Legal Judgment Prediction

Arnaav Anand
Luddy SICE, IUB

Nischal Krupashankar
Luddy SICE, IUB

Anirudh Penmatcha
Luddy SICE, IUB

12/08/2023

project-arnanand-nbangal-anpenma

Abstract

Predicting legal judgments involves the task of forecasting the outcome of a court case based on the provided case information. Existing research in this field has predominantly focused on applying neural models to Chinese legal texts, while English language counterparts have primarily relied on feature-based models like bags of words and topic modeling. There have also been recent developments in neural models for English data sets as well. With our data set, we aim to produce legal judgments with a rather simple neural network and a bag of words model and assess the outcomes.

Our primary aim is to evaluate the performance of neural models and word-based models on the data set, incorporating a two-fold maneuver. The first aspect involves a binary classification on the training set, determining whether a case comprises a violation or not. The second aspect focuses on predicting the case outcome within the legal context and delivering a verdict on test cases. We then compare and contrast the various methods applied on each data set we work with.

Keywords

judgment, prediction, neural network, binary classification, verdict

1 Introduction

In this study, we intend to explore and evaluate the efficacy of distinct approaches to mining legal texts by leveraging a diverse and carefully curated data set. Our approach involves a dual-pronged strategy: first, we aim to perform binary classification on the training data set to discern cases that involve legal violations from those that do not. Subsequently, we will delve into predicting the ultimate case outcome within the legal parameters and offer a verdict on a separate set of test cases. This research seeks to provide a comprehensive comparative analysis of the performance of neural models and traditional word-based models on our data set, shedding light on their strengths and limitations. Ultimately, the insights gained from this study have the potential to inform and improve the legal decision-making process, benefiting both legal practitioners and the broader legal system.

As for the data set, we will try and implement our models with a multitude of data sets if our initial source proves liable. While there are many sources on Supreme Court rulings, there

is a significant dearth of those that can be leveraged by NLP models or any word-based models. The availability of well-annotated data sets related to legal documents from the Supreme Court of the United States (SCOTUS) for public use is currently quite limited. To start with, we use a simple SCOTUS data set from Kaggle. There are a total of 3,304 cases retrieved from SCOTUS, spanning the years 1955 to 2021. ¡about echr and ildc¿

Previous work

There have been many attempts at extracting legal information from public domain sources, but only a few done on English data sets with a complete implementation of legal prediction. Alteras et al.[1] in 2016 leveraged a data set from the European Court of Human Rights (ECHR) to produce an empirical analysis using N-grams, resulting in a 79% average accuracy and highlighting the facts of the case as the most crucial element in this context. In 2019, Chalkidis et al.[2] worked on several more data points from the same ECHR data set and implemented a neural network that considerably outperformed previous bag-of-words models. Feng et al.[3] surveyed a state-of-the-art, highlighting the work done on the topic and shedding light on several considerations in the NLP implementation. In 2018, Xiao et al.[6] worked on a Chinese data set with over 2.6 million data points using both TF-IDF + SVM and Convolutional Neural Network (CNN). Medvedeva et al.[5] also attempted a neural network model in the same domain which resulted in a low accuracy of 75% but showed that predicting decisions for future cases based on the cases from the past negatively impacts performance. We make our foray into implementing the various approaches on the Kaggle data set, which has been utilized by Kachroo et al.[4].

2 Methods

2.1 Data Sets

We have chosen a variety of data sets for our models to work on. The 3 data sets originate from different continents to account for differences in legal procedures as well as how the facts are presented. Each one has a different prediction task based on the classification label.

The first data set is extracted from the Supreme Court of the United States (SCOTUS), containing 3304 cases from 1955 to 2021. It contains paragraphic facts about the case along with whether the first party i.e. the defendant won the case or not. The prediction task with the testing data here would be to determine whether the defendant would win based on the given facts. Figure 1 shows the most common words occurring as both unigrams and bigrams.

The second data set is obtained from the European Court of Human Rights (ECHR), containing 11478 JSON files of cases. The JSON files had to be converted into a pandas dataframe for pre-processing. Each case contained a list of violated articles/paragraphs, which we converted into a binary column with 1 representing that at least one article/paragraph was violated and 0 representing no violations (arising from an empty list). Figure 2 shows the most common words occurring as both unigrams and bigrams.

