A PROJECT REPORT

on

INSAAF INSIGHT: A LEGAL ASSISTANT

Submitted by

Ishaan Bhalla 20CSU364

Priyanshu Mishra 20CSU360

Mukul 20CSU355

Department of CSE

The NorthCap University

Gurgaon

A PROJECT REPORT
on
INSAAF INSIGHT: A LEGAL ASSISTANT

submitted in partial fulfillment of the requirement for the award of the degree

of

Bachelor of Technology
in Computer Science Engineering and Information Technology

by

Ishaan Bhalla 20CSU364
Priyanshu Mishra 20csu360
Mukul 20csu355

Under supervision of
Dr. Nitin Malik
Professor & Dy. Dean (PhD & RDIL)

Department of CSE & IT
The NorthCap University, Gurgaon
May 2024

CERTIFICATE

This is to certify that the Project Synopsis entitled, “Insaaf Insight” submitted by Ishaan Bhalla(20CSU364), Priyanshu Mishra(20CSU360), Mukul(20CSU355) to The NorthCap University, Gurugram, India, is a record of bona fide synopsis work carried out by them under my supervision and guidance and is worthy of consideration for the partial fulfilment of the degree of Bachelor of Technology in Computer Science and Engineering of the University.

Dr. Nitin Malik

(Professor & Dy. Dean (PhD & RDIL)

Date: 16/05/2024

ACKNOWLEDGEMENT

An undertaking is never a result of a solitary individual; rather it bears the engravings of various individuals who specifically or by implication helped in finishing that venture. We would bomb in my obligations on the off chance that we don't let out the slightest peep of gratitude to every one of the individuals who helped us in finishing this task of our own. Before we start with the details of my projects, we would like to add a few heartfelt words for the people who were part of our project in numerous ways, the people who gave us their immense support right from the initial stage. As a matter of first importance, we are amazingly appreciative of Dr. Nitin Malik for his direction, smooth feedback, and tutelage throughout this task. We also heartily thank our friends who greatly helped us in our project work without them we would never have gained the actual problem set solutions that we faced.

Ishaan Bhalla 20CSU364
Priyanshu Mishra 20CSU360
Mukul 20CSU355

ABSTRACT

In a time when legal intricacy is often intimidating, "Insaaf Insight" shows itself to be a ground-breaking platform that simplifies the acquisition of legal knowledge. By use of state-of-the-art artificial intelligence (AI) technologies, especially the Llama language model, "Insaaf Insight" provides clients with prompt access to precise, legally relevant information. This site tries to democratise legal knowledge by making it freely available and understandable to everyone, regardless of legal expertise.

The user-friendly interface and large library of statutes and legal precedents in Insaaf Insight enable users to quickly traverse the intricate legal system. Because the platform's AI-powered answers are based on a sizable dataset that was provided during the project's development, the data offered is guaranteed to be accurate and reliable. Furthermore according to stringent data privacy and security regulations, Insaaf Insight uses encrypted methods and safe databases to safeguard user data.

Comprehensive legal materials combined with cutting-edge AI raises legal literacy and enables people to decide with knowledge what their legal rights and responsibilities are. Leading the way in legal innovation, Insaaf Insight is creating a better educated and empowered society by bridging the gap between complex legal concepts and common knowledge.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

Navigating through the labyrinthine complexities of legislation and legal rights can be daunting. Our project “Insaaf Insight” hopes to make sense of them. In particular, we provide a one-stop platform for legal inquiries as well as a library of relevant legal precedence, hoping to democratize the law.

With the secure and extensive capabilities of the Llama language model, our application has two core features: the inquiry-based system and the curated list of precedence inquiries. The former allows the users to ask queries relevant their legal rights, and provides inquiries in turn; the latter allows them to also access historical judgements before the law.

Our project empowers individuals with the knowledge needed to confidently navigate the legal landscape, ensuring that legal information is both accessible and understandable.

1.1. Background

The Indian legal system represents a complicated set of rules, regulations, precedents. For most laymen it is often hard to navigate. "Insaaf Insight" meets this complexity head on. The hope is that ordinary people can have easy access to clear and layman terms explanations of legal knowledge, without the necessity to consult an actual lawyer. Just as the internet has revolutionized the way we gather information, so "Insaaf Insight" is about to revolutionize the way in which our knowledge of law is formulated. Combining innovative AI technology with a comprehensive legal database, we offer users live legal responses which are both accurate and timely. And offer them access to a huge library of legal precedent.

Our mission is to democratize legal knowledge – make it available to everyone, regardless of any background or legal talent. We believe that knowing one’s legal rights is not a luxury; it is rather a fundamental right. “Insaaf insight” removes roadblocks to legal data for its users. Our mission is to empower them with knowledge rather than relying on an intermediary for accessing that meaning.

We feel that the information we have gathered is enough for us to begin and that we have thought about the software’s functioning and purpose. In light of this, we anticipate that legal information systems will become even more integrated and available. We have a once-in-a-lifetime opportunity to develop premium, user-friendly legal services, and we would like Insaaf Insight to lead the way. While working on the project, we claim not to be better than legal agencies or qualified attorneys. We claim that we are distinct and that our goal is not only to replace formal legal representation but also to enhance the way they communicate with the law.

Considering how digital innovations have altered areas such as healthcare, education, and finance, increasing service accessibility and efficiency. Similarly, the legal profession is on the verge of a major transformation. The market for easily available legal information is large, with chances to improve legal literacy and empower people across the board.
As corporate executives and individuals start to wonder, "How can we ensure that legal knowledge is within everyone's reach?" and "What role does technology play in simplifying the legal landscape?" The need for tools such as "Insaaf Insight" becomes clear. These questions emphasize the need of democratizing legal knowledge and the disruptive impact that "Insaaf Insight" is positioned to provide.

1.2. Feasibility Study

Selecting the right technologies is crucial for the success of "Insaaf Insight." Our project involves real-time legal inquiries and access to a comprehensive library of legal precedents. By leveraging advanced AI integration, using the Llama language model for natural language processing, we ensure accurate and contextually relevant responses to user inquiries. The main technologies that will be used include:

1. Secure and Scalable Database
2. Interactive Query System
3. Comprehensive Precedent Library
4. API Integration

These technologies together create a robust platform for legal information. The secure and scalable database guarantees the safe storage of data and efficient retrieval, while the interactive query system facilitates easy user interaction. The comprehensive precedent library offers valuable legal references, and API integration broadens the platform's access to a wide array of legal information, enhancing its value to users.

1.3. Results

1. **Seamless User Experience:** The platform will offer a user-friendly interface for legal inquiries, ensuring a seamless experience. Users can interact with the AI-driven system to ask questions and receive precise legal information.
2. **Comprehensive Legal Information Access:** The entire system will provide access to a vast database of legal precedents and statutes. Users can search and reference past judicial decisions, enhancing their understanding and application of the law.
3. **Efficient Data Management and Retrieval:** The system will utilize a secure and scalable database to store and manage legal information. This ensures efficient data retrieval and management, supporting the platform's overall performance.
4. **Advanced AI Integration:** During the project development, a Large Language Model (LLM) was fed extensive legal data. This allows the AI to generate accurate and contextually relevant legal information based on the dataset provided.

Additionally, the survey conducted indicated the following benefits of using "Insaaf Insight" for accessing legal information:

- **Accuracy and Reliability:** Users appreciate the accuracy and reliability of the AI-driven responses, which provide clear and precise legal information.
- **Accessibility and Convenience:** The platform's user-friendly interface makes legal information easily accessible to a broad audience, including those without legal expertise.
- **Enhanced Legal Literacy:** By providing comprehensive legal information and resources, "Insaaf Insight" contributes to improving legal literacy among users.
- **Secure Data Handling:** Users value the secure handling of their data, ensuring privacy and protection of sensitive information.

Chapter 2

STUDY OF EXISTING SOLUTIONS

As per the our surveys and research, various applications exist in the market that offer users access to legal information and services. Some of the most notable applications are:

- MyAdvo
- LawRato
- IndiaFilings

All these applications provide users with legal services such as consultations, document drafting, and legal advice. They typically operate on a centralized platform where users can request services, and these services are delivered by legal professionals through the application.

2.1. Comparison with Existing Legal Information Platforms

The following are some general differences between the available apps in the market and Insaaf Insight:

Existing Apps	**Insaaf Insight**
An entity or company holds the authority to make major decisions regarding the platform and provides necessary services.	Utilizes advanced AI and a comprehensive legal dataset to autonomously provide legal information, reducing reliance on centralized human decision-making authority.
Information is stored on centralized servers.	Information is stored in a secure and scalable database, ensuring efficient data retrieval and management.
No integration with advanced AI for legal inquiries.	Uses advanced AI (Llama language model) trained on extensive legal data to provide accurate and contextually relevant legal information.
No comprehensive database of legal precedents.	Offers a searchable database of legal precedents, allowing users to reference past judicial decisions easily.
Searched done manually by the people	Search of Keyboard done automatically by the LLM's
Search about precedence is done manually.	This is done by LLM.
Less accurate and precision.	Better accuracy and precision

fig 2.1.a (Comparison between Existing Apps & Insaaf Insight)

2.2. Key Advantages of Insaaf Insight

- Enhanced Accessibility: Insaaf Insight provides a user-friendly interface, making legal information easily accessible to a broad audience, including those without legal expertise.
- Improved Legal Literacy: By offering comprehensive legal information and resources, Insaaf Insight helps improve legal literacy among users.
- Accurate and Reliable Information: The AI-driven system provides accurate and reliable responses to legal inquiries, ensuring users receive clear and precise legal information.
- Data Privacy and Security: Insaaf Insight maintains data privacy and security by using secure database systems with encryption protocols, ensuring that user data is protected from unauthorized access and breaches. Regular security audits and compliance with data protection regulations further enhance the platform's security measures.

Chapter 3

GAP ANALYSIS

3.1. Current State

- **Interactive Legal Query System**: Insaaf Insight operates a foundational AI-driven system that utilizes the Llama language model to provide basic legal information. This system is currently in a nascent stage, offering essential, somewhat scratchy responses to user inquiries and is not yet fully functional.

- **Comprehensive Legal Database**: The platform hosts a substantial but limited database of legal precedents and statutes, which is searchable and accessible to users. This database is foundational and supports the initial functionality of the query system.

- **User Interface and Accessibility**: The platform features a user-friendly interface that is operational and accessible to users, facilitating straightforward interactions with the AI system, albeit with limited features.

3.2. Final State (Vision)

- **Expansion of Legal Database**: The goal is to significantly expand the legal database to include comprehensive coverage of national and international laws, catering to a broader range of legal inquiries from users worldwide.

- **Advanced Data Analytics**: Implementing sophisticated analytics to provide insights into user interaction patterns, legal trends, and common inquiries, thereby enhancing the personalization and accuracy of responses.

- **Enhanced User Experience**: Enhancements to the user interface to include multilingual support, thus making the platform accessible to non-English speaking users. Further development of mobile applications to provide on-the-go access to legal information.

- **Increase the Precision**: Use Ensemble Method, combining multiple retrieval models in a weighted manner to leverage the strengths of various approaches (e.g., combining VSM, BM25, and semantic models).

- **Community and Professional Collaboration**: Developing a more vibrant community platform by integrating forums directly within the site and collaborating with legal professionals for expert insights and updates to the legal database.

- **Comprehensive Security Enhancements**: Strengthening data protection measures to ensure user data privacy and security, including compliance with data protection regulations.

3.3. GAPs Identified

- **Database Expansion**: There is a need to broaden the scope of the current database to cover more diverse legal areas and jurisdictions to cater to a global audience.

- **Data Analytics**: Currently, the platform does not utilize data analytics extensively to refine and personalize user experiences based on interaction data.

- **User Interface Enhancements**: The interface requires upgrades to support additional languages and improve mobile access, essential for reaching a broader audience.

- **Community Engagement and Professional Collaboration**: While initial steps have been taken to establish a user base, more active strategies are needed to foster a dynamic community and engage with legal professionals actively.

