Estimating Case Outcomes from a Dataset Created Using Turkish Constitutional Court Documents

Salih Talha Akgün
Computer Engineering
Istanbul University - Cerrahpasa
Istanbul, Turkey
akgunsalihtalha@gmail.com

Serdar Raşit Fırlayiş
Computer Engineering
Istanbul University - Cerrahpasa
Istanbul, Turkey
serdarfirlayis@gmail.com

***Abstract*—Legal processes take an important place in the lives of individuals and institutions. With the widespread use of computers, the transfer of texts obtained from legal processes to digital media is also accelerating. This allows us to obtain useful results with natural language processing methods on legal texts. In many languages, studies such as text classification, text generation, question answering, and finding similar texts are carried out. The positive effects of these studies on the legal systems of countries will gain more importance in the future. However, NLP studies with Turkish Law texts are not sufficient. In this article, the Turkish Legal Documents Corpus (TLDC) was created from individual application cases of the Constitutional Court of the Republic of Turkey to contribute to the studies in the field of NLP in Turkish Law. Visualizations (WordCloud, Correlation) were made to understand the data set better. To estimate the case outcomes, binary classification was made. Classes were "violation" and "no violation" of the case outcomes. TFIDF is used for vectorizing the text. After vectorization, Logistic Regression and Naive Bayes models are trained. In addition, the neural network was trained by performing feature extraction with LABSE and Glove models.**

***Index Terms*—Turkish Law Dataset, Natural Language Processing, NLP in Law, Machine Learning, Deep Learning**

I. Introduction

Artificial intelligence uses data to mimic human thought processes, decision-making, and problem-solving abilities. Natural Language Processing (NLP) is one of the fields of artificial intelligence. NLP aims to understand natural languages by computers, produce text from these languages, summarize existing texts, classifying texts and perform various language-related tasks.

Today, the amount of text in the digital environment is gradually increasing due to the increase in the use of social media, the public sharing of the data of the institutions, and the fact that many of our daily works are carried out on the computer environment. Thanks to the creation of many text data in areas such as Politics, Economics, and Health, the impact of research and studies on NLP is expanding. In line with the needs in these areas, field-specific NLP tasks are also being developed. Artificial intelligence models are developed based on data produced in various fields. The field of law, where text data is used and produced extensively, is also an area of increasing importance for NLP.

Artificial intelligence in law is changing the legal field in various ways. The delay of cases and the prolongation of the judgment duration of the results due to various reasons such as lack of judges, lack of technology use in legal services, and legal gaps is an ongoing problem in the legal systems of many countries.

Artificial intelligence systems that can be created using legal texts can reduce the workload in courts. It can enable defendants and plaintiffs to anticipate judgment outcomes. It can help institutions to make their legal expenses predictable and reduce their legal process expenses. In this way, individuals and institutions can reduce legal costs and save time.

In the field of law, specialized NLP models in different languages have been developed and legal data sets have been published. As a result of the studies, systems such as estimating the outcome of the case, answering questions about legal texts, and detecting similar cases have been developed.

In this article, TLDC (Turkish Legal Documents Corpus) was created from $\sim$10k individual application cases published by the Constitutional Court of the Republic of Turkey. Correlation of case results visualized in TLDC. WordCloud was created from TLDC. Binary classification was made on the results of the case as "violation", and "no violation" with Logistic Regression, Naive Bayes, LABSE[1], and Glove[2] methods.

II. Related Works

The texts in the field of law cannot be used directly for NLP studies. For the data to be used in model training, certain processes are required. Some judicial authorities share their legal texts publicly. In this way, publicly available data for NLP studies are obtained. In this section, there is a literature review of NLP studies in the field of Law.

Research made by Malik et al. introduces the ILDC (Indian Legal Documents Corpus) dataset. Court Judgment Prediction and Explanation (CJPE) task is explained in the paper. This task aims to automatically predict an explainable outcome of a case[3].

Research made by Mumcuoğlu et al. uses DT (Decision Tree), RF (Random Forest), SVM (Support Vector Machine), and other algorithms to predict Turkish Courts outcomes. This article is one of the first comprehensive NLP studies

with Turkish Law texts[4]. Sert et al. worked on Turkish Constitutional Courts to predict outcomes as ”violation” or ”no violation”[5]. But both of them did not release the dataset as raw text.

Chalkidis et al. worked with European Court of Human Rights(ECHR) to release a new English legal judgment prediction dataset. They also made binary violation classification, multi-label classification, and case importance prediction models using BERT model[6].

Vold et al. worked with transformer-based models to answer retrieval for legal questions[7].