The third data set is procured from legal proceedings heard in various court echelons in the Indian Legal Data Corpus (ILDC). The data set contains 34,816 annotated cases split into train, test, and validation sets and the target variable here is the same binary label indicating whether at least one article was violated following the verdict. Figure 3 shows the most common words occurring as both unigrams and bigrams.

Figure 2 shows how the words in each data set are distributed, and provides an overview of frequently occurring words in the facts of the cases. It serves as a good representation of how diverse the data sets are as well.

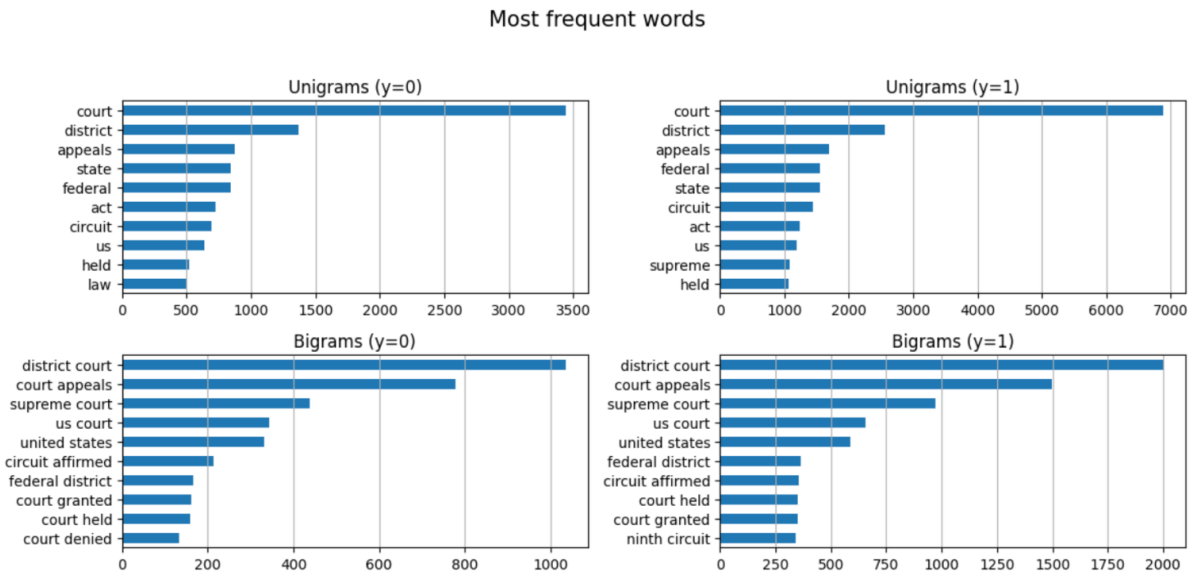

Figure 1: Word distributions in the SCOTUS data set.

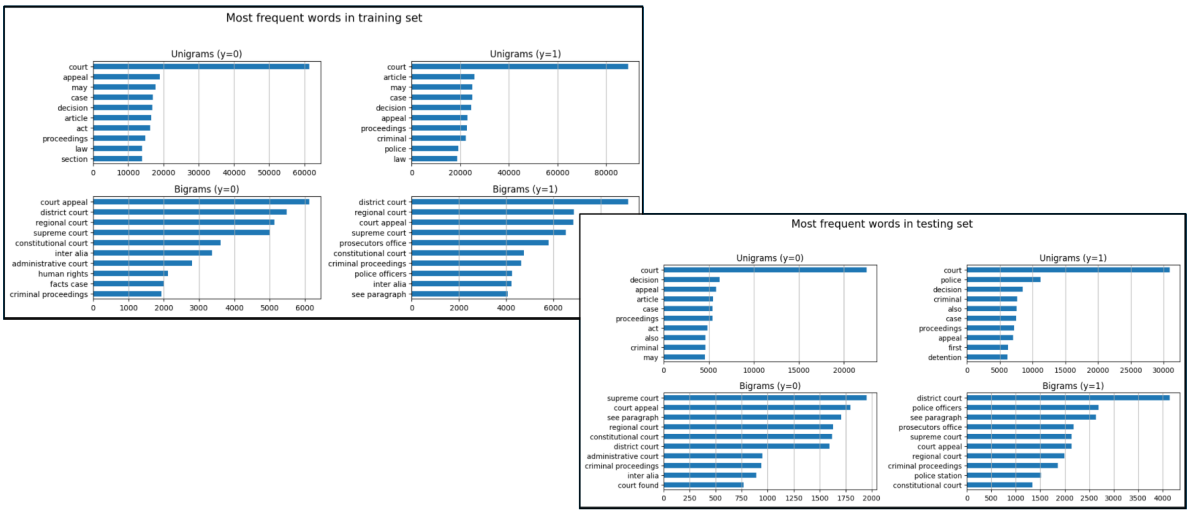

Figure 2: Word distributions in the ECHR data set.

