

Nyaay Sakha – Your Legal AI Assistant

M S Pavithra¹, Ankur Chaudhary², Chandru V³ and V N Manjunath Aradhya⁴

¹Associate Professor, Department of CSE-Cyber Security, ATME College of Engineering Mysuru, India
Email: pavithrangsjanya@gmail.com

²Assistant Professor, Department of Information Technology, Noida Institute of Engineering and Technology, Greater Noida
Email: ankurchaudhary849@gmail.com

³Student, PES University Bangalore, India
Email: chandruvanniaraj@gmail.com

⁴Department of Computer Applications JSS Science and Technology University Mysuru, India
Email: aradhya@sjce.ac.in

Abstract— One of the biggest problems facing the country's court systems is the growing backlog of cases. These inefficiencies affect attorneys as well as legal professionals. Innovative approaches that might increase productivity without sacrificing legal correctness and equity are needed to address these issues. The system extracts and processes legal materials, summarizes important concerns, and suggests relevant earlier occurrences. The artificial intelligence (AI) system uses advanced natural language processing (NLP) techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity to quickly evaluate large volumes of legal documents and identify cases with similar criteria. Legal decision-making can be made more quickly, accurately, and efficiently with the application of artificial intelligence (AI).

Index Terms—NLP, AI, TF-IDF, Summarization.

I. INTRODUCTION

The judicial system has an important function for main-taining justice, however, a growing backlog of cases creates noticeable delays in the progress of the court. Delays impact the efficiency of the judicial system and timely access to justice. With an ever growing number of legal documents, legal professionals sometimes struggle to complete the tedious steps of researching, reviewing case files, and locating an applicable precedent. This study utilizes artificial intelligence (AI) to automate key elements of case analysis in order to assist legal practitioners with their challenges. The system can successfully extract details, provide summaries of legal documents, and identify suitable past cases through the use of Natural Language Processing (NLP) techniques like TF-IDF and cosine similarity. In addition, AI facilitated prediction models can produce insight that provide information to assist legal earlier decision making, greatly reduce the amount of manual hours spent reviewing documents, and thus ease the litigation and resolution process. In making case analysis easier, the project also improves workflow efficiency, speed of decision-making, and speed of litigation resolution. Artificial intelligence (AI) is intended as a strong ancillary product for lawyers, judges and researchers - not a replacement for legal professionals. In this way, AI promises greater access to justice while mitigating court delays. The judicial system plays an essential role in maintaining justice, despite of this the growing backlog of cases has caused major delays in court processes.

Access to prompt justice is also impacted by these delays, in addition to the effectiveness of the judicial system. Legal practitioners sometimes find it difficult to complete the time-consuming tasks of researching, reviewing case files, and locating pertinent precedents as the volume of legal papers keeps increasing. By automating crucial components of case analysis, this research uses artificial intelligence (AI) to help legal practitioners overcome these obstacles. The system effectively extracts information, creates summaries of legal texts, and suggests pertinent prior cases using Natural Language Processing (NLP) techniques including TF-IDF and cosine similarity. Furthermore, AI-powered prediction models greatly reduce the amount of manual labor needed for document review by offering insights that assist legal research and decision-making. By making these important tasks simpler, the initiative increases workflow efficiency, decision-making speed, and case resolution speed. By serving as a powerful supplemental tool for lawyers, judges, and researchers rather than replacing legal specialists, artificial intelligence (AI) guarantees improved legal accessibility while cutting down on judicial delays.

II. LITERATURE REVIEW

Jujjavarapu Sujana Chowdary, Dr. Beebi Naseeba , Karaka Rupasree , Dr. Nagendra Panini, Eswara Venkata Sai Raja , Prof. Morampudi Mahathi ,AI based Attorney Using

NLTK Library proposes a solution to suggest sections and articles based on the scenario given through text or submitting a pdf file. It uses NLP techniques to preprocess the data. Majorly it uses NLTK library to do the cleaning of data and matches the extract with available law dataset and retrieves the matching articles and sections. The only achieved precision of 0.144 and 0.07 percent accuracy.

