

Review and Approaches to Develop Legal Assistance for Lawyers and Legal Professionals using Artificial Intelligence and Machine Learning

Suma R.¹, Nayan Prakash Lal², Nethravathi K.³, Poreddy Sasi Vardhan Reddy⁴, N.V. Ajay Kumar⁵

¹ Associate Prof. and Head, Dept. of Computer Science and Engineering, T John Institute of Technology, Bengaluru
^{2,3,4,5} Department of Computer Science and Engineering, T John Institute of Technology, Bengaluru

Abstract - Legal assistance is an automated approach to providing accurate references, predictions and judgements to lawyers and legal professionals for case preparation using Machine Learning and Artificial intelligence. It reduces the time spent by the lawyer in the preparation of a case study for the client's representation in court. It also improves the overall quality of the case study by providing better references and best case-related judgments. This also helps the judiciary in its decision-making process. This research work is a critical analysis of various AI and Machine Learning approaches used to provide legal assistance to judicial systems all over the world.

Key Words: Artificial Intelligence (AI), Machine Learning, Courts of India, Lawyers, IPC, Legal, Judgements, Natural Language Processing (NLP), Ontologies etc.

1. INTRODUCTION

The law system in India is having a slow pace when it comes to the resolution of cases, Supreme Court of India alone has a list of 71,411 pending cases as of August 2nd, 2022, out of which 56,365 are civil cases and 15,076 criminal cases. In the High courts of India, the average hearing time for one case is only about 300 to 500 seconds with the next hearing to be conducted after 16 to 80 days. This increases the pressure on lawyers and Legal Professionals in preparation for their case studies for cases and trials. Even small negligence can be the difference between winning and losing the case. Thus, sometimes justice could be denied because of Human error. The only present option is to stretch the time of duration of the case to get proper evidence and prevent misjudgments.

The majority of work in the legal system of India is still being done manually which includes data collection for a case study, referencing, searching for the right judgements to refer for the cases etc. [11] In the District and Talukas courts of India more than 31 million cases are pending. In the High Courts of India, the tally reaches up to 4.5 million pending cases.

We can use AI/ML to assist lawyers in their case studies and reduce human error and inefficiency, resulting in less pressure on lawyers and advocates, better infographics for

judgement, and less time spent dealing with and finishing a case or trial. This study is a critical review of existing research and development for providing legal assistance using Artificial Intelligence and Machine Learning techniques. Machine Learning and Artificial Intelligence have transformed every major field's working model. AI and ML have automated the work that was previously done manually, it saves time, increases quality, and improves overall efficiency. The AI-based prediction model will help the Lawyer and Advocate decide the level of expertise that is needed in a case.

Following are the AI and Machine learning techniques that are used for building a legal assistant.

- Text Analysis
- Sentimental Analysis
- Neural Network Model
- Decision Tree
- Clustering

1.1 Motivation

The motivation for this research is to identify optimal techniques that could aid in the development of an AI-powered Assistant for legal use, which could speed up the process of case completion while also providing better representation to clients.

As a country with a population of over 1.3 billion people, it is critical to optimize the resources available in the legal system so that everyone can exercise the fundamental rights guaranteed by the Indian Constitution.

1.2 Problem Statement

There are no mainstream automation strategies implemented in today's Indian legal system that helps lawyers and legal professionals prepare for the representation of their client. Due to the high number of cases that are pending in Indian judicial courts, this problem amplifies furthermore and could lead to underrepresentation due to human limited capabilities, increasing time in resolution of cases and making the legal system out of reach for the majority of the citizens.

1.3 Programming Language

Python is used for the development of Machine learning and Artificial Intelligence project because of its massive libraries that are useful for data analysis, data manipulation, easier implementation of Machine learning and AI models etc. Python frameworks are referred in this research work for the implementation of different techniques used by a legal assistant.

2. Text Analysis

Text analysis is a process to analyze text and find out the indented of the text. This process helps the machine understand the true meaning of the text and can generalize the text into different categories. [12] Some models use NLP for legal document classification. The techniques used in those models are Tokenization, Lowercasing, Punctuation removal, Part of Speech tagging, Numeral normalization, Name normalization, Lemmatization, and Non-ASCII character removal. [3] After information extraction for text analysis, we need to perform normalization. The normalization process finds the guessed entities, and aliases for those guessed entities and groups those entities together. Aliases are entities that are misspelt to a certain tolerance level from the root entity or are written in the scripts that define two entities as aliases.

