Prediction Model for Legal Judgments Using Deep Learning Techniques*

Hussain Mustansir
School of Science and Engineering
Habib University
Karachi, Pakistan
hm08436@st.habib.edu.pk

Muhammad Anas
School of Science and Engineering
Habib University
Karachi, Pakistan
ma08458@st.habib.edu.pk

Muhammad Ansab Chaudary
School of Science and Engineering
Habib University
Karachi, Pakistan
mc08077@st.habib.edu.pk

Abdul Samad
School of Science and Engineering
Habib University
Karachi, Pakistan
abdul.samad@sse.habib.edu.pk

Sandesh Kumar
School of Science and Engineering
Habib University
Karachi, Pakistan
sandesh.kumar@sse.habib.edu.pk

***Abstract*—Predicting outcomes of legal cases involves predicting potential outcome of a criminal case in legal terms, fines and punishments based on the given scenario and available witnesses. This study presents a predictive model that uses deep learning based on Pakistani Criminal Law Suits to predict potential fines, charges and legal sections based on a crime scenario. It is trained on transfomer models involving BART, T5 Base, Distil GPT-2, Facebook-opt-125m. The dataset used consists of past judgments obtained from websites of Supreme Courts and High Courts of Pakistan, which were cleaned and preprocessed for training. The model has the potential to provide a practical aid for legal practitioners, ordinary citizens and juries trying Criminal matters. Results revealed that Bart was better than other models in terms of producing a judgment similar to original judgment.**

***Index Terms*—BART, T5 Base, Distil GPT-2, Facebook-opt-125m, Deep Learning, Judgment Prediction**

I. Introduction

The legal profession often requires specialized expertise to predict case outcomes based on historical precedents. This research introduces an AI-based prediction model designed to support Pakistan's legal system by predicting judgments for criminal cases using advanced deep learning techniques. A typical judgment from a law court in a criminal trial includes details such as the criminal activity scenario, statements from prosecution witnesses, arguments presented by the counsels of both parties, the judge's remarks, and the final judgment comprising charges and punishment. Past judgments serve as essential resources for understanding crime scenarios and their corresponding legal outcomes.

With advancements in deep learning-based natural language processing (NLP) algorithms, this technology has found widespread application in various legal tasks. Notable examples include automated legal text generation and natural language-based case retrieval, both of which leverage the capabilities of deep learning, delivering impressive results. Among the most prominent applications of artificial intelligence in the legal domain is legal judgment prediction, particularly through natural language processing techniques. These tasks typically encompass subtasks such as crime identification, determination of relevant laws and regulations, and prediction of criminal sentences. By analyzing legal materials, machine learning algorithms are employed to develop models capable of making these predictions.

Our model leverages this rich data to predict judgments based on specific crime scenarios, thereby providing a crucial tool for legal practitioners. By automating this process, the model minimizes the need for exhaustive interactions with Pakistan Law Diaries and other legal resources . It can analyze case information, predict relevant charges, identify applicable laws, and recommend sentences efficiently. The model is designed to complement the efforts of legal counsels and juries, offering them actionable insights to make informed decisions.

The dataset for this research has been compiled from various credible sources, including judgments and legal documents from the **Sindh High Court** [1], **Peshawar High Court** [2], and **Lahore High Court** [3], alongside references from the **Pakistan Penal Code** [4]. To supplement the dataset, additional legal datasets were obtained from **Kaggle** [5], providing a robust foundation for training and evaluation. This diverse collection ensures that the model captures the nuances of Pakistan's legal landscape and reflects a wide spectrum of criminal cases.

The project employs state-of-the-art models, including **BART**, **T5 Base**, **Distil GPT-2**, and **Facebook-opt-125m**, for natural language processing and judgment prediction tasks .

Identify applicable funding agency here. If none, delete this.