Donny Kurniawan, Siti Elda Hiererra AI Legal Companion: Enhancing Access to Justice and Legal Literacy for the Public purpose of the paper is to integrate AI in the legal sector

.It proposes a web application integrated with RAG models to provide information regarding articles and sections. They aim to make people more legally sound regarding laws and rules. The system is evaluated on the basis of user experience and accuracy. Overall the model got accuracy of 2 on the scale of 1-5 and user experience as 4 on scale of 1-5 after the feedback got from an legal practitioner. The focus is more on the user experience and accuracy can be increased by fine tuning the model.

Kishan Kanhaiya, Kamlesh Gautam, Naveen ,Pramod Singh Rathore, Arpit Kumar Sharma AI Enabled-Information Retrieval Engine (AI-IRE) in Legal Services: An Expert-Annotated NLP for Legal Judgements The paper proposes a methodology to extract key attributes from the given document which helps the law practitioners to make a decision which law article to access pertaining to the given document. The data is collected for law firm and reliable databases .The key attributes mainly focused to extract are appellant number, court name, judge name, appellant, respondents, and counsellor name for appellant and respondents. Different ML classification models were used to predict the class label . overall accuracy achieved was 91% in every model. There was a regex method if prediction was not done by model and flagging mechanism to control the outcome for repetitive classes. At the last NER was used to extract the key content from the document.

Surisetty Hima Varshini, Gottimukkala Sarayu Varma, Priyanka Prabhakar, Peeta Basa Pati Comparative Analysis of Legal Outcome Prediction with Detailed and Summarized Text The work in the paper aims to do comparative study on the Legal Outcome Prediction using pretrained models. Legal Outcome Prediction(LOP) refers to the process of predicting the outcome of a case by analysing the case file with usage of technology. This work employs four deep learning models, namely, BERT, LegalBERT, DistilBERT and RoBERTa for the LOP task on Indian case documents. Each of the case file judgements had one of the following four outcomes:(i) appeal allowed, (ii) appeal dismissed,(iii) petition dismissed, and (iv) petition allowed.Each models were built using complete data, extractive and abstractive summarized data. On the whole RoBERTa model performance was better compared to other models.It achieved F1 score of 0.47 on complete and extractive data.

Angelina George, Alphonsa Jose, Ashik F, Priyanka Prabhakar, Peeta Basa Pati, Shantipriya Parida Enhancing Legal Decision Making: WRIT Case Outcome Prediction with LegalBERT Embeddings and AdaBoost Classifier proposes an work to overcome difficulties and challenges from traditional way of judgement procedure.The paper is concerned only with the prediction of writs. It introduces technique of word embeddings with text embeddings from LLM's with statistical classifiers to effectively and precisely forecast the outcomes for writ petitions.It predicts whether petition is allowed or dismissed and partly allowed.Different classification models are used for this study. This approach delivers an accuracy of 54% and a F1-score of 0.42 with LegalBERT embedder and AdaBoost classifier.

Prameela Madambakam, Shathanaa Rajmohan : A Study on Legal Judgment Prediction using Deep Learning Techniques This paper reviews state-of-the-art Legal Judgment Prediction (LJP) algorithms published between 2018 and 2022, focusing on Deep Learning (DL) techniques applied to criminal and civil cases. A taxonomy is proposed, categorizing works into text classification for criminal cases and Legal Reading Comprehension (LRC) for civil cases, highlighting that most research addresses criminal cases due to the complexity of heterogeneous inputs in civil cases. The study emphasizes the superiority of DL models over feature-based ones in accuracy but notes their limitations in providing reasoning for predictions, which could be improved with attention mechanisms. It also identifies challenges, including the lack of standard Indian datasets, and stresses the importance of developing models capable of generating justifications for their predictions to gain trust in legal practice.

Allan Almeida, Shubh Shah, Neha Agarwal, Neer Bhavesh Khandor LEGAL AI: A Tool to help Lawyers Optimize their Performance proposes “Legal AI” which would solve critical challenges such as document analysis, translation and also extract key attributes from the document. It uses LLMs to address the enquiries based on the content of input documents. The system also ensures privacy so that any legal document can be entered without any fear of breach of data from the system. Legal AI becomes a viable solution to help the law practitioners and lawyers to optimize their process of taking decision.