2.2 Data Pre-processing

The data sets we implemented were extracted from the source to be presented in NLP tasks, simplifying our data pre-processing phase. However, the data still needed to be tokenized and cleaned for further analysis. Depending on our implementation, we needed to perform crucial steps before training our model, such as removing punctuations, special characters, stop words, etc.

Since our research aimed to determine the effectiveness of diverse computational techniques in the realm of legal judgment prediction, we employed many distinct approaches. To begin with, we planned to start with a few such approaches - Naive Bayes Classifier, k-nearest neighbors classifier, and a simple neural network. Each approach served a specific purpose in the context of our study.

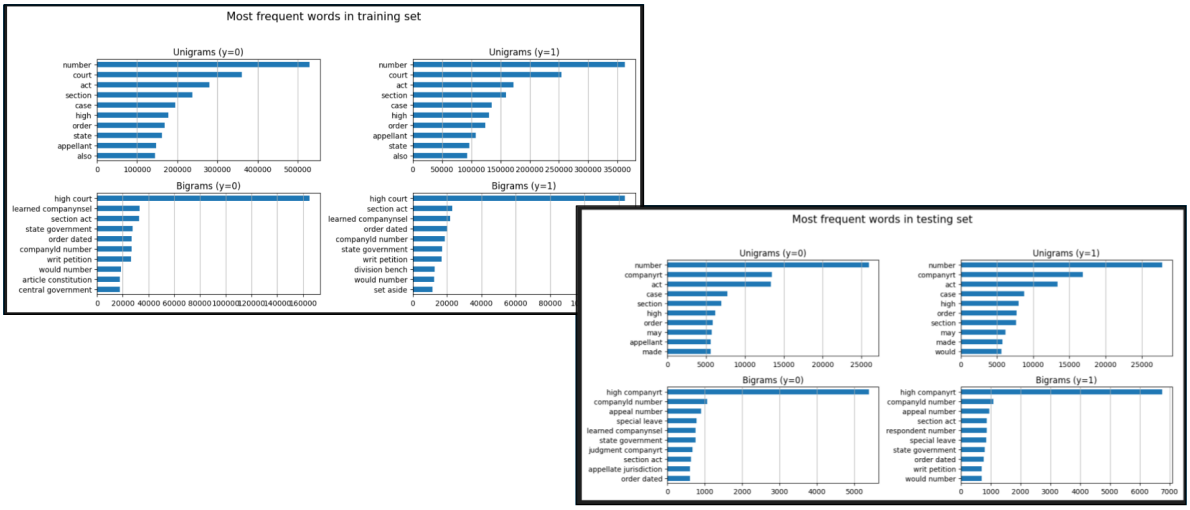

Figure 3: Word distributions in the ILDC data set.

Figure 4: Word Clouds for SCOTUS, ECHR, and ILDC respectively.

2.3 Naive Bayes Classifier

Naive Bayes is a probabilistic classification algorithm well-suited for text-based tasks. We utilized Naive Bayes for binary classification on the training dataset, distinguishing cases involving legal violations from those that did not. Its foundation on Bayes' theorem, incorporating the independence assumption, made it a valuable tool for textual analysis.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1}$$

where:

$P(A|B)$ is the probability of event A given that event B has occurred,
$P(B|A)$ is the probability of event B given that event A has occurred,
$P(A)$ is the probability of event A,
$P(B)$ is the probability of event B.

The vectorized data was fed to the Bayes classifier model, but we also compared the model's efficacy in handling unigrams and bigrams. Since the data set was not very massive, this computationally heavy task was achievable.

2.4 k-Nearest Neighbors Classifier

The k-nearest neighbors (kNN) classifier is a non-parametric, instance-based learning algorithm. In our study, kNN was employed to assess the proximity of cases in the feature space, determining legal judgment outcomes based on the similarity of their factual contexts. Both Naive Bayes and kNN used training data that had been processed through a vectorizer for document matching.

