International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

# Digital Legal Aid Assistant For Marginalized Communities in India

**Mrs. Hilda Kanaga V[1], Aishwarya B[2], Ashika T[3], Sandhiya R[4], Synthavi R A[5]**

[1] *Assistant Professor ,Department of Information Technology, Meenakshi Collegeof Engineering , K K Nagar , Chennai.*
[2,3,4,5] *UG Student , Department of Information and Technology , Meenakshi Collegeof Engineering , K K Nagar , Chennai.*

---

**Abstract -** *Marginalized communities often face barriers in accessing legal assistance due to various factors including language barriers, lack of resources, and limited accessibility to legal services. In this paper, we propose the design and development of a comprehensive legal chatbot tailored to address the needs of marginalized communities. Our chatbot incorporates voice recognition capabilities and language selection features to enhance accessibility for users with diverse linguistic backgrounds. Drawing upon advances in natural language processing (NLP) and machine learning, we outline the architecture, functionalities, and ethical considerations of the chatbot. Additionally, we discuss the importance of community engagement and user-centered design in ensuring the effectiveness and inclusivity of the chatbot. Through the integration of voice recognition, language selection, and a user-centric approach, this paper aims to empower marginalized communities by providing them with accessible and reliable legal assistance.*

***Key Words***: **Artificial Intelligence (AI), Machine Learning, Courts of India, Lawyers, IPC, Legal, Judgements, TF-IDF Algorithm, SVM Algorithm , Random Forest etc.**

## 1. INTRODUCTION

In many parts of the world, marginalized communities face significant challenges in accessing legal assistance and understanding their rights. Language barriers, financial constraints, and geographical limitations often exacerbate these difficulties, leading to a lack of access to justice. As technology continues to advance, there is a growing opportunity to barriers and improve legal access for marginalized populations.

One promising approach is the development of chatbot applications that provide legal information and assistance in a user-friendly and accessible manner.

Chatbots, powered by artificial intelligence (AI) and TF-IDF ,Support Vector Machine,Random Forest technologies, have the potential to interact with users conversationally, understand their queries, and provide relevant information and guidance. By integrating voice recognition technology and offering multi-language support, these chatbots can further enhance accessibility and cater to the diverse needs of marginalized communities.

In this paper, we present the concept of a chatbot application designed specifically for marginalized communities, aiming to bridge the gap in legal access. The chatbot will serve as a virtual legal assistant, offering information on various legal topics, guiding users through legal processes, and connecting them with relevant resources and services. Through a user-friendly interface and intuitive interaction, the chatbot seeks to empower individuals to navigate the legal system more effectively and assert their rights.

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

- TF-IDF
- Support Vector Machine
- Random Forest.

### 1.1 MOTIVATION

The motivation of this research is to empower marginalized communities in India by providing accessible legal aid. With language selection and categorized chatbot support for civil, immigration, LGBTQ, and socio-economic issues, the project ensures equitable access to justice.

One of the critical components of this project is its emphasis on language selection. India is a linguistically diverse country, with numerous regional languages spoken across its vast landscape. By offering virtual lawyer consultations and centralizing legal resources, it reduces barriers to legal representation.

### 1.2 PROBLEM STATEMENT

Financial obstacles prevent marginalized communities from accessing legal representation. To empower them in advocating for systemic reforms, structural biases within the legal system must be addressed. Leveraging technology to disseminate legal information can bridge gaps for those with limited access. Language barriers hinder communication, presenting challenges for clients and legal professionals alike.

### 1.2 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. TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY

TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistic used in information retrieval and text mining to evaluate the importance of a word in a document relative to a collection of documents. In the context of a legal aid chatbot, TF-IDF can play a crucial role in analyzing and categorizing legal documents, assisting users in finding relevant information, and improving the overall effectiveness of the chatbot. Here's how TF-IDF works and its application in a legal aid chatbot.

Term Frequency (TF) measures the frequency of a term (word) within a document. It reflects how often a word appears in a document relative to the total number of words in that document.

Inverse Document Frequency (IDF) measures the importance of a term across a collection of documents. It assigns higher weights to terms that appear less frequently across all documents in the corpus.