We have created a dataset called TLDC and used statistical models and transformer-based models to make predictions on TLDC. We are going to publish TLDC for further work that will be done on this topic after we have anonymized TLDC dataset.

III. Methodology

We created the TLDC dataset by collecting individual applications published by the Republic of Turkey Constitutional Court. As a result of the creation of the TLDC data set, we aim to predict the outcome by using the relevant parts of the case.

A. Creation of TLDC Dataset

The website of the constitutional court was examined to create the data set. As a result of the examination, HTML data were collected by making GET requests to get the website's content. Afterwards, this data was parsed using the BeautifulSoup library in Python, and the sections related to the case text and the case result is separated using the HTML tags. The results of the lawsuit are grouped as a violation and no violation. During grouping, results such as Explicit Lack of Basis were also considered as no violations. The text of the case is divided into titles. Regex Strings are used for separation. Segmentation processes of the case text into titles are shown in Figure 1 as a pipeline. As the result, we split our documents by their titles. These titles are given in Table 1.

TABLE I
Translation of Titles of the TLDC

Turkish	English
Link	Link
Başvuru Konusu	Subject of Application
Başvuru Süreci	Application Process
Olay ve Olgular	Event and Facts
İlgili Hukuk	Related Law
İnceleme ve Gerekçe	Review and Rationale
Hüküm	Judgment
Hak	Right
Müdahale İddiası	Alleged Intervention
Sonuç	Result
Giderim	Compensation
Başvuru Tarihi	Application Date
Karar Tarihi	Judgement Date

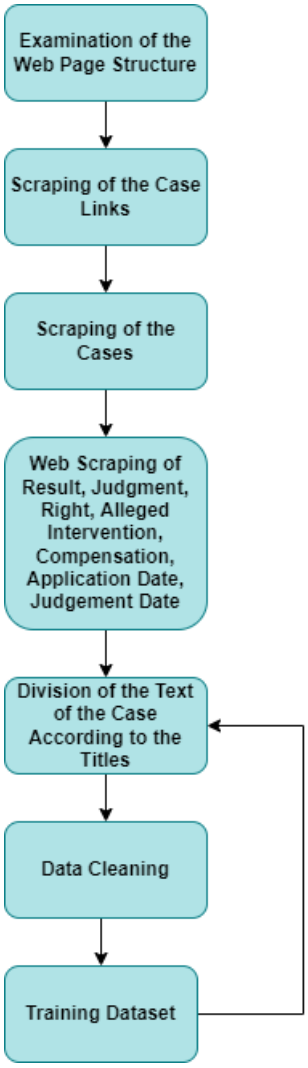

Fig. 1. Pipeline of Creating the Dataset

B. Visualization of the Dataset

Python was used to visualize the dataset and extract semantic results. Pandas and Seaborn libraries are used for visualization.

C. Outcome Prediction

The following models were used to find the baseline accuracy with the statistical models for the TLDC data set. TFIDF was used to vectorize the texts.

- Logistic Regression
- Naive Bayes

Also, the following models are used for feature extraction and a neural network training is performed on them, and the results are observed.

- LABSE
- Glove

SparkNLP library was used for training these models. This library enables rapid creation and training of NLP pipelines.

IV. Measurements

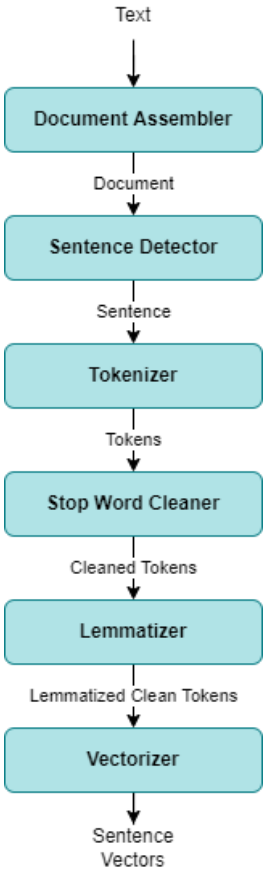

Fig. 2. Preprocessing Pipeline

Individual applications made to the Constitutional Court of the Republic of Turkey are collected from the internet. The raw text of the TLDC formed after these processes. Finally, raw text is separated according to the headers related to the application.

After the separations were done, the first words of all columns of each data were visually checked to ensure that the dataset was separated as we wanted. All columns are initialized by numbering or lettering. Lines that do not comply with this were checked and the possibility of this problem occurring in other lines was evaluated. If it is not a general problem, it was fixed manually, if it is a general problem, the cause of the problem was investigated. If the cause of the problem is not found, this data is removed from the TLDC dataset.

