Predicting Legal Outcomes from Case Summaries: A Data-Driven Approach

G Meher Pranav Narayan
Department of Computer Science & Engineering
Amrita School Of Computing,Bengaluru
Amrita Vishwa Vidyapeetham, India
BL.EN.U4CSE22223
bl.en.u4cse22223@bl.students.amrita.edu

Saurabh Kiran Mahuli
Department of Computer Science & Engineering
Amrita School Of Computing,Bengaluru
Amrita Vishwa Vidyapeetham, India
BL.EN.U4CSE22253
bl.en.u4cse22253@bl.students.amrita.edu

Ojas
Department of Computer Science & Engineering
Amrita School Of Computing,Bengaluru
Amrita Vishwa Vidyapeetham, India
BL.EN.U4CSE22281
bl.en.u4cse22281@bl.students.amrita.edu

***Abstract*—The project undertaken uses various Natural Language Processing techniques and multiple different strategies to convert text into numbers and aims to establish a relation between the closing argument and the result of the case. Machine Learning models are chosen to predict the outcome of case, providing the insights into which elements of closing arguments are most influential. The findings offer can a great judgment predictions by providing a data-driven understanding in the hall of justice.**

I. Introduction

The project applies Natural Language Processing techniques to the legal text to change it into numerical representations. It, therefore, is a kind of project which would apply different ways of changing text into numbers in an effort to come up with a clear relationship relating the content of closing arguments with the results of legal cases. In these regards, the analysis will employ advanced NLP methods—word embeddings, vectorization—to underscore the latent semantics of arguments. On that account, this project quantitatively links these textual features to case results in a bid to uncover patterns and insights that could turn out to predict verdicts on the linguistic characteristics of closing statements. The new approach will be applied in case outcome prediction, which will turn out to be of great importance to the field of legal analytics.

II. Literature Survey

Varshini et al. [4,1] have compared the legal outcome prediction using language models and further evaluated their effectiveness of prediction on BERT, LegalBERT, DistilBERT, and RoBERTa. Comparative effectiveness of language model on legal outcome prediction using case file text. Additional legal features could form future research scope for improved prediction accuracy. S et al. [4,2] have constructed an accurate legal case outcome prediction model which comes under the domain of AI to build a legal model. The uniqueness is that it states the lack of explainability usually reflected in legal prediction models by pointing out the dissimilarity between models and human judges. The application of this model is limited only to an English version, large data sets can be trained on the model, and its application is therefore limited to small jurisdictions. In the future scope of the legal outcome prediction model, different choices for encoding can be looked. Almuslim and Inkpen [3] discusses Natural Language Processing and Artificial Intelligence for judicial outcome predictions of Canadian appeal cases. Experiments show that very high accuracy in appeal court outcome predictions can be achieved, above all by methods of Deep Learning, thereby performing comparably to previous studies in other languages and legal systems. Liu et al. [5,4] describes how AI has advanced and its use in the judiciary sector. An important task to advance the accuracy in legal verdict prediction is the analysis of factual descriptions of real cases for text feature extraction. While many current methods are based on deep learning techniques, further increase in the accuracy rate is still required. Raghupathi et al. [3,5] seeks to apply big data analytics in the analysis of pharmaceutical patent validity cases and is intended to provide insights for stakeholders in making informed decisions by reducing costs associated with litigation. Kwatra and Gupta [6] had considered the fact that the amount of data in the internet is ever-increasing, hence too vast for the users to manually summarize it. The authors then introduces a solution called automatic text summarization, which falls into two key categories. Extractive summarization is where important sentences are extracted directly from the original document. The actual data remains the same. Abstractive Summarization is where the rewords and expresses the extracted sentences in a more comprehensible semantic form, hence addressing the limitation of the extractive methods. Various abstractive text summarization techniques were discussed, and the need is focused on to evaluate those techniques with a parametric approach. Lam et al. [4,7] calls for the use of intelligent agent technologies as a practical solution to support refugees with legal aid. The authors implemented a digital legal assistant that uses AI, NLP and a reasonably accurate database in order to provide legal information to refugees through mobile devices. This chatbot interface communicates by means of a secure link and taps the database containing the legal documentation. The authors have implemented a Learning model to enhance the performance of the chatbot through Deep Q-Network to better its understanding regarding the user intention and responses. Rebolledo-Mendez et al. [3,8] discusses the necessity of data mining for the retrieval of useful insights from massive datasets. The work points out the design of the challenges facing effective data mining, dealing with the volume, variety, and velocity of data calls for advanced techniques and tools. Authors participates in this ongoing discussion with new findings and methodologies that can help in the exercise of better data mining practice to deal with the challenges at hand. Modi and Oza [9] presents a study on automated text summarization of web pages that highlights the challenge of inefficiency users face in locating information spread across multiple web pages. Authors explain the differences between extractive and abstractive summarization and focuses on abstractive summarization to produce summaries similar to humans. This study proposes a deep neural network model that further polishes the sequence-to-sequence approach by finding topic-related keywords that could potentially enhance the quality of the summary. Authors points out the issue with codec attention in pre-trained models as a step towards improving the summarization process. Liaqat et al. [4,10] has identified the need for better reduction techniques that must incorporate abstractive methods. The dataset was sourced from news websites however, it had extensive preprocessing for the removal of non-Hindi words and noise that would impact the integrity of the text content under analysis. TF-IDF is used to quantify the relevance of words, while cosine similarity is used to gauge the similarity between sentences so that semantically related sentences shall be effectively clustered. A modified PageRank algorithm shall be applied in selecting the sentences for summarization, so as to award importance to such sentences within the clusters [10].

