



RAJASTHAN POLICE

HACKATHON 1.0

Problem Statement 4:
Analysis of FIR using AI/ML for proper Act and Section

TABLE OF CONTENTS

- INTRODUCTION
- MOTIVATION
- APPROACH
- LIBRARIES
- DATASET
- BERT MODEL
- TRAINING
- TESTING
- MODEL DEPLOYMENT
- FEATURES
- FUTURE SCOPES

INTRODUCTION:

- The problem statement, "Analysis of FIR using AI/ML for proper Act and Section," addresses the need for an efficient and accurate system to analyze First Information Reports (FIRs) using Artificial Intelligence (AI) and Machine Learning (ML) techniques to determine the appropriate legal Act and Section.
- This initiative stands at the intersection of legal expertise and technological innovation, poised to redefine how we approach the crucial task of deciphering legal complexities within First Information Reports.
- FIRs are processed, ensuring a more efficient, objective, and precise categorization of offenses.

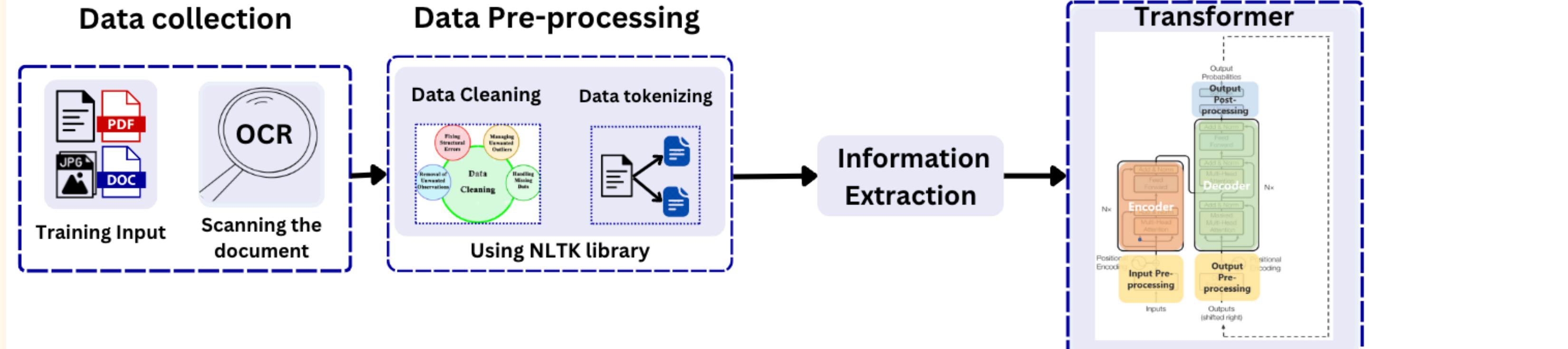
MOTIVATION:

- Handling FIRs can be a complex and time-consuming process due to the multitude of available IPC sections.
- The decision-making involved in choosing the appropriate section is intricate and prone to inaccuracies. Given the extensive range of IPC sections, determining the most suitable one poses a considerable challenge.
- Ensuring the proper categorization of Acts and Sections is crucial for legal compliance. AI/ML algorithms can help automate this process
- It also represents an opportunity to innovate and apply cutting-edge technologies. This project has allowed us to explore and contribute to advancements in artificial intelligence and machine learning, showcasing our skills and creativity.