XURAN WANG, XINGUANG ZHANG, VANESSA HOO, ZHOUEHANG SHAO, AND XUGUANG ZHANG LegalReasoner: A Multi-Stage Framework for Legal Judgment Prediction via Large Language Models and Knowledge Integration This paper introduces LegalReasoner, a novel multi-stage framework that leverages large language models (LLMs) and integrates domain-specific knowledge for enhanced legal judgment prediction. The approach encompasses four key stages: 1) legal knowledge infusion, where we pre-train an LLM on a vast corpus of legal literature using contrastive learning techniques; 2) case-law retrieval, employing a graph neural network to identify relevant precedents and statutes; 3) multi-hop reasoning, utilizing a transformer-based architecture with a hierarchical attention mechanism to navigate complex legal arguments; and 4) judgment synthesis. The framework demonstrates substantial improvements over state-of-the-art baselines, achieving an average accuracy increase of 7.8% across all datasets. LegalReasoner represents a significant step forward in the field of legal AI, offering a powerful and interpretable framework for legal judgment prediction.

Rungpilin Anantathanavit, Jomchai Chongthanakorn, Warisa Kongsantinart, Pisit Praiwattana, Tipajin Thapisutikul Utilizing AI and Natural Language Processing to Predict Supreme Court Decisions in Thailand proposes solution to forecast Thai supreme court decisions based on the past proceedings of the court. The methodology consists of processing legal texts, using machine learning algorithms to forecast court outcomes and extract pertaining features through NLP techniques. It uses a learning model called Wangchan Bert which is trained on the Thai laws and articles. The most accurate model will be the Wangchan BERT model, with an accuracy of 0.746. Besides the WangChanBERT, using the LogisticRegression and RandomForest with the TF-IDF vectorizer also performs high accuracy at 0.665.

Ritika Singh, Bura Vijay Kumar, Kanchan Kishor Rawat, Shaik Vaseem, Samta Kathuria LCD, Vikrant Pachouri Intervention of AI in Prediction of Judgement of Cases The paper aims to focus on the application of AI techniques in the prediction of legal judgement. The study further discusses ML algorithms and their functions in the process of court proceedings, the legal ecosystem is increasingly taken over by AI-based technology. The methodology proposed is to collect judicial data preprocess them and do feature selection and apply ML model to predict the court judgement. The CNN+LSTM model is referred that has the highest accuracy potential and has helped the judiciary to dispose of matters more effectively in comparison to human based work, the techno-based programming has made evidence collection easier, with no privacy concern issues, and is a blessing in disguise for the legal world, its further contribution has been stated in the paper.

III. PROPOSED SYSTEM

An AI-based solution is proposed to address the challenges the legal system faces such as can predict case, recommend similar cases for legal research, and summarize documents to extract meaningful key points in the document. By applying Machine Learning (ML) and Natural Language Processing (NLP) techniques, this system can significantly improve the effectiveness of legal research and decision-making. The integration of these AI capabilities into a web-based user interface (UI) ensures accessibility, usability, and seamless interaction for legal professionals and practitioners.

The system simplifies legal procedures with features such as automated summarization, which distills lengthy rulings for easy comprehension; similar case recommendation, which retrieves pertinent prior cases using NLP techniques like TF-IDF and cosine similarity; case prediction, which evaluates legal documents to estimate

potential judgments. These characteristics shorten the time needed for case preparation, increase decision-making

The main purpose of this project is to create an AI solution that addresses the increasing delays in the court system. As cases continue to climb, efficiency is becoming more difficult for courts and legal professionals due to the huge traditional caseload. For example, reading and deciding on evidence, complex legal documents, and thorough research are becoming challenging due to the enormous amount of data. It will take a long time before legal professionals can respond to the litigant or determine the case where there is a delayed justice affecting good will in the public's trust of the courts and the system itself. To tackle this task, the proposed system utilizes AI and machine learning to automate and enhance a number of core functions in the workflow of the legal process. The system is capable of being a tool to assist legal professionals by automating the extraction and analysis of text information from case documents as either PDF's or text. So rather than taking the time to read and go through the case files, the AI would rapidly process and analyze the materials using natural language processing.