The Natural language processing required for text analysis of legal documents are: -

- Tokenization
- Stop words removal
- Stemming
- Part of speech tagging

In Figure 1 We can see the sequential process of text analysis. First, the text is translated into a common language, in this case, Hindi text (legal text can be in any of the 22 scheduled languages in India) is translated into English. After the translation process, tokenization takes place. The list of tokens generated then goes for stop word removal. After the removal of stop words stemming from the text is performed. The root words then go for part of speech tagging. Now the text is ready for the sentimental analysis.

[15] In Figure 2 we can see different levels of ontology and NLP techniques to transform semi-structured documents into structured RDF triples and we described suitable algorithms that can able to segment and extract the required information from legal documents. The experimental section has shown very encouraging results

यह एक शोध परियोजना है

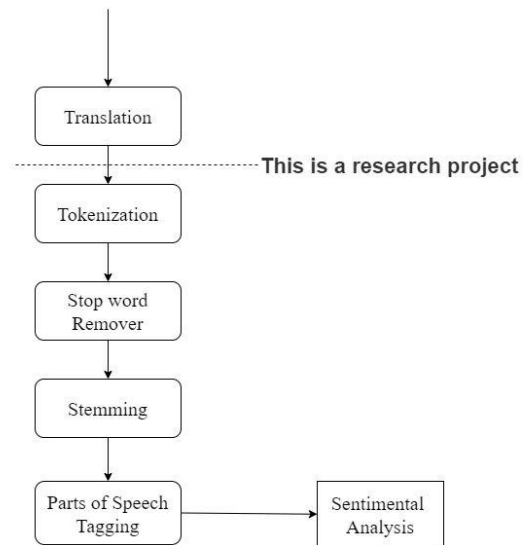


Figure 1. Sequential use of Natural Language Processing Techniques.

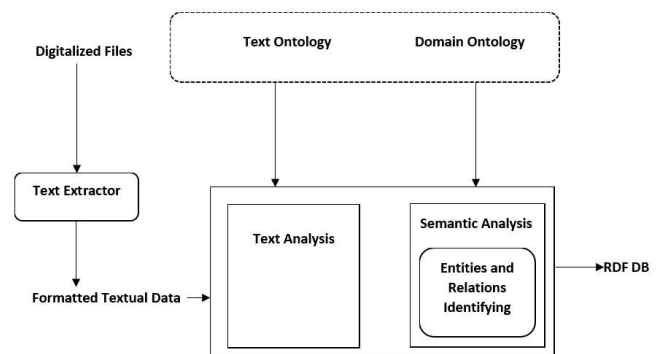


Figure 2. Different levels of Ontologies

2.1 Tokenization

Tokenizing is a pre-processing process which is used to split the text into small chunks or blocks such as a word, term, letters etc. Each chunk or block is called a token. Tokenization is the most basic step for Natural Language Processing. It is important because we can identify different kinds of words in a text. it will help in further processing the text to analyze the meaning of the text. Further processing of tokens can lead to many Natural Language applications such as sentimental analysis, removing stop words etc.

2.2 Stop Words

The set of commonly used words in a text of any given language is called stop words. Stop words in data mining and text filtering are used to remove common occurring words to focus on important or special words. Some of the stop words in English are: - the, is, are, if, can etc.

2.3 Stemming

Stemming is a form of data pre-processing in which an inflicted word or derived word is filtered back to its base word. The base word is also known as the "stem word" thus deriving the word Stemming. Stemming is important in data processing and analysis because sometimes many variations of the stem words may be present in a text file including suffixes and prefixes but the meaning of the word is the same in every context [19]. It would be a waste of processing power and time to consider each word with the same stem as a separate entity.

2.4 Parts of Speech Tagging

We can use NLTK (NLTK stands for natural language toolkit is a python package used for natural language processing) for tokenizing the text as per the part of speech that they belong to. This type of tokenizing can hugely help the machine to understand not only the literal meaning of the text but also the true intended meaning of the text.

2.5 Sentimental Analysis

Sentimental analysis is a process in which we can deduce how the text is been presented and what sentiments the text shows. For criminal legal text, the system can deduce which category the case belongs based on the sentiments of the legal text. Analysis of the sentiments could further help in the grouping of the case with other cases with the same sentiments.

3. Models

[4] In Figure 3 Machine Learning is used as a support system for the jury. The prosecution and defence and judge are given as an Argument framework that will be used as input for the Machine learning model as well as be a critical part of jury decision-making.