These models were selected for their robust text generation and comprehension capabilities, enabling accurate predictions tailored to Pakistani criminal law. Additionally, the implementation of **knowledge distillation** enhances computational efficiency by transferring knowledge from more complex models to lighter ones, ensuring accessibility in resource-limited settings .

By integrating deep learning algorithms with a well-curated dataset, this system addresses challenges like delayed judgments and case backlogs, while also empowering individuals without formal legal knowledge . The model provides actionable insights and intelligent guidance, significantly contributing to the modernization and digitization of judicial processes in Pakistan .

II. Research Question and Problem Statement

The Research Question for this study is: *Can deep learning models predict legal judgments for criminal cases in Pakistan?* This question is related to an issue in legal domain where increasing number of criminal cases are being filed in Pakistan. The research explores a deep learning model that can predict a potential judgement for a criminal case based on the given scenario and available witnesses to aid common citizens, lawyers and other people in the domain of law in Pakistan. With the significant growth in Deep Learning, there is an opportunity to develop a model that can predict criminal judgments prediction based on historical data. Analysing Criminal Judgements in general, especially in Pakistan's legal system, a significant direction of judgements are based on the crime committed, evidences and statements of prosecution witnesses, therefore deep learning model will be very effective on predicting criminal judgements in domain of Pakistani law where many nuances are not involved as in the counterpart of Judgements of Family Cases.

Current models for criminal judgement predictions are limited as such they address the problem of classifying whether the prediction will be *guilty* or *not guilty* based on the given scenario. Moreover, there are very few models out there that address the problem of predicting potential judgement of Criminal Cases within the legal system of Pakistan. The problem at hand is to develop a model that can understand and analyze complex legal texts, recognize patterns in legal reasoning, and produce reliable predictions in a country with a distinct legal framework and societal context like Pakistan.

This research by training deep learning models on past criminal case judgments in Pakistan, aims is design a Transformer based model that can predict the outcomes of criminal cases, considering factors such as case facts, witnesses statements and other relevant information. The study also evaluates the feasibility, accuracy, and reliability of such models in predicting criminal judgements. The findings of this study can be used to inform legal assist legal professionals, researchers, common people and juries in streamlining the judicial process in the legal domain of Pakistan.

III. Literature Review

We explored a research paper [6] published in 2022 on the use of deep learning models to predict the outcomes of legal appeals in Brazilian federal courts, specifically within the 5th Regional Federal Court. In this study, three models were trained: ULMFiT (which uses LSTMs), BERT, and Big Bird. The input for these models was the full text of first-instance court decisions, which were processed without any preprocessing other than text splitting and removal of web scraping artifacts. The ULMFiT model used a tweaked Long Short-Term Memory (LSTM) network, while the BERT model combined a Transformer with an LSTM to handle sequence length limitations. The Big Bird model utilized a sparse attention mechanism to manage longer texts up to 7,680 tokens. The outputs of these models were binary classifications indicating whether the appellate panel would "affirm" (keep it binding) or "reverse" (amend) the lower court's decision.

The study found that all models outperformed human experts (in this case, 22 legal experts volunteered), with the ULMFiT bidirectional model achieving the highest Matthews Correlation Coefficient (MCC) of 0.3688, compared to 0.1253 from human experts.

In 2017, Chinese researchers developed a model to predict appropriate criminal charges based on textual descriptions of facts. First, the text and sequences were analyzed using the Bi-GRU model, which was utilized to extract pertinent information from the descriptions. To estimate the charges, a Support Vector Machine (SVM) model was used once the data had been cleansed and arranged. When given clean, factually correct input data, the SVM obtained a precision score of 93%. However, the precision fell to 82.12% when the full text was fed into the model without any fact extraction. The paper emphasized the importance of fact extraction and data cleansing. By contrasting the model's charge predictions with the actual rulings in the instances, the degree of precision was determined. This study suggests the use of the Bi-GRU model for effectively extracting relevant information from legal texts [7].