A distinctive element of this project is intelligent case summarization (e.g. through LexRank). It achieves this by applying various summarization techniques (such as LexRank or transformer-based models like LegalBERT) that allow a significant volume of text to be encapsulated to create a summary version which retains the most critical and informative pieces of information and allow both lawyers and judges to understand the main facts and issues without having to read the entire document. This alleviates the burden of time and cognitive energy to read and understand the entire document.

Another equally important component of the application is the case recommendation module, which is based on Term Frequency-Inverse Document Frequency (TF-IDF). The cosine similarities between the input case and a library of existing legal cases are used, the module will then recommend the top five most similar cases. Legal practitioners rely on comparative analysis to argue their case. If they can compare their case to precedent cases and apply arguments based upon the precedent case it saves a lot of time on manual legal research.

The system also includes an outcome predicting feature, using supervised machine learning. The outcome prediction is achieved through a Support Vector Machine (SVM) model that we train on previously labelled legal cases. The outcome will either be accepted or rejected, allowing lawyers to articulate what the likely outcome will be in a currently indeterminate case. The prediction feature allows legal practitioners to gain an early indication of the likely decision, this enables them to adjust their strategy and ensure they are best prepared for hearings.

Ultimately, this project brings an innovative plan to the legal ecosystem by integrating artificial intelligence in the legal space will process efficiently, provide improved accuracy and enhance speed in decision-making. The project eliminates manual reliance on routine legal tasks like summarization and research, provides legal professionals with supporting smart tools to increase the quality of legal services through timeliness and efficiency. Most importantly, it advances the overall goal of providing justice more promptly and effectively.

A. System Architecture

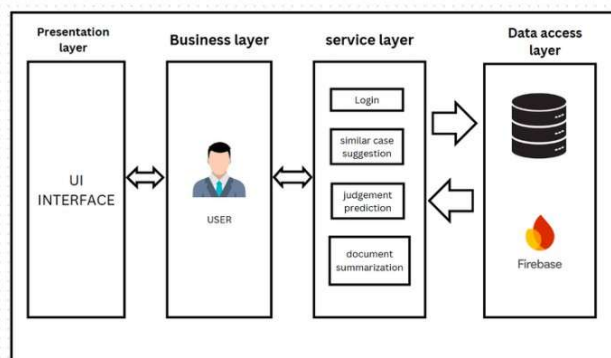


Fig. 1. system architecture

The architecture diagram illustrates the overall structure and interactions within the system. The system is divided into four layers i.e, presentation ,Business ,Service and Data access Lay-ers. The presentation layer comprises of user interface through which the user interacts with the system. Business layer consists of users of

the system. users are legal practitioners. The service layer consists of functionalities which are provided by the system which are user registration, Login and profile management. It also includes similar case recommendation, document summarization and possible case prediction. Lastly, the data access layer which comprises of database used by the system. The Firebase is used to store the information of the users and also to store their login credentials.

B. Process Flow

The process of the Legal Companion System is illustrated in the picture, starting with the user registering and login, getting into the system. Once logged in, the user is required to enter the details of their case, which the system pre-processes by performing vectorization (the process of converting a piece of text into numerical representations of text data and therefore creating meaning from words so that it can be analyzed computationally), stopword removal is the process of removing common words, or stop words, which do not add much significant meaning, and tokenize: the process of cutting text into smaller parts. After pre-processing the system summarizes the incoming text into its important parts. The system also finds and recommends relevant prior court cases, using case recommendation techniques. for example: TF-IDF, at which point the technology predicts the case outcome based on previous cases. Finally, the process completes with the user having had the benefit of obtained useful information and having improved efficacy of legal research and the supporting usability of case analysis.