The TF-IDF score for a term in a document is calculated by multiplying its Term Frequency (TF) by its Inverse Document Frequency (IDF).

The features extracted from legal documents using TF-IDF for text analysis are as follows: -

- Document Retrieval
- Keyword Extraction
- Topic Modeling

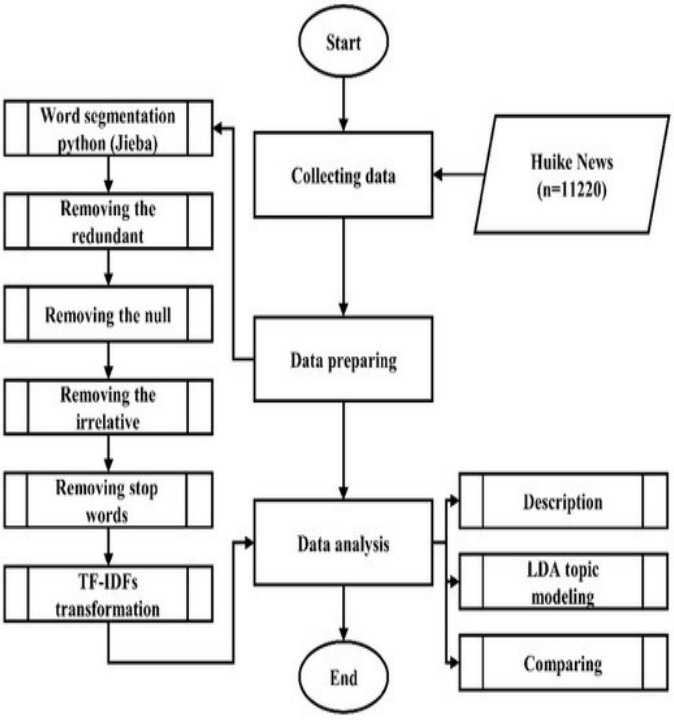


Figure 1. Data processing flowchart.

### 2.1 DOCUMENT RETRIEVAL

TF-IDF can be used to rank legal documents based on their relevance to a user query. Documents containing terms that are both frequent within the document and rare across the entire document collection are considered more relevant. TF-IDF is widely used in information retrieval systems like search engines, document clustering, and recommendation systems. It helps in effectively identifying relevant documents by considering both the local importance of terms in documents and their global rarity across the entire corpus.

### 2.2 KEYWORD EXTRACTION

Keyword extraction techniques aim to distill the most salient information from a given text, aiding in summarization, categorization, and search. These methods often leverage statistical and linguistic analysis to identify words or phrases that carry significant meaning within the context of the document. One common approach is based on frequency analysis, where terms appearing more frequently are considered more important. Additionally, algorithms such as TF-IDF (Term Frequency-Inverse Document Frequency) assess the relevance of terms by considering their occurrence not only within the document but also across a collection of documents.

Another method involves graph-based algorithms like TextRank, which treat the text as a network of interconnected terms and rank them based on their centrality within the network. Keyword extraction plays a crucial role in various applications, including document summarization,information retrieval, content recommendation, and search engine optimization (SEO), facilitating efficient access to and understanding of large volumes of textual data.

## 2.3 TOPIC MODELING

Topic modeling is a powerful technique used in natural language processing to uncover the latent themes or topics present in a collection of documents. While TF-IDF is traditionally associated with document retrieval and term weighting, it can also be leveraged for topic modeling. In the context of TF-IDF, topic modeling involves identifying clusters of terms that frequently co-occur across documents, indicating their association with specific topics or concepts.

Once the TF-IDF matrix is decomposed, the resulting topic matrix can be interpreted to extract the underlying topics present in the corpus. Terms with high weights within a topic indicate their importance in defining that particular topic. By examining these top- weighted terms, human interpreters can assign meaningful labels to each topic, providing insights into the main themes represented in the document collection.