Another Chinese study in 2018 introduced the TOPJUDGE framework, which consisted of the computation of all the dependent subtasks i.e. fact extraction, fact classification, model initialization, model prediction, etc. This research uses CNN (Convolutional Neural Network) for fact extraction and encoding. The LSTM (Long-Short Term Memory) model is used for the prediction of charges based on the cases. This TOPJUDGE framework was able to achieve a 94-95% accuracy in predicting appropriate charges based on their factual description [8].

Exploring another research [9] on predicting prison terms for theft cases used a dataset of about 41,000 judgment documents from China Judgments Online. The study treated prison term prediction as a regression problem, meaning it aimed to find a relationship between various factors and the length of the sentence. It used both linear regression (LR) and neural network (NN) models, particularly focusing on multiple linear

regression to handle the many independent variables that can affect sentencing and final judgment. A key part of the study was using a Gated Recurrent Unit (GRU) sequence encoder for feature extraction, which achieved an accuracy of 99.45%. This level of accuracy shows that the model can identify complex patterns in legal texts that traditional methods might miss. When combining GRU with neural networks, the study achieved a mean absolute error (MAE) of just 3.2087 months, with accuracy rates of 72.54% and 90.01% for prediction errors within three and six months, respectively. These results highlight how advanced neural network models like GRU can effectively capture important details in data, making significant improvements in predicting prison sentences.

A recent Chinese paper published in 2022 [10] introduced a model called KD-BERT, which is based on BERT knowledge distillation, aiming to enhance prediction accuracy, inference speed, and reduce memory consumption. The inputs to the KD-BERT model are legal texts, including case facts, statutes, precedents, and arguments. As per the format of standard judgments, the texts are typically unstructured and require preprocessing to be suitable for the model. The preprocessing steps involve tokenization (breaking down the text into individual words or tokens), normalization (converting all text to lowercase and removing punctuation), and embedding (transforming the tokens into numerical vectors that the model can process). The refinement process includes removing stop words (common words that do not contribute to the meaning, such as "and" and "the"), stemming and lemmatization (reducing words to their base or root form, like changing "running" to "run"), and contextual embeddings (using BERT to generate embeddings that capture the context of each word within the legal text). The outputs/predictions include crime prediction (type of crime as per law), and sentencing recommendations (appropriate punishments based on statutes and court precedents). The outputs are usually expressed as scores (numerical values showing the degree of confidence in each prediction) and labels (categorical labels indicating the type of crime). The KD-BERT model outperformed the conventional BERT models in the evaluation, obtaining the highest F1-score, which indicates improved prediction precision and recall, in addition to a noticeably faster inference speed.

In this context, the literature reviewed shows a clear trend towards utilizing advanced neural networks for improving prediction accuracy. However, this study differentiates itself by focusing on criminal judgments from Pakistan's legal system, using a rich dataset and applying these models to a specific regional context. The literature also indicates that while high accuracy can be achieved with various models, the selection of the right model and dataset is crucial for practical applicability.

Study	Model	Dataset	Accuracy
[1] 2022	ULMFiT, BERT, and Big Bird.	Brazilian Courts Appeal Dataset (BrCAD-5)	Matthews Correlation Coefficient (MCC) of 0.3688
[2] 2017	Bi-GRU, SVM	China Judgements Online	Micro-F1: 90% Micro-F1: 80%
[3] 2018	TOPJUDGE framework	Criminal cases in the civil law system of China.	94-95%
[4] 2020	GRU, Bi-GRU, LSTM, and Bi-LSTM	Criminal cases from the Hong Kong judiciary.	90.01%
[5] 2022	KD-BERT	Chinese AI & Law (CAIL)	91.2%

Fig. 1. Literature Review Summary and Key Statistics

IV. Dataset

A. Collection of Data

For this study, we used a dataset of judgments between 2007 and 2024 found on Kaggle, which we refined to include only judgments related to criminal cases of various types. Moreover, our dataset also included judgments extracted from online court portals (Supreme Court and High Court of Pakistan), which were cleaned using Python libraries such as `BeautifulSoup`, `Tesseract`, and `Pandas`. The data cleaning process involved filtering irrelevant information, including omitting orders and appeals, and ensuring that only judgments with proper punishments were included. This helped ensure the dataset was more focused and aligned with the objectives of our model. Therefore, we used *Llama3.2-8b* for further cleaning and extracting key criminal case information.