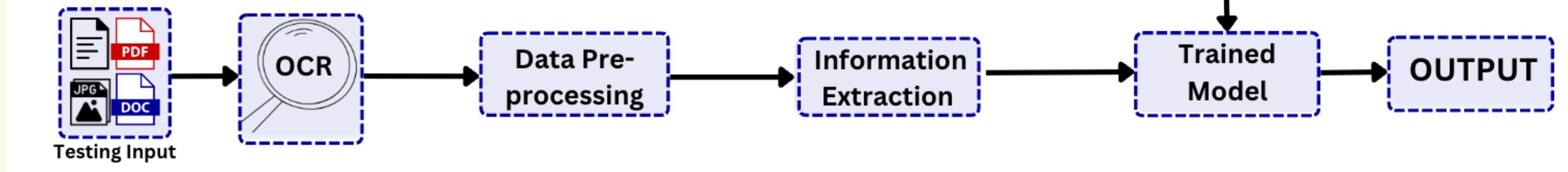
APPROACH:

HOW WE APPROACHED TO THIS PROBLEM?

Model Training:



Testing of Model:



LIBRARIES:

Libraries For Training Phase:

```
# Import necessary libraries
import tensorflow as tf
from transformers import BertTokenizer, TFBertForSequenceClassification
from tensorflow.keras.optimizers import AdamW
from tensorflow.keras.losses import SparseCategoricalCrossentropy
from tensorflow.keras.metrics import SparseCategoricalAccuracy
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np
import json
```

LIBRARIES:

PyMuPDF:

- Usage: It enables you to work with PDF files, such as reading, writing, and extracting information from PDF documents.

tensorflow:

- Usage: It provides tools for various machine learning tasks, including neural network development and deployment.

transformers:

- Usage: It allows you to easily use and fine-tune pre-trained language models for tasks like text classification, named entity recognition, and more.

pytesseract:

- Usage: It's handy for tasks where you need to extract text content from images or scanned documents.

LIBRARIES:

Libraries Used For Deployment:

```
from flask import Flask, render_template, request
from pyngrok import ngrok
import pandas as pd
import tensorflow as tf
from transformers import BertTokenizer, TFBertForSequenceClassification
from sklearn.preprocessing import LabelEncoder
import numpy as np
import os
import fitz # PyMuPDF
import easyocr
from googletrans import Translator
from summa import summarizer
from datetime import datetime
import csv
```

LIBRARIES:

pandas:

- Usage: It provides data structures like DataFrames, making it easy to manipulate and analyze structured data.

pdf2image:

- Usage: It's useful when you need to work with the content of PDF files by converting them into images.

googletrans:

- Usage: It allows you to integrate Google Translate functionality into your Python applications for language translation.

summa:

- Usage: It can be used for text summarization and identifying important phrases within a body of text.

LIBRARIES:

1. flask-ngrok:

- Usage: It's helpful for development and testing when you need to make your local Flask application accessible online.

2. flask-CORS:

- Usage: It enables web applications running at one origin (domain) to request resources from a different origin, which is typically restricted by web browsers.