[7] The following model uses CNN for the classification of law tex. This model uses the K-fold cross-validation technique. K=4 then it means that the model will perform training and testing 4 times. 75% of the data can be taken as a training set and 25% of the data can be taken as a test set. The last layer includes the IPC section into which the law cases are classified. The accuracy of the model is 85%.

KNN can be used as a crime prediction tool that can help law enforcement. The prediction accuracy is significantly increased. [17] There is a model that uses the KNN system to traverse through the crime data to find different reasons that led to crime. Mean Absolute Error and Root Mean Square Error are used to measure the errors in the set of predictions. For the training dataset, 80% of the data is used and for the Test dataset, 20% of the data is used.

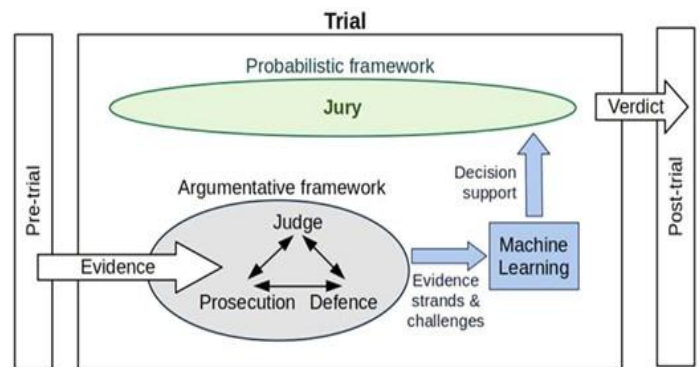


Figure 3. Machine Learning Support Model for the generation of Verdict. Image ref. [4]

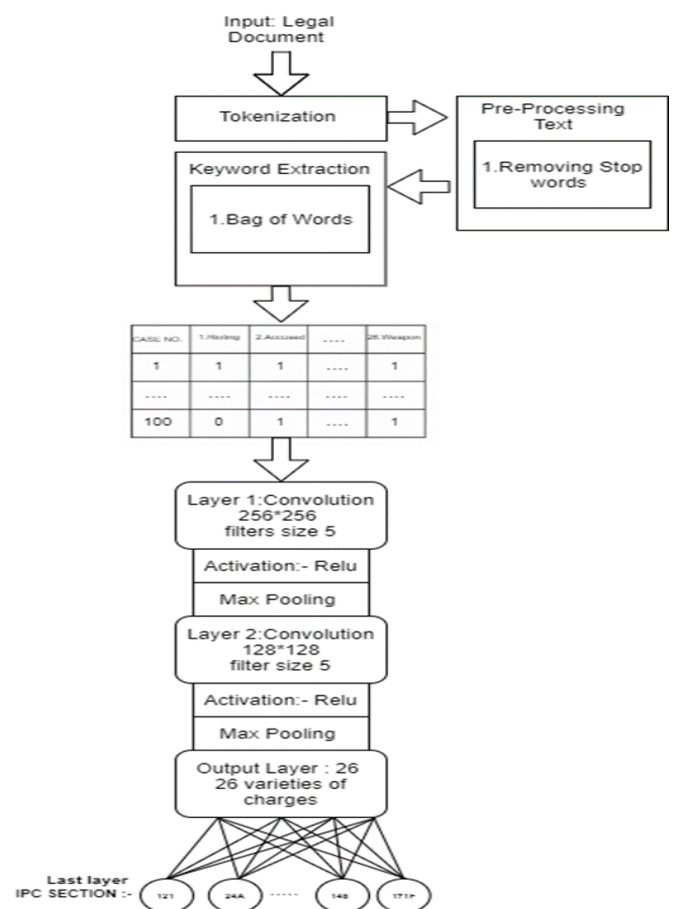


Figure 4. CNN model with NLP extraction and Classification of text. Image Ref. [7]

4. Decision Tree

Decision tree can be used to find out the "guilty" and "not guilty" status of similar kinds of cases which can be used as a winning percentage calculator, better referencing, finding out missing points in those cases etc. Figure 5. Shows the classification of a court case into different groups and in those groups finding out the verdict as guilty or not guilty. [11] Further derivation of the decision tree will help build an

ontology-based intelligent Question Answering model, one of its use cases can be online learning where the answering can be made more intelligent by using semantic query processing.