To ensure the highest level of data quality, we manually reviewed and validated the final set of 1,004 judgments generated by the `Llama3.2-8b`. This review process involved checking the accuracy of the information, filtering out irrelevant details, and ensuring that all judgment records were properly formatted for model training. The final dataset was carefully curated to ensure that only relevant criminal case judgments, with clearly defined punishments and legal sections, were included for further processing.

The dataset included relevant details such as the crime committed, the judgment passed, sections under which the case was filed, and penalties imposed. Data categories covered both text data, such as case descriptions, penal code sections, witness statements, and outcomes, as well as structured data.

Our model was trained on the facts extracted from the dataset. The sample input for the model would be a scenario such as:

> "Mr. X was robbed by Mr. Y at gunpoint. Mr. Y took Mr. X's mobile phone, laptop, and cash while threatening to shoot him in case of non-cooperation. Witnesses reported seeing Mr. Y fleeing the scene in a red car, and one witness stated that they heard Mr. Y yelling at Mr. X during the robbery."

The model classified this text, extracted the relevant keywords, and predicted the appropriate charges based on the information from the dataset.

A sample input scenario and witness statement to be given as input for the model is :

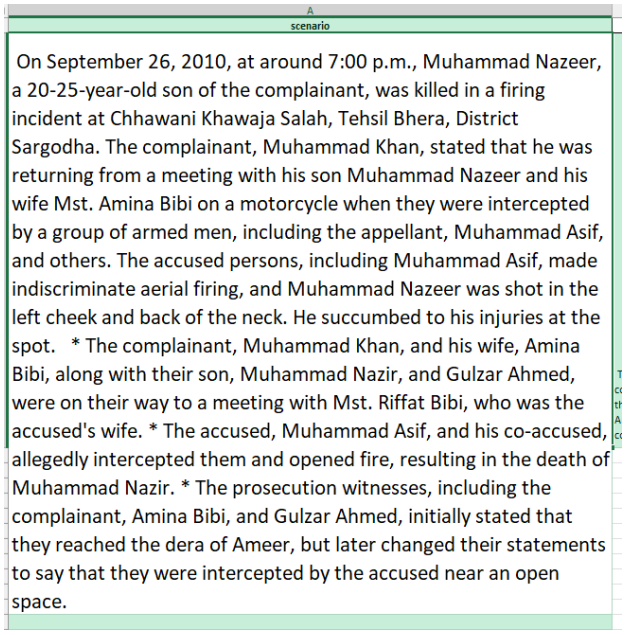

Fig. 2. A sample overview of the Crime committed

Moreover, the witnesses statements plays a cruicial part in prosecution of criminal cases in law domain of Pakistan, therefore understanding its importance we extracted the prosecution witnesses statements from the judgement in a separate column for verification of input.

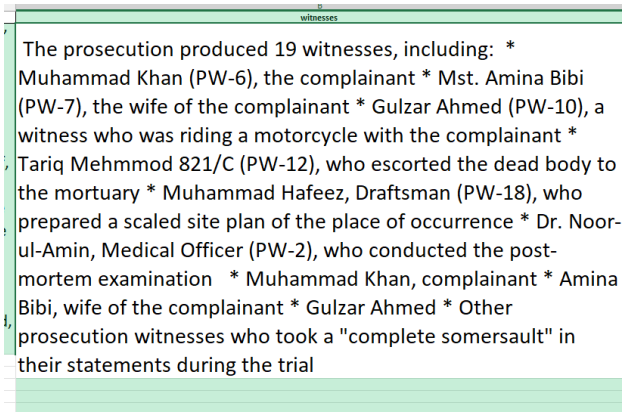

Fig. 3. A sample Witness Statement

Possible enhancements considered for the dataset included the addition of metadata, such as the judicial opinions of the judges, to further improve the accuracy of predictions. However, for the scope of this project, the dataset was primarily used as described above to ensure the model's robustness in predicting criminal charges.

