXiv preprint arXiv:1906.02059.

Chen, L., Chen, P., Lin, Z. (2020). Artificial intelligence in education: A review. Ieee Access, 8, 75264-75278.

Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794, Association for Computing Machinery, New York, United States.

Colaner, N. (2021). Is explainable artificial intelligence intrinsically valuable? AI & SOCIETY, 37, 231-238.

Collenette, J., Atkinson, K., Bench-Capon, T. J. (2020). An Explainable Approach to Deducing Outcomes in European Court of Human Rights Cases Using ADFs, In: COMMA, ed. Prakken, H., Bistarelli, S. and Santini, F., 21-32, IOS Press.

Das, A., Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. arXiv preprint arXiv:2006.11371.

Di Vaio, A., Palladino, R., Hassan, R., Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. Journal of Business Research, 121, 283-314.

Dong, W., Huang, Y., Lehane, B., Ma, G. (2020). XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. Automation in Construction, 114, 103155.

Doshi-Velez, F., Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Fletcher, S., Islam, M. Z. (2019). Decision tree classification with differential privacy: A survey. ACM Computing Surveys (CSUR), 52(4), 1-33.

Gan, L., Li, B., Kuang, K., Yang, Y., & Wu, F. (2022). Exploiting Contrastive Learning and Numerical Evidence for Improving Confusing Legal Judgment Prediction. arXiv preprint arXiv:2211.08238.

Ghorbani, A., Wexler, J., Zou, J. Y., Kim, B. (2019). Towards automatic concept-based explanations. Advances in Neural Information Processing Systems, 32.

Ghosh, S., Dasgupta, A., Swetapadma, A. (2019). A study on support vector machine based linear and non-linear pattern classification. In: 2019 International Conference on Intelligent Sustainable Systems (ICISS) (pp. 24-28). IEEE.

Goertzel, B., Pennachin, C. (2007). The Novamente artificial intelligence engine. Artificial general intelligence, 63-129.

Górski, Ł., Ramakrishna, S. (2021, June). Explainable artificial intelligence, lawyer's perspective. In: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law (pp. 60-68).

Gorski, L., Ramakrishna, S., Nowosielski, J. M. (2020). Towards grad-cam based explainability in a legal text processing pipeline. arXiv preprint arXiv:2012.09603.

Gunning, D., Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. AI magazine, 40(2), 44-58.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G. Z. (2019). XAI—Explainable artificial intelligence. Science robotics, 4(37).

Guo, X., Zhang, H., Ye, L., Li, S. (2021). TenLa: an approach based on controllable tensor decomposition and optimized lasso regression for judgement prediction of legal cases. Applied Intelligence, 51, 2233-2252.

Ivanovs, M., Kadikis, R., Ozols, K. (2021). Perturbation-based methods for explaining deep neural networks: A survey. Pattern Recognition Letters, 150, 228-234.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology, 2(4).

Jiang, H., He, Z., Ye, G., Zhang, H. (2020). Network intrusion detection based on PSO-XGBoost model. IEEE Access, 8, 58392-58401.

Katz, D. M., Bommarito, M. J., Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. PloS one, 12(4).

Kaur, A., Bozic, B. (2019). Convolutional Neural Network-based Automatic Prediction of Judgments of the European Court of Human Rights. In: AICS, pp 458-469

Kenny, E. M., Ford, C., Quinn, M., Keane, M. T. (2021). Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies. Artificial Intelligence, 294, 103459.

Knox, J. (2020). Artificial intelligence and education in China. Learning, Media and Technology, 45(3), 298-311.

Labin, S., Segal, U. (2021). AI-driven contract review: A product development journey. In Research Handbook on Big Data Law, 454-466, Edward Elgar Publishing.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., Müller, K. R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. Nature communications, 10(1), 1096.

Letham, B., Rudin, C., McCormick, T. H., Madigan, D. (2012). Building interpretable classifiers with rules using Bayesian analysis. Department of Statistics Technical Report tr609, University of Washington, 9(3), 1350-1371.