- **Security Measures**: As the platform scales and handles more sensitive information, continuous enhancements to security protocols and compliance measures will be vital.

Chapter 4

PROBLEM STATEMENT

In the legal domain, the accessibility and analysis of relevant statutes and case precedents are fundamental to informed decision-making and effective legal reasoning. Traditional methods of legal research are often time-consuming and require extensive manual effort, which can delay legal proceedings and decision-making processes. Despite advancements in technology, the legal profession continues to face significant challenges in efficiently navigating the vast amounts of legal texts and case law.

The Insaaf Insight project addresses these challenges through two innovative solutions:

- **Chat-based Legal Assistant:** The first component of the project aims to revolutionize how legal professionals and the public access and interact with statutory law, particularly the Indian Penal Code (IPC) and other significant documents. The assistant utilizes advanced natural language processing (NLP) techniques and large language models (LLMs) to offer real-time, conversational legal guidance. This system is designed to understand and respond to queries by embedding and indexing vast amounts of legal text, thus providing accurate and contextually relevant information. However, the challenge lies in accurately processing, embedding, and retrieving legal texts in a way that faithfully represents their content and relevance to various legal inquiries.

- **AI-driven Precedence Library:** The second component focuses on enhancing the ability to identify and retrieve relevant case precedents quickly and accurately. This system leverages text analytics and machine learning models, such as BM25 and Doc2Vec, to automate the search for pertinent legal precedents based on textual similarity. The primary challenge is developing an effective system that can interpret the nuances of legal language and deliver highly relevant precedents that can aid in legal reasoning and decision-making.

Both components of the Insaaf Insight project are designed to enhance the accessibility, efficiency, and accuracy of legal research. By integrating these systems, the project seeks to provide a comprehensive tool that supports legal professionals in delivering faster and more informed legal services. However, the effectiveness of these systems hinges on the successful application of complex AI models and NLP techniques to a domain as intricate and sensitive as law, presenting a unique set of technical and ethical challenges.

Chapter 5

OBJECTIVES

5.1. General Objectives for "Insaaf Insight":

1. **Enhance Legal Research Efficiency**: Significantly reduce the time and effort required for legal professionals and the public to access and interpret legal information.
2. **Improve Accessibility of Legal Resources**: Make legal texts, especially complex statutes and case precedents, more accessible and understandable to non-experts.
3. **Foster Informed Legal Decision-Making**: Provide users with reliable, contextually relevant legal information to support more informed decision-making in legal contexts.
4. **Innovate Legal Information Delivery**: Introduce cutting-edge technology into the legal field to modernize how legal information is processed and delivered.

5.2. Specific Objectives for the Chat-based Legal Assistant:

1. **Develop a User-Friendly Interface**: Create an intuitive chat-based interface that allows users to interact easily with the legal assistant.
2. **Implement Advanced NLP Techniques**: Utilize state-of-the-art NLP models like Llama 3 8b to process and understand complex legal language in real-time.

3. **Ensure Real-Time Response Generation**: Design the system to provide immediate and accurate responses to user inquiries based on legal texts.
4. **Contextualize User Interactions**: Enable the system to understand the context of user queries and provide responses that are not only legally accurate but also tailored to the specific needs of the user.
5. **Maintain High Standards of Privacy and Security**: Ensure that all user data and interactions are handled with the highest standards of privacy and security.

5.3. Specific Objectives for the AI-driven Library of Legal Precedents:

1. **Automate Case Law Retrieval**: Develop algorithms to automate the retrieval of relevant case precedents quickly and accurately.
2. **Incorporate Comprehensive Text Analysis**: Apply text analysis methods such as TF-IDF and Vector Space Models to enhance the relevance and accuracy of search results.
3. **Evaluate System Performance**: Measure the performance of the retrieval system using metrics such as precision, recall, and relevance to ensure it meets the needs of the user.
4. **Enable Seamless Integration**: Ensure that the system integrates smoothly to provide a unified resource for legal research.

Chapter 6

TOOLS/ PLATFORM USED

6.1. For the Chat-based Legal Assistant:

1. **Programming Languages and Frameworks:**
   - Python: Used for scripting and automating tasks such as text processing, data loading, and embedding.
   - Java: Employed for backend services, integrating with databases, and possibly interfacing with the chat system.
2. **Natural Language Processing (NLP) Libraries:**
   - Hugging Face Transformers: Provides pre-trained models and utilities for embedding generation and natural language understanding, for Llama-3-8b.
   - FAISS (Facebook AI Similarity Search): Used for efficient similarity search and retrieval of documents based on embeddings.
3. **Data Handling and Storage:**
   - Pandas and NumPy: For data manipulation and numerical operations within Python.
4. **IDE, Design &collaboration platforms**

- VS Code: A source-code editor developed by Microsoft for Windows, Linux, macOS and web browsers.
- Miro: Digital collaboration platform designed to facilitate remote and distributed team communication and project management.

6.2. For the AI-driven Library of Legal Precedents:

1. **Text Processing and Machine Learning Libraries:**
   - Scikit-learn: For implementing machine learning algorithms like TF-IDF, and for general data processing and model evaluation.
   - Gensim: For more advanced document similarity models such as Doc2Vec.
   - Rank_bm25: A library specific for implementing the BM25 algorithm for ranking and retrieval tasks.
2. **Data Preparation and Visualization Tools:**
   - Matplotlib and Seaborn: For data visualization to analyze the dataset and results of model evaluations.
   - Jupyter Notebook: For interactive development and testing of data processing and machine learning models.
3. **Information Retrieval Models:**
   - Vector Space Model (VSM): Utilized to represent documents and queries as vectors in a multi-dimensional space, facilitating the ranking of documents based on their relevance to queries using cosine similarity.
4. **IDE, Design &collaboration platforms**

- PyCharm: IDE used for programming in Python. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems, and supports web development with Django.
- Miro: Digital collaboration platform designed to facilitate remote and distributed team communication and project management.

Chapter 7

DESIGN METHODOLOGY

Design methods are procedures, techniques, aids, or tools for designing. They offer a number of different kinds of activities that a designer might use within an overall design process. Conventional procedures of design, such as drawing, can be regarded as design methods, but since the 1950s new procedures have been developed that are more usually grouped together under the name of "design methods". What design methods have in common is that they "are attempts to make public the hitherto private thinking of designers; to externalise the design process.

Design methodology is the broader study of method in design: the study of the principles, practices and procedures of designing.

Design Methodology used: -

AGILE SCRUM

Most Large language models or AI books contain very little about design methodology. For our project, Agile Scrum is ideal due to its adaptability and iterative nature. The project begins by gathering requirements from stakeholders, including legal experts and potential users, to create a prioritized product backlog. This backlog guides the team on what features and tasks are most important. During each sprint, which typically lasted 2-3 hours a week, the team selected high-priority items from the backlog to focus on. Daily stand-up meetings were held to discuss progress, address obstacles, and ensure everyone is aligned.

Agile Scrum’s flexibility allows Insaaf Insight to adapt quickly to changes in legal requirements and user feedback. This adaptability is crucial for a legal information platform, where accuracy and relevance are paramount. The iterative approach also integrates continuous testing throughout the development process, allowing the team to identify and resolve issues early. This leads to a more stable and reliable platform. Overall, Agile Scrum supports the dynamic nature of the Insaaf Insight project. By promoting ongoing improvement and fostering a user-centric development process, it ensures that the platform not only meets but exceeds user expectations. The methodology’s emphasis on collaboration, regular feedback, and adaptability makes it well-suited for the ever-evolving landscape of legal information, ultimately empowering users with the knowledge they need to navigate legal complexities confidently.

Chapter 8

METHODOLOGY

8.1. Chat Based Legal Assistant:

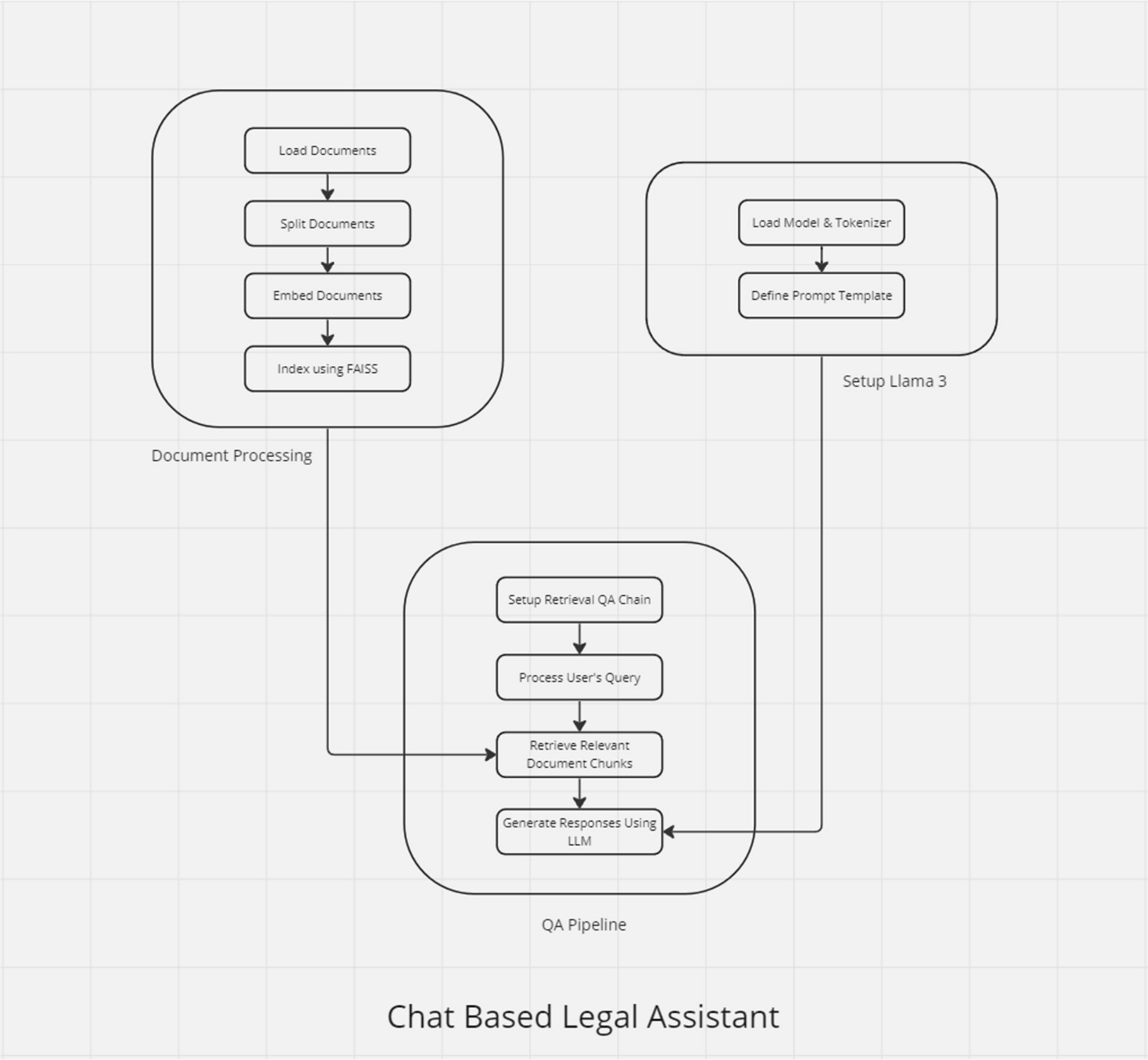


fig 8.1.a (Flow of Chat-based Legal Assistant)

Implements a chat-based legal assistant that utilizes a combination of natural language processing (NLP) techniques and large language models (LLMs) to provide guidance on Indian law, leveraging the Indian Penal Code (IPC) and 14 additional documents. Here's a detailed breakdown of how it works:

8.1.1 Steps Involved in the implementation of Chat-Based Legal Assistant

Embedding and Indexing the Documents

1. **Dependencies Importation**: Initially, it imports necessary libraries and modules for handling text data, loading documents, embedding text, and indexing.