B. Rationale of Dataset / online Judgements

The selection of this dataset was motivated by the availability of reliable, publicly accessible online resources in Pakistan. Given the lack of comprehensive and standardized legal datasets, especially for criminal cases, the Supreme Court and High Court of Pakistan's online portals were among the only credible sources of data.

Considering that this dataset was sourced from some of the most reliable and authoritative sources available in Pakistan, it provided a unique and invaluable resource for our study. No other publicly available dataset in Pakistan offered a comparable level of accuracy and comprehensiveness, making this dataset the optimal choice for training our criminal case prediction model.

Source	Entries (original)	Final Entries (Human Verified)	Discarded %
High Court of Sindh	300	102	29
Lahore High Court	500	183	27.3
Peshawar High Court	700	605	11.57
Supreme Court of Pakistan	200	58	34.4
Online Dataset of Supreme Court Judgements	300	55	54.5

Fig. 4. Summary of Dataset

V. Models and Experiments

A. Models Selection and Overview

Our project primarily focuses on text generation, where we aim to generate textual outputs from given text inputs. To accomplish this, we employed transformer-based models, such as BART, GPT-2, and T5-Base, which have demonstrated exceptional performance in various text generation tasks. Specifically, we used transformer models to predict judgments, a task requiring the generation of coherent and contextually relevant text based on input information. By leveraging the attention mechanism, our models could effectively capture long-range dependencies within the input text, enabling it to produce accurate and insightful judgments.

Initially, we attempted to train a Llama 3.2-1B model for our judgment prediction task. However, due to the model's substantial size and our limited computational resources, we encountered challenges related to RAM constraints and processing power. We also attempted to train models like BERT and 4bit quantized version of Llama-2-7B. However, they did not yield satisfactory results or raised computation error and we concluded they were not suitable for our specific task.

Therefore, we went ahead and trained BART, T5-Base, Distil-GPT2, and facebook-opt-125m as they were smaller in size and compatible with our computational power.

All models utilized in this analysis were Open-Source, loaded using the Hugging Face library, ensuring a standardized and efficient setup process. Performance metrics and model weights were meticulously tracked and stored using Weights & Biases (W&B), facilitating comprehensive monitoring and reproducibility of the experiments.

B. Evaluation Metric

We primarily used the ROUGE-1 score as our evaluation metric for the judgment prediction task. This metric measures the overlap of single words between the model-generated and reference judgments, effectively assessing the model's ability to generate accurate and relevant text. ROUGE-1's simplicity, computational efficiency, and interpretability make it a suitable choice for our task, especially when precise keyword and phrase matching is crucial.

While we considered BERT Score, it may not be the most appropriate metric for our specific task. BERT Score, though effective for general text generation, might penalize variations in phrasing or sentence structure that do not significantly impact the overall meaning of a judgment. Additionally, its computational cost can be prohibitive for large-scale evaluation.

Given these considerations, ROUGE-1 emerges as a more suitable and efficient metric for evaluating the quality of our generated judgments, focusing on the core content and keywords that are essential for accurate prediction.

C. Training Parameters and Model Performance

Models	Epochs	Batch Size	Roguel Score
BART	30	4	0.473
T5-Base	5	4	0.401
Distil-GPT2	33	5	0.390
Facebook opt-125m	10	3	0.315

Fig. 5. Model Parameters and Evaluation

Initial training experiments involved a range of epochs and batch sizes, resulting in varying Rogue scores. Due to computational constraints, further exploration of hyperparameter tuning was limited.