V. Results

TLDC data set, which was created using individual applications made to the Constitutional Court of the Republic of Turkey, is used to carry out visualization and result estimation studies.

We made visualizations to better understand the TLDC dataset. These visualizations will enable us to make more consistent interpretations of the results we get while working on a model and to reach results faster throughout the model development phase. We started the visualization by extracting the distribution of the results of the data set we created. Then we examined the correlation of the results with each other. Finally, we tried to get an idea about the text of the dataset by extracting WordCloud.

TABLE II
Classification Results for Violation

	Precision	Recall	F1	Accuracy
Naive Bayes	0.65	0.64	0.65	0.67
Logistic Regression	0.74	0.74	0.74	0.75
Glove	0.77	0.64	0.70	0.71
LABSE	0.75	0.74	0.75	0.74

TABLE III
Classification Results for No Violation

	Precision	Recall	F1	Accuracy
Naive Bayes	0.68	0.69	0.68	0.67
Logistic Regression	0.76	0.76	0.76	0.75
Glove	0.67	0.79	0.73	0.71
LABSE	0.72	0.73	0.72	0.74

The word cloud done on our data set is seen in Figure 3. This figure shows that the most used words, apart from stopwords, are words related to law, as expected.

Fig. 3. Word Cloud

To examine the outcome distribution of the dataset, we looked at the number of outcomes. Some cases can be closed with more than one outcome. For these cases, we have considered each outcome separately, for visualization purposes. We then showed them in a bar graph in Figure 4.

There is no limitation for decisions can be made. Table 4 shows the translation of decisions.

The correlation of the case results with each other was calculated. Then, this correlation matrix is translated into a heat map, which is shown in Figure 5.

Conclusions can be made with this heatmap. For example, in cases where there was a decision of violation, fewer decisions were made as explicit lack of basis or no violation. These results do not contradict what was expected.

In Figure 6, the duration of the cases according to the application years given in the graph. Based on this graph, we

TABLE IV
TRANSLATION OF DECISIONS

Turkish	English
Anayasal ve Kişisel Önemin Olmaması	Lack of Constitutional and Personal Importance
Açıkça Dayanaktan Yoksunluk	Clearly Lack of Basis
Başvuru Yollarının Tüketilmemesi	Not Exhausting Application Paths
Başvurunun Reddi	Refusal of Application
Düşme	Drop
Kişi Bakımından Yetkisizlik	Person's Incompetence
Konu Bakımından Yetkisizlik	Subject Incompetence
Mahkemenin Yetkisizliği	Incompetence of the Court
Süre Aşımı	Timeout
Yer Bakımından Yetkisizlik	Location Incompetence
Zaman Bakımından Yetkisizlik	Time Incompetence
İdari Redde İtirazın Reddi	Objection to Administrative Refusal
İhlal	Violation
İhlal Olmadığı	No Violation
İncelenmesine Yer Olmadığı	No Need for Review

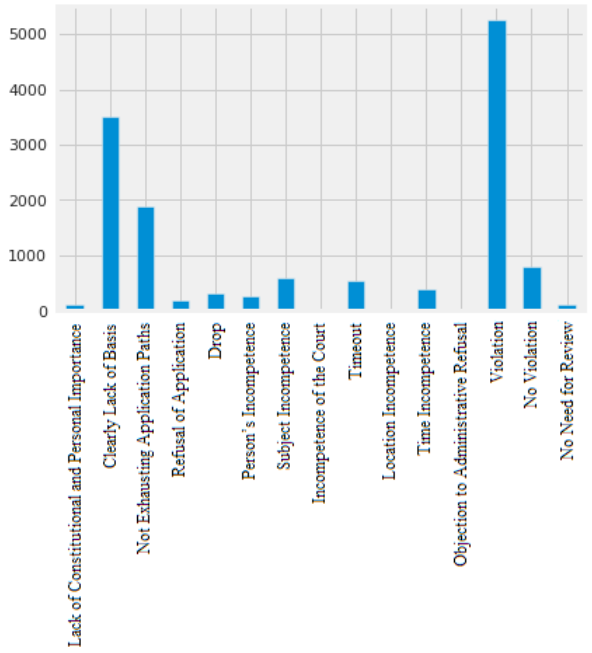

Fig. 4. Count of Results

can say that the decision time of a case is getting shorter and shorter.