Li, S., Zhang, H., Ye, L., Guo, X., Fang, B. (2019). Mann: A multichannel attentive neural network for legal judgment prediction. IEEE Access, 7, 151144-151155.

Lin, Y. S., Lee, W. C., Celik, Z. B. (2021). What do you see? Evaluation of explainable artificial intelligence (XAI) interpretability through neural backdoors. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (pp. 1027-1035).

Long, S., Tu, C., Liu, Z., Sun, M. (2019). Automatic judgment prediction via legal reading comprehension. In: Chinese Computational Linguistics: 18th China National Conference, 558-572, Springer International Publishing.

Loureiro, S. M. C., Guerreiro, J., Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. Journal of business research, 129, 911-926.

Lundberg, SM., Lee, SI. (2017). A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems, 30, 4768-4777.

Ma, L., Zhang, Y., Wang, T., Liu, X., Ye, W., Sun, C., Zhang, S. (2021). Legal judgment prediction with multi-stage case representation learning in the real court setting. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 993-1002).

Mangalathu, S., Hwang, S. H., Jeon, J. S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures, 219, 110927.

Mantovani, R. G., Horváth, T., Cerri, R., Vanschoren, J., De Carvalho, A. C. (2016, October). Hyper-parameter tuning of a decision tree induction algorithm. In: 2016 5th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 37-42). IEEE.

Meço, G., Çoştu, F. (2022). Eğitimde Yapay Zekânın Kullanılması: Betimsel İçerik Analizi Çalışması. Karadeniz Teknik Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Dergisi, 12(23), 171-193.

Mumcuoğlu, E., Öztürk, C. E., Ozaktas, H. M., Koç, A. (2021). Natural language processing in law:

Prediction of outcomes in the higher courts of Turkey. Information Processing & Management, 58(5), 102684.

Mumford, J., Atkinson, K., Bench-Capon, T. (2021). Machine learning and legal argument. In: CEUR Workshop Proceedings (Vol. 2937, pp. 47-56).

Nanfack, G., Temple, P., Frénay, B. (2022). Constraint Enforcement on Decision Trees: A Survey. ACM Computing Surveys (CSUR), 54(10s), 1-36.

Nie, W., Zhang, Y., Patel, A. (2018). A theoretical explanation for perplexing behaviors of backpropagation-based visualizations. In: International Conference on Machine Learning, PMLR, 3809-3818.

Nikam, SS. (2015). A comparative study of classification techniques in data mining algorithms. Oriental Journal of Computer Science and Technology, 8(1), 13-19.

Niklaus, J., Chalkidis, I., Stürmer, M. (2021). Swiss-judgment-prediction: A multilingual legal judgment prediction benchmark. arXiv preprint arXiv:2110.00806.

Niklaus, J., Stürmer, M., Chalkidis, I. (2022). An Empirical Study on Cross-X Transfer for Legal Judgment Prediction. arXiv preprint arXiv:2209.12325.

Ribeiro, M. T., Singh, S., Guestrin, C. (2016). Model-agnostic interpretability of machine learning. arXiv preprint arXiv:1606.05386.

Rodríguez-Pérez, R., Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. Journal of computer-aided molecular design, 34, 1013-1026.

Rokach, L. (2016). Decision forest: Twenty years of research. Information Fusion, 27, 111-125.

Roll, I., Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26, 582-599.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling, 406, 109-120.

Semo, G., Bernsohn, D., Hagag, B., Hayat, G., Niklaus, J. (2022). ClassActionPrediction: A Challenging Benchmark for Legal Judgment Prediction of Class Action Cases in the US. arXiv preprint arXiv:2211.00582.

Sert, M. F., Yıldırım, E., Haşlak, İ. (2022). Using artificial intelligence to predict decisions of the Turkish constitutional court. Social Science Computer Review, 40(6), 1416-1435.

Shahid, R., Bertazzon, S., Knudtson, M. L., Ghali, W. A. (2009). Comparison of distance measures in spatial analytical modeling for health service planning. BMC health services research, 9(1), 1-14.