The models were trained on a dataset comprising 1004 judgments, with an 80-20 split for training and testing, respectively. Input data was preprocessed using the respective model's tokenizer, limiting token sequences to a maximum length of 512. After training, the models generated human-readable text for the judgments by decoding the token sequences. Training each model on a T4 GPU took approximately one hour.

BART and T5-Base performed the best, giving a rogue score of more than 0.400. The score compares the similarity between the original and predicted judgements. While the judgement may not have been exact same, the sentences and fines for each crime were predicted correctly.

As seen above, BART and T5-base models, with their encoder-decoder architectures and larger model sizes, outperformed DistilGPT-2 and facebook-opt-125M. These larger models, pretrained on extensive datasets, could better capture complex language patterns and generate more coherent text.

DistilGPT-2 andfacebook-opt-125M, being smaller and decoder-only models, were limited in their ability to handle intricate language nuances, especially after the distillation process.

D. Results Validation

Given the limited size of the training set (approximately 200 rows), we were able to manually verify all generated outputs. While the generated text wasn't an exact match, the models accurately predicted sentences and fines for approximately 80% of the judgments, indicating strong prediction capabilities.

VI. Results and Discussions

Following the above discussion two transformer based models generated judgements that were closer in stream to the original judgements i.e BART and T5-base models. We also observed that BART outperformed T5-base in terms of ROUGE-1 score. The best ROUGE-1 score was obtained by BART model. To fine-tune the BART and T5 models, we experimented with various hyperparameters to optimize the model's performance. The key hyperparameters we adjusted include:

Number of Epochs $(num - epochs)$: We increased the number of epochs and trained our trained model again on same configurations to fine tune it.

Number of beams $(num - beams)$: In beam search, this parameter determines how many different sequences the model considers during decoding. Increasing the number of beams enhances the variety in generated judgements. N-gram size: Adjusting the n-gram size helps minimize repetitive patterns in the judgements by discouraging the overuse of identical n-grams.

Batch size: We reduced the batch size from 16 to 4 to manage memory constraints and ensure more stable and effective training.

After fine-tuning these parameters, we retrained the mBART model. The resulting judgements demonstrated significant improvements in terms of relevance to the input.

A sample judgement generated by BART model is as follows:

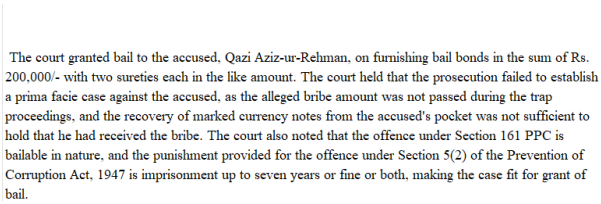

Fig. 6. : Sample Judgement Generated by BART Model

Whereof the original judgement was:

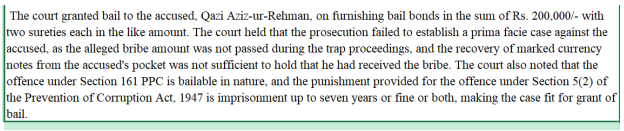

Fig. 7. : Original Judgement

However in some cases, the BART model was not accurate in some information as such it was generating information that was not present in the original judgement. This could be due to the limited size of the training data and the complexity of the legal language. The model may have learned patterns that do not accurately reflect the legal context, leading to the generation of incorrect information.