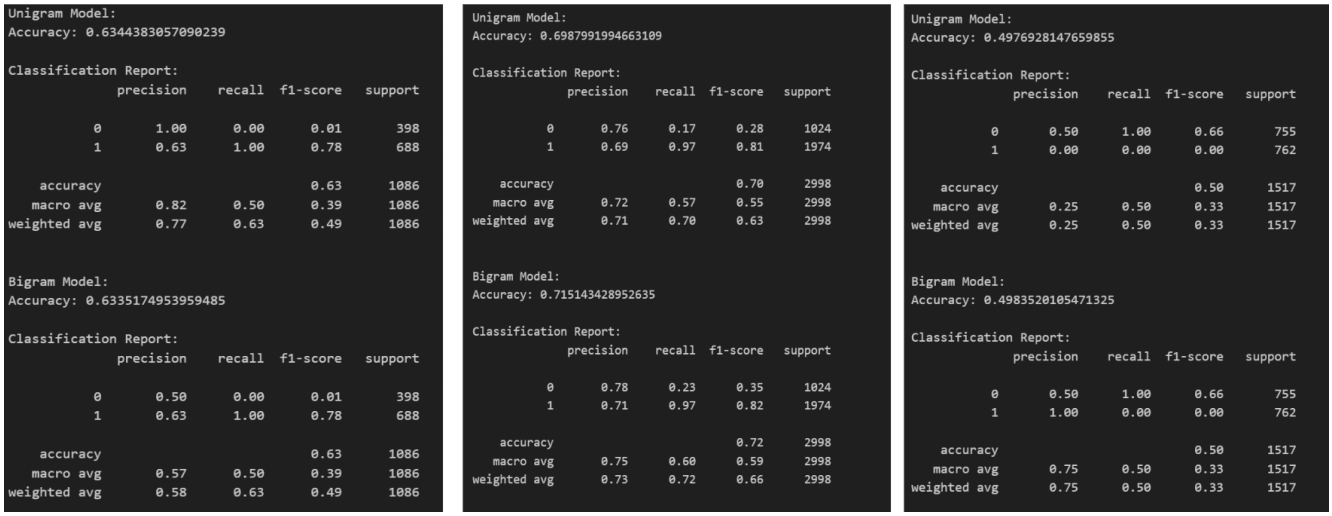

Figure 5: Naive Bayes Classifier Output on the three datasets

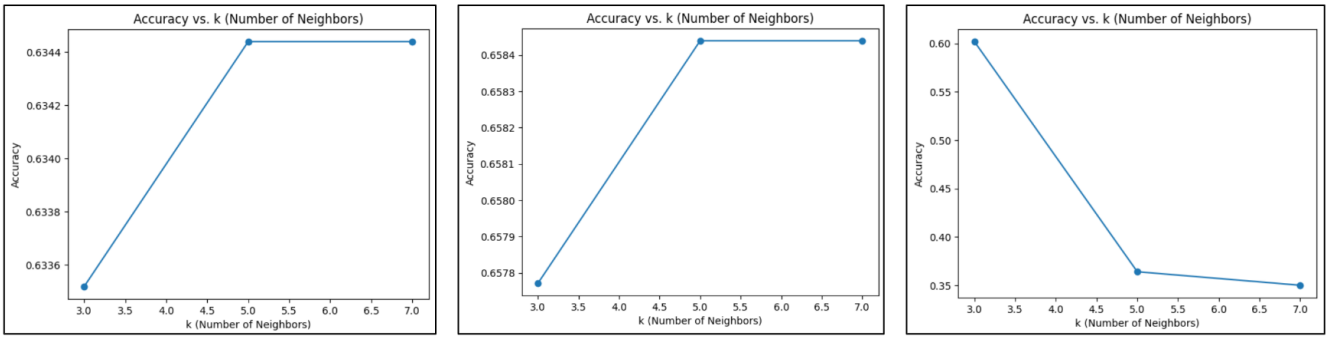

Figure 6: k - Nearest Neighbours Output on the three datasets

2.5 Artificial Neural Network

TF-IDF vectorizing and universal encoding served as crucial components for near-duplicate detection within our text dataset. This technique helped identify cases with similar factual contexts, contributing to the exploration of patterns and influences on court decisions. In the same pipeline, a simple neural network model underwent training on the vectorized dataset to classify and predict case outcomes.