2. **Dataset and Vector Store Setup**: It specifies paths for the dataset and the location to save the FAISS index.

3. **Embed Function**: This function handles the loading, splitting, embedding, and indexing of legal documents. It follows these steps:

   - **Document Loading**: Utilizes **DirectoryLoader** to load all **.pdf** documents from the specified dataset directory.

   - **Document Splitting**: Employs **RecursiveCharacterTextSplitter** to split documents into manageable chunks, facilitating more efficient processing and embedding.

   - **Embedding**: Uses **HuggingFaceEmbeddings** to generate embeddings for each document chunk. These embeddings represent the semantic content of the chunks in a high-dimensional space.

- **Indexing with FAISS**: Creates a FAISS vector store from the document embeddings for efficient similarity search. This vector store is then saved for later retrieval.

Setting Up the LLM Used in Chat-based Legal Assistant

1. **Model and Tokenizer Loading**: Loads a pre-trained LLM (**meta-llama/Llama-2-7b-chat-hf**) and its corresponding tokenizer. This model is configured for text generation, specifically for handling conversational contexts.

2. **Custom Prompt Template**: Defines a custom prompt template that instructs the model on how to handle user queries, including how to incorporate context into its responses.

3. **Retrieval QA Chain Creation**: Constructs a retrieval-based question-answering chain (**RetrievalQA**) that integrates the LLM with document retrieval. This chain:

   - Uses the FAISS index for document retrieval, identifying relevant legal texts based on the user's query.

   - Applies the custom prompt template, ensuring the model's responses are aligned with the instructions.

   - Generates responses based on both the retrieved context and the large language model's capabilities.

Execution Flow

1. **Document Processing**: The **embed_all** function is called to process the legal documents, creating an indexed vector store.

2. **QA Pipeline Setup**: The **qa_pipeline** function initializes the entire QA system, encompassing:

    - Embedding loading and FAISS index preparation.

    - LLM and custom prompt loading.

    - Assembly of the retrieval QA chain.

How the Chat-based Assistant Works

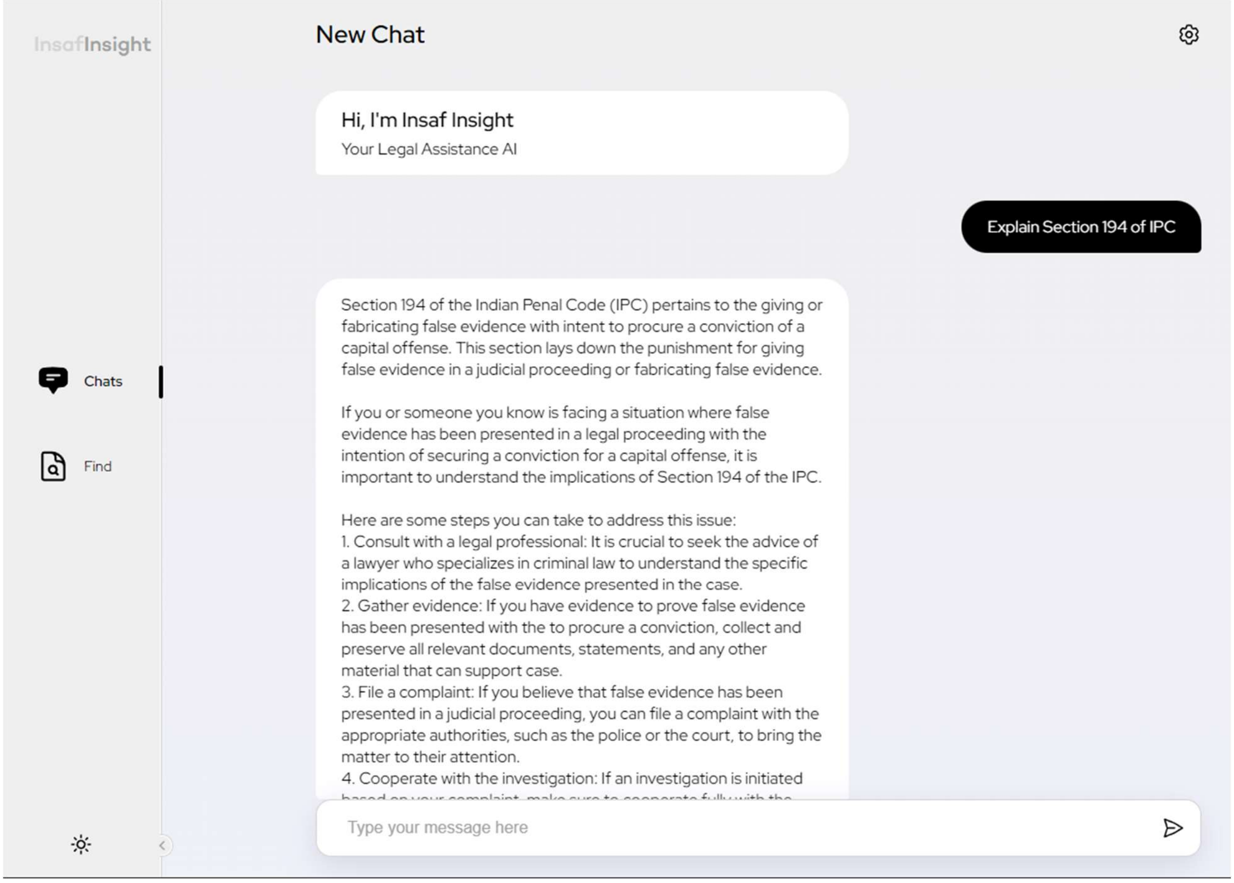
- When a user inputs a query, the system retrieves relevant document chunks from the FAISS index based on semantic similarity to the query.

- The retrieved context and the user's question are fed into the LLM, using the custom prompt template to guide the model's response.

- The model generates a response, leveraging both the large language model's knowledge and the specific legal context retrieved, providing accurate and contextually relevant legal guidance.

This approach combines the strengths of retrieval-based and generative AI models to create a sophisticated legal assistant capable of addressing complex queries with contextually relevant information.

8.1.2 Chat Window



Hi, I'm Insaf Insight

Your Legal Assistance AI

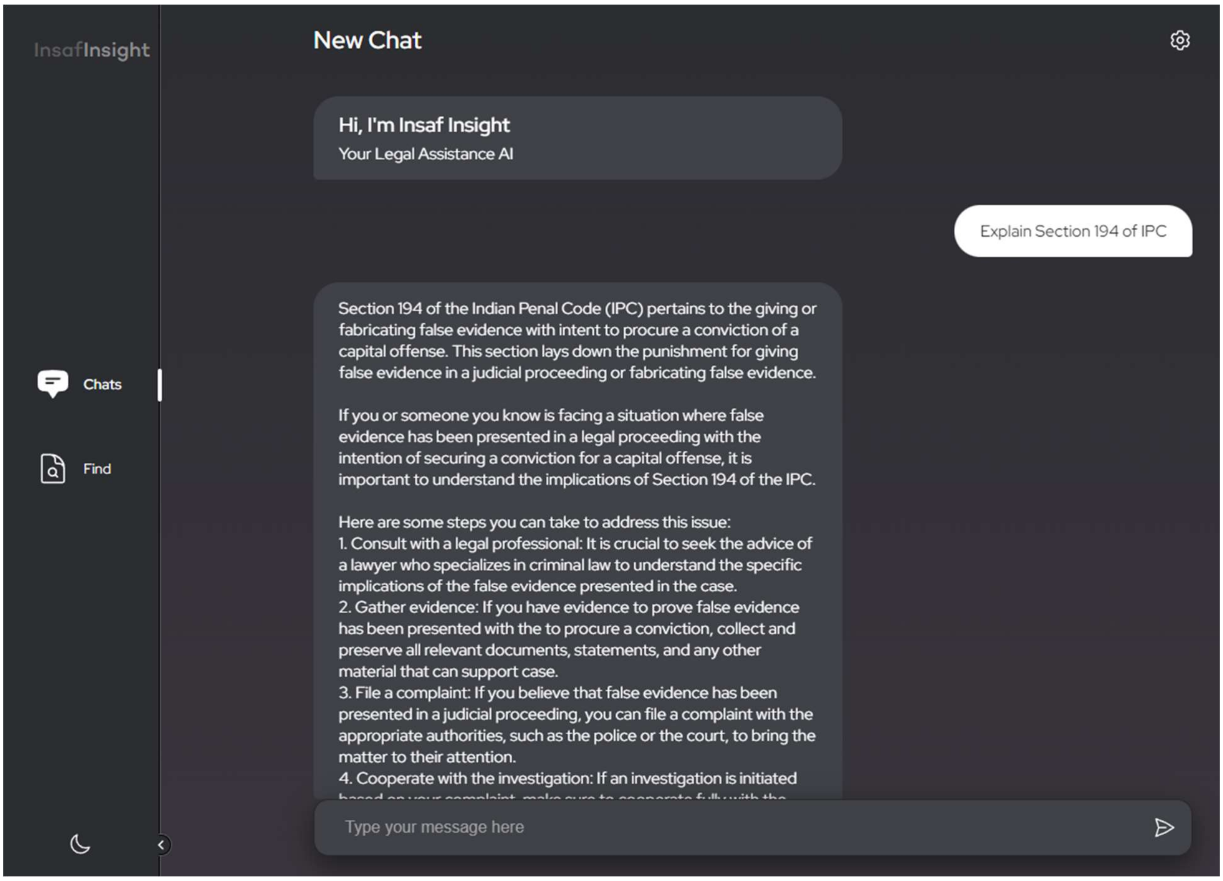
Explain Section 194 of IPC

Section 194 of the Indian Penal Code (IPC) pertains to the giving or fabricating false evidence with intent to procure a conviction of a capital offense. This section lays down the punishment for giving false evidence in a judicial proceeding or fabricating false evidence.

If you or someone you know is facing a situation where false evidence has been presented in a legal proceeding with the intention of securing a conviction for a capital offense, it is important to understand the implications of Section 194 of the IPC.

Here are some steps you can take to address this issue:
1. Consult with a legal professional: It is crucial to seek the advice of a lawyer who specializes in criminal law to understand the specific implications of the false evidence presented in the case.
2. Gather evidence: If you have evidence to prove false evidence has been presented with the to procure a conviction, collect and preserve all relevant documents, statements, and any other material that can support case.
3. File a complaint: If you believe that false evidence has been presented in a judicial proceeding, you can file a complaint with the appropriate authorities, such as the police or the court, to bring the matter to their attention.
4. Cooperate with the investigation: If an investigation is initiated based on your complaint, make sure to cooperate fully with the

Fig 8.1.2.a (Chat window, requesting information on a particular IPC Section, light mode)