III. Methodology

To solve question [1] we have used a module called the linear regression from the sklearn library and trained our training dataset by linear regression and also ensured the predicted values of linear regression. In question [2] we have calculated the Mean Square Error (MSE) , Root Mean Square Error (RMSE) , Mean Absolute Percentage Error (MAPE) and R2 score , we have used the inbulit functions from the sklean.metrics to calculate the above error measures. This error calculation was done only on one single feature vector and output was tested. In [3] a similar exercise was repeated but with large number of attributes and error calculations were noted. In [4] the k-means clustering on the data was performed and the number of clusters were fixed at 2 then further the model is fit into the training data. There are labels created which indicate a point belonging to a respective cluster and the cluster centers are also mentioned. In [5] we used the inbuilt modules or functions of sklearn called Silhouette Score , Calinski harabasz Score and Devies Bouldin Score and fed the parameters as training data and the cluster labels. In [6] we have performed k-means clustering for different k values ranging from 2 to 20 and calculated the Silhouette Score, Calinski Harabasz Score and Devies Bouldin Score all the k values and stored them in an empty list and printed them in an ordered manner. In question [7] using matplotlib we have plotted a line plot for different k values and noted the distorsions on increase with number of clusters. Observed the optimal k value from the elbow plot.

IV. Results

The linear regression method

Following a single feature's training, the model displayed the following metrics: MPE: 8.82451e+14 R2 RMSE: 0.80 MSE: 0.64 Points: 0.4734 These measurements showed that, with some degree of variation from the real values, the model was able to forecast the target variable with a moderate degree of success. Multivariate Regression: Including more attributes slightly enhanced the model's performance but left it with potential for improvement. Classifier using K-Nearest Neighbors (KNN):

With an accuracy of just 22%, the KNN model was far from perfect. Because of the dataset's high dimensionality or the characteristics of its features, it is possible that KNN is not the optimal algorithm for it, as indicated by the poor accuracy. Conclusions: Given the results, we concluded that additional research into feature selection, dimensionality reduction, or model tweaking was required to raise the accuracy of the classification. K-Means Grouping:

Score for Silhouette: 0.4112 130.255 is the Calinski-Harabasz score. Index for Davies-Bouldin: 1.68 These findings suggested that although the clusters were not perfectly segregated, they were quite well-formed. Better clustering quality would be shown by a lower DB Index or higher scores in Silhouette and Calinski-Harabasz. Ideal Grouping (Elbow Plot): Plotting distortions for various values of k revealed that the ideal number of clusters was approximately k = 18–19. This decision represented the underlying structure of the dataset and struck a compromise between minimal distortion and processing efficiency.

With a 22% accuracy rate, the KNN classifier's performance was subpar, indicating the necessity for feature engineering or an alternative classification approach. Predictions from the linear regression model differed from the actual values, indicating a reasonable fit. The model was slightly improved by the multivariate technique, but only 47% of the variance in the target variable could be explained by the model, as indicated by the R2 score of 0.4734, which was still below a tolerable level. Moderate separation between clusters was suggested by the K-means clustering study. The elbow plot and evaluation scores, however, suggested that more clusters (k = 18–19) were required for best performance.

REFERENCES

[1] S. H. Varshini, G. Sarayu Varma, P. Prabhakar and P. B. Pati, "Comparative Analysis of Legal Outcome Prediction with Detailed and Summarized Text," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-6, doi: 10.1109/I2CT61223.2024.10544203.

[2] U. S, A. S, K. J and S. R, "Building accurate legal case outcome prediction models," 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1-6, doi: 10.1109/ICAECA56562.2023.10200651.

[3] I. Almuslim and D. Inkpen, "Legal Judgment Prediction for Canadian Appeal Cases," 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2022, pp. 163-168, doi: 10.1109/CDMA54072.2022.00032.

[4] L. Liu, D. An, Y. Wang, X. Ma and C. Jiang, "Research on Legal Judgment Prediction Based on Bert and LSTM-CNN Fusion Model," 2021 3rd World Symposium on Artificial Intelligence (WSAI), Guangzhou, China, 2021, pp. 41-45, doi: 10.1109/WSAI51899.2021.9486374.

[5] V. Raghupathi, Y. Zhou and W. Raghupathi, "Legal Decision Support: Exploring Big Data Analytics Approach to Modeling Pharma Patent Validity Cases," in IEEE Access, vol. 6, pp. 41518-41528, 2018, doi: 10.1109/ACCESS.2018.2859052.

[6] C. Kwatra and K. Gupta, "Extractive and Abstractive Summarization for Hindi Text using Hierarchical Clustering," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICSES52305.2021.9633789.

[7] J. Lam, Y. Chen, F. Zulkernine and S. Dahan, "Detection of Similar Legal Cases on Personal Injury," 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 2021, pp. 639-646, doi: 10.1109/ICDMW53433.2021.00084.

[8] J. D. Rebolledo-Mendez, F. A. Tonatiuh Gomez Briones and L. G. Gonzalez Cardona, "Legal Artificial Assistance Agent to Assist Refugees," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 5126-5128, doi: 10.1109/BigData55660.2022.10020976.

[9] S. Modi and R. Oza, "Review on Abstractive Text Summarization Techniques (ATST) for single and multi documents," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2018, pp. 1173-1176, doi: 10.1109/GUCON.2018.8674894.

[10] M. I. Liaqat, I. Hamid, Q. Nawaz and N. Shafique, "Abstractive Text Summarization Using Hybrid Technique of Summarization," 2022 14th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, 2022, pp. 141-144, doi: 10.1109/ICCSN55126.2022.9817599.