Simonyan, K., Vedaldi, A., Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Singh, A., Yadav, A., Rana, A. (2013). K-means with Three different Distance Metrics. International Journal of Computer Applications, 67(10), 13-17.

Socatiyanurak, V., Klangpornkun, N., Munthuli, A., Phienphanich, P., Kovudhikulrungsri, L., Saksakulkunakorn, N., Tantibundhit, C. (2021). Law-u: Legal guidance through artificial intelligence chatbot for sexual violence victims and survivors. IEEE Access, 9, 131440-131461.

Somvanshi, M., Chavan, P., Tambade, S., Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine. In 2016 international conference on computing communication control and automation (ICCUBEA) (pp. 1-7). IEEE.

Spinner, T., Schlegel, U., Schäfer, H., El-Assady, M. (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE transactions on visualization and computer graphics, 26(1), 1064-1074.

Stevenson, D., Wagoner, N. J. (2015). Bargaining in the shadow of big data. Fla. L. Rev., 67, 1337.

Strickson, B., De La Iglesia, B. (2020, March). Legal judgement prediction for uk courts. In Proceedings of the 3rd International Conference on Information Science and Systems (pp. 204-209).

Tideman, L. E., Migas, L. G., Djambazova, K. V., Patterson, N. H., Caprioli, R. M., Spraggins, J. M., Van de Plas, R. (2021). Automated biomarker candidate discovery in imaging mass spectrometry data through spatially localized Shapley additive explanations. Analytica Chimica Acta, 1177, 338522.

Tritscher, J., Ring, M., Schlr, D., Hettinger, L., & Hotho, A. (2020). Evaluation of post-hoc XAI approaches through synthetic tabular data. In Foundations of Intelligent Systems: 25th International Symposium, ISMIS 2020, Graz, Austria, September 23–25, 2020, Proceedings (pp. 422-430). Springer International Publishing.

Turan. T., Turan, G., Köse, U. (2022). Uyarlamalı Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi ve Yapay Sinir Ağları ile Türkiye’deki COVID-19 Vefat Sayısının Tahmin Edilmesi. Bilişim Teknolojileri Dergisi 15(2), 97-105.

Weber, T., Wermter, S. (2020). Integrating intrinsic and extrinsic explainability: The relevance of understanding neural networks for human-robot interaction. arXiv preprint arXiv:2010.04602.

Xiao, C., Zhong, H., Guo, Z., Tu, C., Liu, Z., Sun, M., Xu, J. (2018). Cail2018: A large-scale legal dataset for judgment prediction. arXiv preprint arXiv:1807.02478.

Xu, Z. (2022). Human judges in the era of artificial intelligence: challenges and opportunities. Applied Artificial Intelligence, 36(1), 2013652.

Yan, G., Li, Y., Shen, S., Zhang, S., Liu, J. (2019, July). Law article prediction based on deep learning. In 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C) (pp. 281-284). IEEE.

Yang, L., Zeng, J., Peng, T., Luo, X., Zhang, J., Lin, H. (2020, October). Leniency to those who confess. Predicting the Legal Judgement via Multi-Modal Analysis. In Proceedings of the 2020 International Conference on Multimodal Interaction (pp. 645-649).

Yu, K. H., Beam, A. L., Kohane, I. S. (2018). Artificial intelligence in healthcare. Nature biomedical engineering, 2(10), 719-731.

Zadgaonkar, A. V., Agrawal, A. J. (2021). An overview of information extraction techniques for legal document analysis and processing. International Journal of Electrical & Computer Engineering (2088-8708), 11(6).

Zeiler, M. D., Fergus, R. (2014). Visualizing and understanding convolutional networks. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13 (pp. 818-833). Springer International Publishing.

Zhong, H., Guo, Z., Tu, C., Xiao, C., Liu, Z., Sun, M. (2018). Legal judgment prediction via topological learning. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 3540-3549).