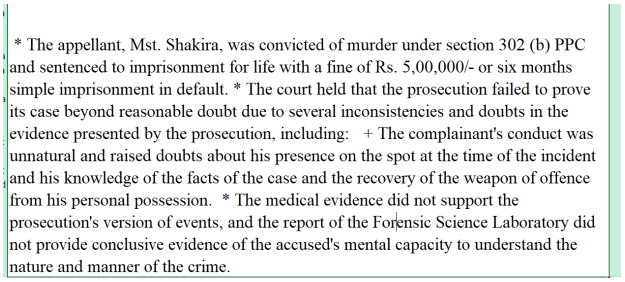

Fig. 8. : Inaccurate Judgement Generated

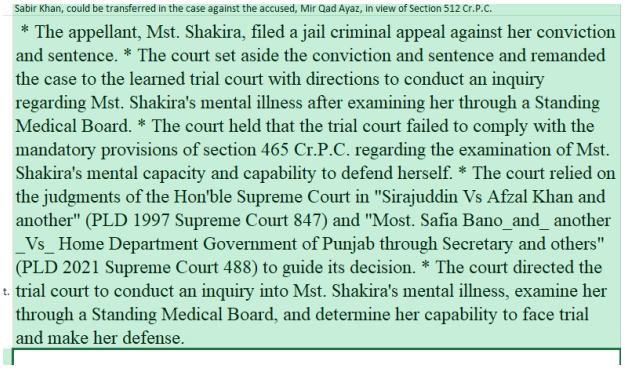

Fig. 9. : Original Judgement

However our model was for most part able to grasp correct sections of Pakistan Penal Code, as such in above example it has rightly learned section 302(b) PPC is related to murder and life prison.

Moreover, for the T5 model the judgement generated was not accurate but was able to generate the correct sections of Pakistan Penal Code.

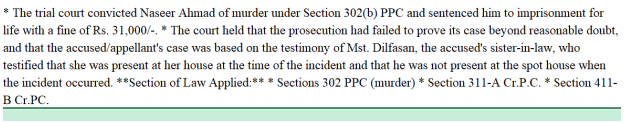

Fig. 10. : Sample Judgement Generated by T5 Model

Whereof, the original judgement was:

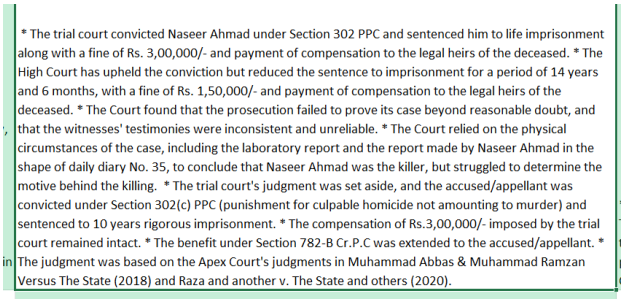

Fig. 11. : Original Judgement

Our results are were able to generate potential judgements in context of criminal law, the judgements contained correct sections. Whereof, with further increase and refinement of data the models will be able to generate more accurate judgements.

VII. Limitations and Future Research

A. Limitations

This study focused on a selection of common criminal case types for training the model. The dataset primarily consisted of commonly encountered crimes, while less common or complex cases, such as unusual crimes or other legal considerations, were underrepresented. This means the performance of the model could be less accurate when dealing with rare or atypical criminal cases. The actual legal cases that judges come across are usually exceptional, with some special circumstances or not-so-well-known laws. For the system to be a successful judicial auxiliary tool, it has to be able to help judges in a wider range of cases, including those involving not-so-common crimes and legal principles. Currently, the predictions of the model are restricted to the most frequent case types, thus limiting its generalization capability.

The dataset used in this research is a curated sample of 1,004 judgments. It is important to note that this is a small subset compared to the millions of legal documents available in Pakistan. There are millions of criminal case judgments spread across the country, and the dataset used does not fully capture the diversity and complexities found in the entire corpus. The model may eventually produce better predictions with more computing resources to analyze a larger dataset, thereby learning from broader legal precedence and nuances.

Many criminal law clauses are by their very nature ambiguous or vague, making it impossible for a machine learning model to determine with consistency and accuracy which of the clauses apply in specific circumstances. Legal language is very often imbued with nuances that call for context, subjective judgment, and interpretation based on a myriad of factors. Due to the ambiguity and imprecision of some legal provisions, the predictions of the model can be influenced. This may decrease the overall accuracy of the predictions. The problem is more serious in those cases where the law is not clearly defined and there is a challenge in giving proper recommendations.