Hi, I'm Insaf Insight
Your Legal Assistance AI

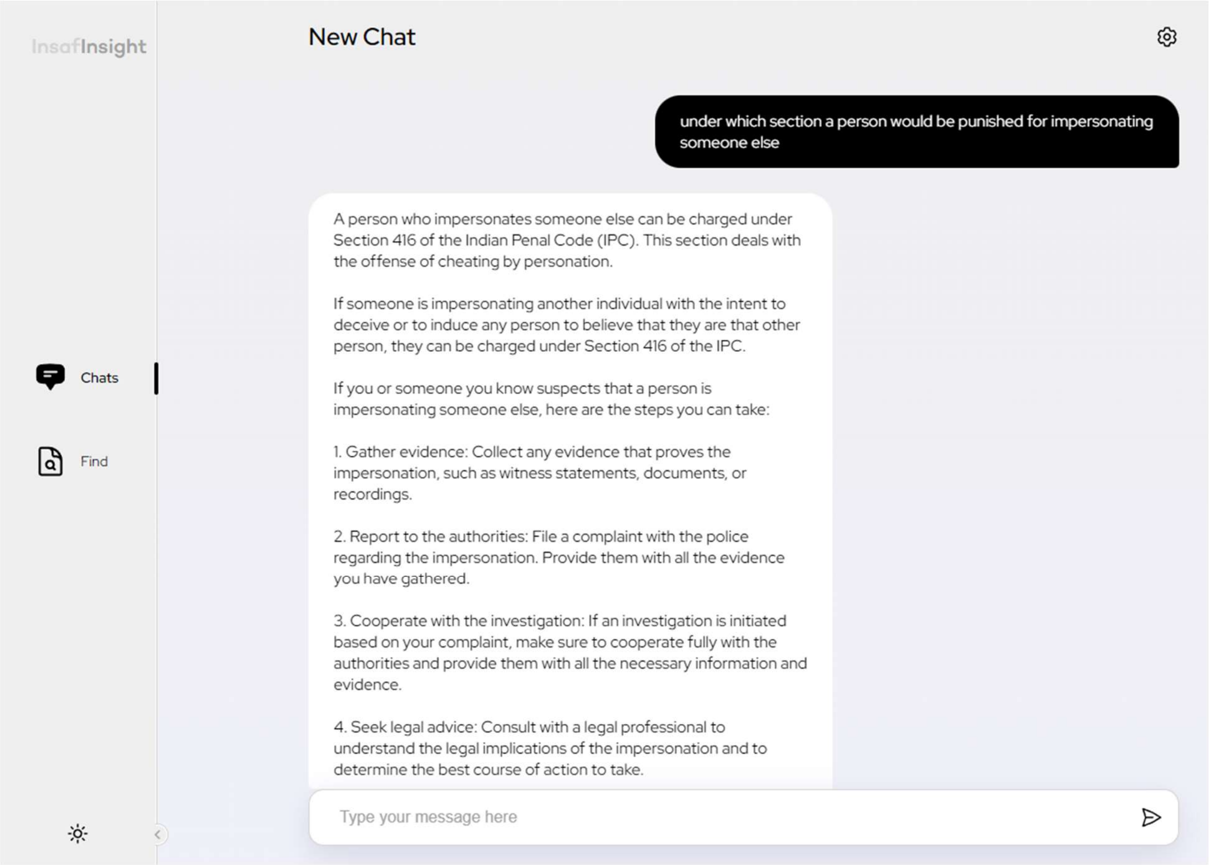
Explain Section 194 of IPC

Section 194 of the Indian Penal Code (IPC) pertains to the giving or fabricating false evidence with intent to procure a conviction of a capital offense. This section lays down the punishment for giving false evidence in a judicial proceeding or fabricating false evidence.

If you or someone you know is facing a situation where false evidence has been presented in a legal proceeding with the intention of securing a conviction for a capital offense, it is important to understand the implications of Section 194 of the IPC.

Here are some steps you can take to address this issue:
1. Consult with a legal professional: It is crucial to seek the advice of a lawyer who specializes in criminal law to understand the specific implications of the false evidence presented in the case.
2. Gather evidence: If you have evidence to prove false evidence has been presented with the to procure a conviction, collect and preserve all relevant documents, statements, and any other material that can support case.
3. File a complaint: If you believe that false evidence has been presented in a judicial proceeding, you can file a complaint with the appropriate authorities, such as the police or the court, to bring the matter to their attention.
4. Cooperate with the investigation: If an investigation is initiated

Fig 8.1.2.b (Chat window, requesting information on a particular IPC Section, dark mode)



under which section a person would be punished for impersonating someone else

A person who impersonates someone else can be charged under Section 416 of the Indian Penal Code (IPC). This section deals with the offense of cheating by personation.

If someone is impersonating another individual with the intent to deceive or to induce any person to believe that they are that other person, they can be charged under Section 416 of the IPC.

If you or someone you know suspects that a person is impersonating someone else, here are the steps you can take:

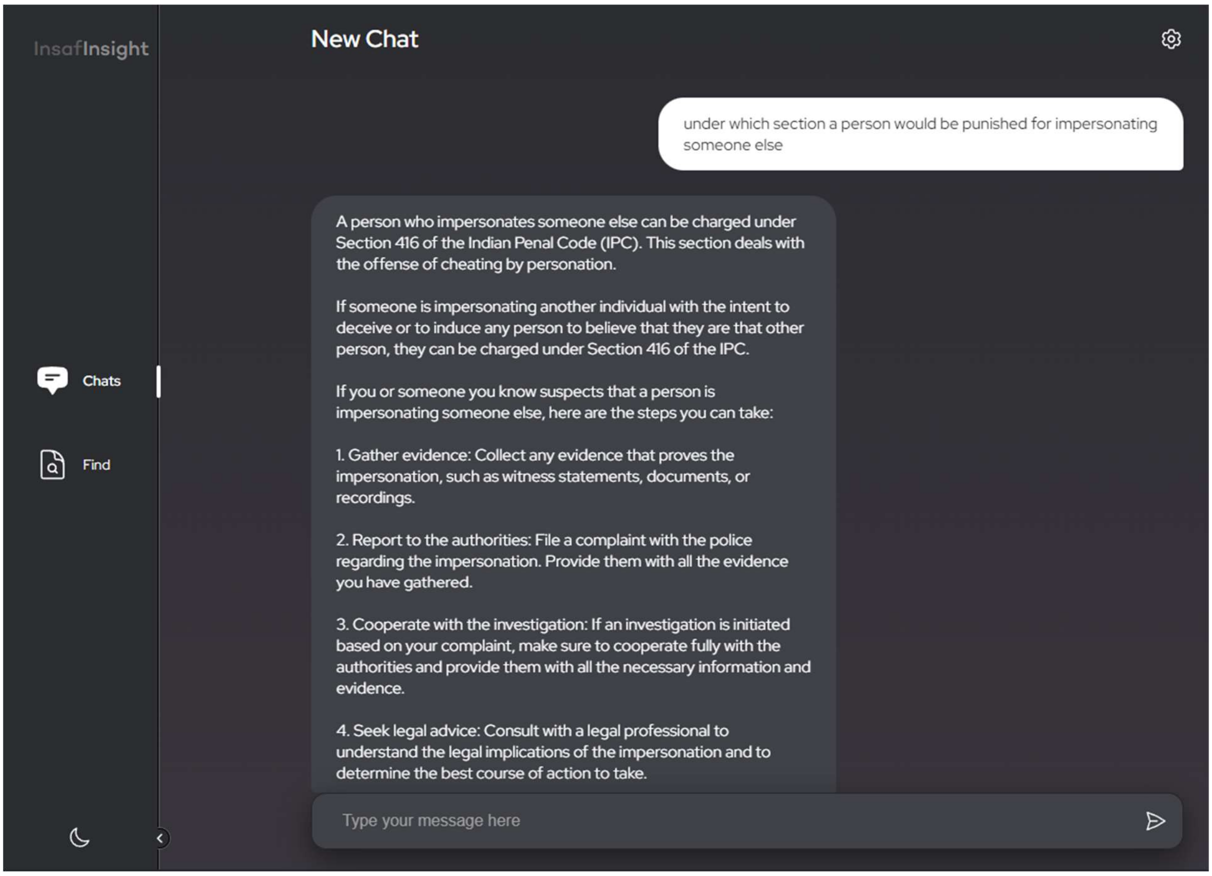
1. Gather evidence: Collect any evidence that proves the impersonation, such as witness statements, documents, or recordings.

2. Report to the authorities: File a complaint with the police regarding the impersonation. Provide them with all the evidence you have gathered.

3. Cooperate with the investigation: If an investigation is initiated based on your complaint, make sure to cooperate fully with the authorities and provide them with all the necessary information and evidence.

4. Seek legal advice: Consult with a legal professional to understand the legal implications of the impersonation and to determine the best course of action to take.

Fig 8.1.2.c (Chat Window, asking for an IPC Section that would be applied for a particular scenario, light mode)



under which section a person would be punished for impersonating someone else

A person who impersonates someone else can be charged under Section 416 of the Indian Penal Code (IPC). This section deals with the offense of cheating by personation.

If someone is impersonating another individual with the intent to deceive or to induce any person to believe that they are that other person, they can be charged under Section 416 of the IPC.

If you or someone you know suspects that a person is impersonating someone else, here are the steps you can take:

1. Gather evidence: Collect any evidence that proves the impersonation, such as witness statements, documents, or recordings.

2. Report to the authorities: File a complaint with the police regarding the impersonation. Provide them with all the evidence you have gathered.

3. Cooperate with the investigation: If an investigation is initiated based on your complaint, make sure to cooperate fully with the authorities and provide them with all the necessary information and evidence.

4. Seek legal advice: Consult with a legal professional to understand the legal implications of the impersonation and to determine the best course of action to take.

Fig 8.1.2.d (Chat Window, asking for an IPC Section that would be applied for a particular scenario, dark mode)

8.2 AI-driven Library of Legal Precedents:

8.2.1 Approach One:

In legal studies, the ability to locate relevant case precedents quickly and accurately is essential for effective legal reasoning and decision-making. This approach leverages modern text processing and machine learning techniques to automate the search and retrieval of pertinent legal documents from a large dataset of case files. The process begins with aggregating and preparing textual data from thousands of court case documents, followed by a series of sophisticated text preprocessing steps to cleanse and normalize the data.

Utilizing advanced natural language processing (NLP) methods, the text is refined through stemming, lemmatization, and removal of stopwords to enhance its suitability for analysis. The cleansed data is then used to train document similarity models—specifically BM25 and Doc2Vec—which serve to rank and retrieve documents based on their relevance to a given legal query. These models are evaluated on their effectiveness through precision and recall metrics, ensuring that the most relevant documents are accurately identified.

This approach not only streamlines the process of legal research but also enhances the accessibility and usability of legal documents, thereby supporting legal professionals in making more informed decisions based on relevant past cases. The combination of text analytics and machine learning opens new avenues for enhancing the efficiency and accuracy of legal research, transforming traditional methods with the power of technology.

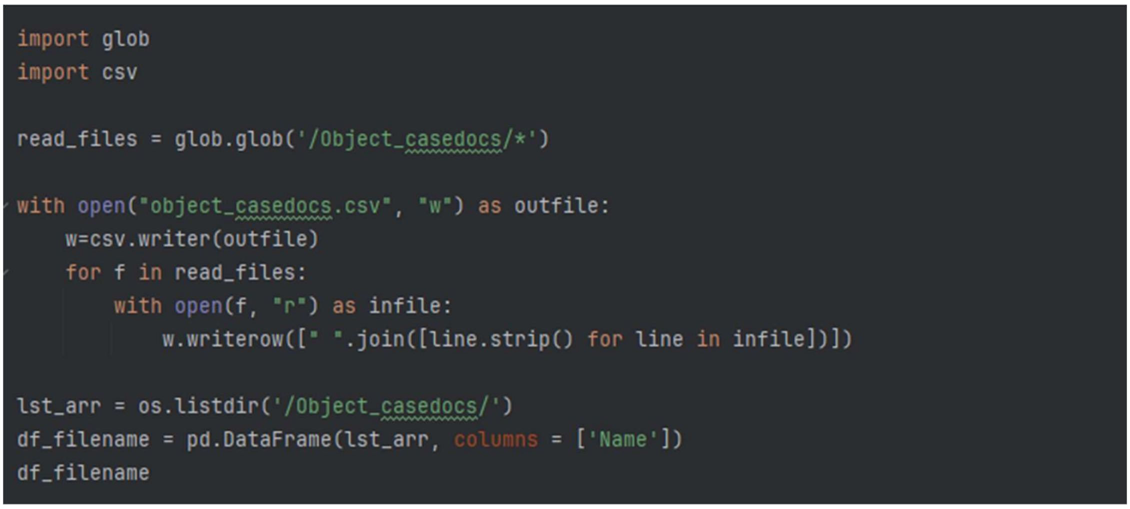
Steps involved in this approach:

Step 1: Environment Setup and Imports

- Importing libraries necessary for handling data (like pandas, NumPy), manipulating files (glob), and visualizing data (matplotlib, seaborn).

Step 2: Data Aggregation

- Combining individual text files containing case documents into a single CSV file for easier processing, storing the content of each document as a row.