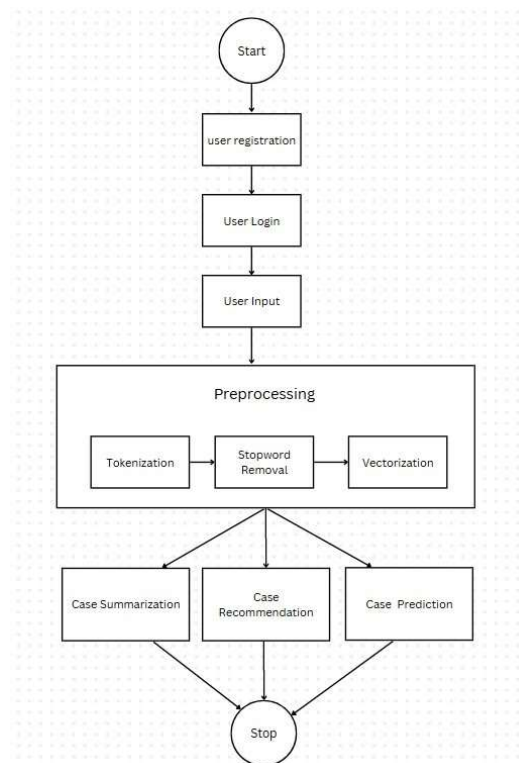


Fig. 2. Process Flow diagram

C. Methodology

The Nyaay Sakha system implements three essential services for representing legal documents: case prediction, similar case recommendation, and summarization. The general procedure we employed to create and execute each of these services is shown in the following sections:

- **Document Upload & Preprocessing:** Users engage with the system by uploading .pdf or .txt legal case documents via the frontend. When it receives a file it is checked for file's presence and type. It saves the file to a specified directory (uploads/) for future processing. The program then uses the extension of the file to extract the content: PyMuPDF (fitz) for .pdf files, Native file reading for .txt files. After the text is extracted it also sanitizes, this is to ensure no empty inputs are processed.

- **Summarization of Legal Document** : To help users more quickly comprehend long case documents we summarize the documents that are extracted sentences using the lextank algorithm (from the sumy library).Using Plain-textParser to parse the document and also using the tokenize function to break the document into sentences. The LexRank algorithm builds a graph of sentence similarities and to find the sentences that are most "central" in the document. A set of top 'n' sentences is returned to represent a short summary. The output is a summary of the document that captures the underlying core legal issues of the document.
- **Similar case Recommendation** : This feature identifies and recommends the cases that, to the best of its abilities, are the most similar to the user uploaded case.The feature loads case files that have been stored in a directory.[Object_casedocs/]Each of the cases including the uploaded case is vectorized with TF-IDF vectorization(TfidfVectorizer from sklearn).Cosine similarity is calculated for the uploaded case and all of the stored cases.The feature returns the top 5 case filenames that are similar to the user uploaded case by similarity score.
- **Case prediction** : To aid you in making legal decisions, the system predicts the likely outcome of a case: Joblib loads a Support Vector Machine (SVM), pre-trained model.A saved TF-IDF vectorizer (training on either plain TF-IDF or training in terms of InCaseLegalBERT features) processes the text of the uploaded document.The SVM model, takes the transformed vector as input and classifies it. The output can be mapped to either: "Accepted" (label for prediction = 1)."Rejected" (label for prediction = 0). This feature, quickly predicts the machine-learned outcome via consideration of the document's contents. This feature is useful to professionals in the legal realm to quickly study precedent cases.

IV. RESULTS

The data is used is trained on different models to compare the accuracy of the prediction of case outcome.Among them SVM model yields the accuracy of 83 %.The summarization and recommendation results have given reasonable accuracy which are based on the LexRank Algorithm and TF-IDF Cosine similarity.