The model training was made on the Google Colab platform. Python, SparkNLP, and PySpark are used in versions 3.7.13, 3.4.4, and 3.0.3, respectively. Model training is done with the SparkNLP library's default hyperparameter tuning. Also, the dataset was shuffled before converting to SparkNLP documents. Dataset split with an 80/20 ratio. This operation was made using the random split method of the DataFrame created with the SparkNLP library, and the seed is given as 100. Afterward, the data set was ready for training after the pipeline in Figure 2, which uses the SparkNLP library. Training results can be seen in Table 2 and Table 3.

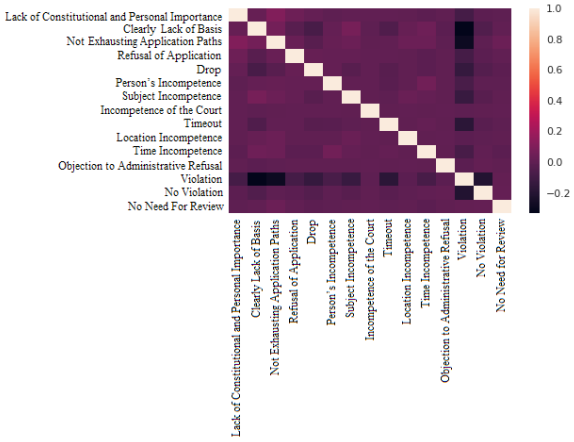

Fig. 5. Correlation of Results in the Same Cases

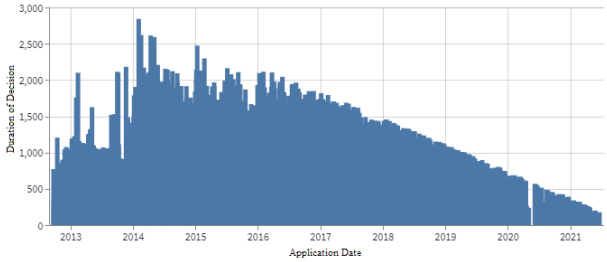

Fig. 6. Duration of Cases Related to Application Date

VI. DISCUSSION AND CONCLUSION

Our study proposes the TLDC dataset. Visualization studies were carried out on this data set, and meaningful results were obtained. SparkNLP pipelines are used to obtain base accuracy with this dataset.

Although the base results obtained using the SparkNLP library are sufficient, they are open to development. These results, which were obtained as a result of quick training in the Colab environment, also indicate that the dataset is suitable for training. Again, good results were obtained by using Logistic Regression and Naive Bayes. It shows that TLDC is a clean dataset.

Previous studies got higher accuracy with their dataset using the LSTM model. But we have used a sentence embedding model that uses BERT called LABSE. Considering the success of transformer-based models in other NLP tasks, we expected to have higher accuracy with LABSE. We believe short training time in the Colab environment is the reason for the low accuracy of LABSE. For future work, we aim to train transformer-based models for longer with better hardware. In addition, models like T5, with customized prefixes, can be trained to deliver a multi-tasking legal assistant model.

VII. ACKNOWLEDGEMENTS

We would like to thank to the Assoc. Prof. Muhammed Ali Aydın and R.A. Ebu Yusuf Güven for their help in the development process of this study under the ISTEC Lab.

REFERENCES

[1] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan and Wei Wang "Language-agnostic BERT Sentence Embedding" arXiv:2007.01852 [cs.CL]

[2] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global Vectors for Word Representation.," in EMNLP, 2014, vol. 14, pp. 1532–1543.

[3] Ashutosh Modi, Vijit Malik, Rishabh Sanjay, Shubham Nigam, Kripa Ghosh, shouvikkumarguha@nujs.edu Guha, Arnab Bhattacharya (Aug 03 2021). ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation, ACL-IJCNLP 2021, Underline Science Inc. DOI: 10.48448/dd6b-s837. Available from: https://underline.io/25670-ildc-for-cjpe-indian-legal-documents-corpus-for-court-judgment-prediction-and-explanation

[4] Emre Mumcuoğlu, Ceyhun E. Öztürk, Haldun M. Ozaktas, Aykut Koç, Natural language processing in law: Prediction of outcomes in the higher courts of Turkey, Information Processing Management, Volume 58, Issue 5, 2021, 102684, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2021.102684

[5] Sert MF, Yıldırım E, Haşlak İ. Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court. Social Science Computer Review. May 2021. doi:10.1177/08944393211010398

[6] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural Legal Judgment Prediction in English. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4317–4323, Florence, Italy. Association for Computational Linguistics. Available from: https://aclanthology.org/P19-1424/

[7] Andrew Vold and Jack G. Conrad. 2021. Using transformers to improve answer retrieval for legal questions. Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law. Association for Computing Machinery, New York, NY, USA, 245–249. https://doi.org/10.1145/3462757.3466102