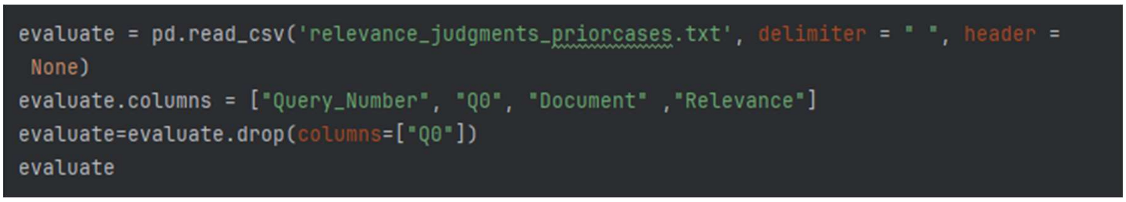


```
import glob
import csv

read_files = glob.glob('/Object_casedocs/*')

with open("object_casedocs.csv", "w") as outfile:
    w=csv.writer(outfile)
    for f in read_files:
        with open(f, "r") as infile:
            w.writerow([" ".join([line.strip() for line in infile])])

lst_arr = os.listdir('/Object_casedocs/')
df_filename = pd.DataFrame(lst_arr, columns = ['Name'])
df_filename
```

Step 3: Data Preparation

- Loads a separate dataset containing relevance judgments for the documents.
- Reads the combined CSV file into a DataFrame, setting appropriate column names.
- Concatenates the DataFrame with filenames and document texts, creating a unified DataFrame.

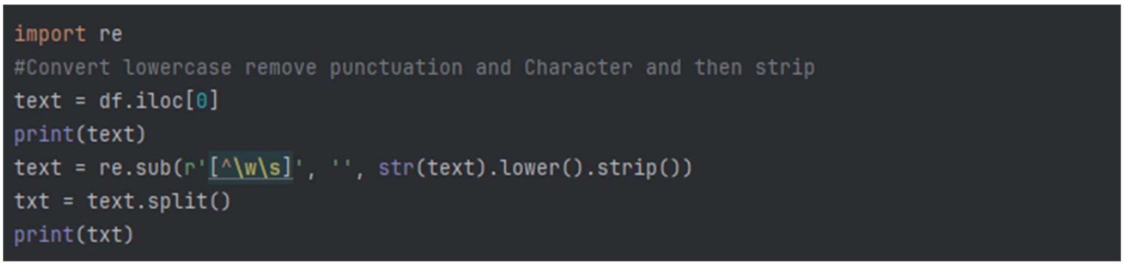
- Checks the total number of documents processed by examining the length of the DataFrame.
- Outputs the shape of the DataFrame, providing insight into the number of rows (documents) and columns (attributes).
- Provides a summary of the DataFrame to review data types and check for missing values.



```
evaluate = pd.read_csv('relevance_judgments_priorcases.txt', delimiter = " ", header =
 None)
evaluate.columns = ["Query_Number", "Q0", "Document" ,"Relevance"]
evaluate=evaluate.drop(columns=["Q0"])
evaluate
```

Step 4: Preliminary Text Processing

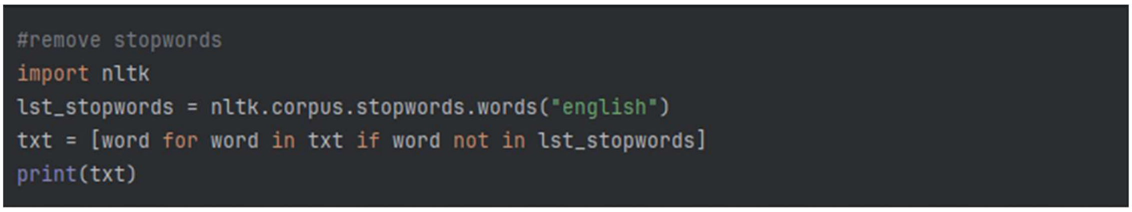
- Performs initial text cleaning on a sample document, including converting to lowercase, removing punctuation, and splitting into words.
- Removes common stop words from the sample text to focus on more meaningful words.



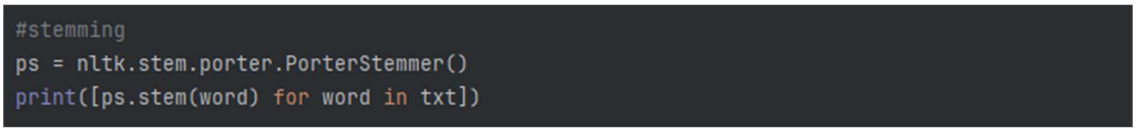
```
import re
#Convert lowercase remove punctuation and Character and then strip
text = df.iloc[0]
print(text)
text = re.sub(r'[^\w\s]', '', str(text).lower().strip())
txt = text.split()
print(txt)
```

Step 5: Applying Text Processing to Entire Dataset

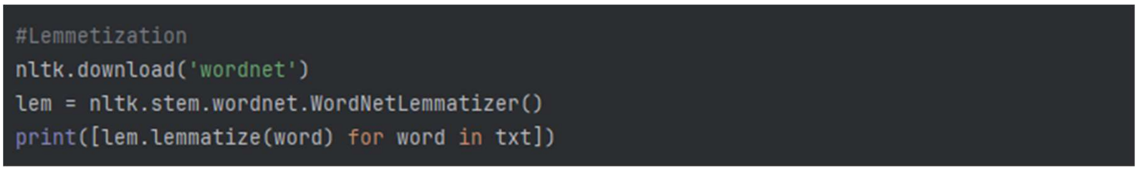
- Defines a utility function that combines cleaning, tokenization, stop words removal, stemming, and lemmatization. Applies this function to all documents in the DataFrame, storing the results in a new column for cleaned text.



```
#remove stopwords
import nltk
lst_stopwords = nltk.corpus.stopwords.words("english")
txt = [word for word in txt if word not in lst_stopwords]
print(txt)
```



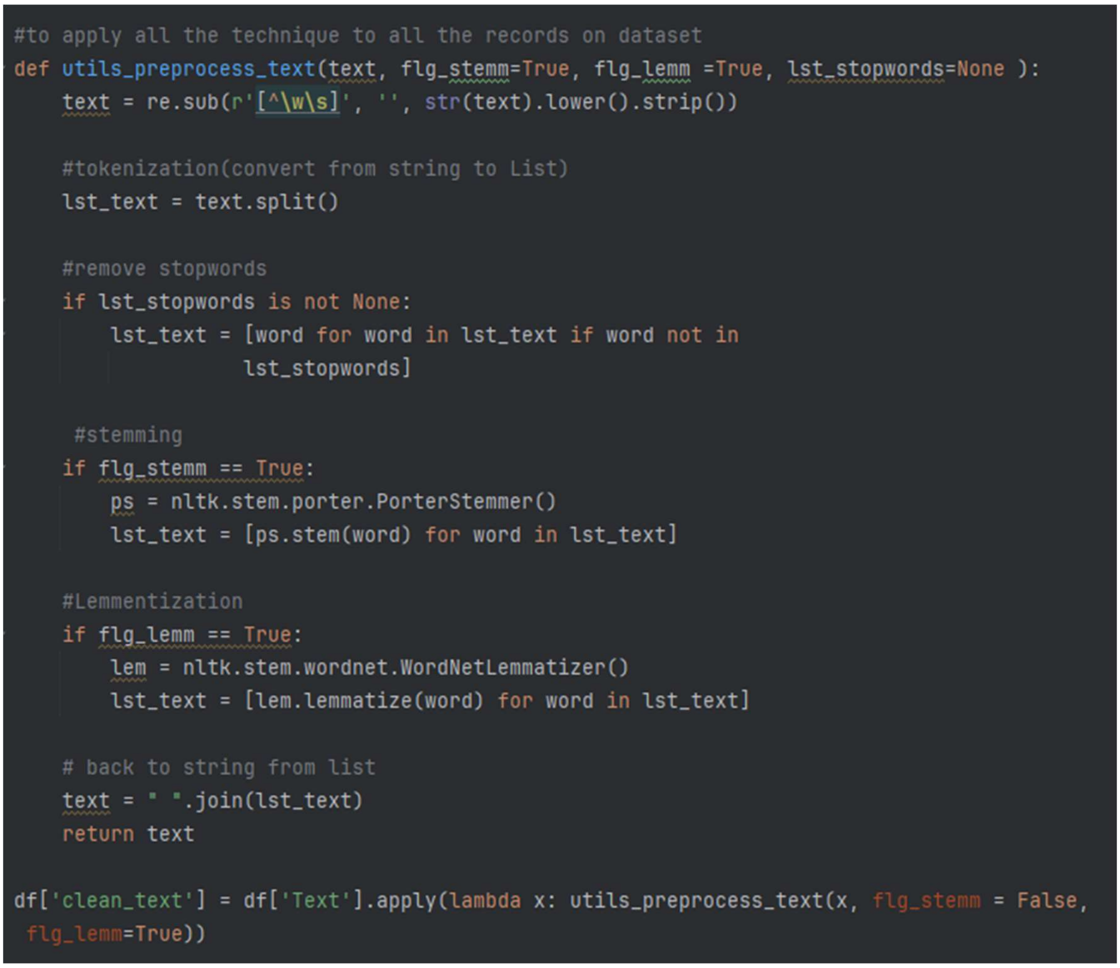
```
#stemming
ps = nltk.stem.porter.PorterStemmer()
print([ps.stem(word) for word in txt])
```



```
#Lemmetization
nltk.download('wordnet')
lem = nltk.stem.wordnet.WordNetLemmatizer()
print([lem.lemmatize(word) for word in txt])
```

Step 6: Further Data Preparation for Modeling

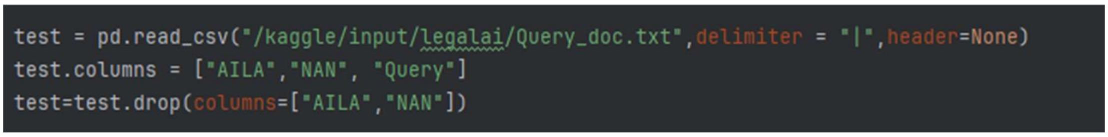
- Prints the DataFrame to confirm the addition of the cleaned text.
- Prepares a training set from the cleaned text.



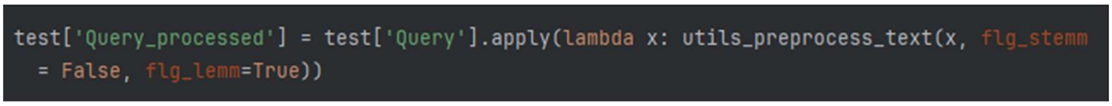
```
#to apply all the technique to all the records on dataset
def utils_preprocess_text(text, flg_stemm=True, flg_lemm =True, lst_stopwords=None ):
    text = re.sub(r'[^\w\s]', '', str(text).lower().strip())

    #tokenization(convert from string to List)
    lst_text = text.split()

    #remove stopwords
    if lst_stopwords is not None:
        lst_text = [word for word in lst_text if word not in
                    lst_stopwords]

     #stemming
    if flg_stemm == True:
        ps = nltk.stem.porter.PorterStemmer()
        lst_text = [ps.stem(word) for word in lst_text]

    #Lemmentization
    if flg_lemm == True:
        lem = nltk.stem.wordnet.WordNetLemmatizer()
        lst_text = [lem.lemmatize(word) for word in lst_text]

    # back to string from list
    text = " ".join(lst_text)
    return text

df['clean_text'] = df['Text'].apply(lambda x: utils_preprocess_text(x, flg_stemm = False,
 flg_lemm=True))
```

Step 7: Additional Libraries and Test Data Preparation

- Imports the texthero library, although not utilized in the visible code.
- Loads and processes a test dataset containing queries for document retrieval.