Criminal sentencing can be highly subjective because of various emotional, ethical, and cultural factors. These are very difficult to capture with machine learning models because they do not know how to account for human emotion and personal judgment. The scope of sentencing for many offenses also varies greatly, and a machine model may not be able to provide context-sensitive recommendations. Whereas this model can predict plausible outcomes of a case, based on legal data it cannot assume the judgment or intuition provided by human judges sitting at the bench. To this date, AI is still unable to put in all those human-centred elements, which have been incorporated within any system to make decisions for sentencing.

B. Future Research

Future research should focus on improving the judgment prediction model by bringing in a broader range of criminal case types, including more rare and complex cases. Enlarging the dataset to include judgments on more diverse cases and especially those on rare or ambiguous crimes may enhance the robustness and predictive power of the model. Further insight would be provided by augmenting the information available along with the case description, such as judicial opinions, considerations involved in sentencing, and previous cases. With this enhanced simulation of the judgment process, the system could ultimately become a more practical and effective tool for supporting real judgments made at trial.

Moreover, future models may incorporate context-sensitive factors such as psychological assessments or social context that often come into play in judicial decision-making. This integration of the elements with a much larger dataset can be the stepping stone to an even more intelligent and accurate legal judgment prediction system that better mirrors the complexity of the real world. Overall, the ultimate dream is to build a system not only that will assist in legal predictions but also contribute towards the humanly greater march of justice by promoting fairness and objectivity in decision-making.

REFERENCES

[1] Sindh High Court, "Official Website of the Sindh High Court," *Sindh High Court Official Website*, https://sindhhighcourt.gov.pk/.

[2] Peshawar High Court, "Official Website of the Peshawar High Court," *Peshawar High Court Official Website*. https://peshawarhighcourt.gov.pk/.

[3] Lahore High Court, "Official Website of the Lahore High Court," *Lahore High Court Official Website*, https://lhc.gov.pk/.

[4] Pakistan Penal Code, "Official Pakistan Penal Code (1860)," *Pakistan Penal Code Official Website*, https://www.pakistani.org/pakistan/legislation/1860/actXLVof1860.html.

[5] Shahsayesha, "Supreme Court of Pakistan Judgment Dataset," https://www.kaggle.com/datasets/shahsayesha/supreme-court-of-pakistan-judgment.

[6] J. de Menezes-Neto and M. B. M. Clementino, "Using deep learning to predict outcomes of legal appeals better than human experts: A study with data from Brazilian federal courts," *PloS One*, vol. 17, no. 7, p. e0272287, 2022.

[7] B. Luo, Y. Feng, J. Xu, X. Zhang, and D. Zhao, "Learning to predict charges for criminal cases with legal basis," *Institute of Computer Science and Technology, Peking University*, China, 2017.

[8] H. Zhong, Z. Guo, C. Tu, C. Xiao, Z. Liu, and M. Sun, "Legal judgment prediction via topological learning," *Department of Computer Science and Technology, State Key Lab on Intelligent Technology and Systems, Institute for Artificial Intelligence, Tsinghua University, Beijing, China*, 2018.

[9] S. Li, H. Zhang, L. Ye, S. Su, X. Guo, H. Yu, and B. Fang, "Prison term prediction on criminal case description with deep learning," *Computers, Materials and Continua*, vol. 66, pp. 1234-1248, 2020.

[10] M. Zheng, B. Liu, L. Sun, and L. Le, "Study of deep learning-based legal judgment prediction in the Internet of Things era," *Computational Intelligence and Neuroscience*, 2022, [Retracted].

[11] Groq, "Groq LLaMA API," Available: https://groq.com/

[12] Weights and Biases, "Weights and Biases," Available: https://wandb.ai/