We endeavored to try these three different models and find one that worked the best for each dataset, contrasting primitive and complex models.

3 Results

After performing Naive Bayes Classifier on the SCOTUS dataset, we were able to achieve 63.44% accuracy using Unigram. With the ECHR dataset, we had got a score of 71.51% using Bigram. And lastly with ILDC it was a score of 49.83% with Bigram. Figure 5 shows the program's output.

Next, we had implemented k - Nearest Neighbours starting with SCOTUS and k set to 5 it gave an accuracy of 63.44%. ECHR using 5 neighbours gave an output accuracy of 65.84%. Finally, ILDC with 3 neighbours has an accuracy score of 60.00%. Figure 6 shows the above discussed outputs.

Finally, with the Artificial Neural Networks we have an accuracy of 58% on SCOTUS, 63% on ECHR and 59.31% on ILDC. Figure 7 shows a snippet of the outputs we obtained.

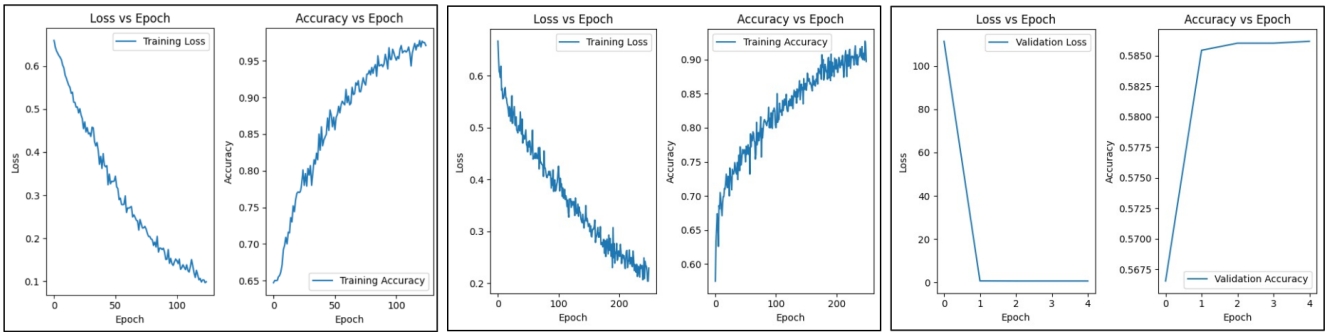

Figure 7: Artificial Neural Network Outputs on the three datasets

4 Discussion

(One line about the results, just summarizing our findings)

Moreover, striking the right balance in hyperparameter adjustments of the neural model training can significantly enhance the ANN's predictive capabilities and overall effectiveness in legal judgment prediction. This was unfortunately hampered in the scope of our project due to space allocation constraints across our virtual environments.

Furthermore, we are working on implementing a BERT uncased classifier model to further carry out the task of legal judgment prediction, and expect it to outperform the other models significantly. This is owing to its ability to infer more context out of the factual statements of each case.

Legal judgment prediction has come a long way, starting from primitive model implementations to more complex neural and BERT models. The availability of more diverse legal data, along with the resources capable of harnessing as much of the data as possible, are crucial in paving the way for more comprehensive and effective legal predictive models to fruition.

5 Author Contribution

The task of finding the data sets and literature survey forming the fundamental basis of our project was collectively done by the three of us. The work was divided based on the model, with all 3 cleaned data sets being trained on the same model simultaneously. Arnaav took care of the Naive Bayes classifier, while Anirudh handled the kNN classifier, and Nischal implemented the ANN model. Furthermore, Arnaav carried out the pre-processing steps and exploratory data analysis.

References

[1] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro, and Vasileios Lampos. Predicting judicial decisions of the european court of human rights: a natural language processing perspective, 10 2016.

[2] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. Neural legal judgment prediction in english, 2019.

[3] Yi Feng, Chuanyi Li, and Vincent Ng. Legal judgment prediction: A survey of the state of the art. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5461–5469. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Survey Track.

[4] Raghav Kachroo and Ayush Kashyap. Supreme court judgement prediction, 2021.

[5] Masha Medvedeva, Michel Vols, and Martijn Wieling. Using machine learning to predict decisions of the european court of human rights. *Artificial Intelligence and Law*, 28(2):237–266, 2020.

[6] Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, and Jianfeng Xu. Cail2018: A large-scale legal dataset for judgment prediction, 2018.