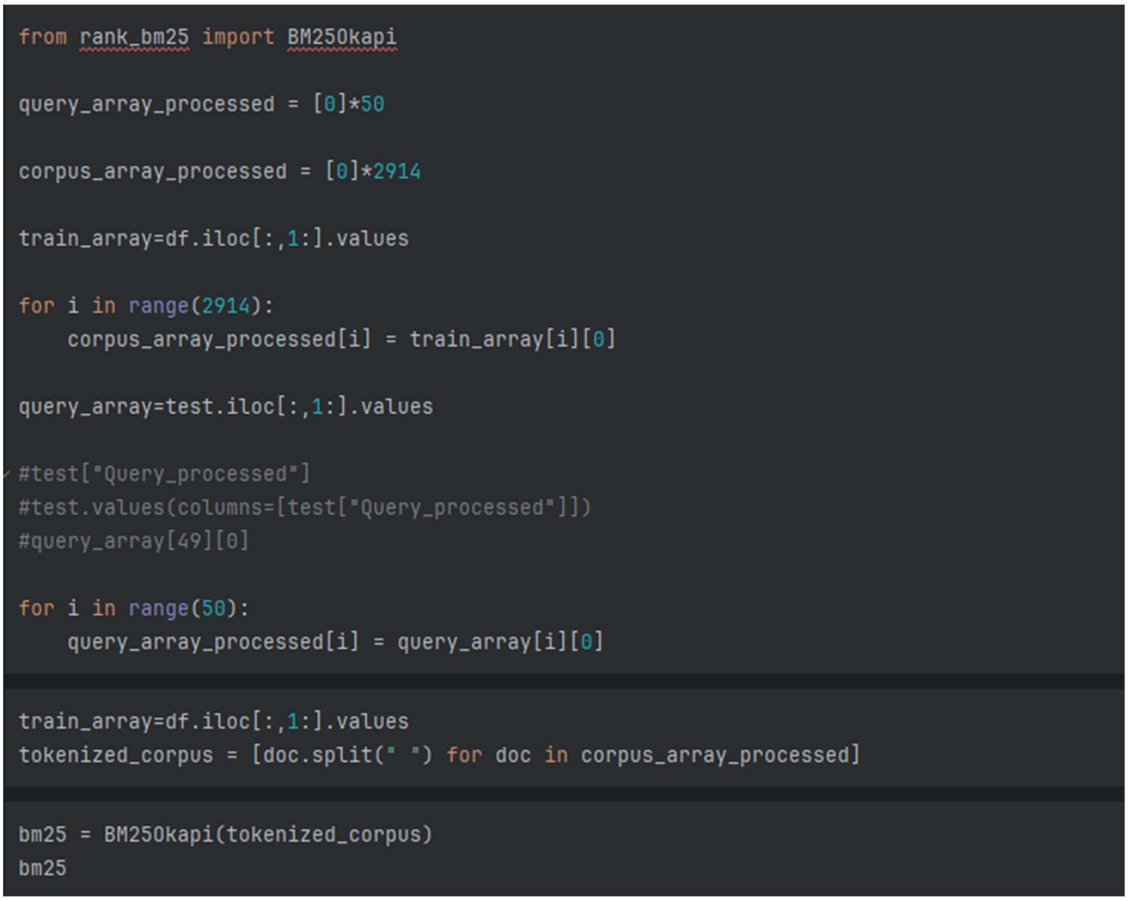
```
test = pd.read_csv("/kaggle/input/legalai/Query_doc.txt",delimiter = "|",header=None)
test.columns = ["AILA","NAN", "Query"]
test=test.drop(columns=["AILA","NAN"])
```



```
test['Query_processed'] = test['Query'].apply(lambda x: utils_preprocess_text(x, flg_stemm
  = False, flg_lemm=True))
```

Step 8: Document Similarity and Retrieval

- Texts are processed for queries.
- Installing and importing the rank_bm25 library, setting up the BM25 model.
- Tokenizes the corpus and queries, applies BM25 to find the most relevant documents for the queries.
- Retrieves and prints the top relevant documents for a sample query.



```
from rank_bm25 import BM25Okapi

query_array_processed = [0]*50

corpus_array_processed = [0]*2914

train_array=df.iloc[:,1:].values

for i in range(2914):
    corpus_array_processed[i] = train_array[i][0]

query_array=test.iloc[:,1:].values

#test["Query_processed"]
#test.values(columns=[test["Query_processed"]])
#query_array[49][0]

for i in range(50):
    query_array_processed[i] = query_array[i][0]
```

```
train_array=df.iloc[:,1:].values
tokenized_corpus = [doc.split(" ") for doc in corpus_array_processed]
```

```
bm25 = BM25Okapi(tokenized_corpus)
bm25
```

Step 9: Evaluation of Retrieval Effectiveness

- Calculating the precision and recall of the BM25 model.

8.2.2 Approach two:

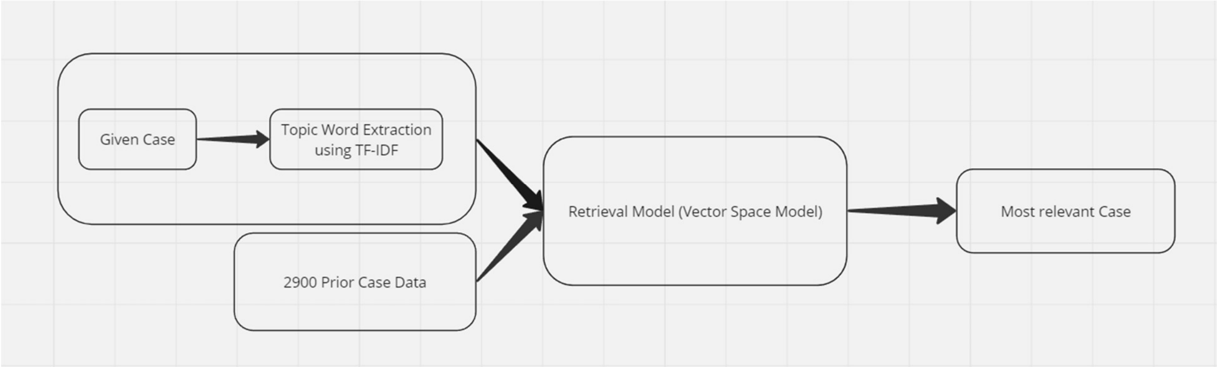

Given Case
Topic Word Extraction using TF-IDF
2900 Prior Case Data
Retrieval Model (Vector Space Model)
Most relevant Case

Fig 8.2.2.a (Flow of Precedence retrieval)

Firstly, the topic extraction module employs the Term Frequency-Inverse Document Frequency (TF-IDF) technique to identify and extract the key topics from a given case. Subsequently, the top 'k' topic words, selected based on their relevance, are formulated into a query. This query is then processed by the retrieval model, known as the Vector Space Model (VSM). The VSM evaluates and ranks the cases by their similarity to the query. The top 'n' cases, which are ranked highest in relevance by the model, are considered to be the most pertinent prior cases.

Step 1: Topic Extraction

The method begins by extracting key topic words from the description of a current legal situation. This is achieved through two main techniques:

TF-IDF (Term Frequency-Inverse Document Frequency): This technique measures the importance of a word within a document in relation to a collection of documents. It helps to highlight words that are distinctive in a given document.

- Term Frequency (TF): This measures the frequency of a word in a document, indicating its importance within that specific document.
- Inverse Document Frequency (IDF): This measures the importance of the word across a set of documents. The idea is that the more documents the word appears in, the less distinguishing it is. Therefore, IDF decreases the weight for commonly used words and increases the weight for words that are not used very much across documents.
- The TF-IDF score is the product of these two metrics, aiming to balance the frequency of terms while considering how common they are across all documents. Words with high TF-IDF scores are selected as topic words because they are both prevalent in a particular document and rare across all documents, making them significant for that document.

Step 2: Query Formation

The extracted topic words are treated as queries. These queries are used to search through the dataset to find relevant cases. This mirrors how a user might input search terms into a system to find relevant information.

Step 3: Information Retrieval Models

Several retrieval models are used to identify and rank prior cases based on their relevance to the query terms:

Vector Space Model (VSM):

- Represents both the query and documents as vectors in a multi-dimensional space, where each dimension corresponds to a term from the document set, weighted by TF-IDF.
- Cosine Similarity: The relevance of a document to a query is determined by the cosine of the angle between their vector representations. A smaller angle or higher cosine value indicates higher similarity.

Step 4: Evaluation

The approach was evaluated using the standard metrics, Mean Average Precision (MAP) and Precision at K.

8.2.3 Results of both Approaches:

Method	P@10	MAP
Approach 1	0.027	0.0984
Approach 2	0.091	0.2432

fig 8.2.3.a (p@10 (Precision at 10) & MAP (Mean Average Precision))

Although these numbers are low, this shows the difficulty of retrieving legal documents due to their high complexity. Accuracy is low but it is much preferable to use an automated system that can retrieve approximately 1 relevant case in the top 10 instead of manual labor to find relevant precedents.

To increase the precision, we can use Ensemble Method, combining multiple retrieval models in a weighted manner to leverage the strengths of various approaches (e.g., combining VSM, BM25, and semantic models).

Chapter 9

CHALLENGES AND ISSUES IDENTIFIED

During the development of the Insaaf Insight project, several challenges and issues emerged, impacting various aspects of the platform. These challenges have required careful consideration and strategic planning to address effectively. Here’s a detailed list of these challenges:

1. **Data Management and Scalability**
   - Challenge: The platform initially struggled with managing and scaling the vast amount of legal data needed for the comprehensive database.
   - Impact: This affected the platform's ability to quickly and efficiently process user queries and provide accurate legal information.

2. **AI Integration and Response Accuracy**
   - Challenge: Integrating the Llama language model effectively and ensuring the accuracy of AI-generated responses was more complex than anticipated.
   - Impact: Initially led to less reliable legal advice, which could undermine user trust and platform credibility.

3. **User Interface Usability**
   - Challenge: Designing a user interface that is intuitive and accessible to a diverse user base, including those without legal or technical expertise.

- Impact: Potential barriers to user engagement and platform adoption, especially among less tech-savvy users.

4. **Legal Data Comprehensiveness**
    - Challenge: Ensuring the legal database is comprehensive and up to date, covering various jurisdictions and legal areas.
    - Impact: Incomplete coverage could limit the usefulness of the platform for certain users or in specific legal scenarios.

5. **Data Privacy and Security**
    - Challenge: Maintaining rigorous data privacy and security standards to protect sensitive user information.
    - Impact: Vulnerabilities could lead to data breaches, negatively affecting user trust and legal compliance.

Chapter 10

PERFORMANCE EVALUATION

1. Precision of the Models

Metrics: Precision at 10, Mean Average Performance

Output:

Method	P@10	MAP
Approach 1	0.027	0.0984
Approach 2	0.091	0.2432

fig 10.a (p@10 & MAP of AI-driven Library of Legal Precedents)

Although these numbers are low, this shows the difficulty of retrieving legal documents due to their high complexity. Accuracy is low but it is much preferable to use an automated system that can retrieve approximately 1 relevant case in the top 10 instead of manual labor to find relevant precedents.

To increase the precision, we can use Ensemble Method, combining multiple retrieval models in a weighted manner to leverage the strengths of various approaches (e.g., combining VSM, BM25, and semantic models).

2. **Preliminary Legal Information Accuracy**

   Metrics: Accuracy of responses based on initial expert reviews.

   Outputs: A compilation of expert feedback on a sample of responses provided by the AI. This helped to determine how well the AI understands and responds to legal queries and where improvements are needed in terms of data accuracy and relevance.

3. **Database Development Progress**

   Metrics: Volume and variety of legal data integrated into the database.

   Tangible Outputs: A development log that lists all legal documents and resources added to the database. This log will serve as a clear indicator of how the database is growing and diversifying to cover more legal areas and jurisdictions.

4. **Data Security Initial Setup**

   Metrics: Implementation of basic security measures and compliance with data protection laws.

   Outputs: A security setup report detailing the implemented security measures, such as encryption protocols, secure access controls, and data anonymization practices. This report will also highlight any compliance checks or initial security audits conducted.