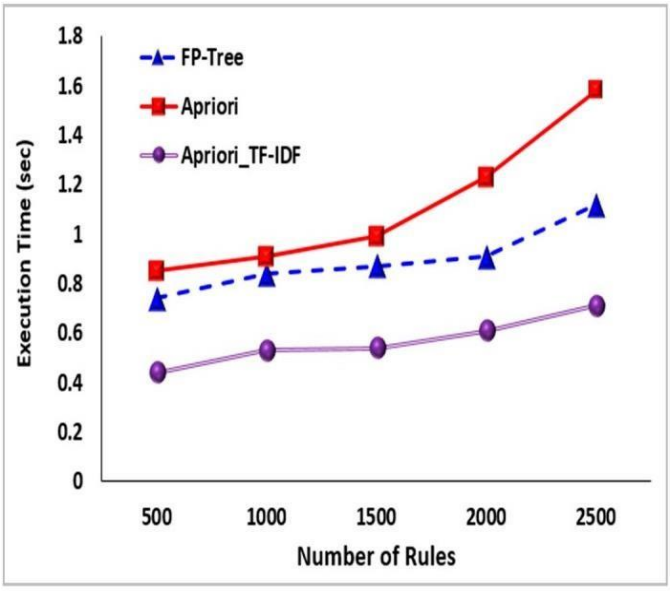


Figure 2. TF_IDF Ranking Score

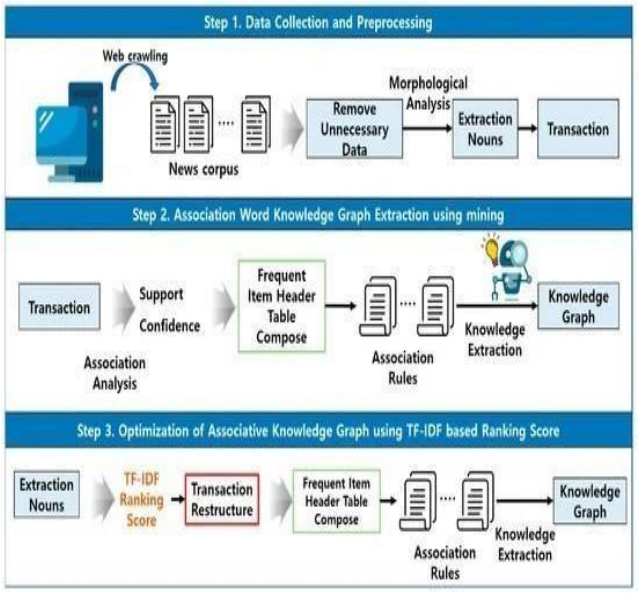


Figure 3. Optimizing process of association graph using TF-IDF ranking score.

Association graphs represent relationships between items in a dataset, where nodes represent items and edges denote associations between them. TF-IDF, a statistical measure, quantifies the importance of terms in documents within a corpus. By integrating TF-IDF ranking scores into association graphs, we refine the graph structure, facilitating more accurate analysis and extraction of meaningful associations.

## 3. SVM (Support Vector Machine)

Support Vector Machines (SVM) is a supervised learning algorithm that can be used for classification and regression tasks. In the context of a legal support chatbot, SVM can be employed for various tasks such as text classification, document categorization, and case prediction. SVM is based on the concept of finding the optimal hyperplane that best separates the data points into different classes in a high-dimensional space. The hyperplane is chosen such that the margin, which is the distance between the hyperplane and the nearest data point from each class (called support vectors), is maximized.

### 3.1 HYPERPLANE AND MARGIN

SVM works by transforming the input data into a higher-dimensional space using a kernel function. In this transformed space, SVM finds the hyperplane that best separates the data points into different classes. The hyperplane is chosen such that it maximizes the margin between the closest points (support vectors) of the classes. The hyperplane in SVM is the decision boundary that separates the classes. It is defined by the equation w*x - b = 0, where w is the weight vector, x is the input vector, and b is the bias term. The margin is the distance between the hyperplane and the nearest data point from either class. SVM aims to maximize this margin.