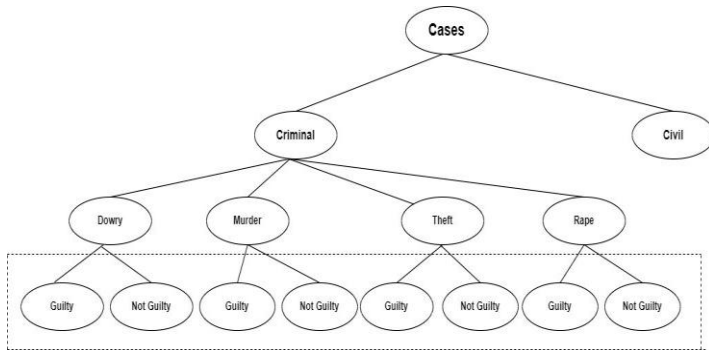


Figure 5. Decision Tree for classification of cases

5. Clustering

Clustering legal judgements could help us to identify the different clusters of judgements that could be used as the base for similar kinds of cases. Due to the high volume of judgements present in the Indian Judicial System with Supreme Court's Judgement has

[5] For the clustering of data, there may come a situation in which it is hard to identify the group to which the particular information belongs. For that purpose, we have to consider the gap in the clusters. The data point that comes between the gap lines is the data that have to be considered again to find out to what group it belongs. In Figure 6. We can see the gap identify what values can be identified as dogs and cats and what values cant' be decided.

[11] There is a Chinese legal judgement prediction system that includes physiological features and the description of facts to predict the judgements. Yinglong Ma, Peng Zhang and Jiangang Ma have proposed a system that uses ontologically driven knowledge blocks to summarize and compute documents. The proposed system uses a KNN-based approach for document classification. Figure 7 is the Illustration of the Chinese legal judgement prediction system that is been modified to be used in the Indian judicial system.

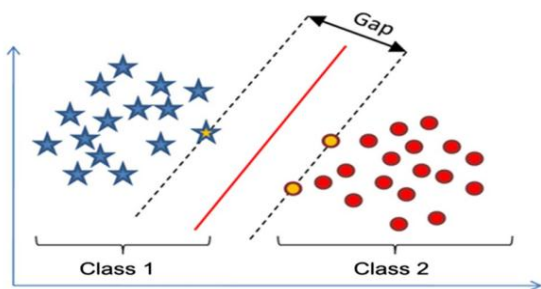


Figure 6. Gap Analysis

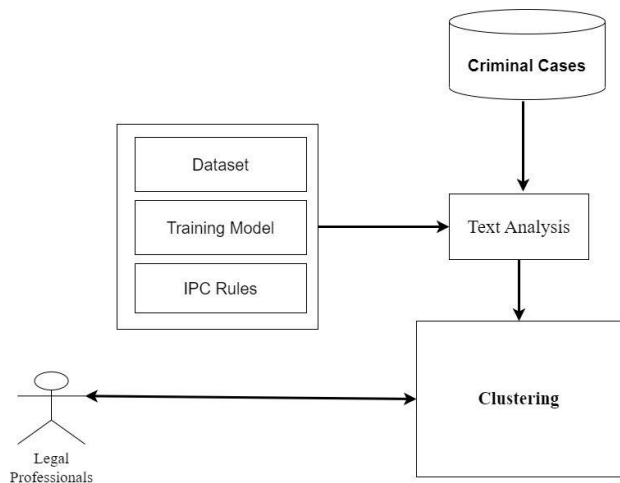


Figure 7. Legal Judgement prediction system

3. CONCLUSION

Legal Assistance is the need of the hour. Research and development are going on in this field to make an automated system. The research mostly focuses on Natural Language Processing for text analysis and Deep learning techniques for clustering of text into different groups. AI and Machine Learning models have proven to be useful in preparation of case studies as well as providing assistance to decision makers such as the jury [4].

There are many models which are currently at the research stage and will soon be applied in real world applications.

The current development mostly focuses on legal text segmentation for further analysis. CNN and KNN models are used for clustering cases into groups [7][17] and Decision Tree is used for classifying the groups to which a case belongs [11]. Few successful models work at the efficiency of 75% to 85%.

REFERENCES

- [1] Masha Medvedeva, Michel Vols and Martijn Wieling." Using machine learning to predict decisions of the European Court of Human Rights". Published on Springer on 26 June 2019 open access, Creative Commons Attribution 4.0 International License. link.springer.com/article/10.1007/s10506-019-09255-y.
- [2] Sandeep Bhupatiraju, Daniel L, Chen and Shareen Joshi." The Promise of Machine Learning for the Courts of India". The Promise of Machine Learning for the Courts of India. National Law School of India Review, 2021, 33 (2), pp.462-474. (hal-03629734)
- [3] Jonathan Brett Crawley and Gerhard Wagner. "Desktop Text Mining for Law Enforcement".2010 IEEE

