Literature Review- DSA Group 1

Introduction:

The topic or area of concern for research is to provide help to the judges the automated prediction of legal judgment cases outcomes based on the supplied data and suggest them previous relevant cases and to predict category of the legal cases using the cutting-edge technology like Natural Language Processing (NLP). The NLP provides the help to build and use the model like Bidirectional Encoder Representations from Transformers (BERT) which is domain specific language model. Utilizing the already provided data set on the Kaggle which is ruling of the Supreme Court of the United States (SCOTUS) which is in public domain knowledge. According to different research papers to predict the verdicts for the judges most of them taking of machine learning and NLP using domain specific language model (Legal-BERT) which is already trained model to generate the legal verdicts. Publications are also taking look at many methods, including neural models, manually created features, cross-task dependencies, and label semantics, highlighting systems that produced cutting edge outcomes. Researchers has built different models like LSTM, ChatGPT-2, Rouge-3, BLEU including BERT and Legal-BERT to check the performance and accuracy of the models related to the text generation problems. For model Evaluation and performance metrics researchers are proposing and comparing evaluation metrics to assess the performance of predictive models. Regarding the gaps major conflicts, for the researcher the challenging part is to strike a balance between model interpret ability and predictive accuracy, Lack of Transparency in Judicial Decision-Making reconciling model outputs with the intricate reasoning of judges poses challenges. The ability of models to generalize beyond the training data and apply to different contexts or jurisdictions is uncertain. After having look at different articles and research papers we get to points the importance of specific features and variables in improving the predictive power of the models, we explore the strengths and weaknesses of each method and highlight the most commonly employed models and we found that BERT is best in that case, also highlight the gaps in knowledge, suggest new methodologies, or propose ways to improve the accuracy and interpret-ability of legal case predictions.

Body:

To thoroughly evaluate the state of the art in LJP and offer insights into its difficulties, techniques, and future directions, several surveys and research have been carried out.

Zhu, Liu, and Sun (2022) provide a thorough summary of automatic LJP in their survey. The authors go over several LJP research subtasks, assessment measures, and systems. They look at many methods, including neural models, manually created features, cross-task dependencies, and label semantics, highlighting systems that produced cutting edge outcomes. The research also discusses difficulties in LJP, such as label imbalance, training data biases, and prison term prediction. In addition, the authors underline the importance of fairness, prejudice reduction, and ethical

considerations, and they suggest future research areas which are much related to our topic in a sense to label the data and specific features.
According to Nishchal, Mohand and Taoufiq (2022) from (IRIT), Toulouse, France, a model to accurately predict the best probable decision of a legal case from the facts is desired. The used the different modeling techniques like Legal-BERT, LSTM-GRU and their combination to evaluate the performance and they found Legal-BERT as the best for the performance.
Cui, Liu, Wang, Chen, and Huang (2021) conduct yet another survey that examines LJP from the viewpoint of datasets, metrics, models, and problems. The authors offer suggestions for improving LJP datasets and neural network models after comparing benchmark datasets and experimental findings from various techniques. They emphasize how crucial it is for LJP systems to have elements for complicated reasoning, admissibility of evidence, and task-specificity.
Tal, Matthew, Nikolaus and Christopher (2022) conduct the researcher for nine text generation models and they found that most human-consistent model tested was GPT-2 and BERT as they are per-trained models on large set of data and provide high performance and accuracy.
Another research paper is conducted by authors Mihai, Ana, and Horia in Romania where they used the Romanian based BERT model to predict the legal jury verdicts on large, specialized corpus. The authors suggest due to low resource languages and for highly specialized tasks, transformer models tend to lag more classical approaches (e.g., SVM, LSTM).
Mohammed Alsayed, Shaayan Syed, and Mohammad Alali, (2021) conducted research in USA using the same data set as we are going to use and performed predictions using the LSTM model and KNN and they suggest to focused on researching which source of Supreme Court cases would provide the most appropriate data for our purposes. In Addition, their research has gaps in terms of model performance and accuracy. By analysing the results of this study, we gained a better understanding of how the dataset could be modified to improve and train more accurate models.
So, the question is how our study is different from the already conducted research and surveys, in terms of that we are using per-trained BERT model to not only for verdict generation but also for classification of Legal cases categories and suggesting the previous verdicts related to problem.

Conclusion:

To conclude, the articles offer a thorough summary of the current state of LJP research, including its obstacles and potential future paths in terms of prediction models. The authors stress the significance of model's accuracies for predicting the verdicts and about ethical issues in LJP and suggest lines of inquiry for next studies to tackle these difficulties. The well-structured studies offer a concise and in-depth summary of the subject.

References

Masala, M., Iacob, R., Uban, A. S., H., Rebedea, T., & Popescu, M. (2021). jurBERT: A Romanian BERT Model for Legal Judgement Prediction. University Politehnica of Bucharest, University of Bucharest, BRD Groupe Societe Generale.

Alsayed, M., Syed, S.& Bodala, H. (2021). JUSTICE: A Benchmark Dataset for Supreme Court's Judgment Prediction. University of Southern California.

Golan, T., Siegelman, M., Kriegeskorte, N., & Baldassano, C. (2023, page46). Evaluating the limits of natural language models for predicting human language judgments. Zuckerman Mind Brain Behavior Institute, Columbia University, Department of Cognitive and Brain Sciences.

Prasad, N., Boughanem, M., & Dkaki, T. (2022, p 43). Effect of Hierarchical Domain-specific Language Models and Attention in the Classification of Decisions for Legal Cases. Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France.

Source: Zhu, X., Liu, Z., & Sun, M. (2020). A Survey of Automatic Legal Judgment Prediction. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020).
Cui, J., Liu, Z., Wang, S., Chen, M., & Huang, Y. (2021). A Survey on Legal Judgment Prediction. Preprint submitted to Elsevier.