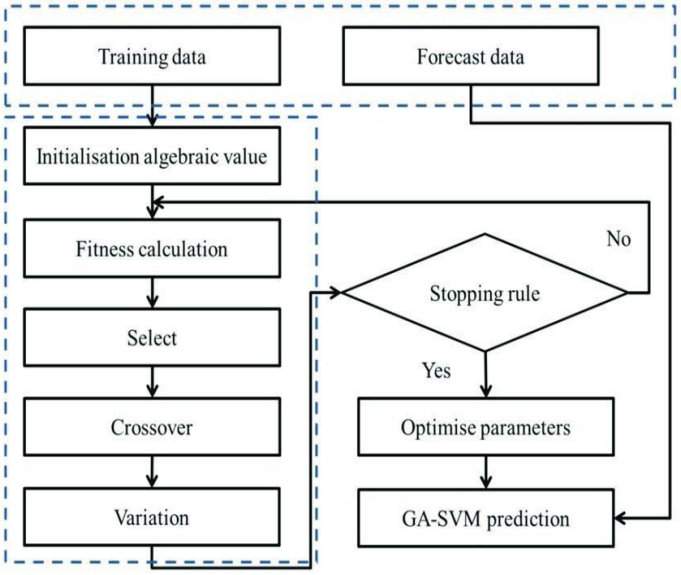


Figure 4. Flowchart of genetic algorithm-support vector machine

### 3.2 TRAINING PROCESS

The training process involves optimizing a cost function that penalizes misclassifications and adjusts the position of the hyperplane to maximize the margin. SVM aims to find the optimal decision boundary that separates the data points into different classes with the largest possible margin.

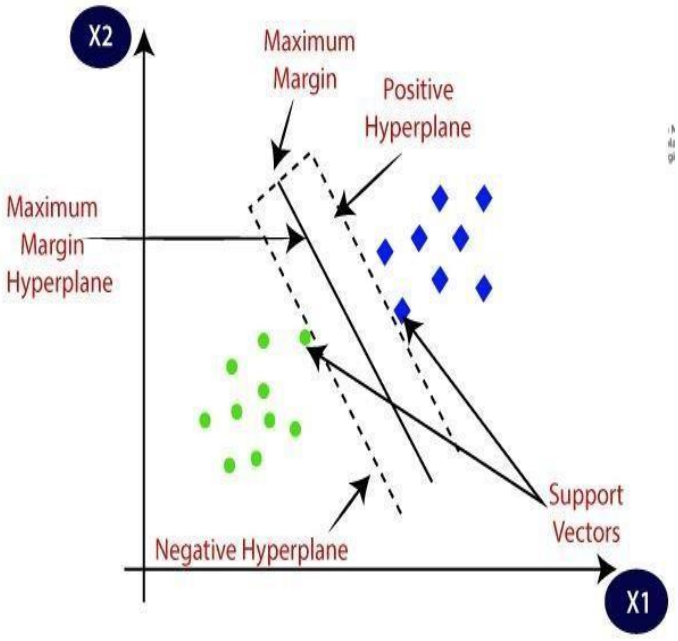


Figure 5. SVM Graph Plot

## 4. RANDOM FOREST

Random Forest is a versatile and powerful ensemble learning method used for both classification and regression tasks. It belongs to the family of tree-based algorithms and is renowned for its robustness and effectiveness in various machine learning applications.

Random Forest is built upon the concept of decision trees, which are simple yet effective models for classification and regression. However, decision trees are prone to overfitting, especially on complex datasets with noise. Random Forest addresses this issue by combining multiple decision trees, thus improving the overall predictive performance and generalization ability.

### 4.1 ENSEMBLE LEARNING

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.

### 4.2 DECISION TREES

Each decision tree in the Random Forest is built independently using a subset of the training data and a random selection of features. Decision trees are constructed by recursively splitting the data based on feature values to create branches and make predictions at the leaf nodes.