- International Conference on Intelligence and Security Informatics. 10.1109/ISI.2010.548476.
- [4] Jane Mitchell, Simon Mitchell and Cliff Mitchell. "Machine learning for determining accurate outcomes in criminal trials". Law, Probability and Risk, Volume 19, Issue 1, March 2020, Pages 43–65, <https://doi.org/10.1093/lpr/mgaa003>
- [5] Riya Sil and Abhishek Roy. "A Novel Approach on Argument based Legal Prediction Model using Machine Learning". Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020). 10.1109/ICOSEC49089.2020
- [6] Huseyin Umutcan Ay, Alime Aysu Oner and Tolga Kaya. "A Machine Learning- Based Decision Support System Design for Restraining Orders in Turkey".2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC). 10.1109/COMPSAC51774.2021.00226
- [7] V Gokul Pillai and Lekshmi R Chandran. "Verdict Prediction for Indian Courts Using Bag of Words and Convolutional Neural Network". Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020). 10.1109/ICSSIT48917.2020.9214278
- [8] Hui Wang, Tieke He, Zhipeng Zou, Siyuan Shen and Yu Li". Using Case Facts to Predict Accusation Based on Deep Learning".2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C). 10.1109/QRS-C.2019.00038
- [9] J Haitham Hmoud Alshibly and Mohammad atwah Alma'aitah. "Artificial Intelligence in Law Enforcement". International Journal of Advanced Information Technology (IJAIT) Vol. 4, No. 4, August 2014. 10.1109/5254.653229
- [10] Kevin D, Ashley." Artificial Intelligence and Legal Analytics (New Tools for Law Practice in the Digital Age) Machine Learning with Legal Texts". University of Florida, Book - Artificial Intelligence and Legal Analytics (New Tools for Law Practice in the Digital Age) Machine Learning with Legal Texts
- [11] Nishant Jain and Gaurav Goel. "An Approach to Get Legal Assistance Using Artificial Intelligence". 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India. June 4-5, 2020 .10.1109/ICRITO48877.2020.9198029
- [12] Mariana Y. Noguti, Eduardo Vellasques, Luiz S. Oliveira. "Legal Document Classification: An Application to Law Area Prediction of Petitions to Public Prosecution Service".2020 International Joint Conference on Neural Networks (IJCNN). 10.1109/IJCNN48605.2020.9207211
- [13] Hemlata Sharma, Aakanksha. "Artificial Intelligence and Law: An Effective and Efficient Instrument".2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). 10.1109/ICRITO51393.2021.9596503
- [14] Vishu Madaan, Subratho Kumar Das, Prateek Agrawal, Charu Gupta and Dhruv Goel. "Fusion of ML models to Identify Sexual Harassment Cases".2021 International Conference on Computing Sciences (ICCS). 10.1109/ICCS54944.2021.00058'
- [15] Flora Amato, Antonino Mazzeo, Antonio Penta and Antonio Picariello. "Using NLP and Ontologies for Notary Document Management Systems".2008 19th International Workshop on Database and Expert Systems Applications. 10.1109/DEXA.2008.86
- [16] Suncana Roksandic, Nikola Protrka and Marc Engelhart. "Trustworthy Artificial Intelligence and its use by Law Enforcement Authorities: where do we stand?" 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO). 10.23919/MIPRO55190.2022.9803606
- [17] Akash Kumar, Aniket Verma, Gandhali Shinde, Yash Sukhdeve and Nidhi Lal. "Crime Prediction Using K-Nearest Neighboring Algorithm".2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). 10.1109/ic-ETITE47903.2020.155
- [18] Saqeeb Abdullah, Farah Idid Nibir, Suraiya Salam, Akash Dey, Md Ashrafal Alam and Md Tanzim Reza. "Intelligent Crime Investigation Assistance Using Machine Learning Classifiers on Crime and Victim Information".2020 23rd International Conference on Computer and Information Technology (ICCIT). 10.1109/ICCIT51783.2020.9392668
- [19] Marijn Schraagen, Floris Bex. "Extraction of Semantic Relations in Noisy User-Generated Law Enforcement Data".2019 IEEE 13th International Conference on Semantic Computing (ICSC). 10.1109/ICOSC.2019.8665497
- [20] Jishitha Kuppala, K. Kalyana Srinivas, P. Anudeep, R. Sravanth Kumar and P. A Harsha Vardhini. "Benefits of Artificial Intelligence in the Legal System and Law Enforcement".2022 International Mobile and Embedded Technology Conference (MECON). 10.1109/MECON53876.2022.9752352