Chapter 11

OUTCOME

The "Insaaf Insight" project, encompassing both a chat-based legal assistant and an AI-driven library of legal precedents, has achieved several significant outcomes since its implementation. These outcomes reflect the project's success in leveraging advanced technology to enhance legal research and accessibility. Here’s an overview of the key achievements:

1. Enhanced Legal Research Efficiency

The project significantly reduced the time required to access and interpret legal information. Legal professionals and students can now retrieve statutes and case precedents in a matter of seconds rather than hours, dramatically increasing productivity.

2. Improved Legal Accessibility

The chat-based legal assistant has made legal information more accessible to the general public, including those without a legal background. This tool simplifies complex legal terminology and provides clear, concise explanations, helping individuals understand their legal rights and responsibilities more effectively.

2. Increased Accuracy and Relevance in Legal Assistance

Thanks to the sophisticated NLP capabilities of the Llama 3 8b model, the chat-based assistant provides highly accurate and contextually relevant responses. This accuracy is further enhanced by continuous learning mechanisms that refine the model's responses based on new information and user feedback.

3. Streamlined Case Law Retrieval

The AI-driven library of legal precedents utilizes advanced text processing algorithms like TF-IDF and vector space models to efficiently sort through vast amounts of data. This system has streamlined the process of finding relevant case precedents, making legal research more efficient and reducing the likelihood of overlooking important cases.

4. Technological Integration and Scalability

The project's infrastructure, designed with scalability in mind, successfully handles increased loads and integrates smoothly with existing legal databases and systems. This has facilitated a more comprehensive and interconnected legal research environment.

5. Contribution to Legal Education and Awareness

"Insaaf Insight" has been integrated into educational programs, serving as a practical tool for law students. The platform's real-time feedback and interactive nature have made it a valuable educational resource, enhancing students' learning experiences by providing direct engagement with real-world legal documents.

6. Broadening the Scope of Legal Technology

"Insaaf Insight" has set a new standard for legal technology, demonstrating the potential for AI and machine learning to transform traditional practices. This has spurred interest in further technological advancements in the legal field, paving the way for future innovations.

7. Economic Impact

By reducing the time and resources spent on legal research, "Insaaf Insight" has also provided economic benefits, making legal services more cost-effective and accessible, particularly for individuals and small firms with limited resources.

8. Comprehensive Legal Precedent Library

Insaaf Insight maintains a robust library of legal precedents, continually updated to reflect current rulings and notable cases. This resource allows users to access and study past legal decisions to better understand how laws have been interpreted and applied.

Having such a resource available enhances users' ability to conduct thorough legal research and supports academic, professional, and personal legal inquiries. This library serves as a critical tool for those preparing legal documents or seeking to understand the complexities of legal applications in various scenarios.

Chapter 12

GANTT CHART

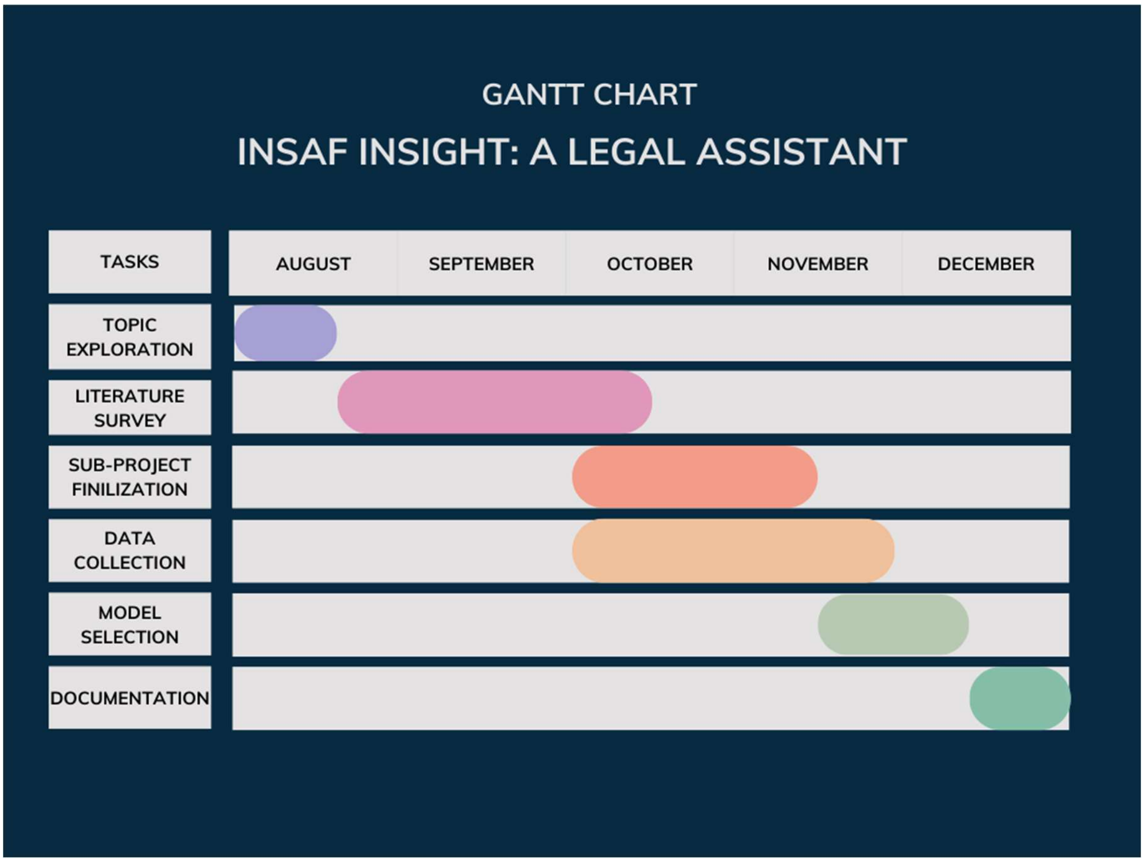


fig 12.a (Gantt Chart)

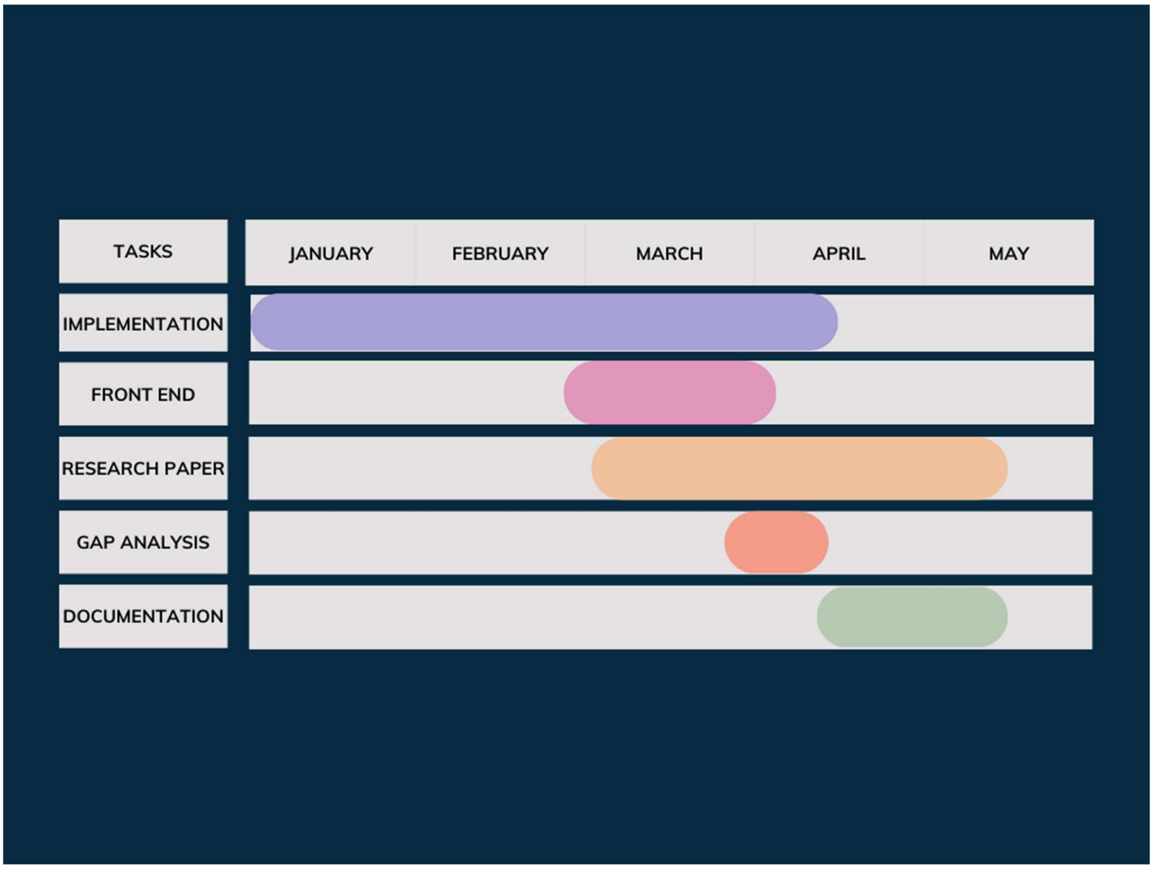
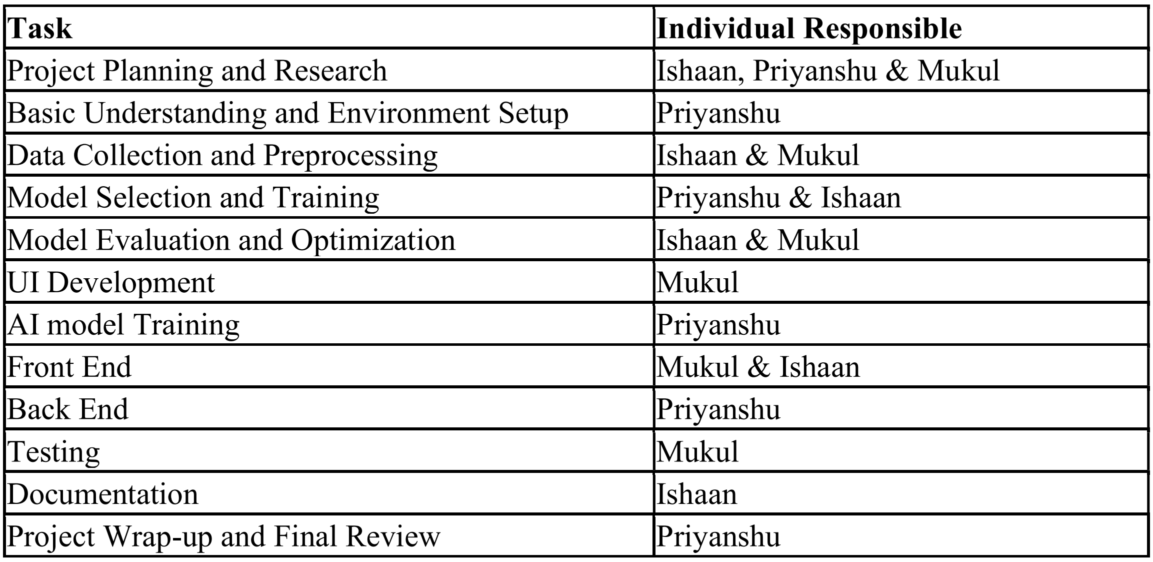


fig 12.b (Gantt Chart)

Chapter 13

RESPONSIBILITY CHART



Task	Individual Responsible
Project Planning and Research	Ishaan, Priyanshu & Mukul
Basic Understanding and Environment Setup	Priyanshu
Data Collection and Preprocessing	Ishaan & Mukul
Model Selection and Training	Priyanshu & Ishaan
Model Evaluation and Optimization	Ishaan & Mukul
UI Development	Mukul
AI model Training	Priyanshu
Front End	Mukul & Ishaan
Back End	Priyanshu
Testing	Mukul
Documentation	Ishaan
Project Wrap-up and Final Review	Priyanshu

fig 13.a (Responsibility Chart)

Chapter 14

REFERENCES

[1] Mohit Sharma. (2023). India's Courts and Artificial Intelligence: A Future Outlook. ResearchGate.https://www.researchgate.net/publication/377062808_India's_Courts_and_Artificial_Intelligence_A_Future_Outlook

[2] Muskan Shokeen, Vinit Sharma. (2023) Artificial intelligence and criminal justice system in India: A critical study, Lawjournals, www.lawjournals.net/assets/archives/2023/vol5issue4/5123.pdf

[3] Bhupatiraju, Sandeep; Chen, Daniel L.; and Joshi, Shareen (2021) "THE PROMISE OF MACHINE LEARNING FOR THE COURTS OF INDIA," National Law School of India Review: Vol. 33: Iss. 2, Article 10. Available at: https://repository.nls.ac.in/nlsir/vol33/iss2/10

[4] Responsible Artificial Intelligence for the Indian Justice System, (2021), Vidhi | Center for Legal Policy, tcg crest | Investing Harmonious Future, https://vidhilegalpolicy.in/wp-content/uploads/2021/04/Responsible-AI-in-the-Indian-Justice-System-A-Strategy-Paper.pdf

[5] Department of Justice, 'End of year review', (Ministry of Law & Justice, 31 December 2020), accessed on 15 March, 2021; and Justice D.Y. Chandrachud, 'Future of virtual courts and access to justice in India', (Nyaya Forum, NALSAR, online webinar 24 May, 2020)

[6] Robertson, Stephen, and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. Foundations and Trends® in Information Retrieval 3.4 (2009): 333-389.