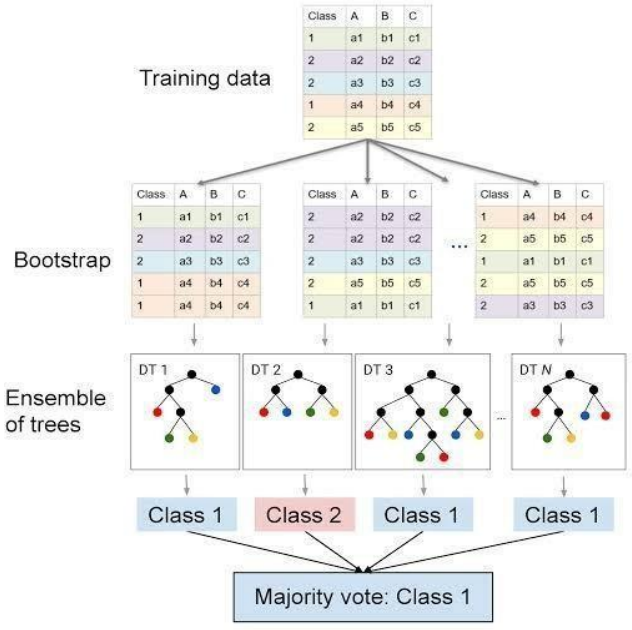


Figure 6. Flow Process of Random Forest

### 4.3 BOOTSTRAP AGGREGATING

Random Forest employs a technique called bagging, where multiple decision trees are trained on different random subsets of the training data with replacement. This sampling with replacement ensures that each decision tree in the forest sees a slightly different perspective of the data, reducing the risk of overfitting.

### 4.4 RANDOM FEATURE SELECTION

At each node of the decision tree, a random subset of features is considered for splitting. This randomness helps decorrelate the trees in the forest and ensures that each tree makes decisions based on different subsets of features.

### 4.5 VOTING OR AVERAGING

For classification tasks, each decision tree in the Random Forest independently predicts the class label of a sample.

The final prediction is determined by majority voting among the predictions of all decision trees. For regression tasks, the final prediction is typically the average of the predictions made by all decision trees.

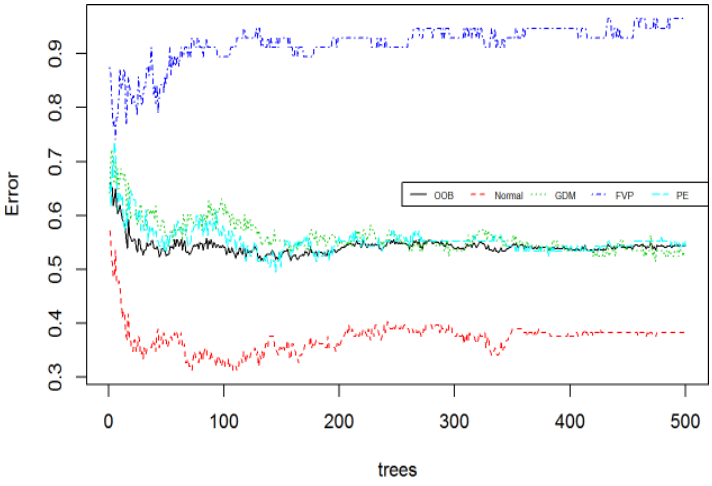


Figure 7. Random Forest Plot Interpretation

## 5. RESULT AND DISCUSSION

The proposed design and development of a comprehensive legal chatbot tailored to serve marginalized communities represents a significant step towards addressing the barriers they face in accessing legal assistance. By incorporating voice recognition capabilities and language selection features, the chatbot aims to enhance accessibility for users with diverse linguistic backgrounds.

The project will involve the development of an architecture that supports voice recognition and multi-language capabilities while ensuring data privacy and security. Ethical considerations will be paramount throughout the development process, with measures in place to safeguard user confidentiality and prevent biases in the chatbot's responses. Community engagement will be central to the project, with marginalized communities actively involved in the design and testing phases to ensure that the chatbot effectively meets their needs and preferences.