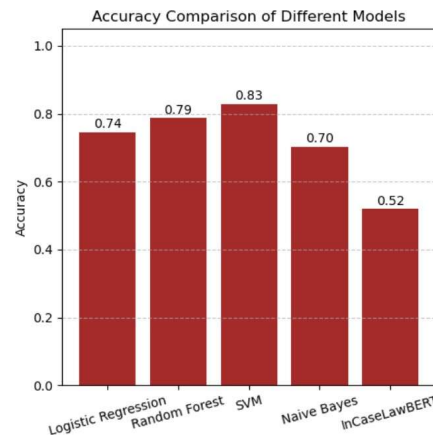


Fig. 3. Prediction Accuracies of different models

V. CONCLUSION AND FUTURE WORKS

The primary objective of this project is to decrease judicial system delays by giving AI-driven solution that speed up case analysis, legal research. Courts and legal experts find it very challenging to manage high caseloads, carry out in-depth investigations, and guarantee prompt case resolutions due to the growing backlog of cases. Artificial intelligence (AI) reduces the time needed for legal research while increasing efficiency and decision-making by automating time-consuming processes. Legal professionals operate more efficiently and provide justice more quickly and efficiently by utilizing AI for intelligent document processing, case summarizing, and predictive analytics. The project can be integrated with an AI chatbot to retrieve information about articles and Sections. As of now pdf or txt is accepted as input further it can be integrated with any optical reader to take input from an image for the functionalities. The project can be expanded for multi language support.The features can be further enhanced with complex algorithms which can be more faster to give the output.

REFERENCES

- [1] Prameela Madambakam, Shathanaa Rajmohan, "AI based Attorney Using NLTK Library", 2022 IEEE Silchar Subsection Conference (SILCON) | 978-1-6654-7100-8/22/\$31.00 ©2022 IEEE, 2022
- [2] Donny Kurniawan, Siti Elda Hiererra, "AI Legal Companion: Enhancing Access to Justice and Legal Literacy for the Public", 2024 International Conference on ICT for Smart Society (ICISS) | 979-8-3503-8965-4/24/\$31.00 ©2024 IEEE, 2024
- [3] Kishan Kanhaiya, Kamlesh Gautam, Naveen, Pramod Singh Rathore, Arpit Kumar Sharma, "AI Enabled- Information Retrieval Engine (AI-IRE) in Legal Services: An Expert-Annotated NLP for Legal Judgements", 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) | 979-8-3503-2579-9/23/\$31.00 ©2023 IEEE, 2023
- [4] Surisetty Hima Varshini, Gottimukkala Sarayu Varma, Priyanka Prab-hakar, Peeta Basa Pati, "Comparative Analysis of Legal Outcome Prediction with Detailed and Summarized Text", 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) | 979-8-3503-9447-4/24/\$31.00 ©2024 IEEE, 2024
- [5] Angelina George, Alphonsa Jose, Ashik F, Priyanka Prabhakar, Peeta Basa Pati, Shantipriya Parida, "Enhancing Legal Decision Making: WRIT Case Outcome Prediction with LegalBERT Embeddings and Ad-aBoost Classifier", 2024 IEEE International Conference on Contemporary Computing and Communications (InC4) | 979-8-3503-8365-2/24/\$31.00 ©2024 IEEE, 2024
- [6] Prameela Madambakam, Shathanaa Rajmohan, "A Study on Legal Judgment Prediction using Deep Learning Techniques", 2022 IEEE Silchar Subsection Conference (SILCON) | 978-1-6654-7100-8/22/\$31.00 ©2022 IEEE, 2022
- [7] Allan Almeida, Shubh Shah, Neha Agarwal, Neer Bhavesh Khan-dor, "LEGALAI: A Tool to help Lawyers Optimize their Performance", 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT) | 979-8-3503-7809-2/24/\$31.00 ©2024 IEEE, 2024
- [8] XURAN WANG, XINGUANG ZHANG, (Member, IEEE), VANESSA HOO, ZHOUEHANG SHAO, AND XUGUANG ZHANG, (Associate Member, IEEE), "LegalReasoner: A Multi-Stage Framework for Legal Judgment Prediction via Large Language Models and Knowledge Integration", IEEE Access Research Article, 2024
- [9] Rungpilin Anantathanavit, Jomchai Chongthanakorn, Warisa Kongsanti-nart, Pisit Praiwattana, Tipajin Thaipisutikul, "Utilizing AI and Natural Language Processing to Predict Supreme Court Decisions in Thailand", 2024 16th International Conference on Knowledge and Smart Technology (KST) | 979-8-3503-7073-7/24/\$31.00 ©2024 IEEE, 2024
- [10] Ritika Singh, Bura Vijay Kumar, Kanchan Kishor Rawat, Shaik Vaseem, Samta Kathuria LCD, Vikrant Pachouri, "Intervention of AI in Prediction of Judgement of Cases", 4th International Conference on Smart Electronics and Communication (ICOSEC) | 979-8-3503-0088-8/23/\$31.00 ©2023 IEEE, 2023