[7] Ogilvie, Paul, and Jamie Callan. Experiments using the Lemur toolkit. TREC. Vol. 10. 2001: 103-108.

[8] Allan Dafoe, 'AI Governance: A research agenda' (Centre for the Governance of AI, University of Oxford, 2018)

[9] AI can improve judicial system's efficiency — full text of CJI Bobde's Constitution Day speech' (ThePrint, 27 November 2019)

[10] In the Indian context, AI ethics would primarily flow from the Constitution. See for AI ethics in India, Anna Roy, Rohit Satish & Tanay Maindru, 'Responsible AI: Approach document for India - part I principles for responsible AI' (NITI Aayog, February 2021)

[11] Lexix plus AI - https://www.lexisnexis.com/en-us/products/lexis-plus-ai.page

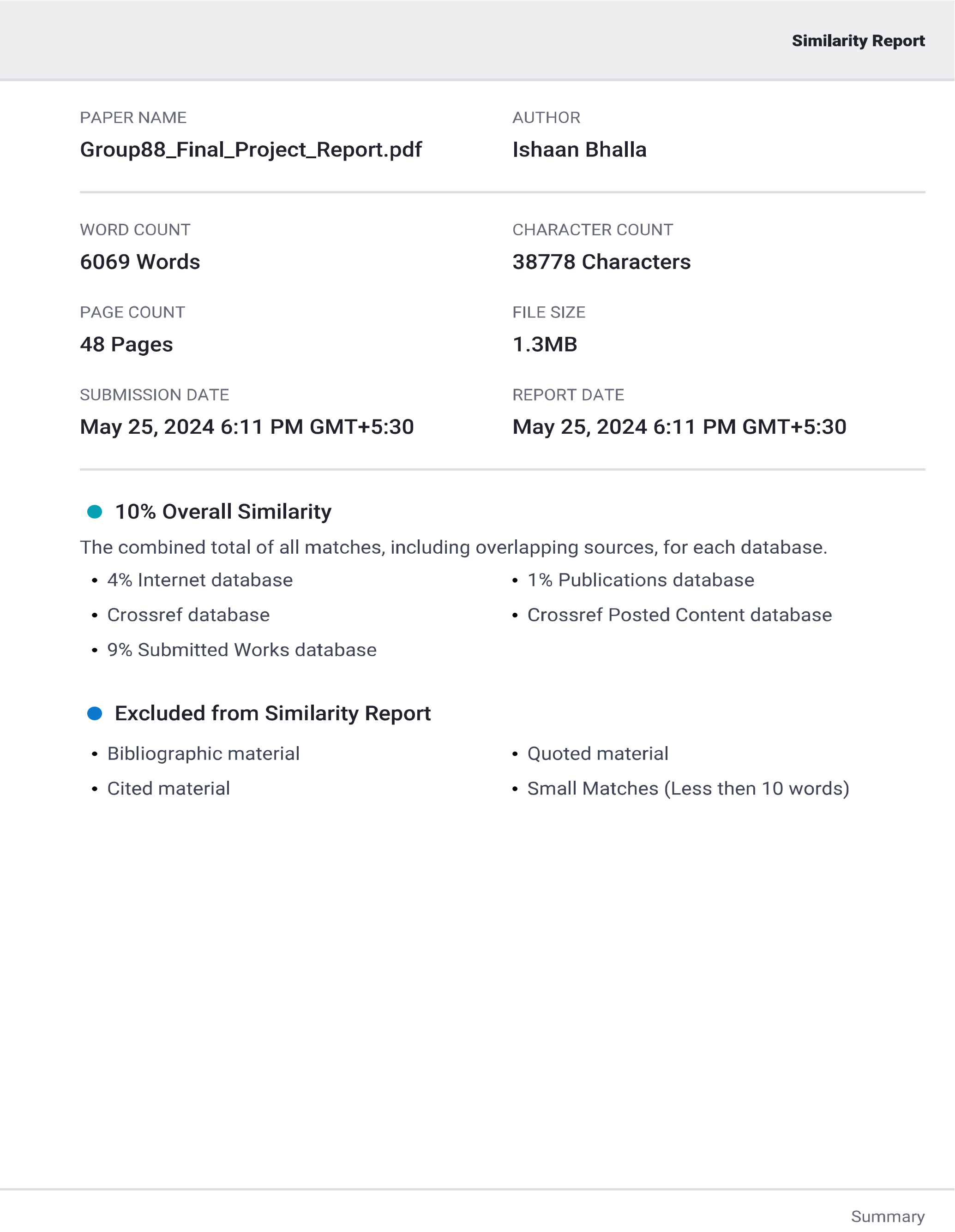
[12] Bhattacharya, P., Hiware, K., Rajgaria, S., Pochhi, N., Ghosh, K., Ghosh, S.: A Comparative Study of Summarization Algorithms Applied to Legal Case Judgments. In: Advances in Information Retrieval – Proceedings of European Conference on Information Retrieval (ECIR). pp. 413–428 (2019)

Chapter 15

HANDWRITTEN COMMENT FROM GUIDE

Chapter 16

Plagiarism Report



Similarity Report

PAPER NAME
Group88_Final_Project_Report.pdf

AUTHOR
Ishaan Bhalla

WORD COUNT
6069 Words

CHARACTER COUNT
38778 Characters

PAGE COUNT
48 Pages

FILE SIZE
1.3MB

SUBMISSION DATE
May 25, 2024 6:11 PM GMT+5:30

REPORT DATE
May 25, 2024 6:11 PM GMT+5:30

● 10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

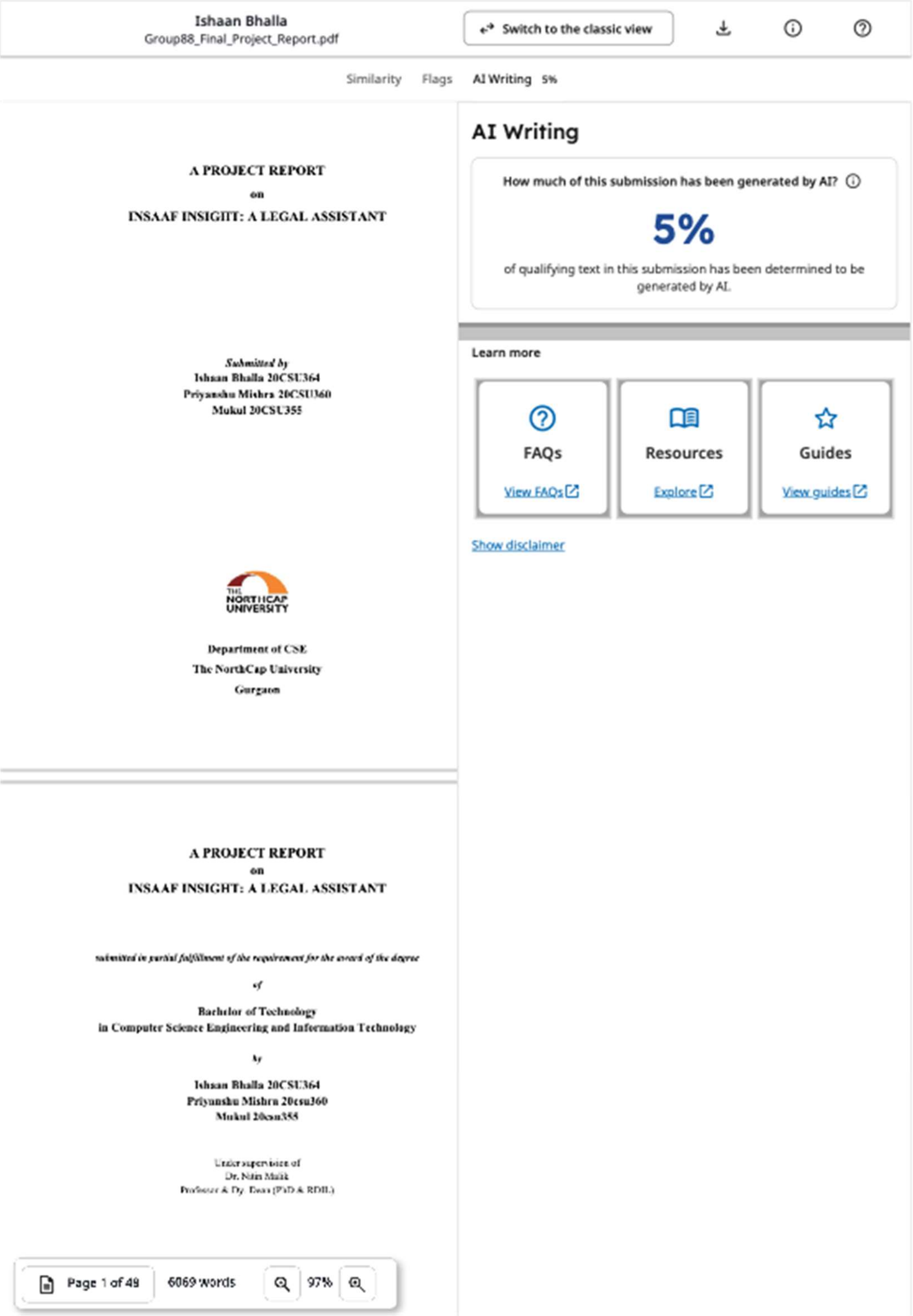
- 4% Internet database
- Crossref database
- 9% Submitted Works database
- 1% Publications database
- Crossref Posted Content database

● Excluded from Similarity Report

- Bibliographic material
- Cited material
- Quoted material
- Small Matches (Less then 10 words)

Summary

fig 16.a (Plagiarism Report)



A PROJECT REPORT
on
INSAAF INSIGHT: A LEGAL ASSISTANT

Submitted by
Ishaan Bhalla 20CSU364
Priyanshu Mishra 20CSU360
Mukul 20CSU355

THE NORTHCAP UNIVERSITY

Department of CSE
The NorthCap University
Gurgaon

A PROJECT REPORT
on
INSAAF INSIGHT: A LEGAL ASSISTANT

submitted in partial fulfillment of the requirement for the award of the degree

of

Bachelor of Technology
in Computer Science Engineering and Information Technology

by

Ishaan Bhalla 20CSU364
Priyanshu Mishra 20csu360
Mukul 20csu355

Under supervision of
Dr. Nitin Malik
Professor & Dy. Dean (PhD & RDIL)

AI Writing

How much of this submission has been generated by AI?

5%

of qualifying text in this submission has been determined to be generated by AI.

fig 16.b (Plagiarism Report)