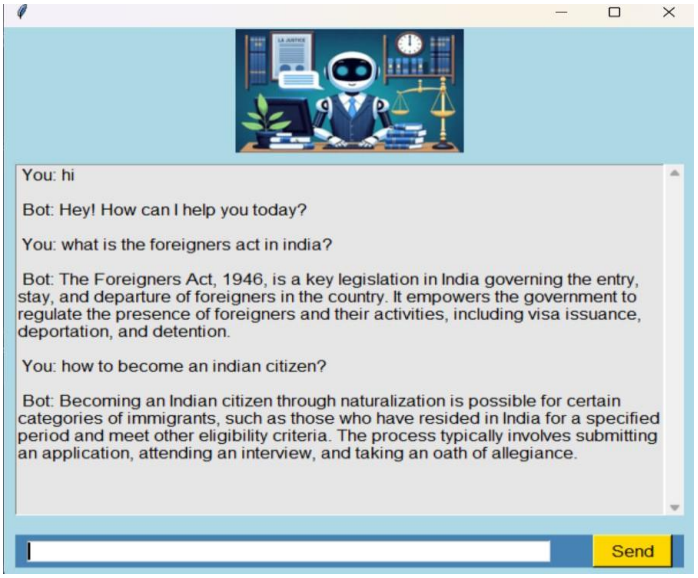


Figure 8. LegalAid ChatBot

Upon completion, the project aims to deliver a user-friendly and inclusive legal chatbot that serves as a reliable resource for marginalized communities seeking legal assistance. By breaking down language barriers and providing accessible support, the chatbot has the potential to promote equity and justice within the legal system, ultimately empowering marginalized individuals to assert their rights and access the legal resources they need.

## 6. CONCLUSIONS

In conclusion, the development of a chatbot application tailored to the needs of marginalized communities represents a significant step towards bridging the gap in legal access. By leveraging voice recognition technology and offering multi-language support, the chatbot addresses key barriers such as language proficiency and literacy, making legal assistance more accessible and more inclusively.

Through a user- friendly interface and personalized interaction, the chatbot empowers individuals to navigate the complexities of the legal system with confidence and clarity.

The potential impact of such a chatbot application is profound. By providing timely and accurate legal information, guiding users through legal processes, and connecting them with relevant resources and services, the chatbot has the potential to empower marginalized communities to assert their rights, address injustices, and seek redress for grievances. Moreover, by promoting legal literacy and awareness, the chatbot can contribute to the overall empowerment and well- being of these communities.

## REFERENCES

[1] An Approach to Get Legal Assistance Using Artificial Intelligence IEEE 2022 Nishant Jain,Gaurav Goel | Transaction Paper |Publisher: IEEE

[2] 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 Attribution4.0InternationalLicense.link.springer.com/article /10.1007/s10506-019- 09255- y.

[3] 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⟩

[4] 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.

[5] 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.

[6] 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

[7] 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

[8] 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 InventiveTechnology (ICSSIT 2020).

[9] 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

[10] Haitham Hmoud Alshibly and Mohammad atwah Al-ma'aitah. "Artificial Intelligence in Law Enforcement". International Journal of Advanced Information Technology (IJAIT) Vol. 4, No. 4, August 2014.
10.1109/5254.653229

[11] 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

[12] 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

[13] Vishu Madaan, Subratho Kumar Das, Prateek Agrawal, Charu Gupta and Dhruv Goel. "Fusion of Sexual Harassment Cases".2021 International Conference on Computing Sciences(ICCS).
10.1109/ICCS54944.2021.00058'

[14] Flora Amato, Antonino Mazzeo, Antonio Penta and Antonio Picariello. "Using NLP and Ontologies for Notary Document Management Systems".2008 19thInternational Workshop on Database And Systems Applications.
10.1109/DEXA.2008.86

[15] 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 electronict echnology (MIPRO). 10.23919/MIPRO55190.2022.9803606

[16] 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

[17] Saqueeb Abdullah, Farah Idid Nibir, Suraiya Salam, Akash Dey, Md Ashraful 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

[18] 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

**BIOGRAPHIES**

Mrs. Hilda Kanaga V
Assistant Professor
Meenakshi College Of Engineering

Aishwarya B
Student
Meenakshi College Of Engineering

Ashika T
Student
Meenakshi College Of Engineering

Sandhiya R
Student
Meenakshi College Of Engineering

Synthavi R A
Student
Meenakshi College Of Engineering