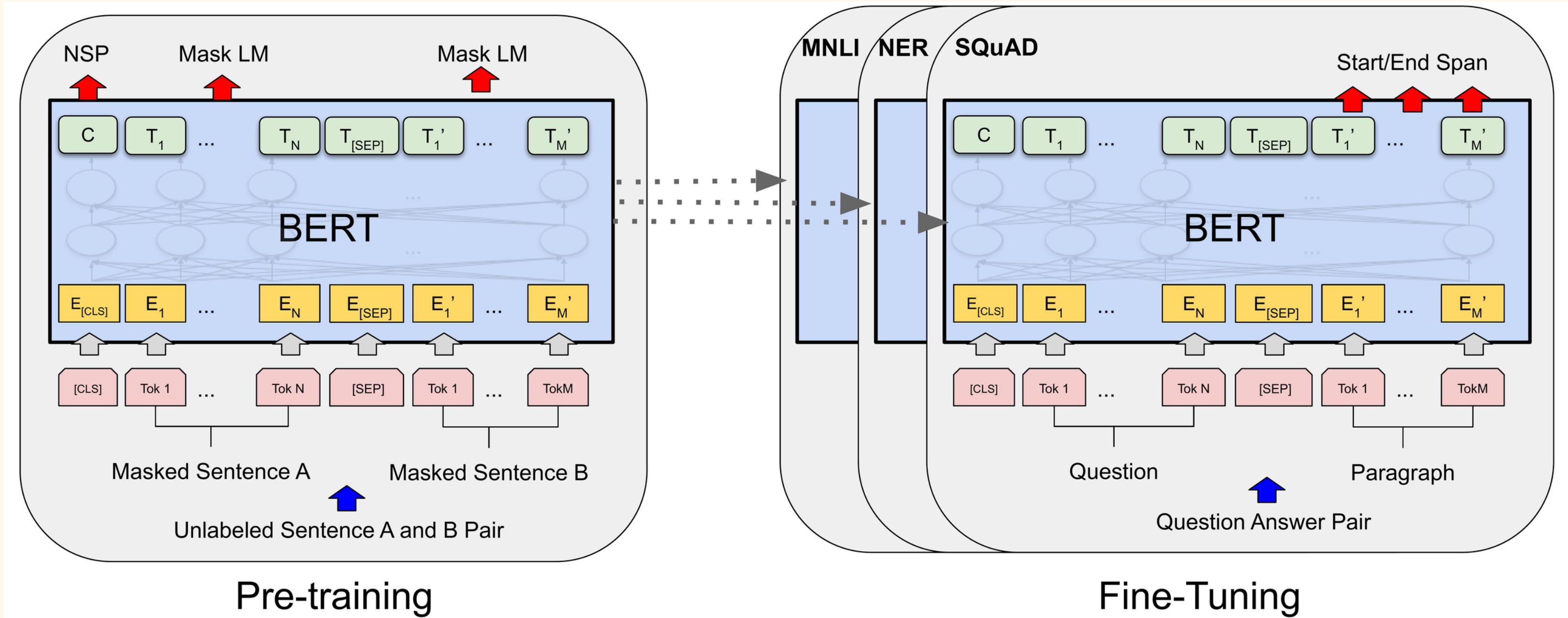
3. pickle:

- Usage: It allows you to save Python objects to a file and later load them back, preserving their state.

DATASET

A	B	C
IPC-Section	Description	Offense
82	Act of a child under seven years of age	Sophie (6 years old): In an unintentional act, Sophie spills paint on a neighbor's property while playing.
82	Act of a child under seven years of age	James (5 years old): Accidentally breaks a neighbor's window with a ball while playing, demonstrating an act of immaturity.
82	Act of a child under seven years of age	Emma (4 years old): Unintentionally damages a neighbor's garden while playing, engaging in an act of immaturity.
82	Act of a child under seven years of age	In a suburban neighborhood, during a community picnic, Sarah's toddler, Emma, innocently spills paint on a public bench.
82	Act of a child under seven years of age	During a family gathering, Mark's five-year-old son, Jake, unknowingly entered a neighbor's property and damaged a bicycle.
83	Act of a child above seven and under twelve of immature understanding	Mark (Age 9): Accidentally damages a neighbor's bicycle while playing, demonstrating an act of immaturity.
83	Act of a child above seven and under twelve of immature understanding	Being under twelve, damages school property unintentionally, showcasing an act due to immaturity.
83	Act of a child above seven and under twelve of immature understanding	Aged seven, unintentionally spills paint on a public bench, engaging in an act exempt from criminal liability.
83	Act of a child above seven and under twelve of immature understanding	At a local bookstore, Emily, an eight-year-old with an immature understanding, unintentionally damages a book.
83	Act of a child above seven and under twelve of immature understanding	Imagine an 11-year-old child, aware of the concept of ownership but not fully grasping the gravity of their actions, accidentally damages a toy.
84	a person who, at the time of doing it, by reason of unsoundness of mind	Due to a temporary episode of severe mental illness, David, at the time, is incapable of forming a criminal intent.
84	a person who, at the time of doing it, by reason of unsoundness of mind	Suffers from a momentary lapse of sanity, causing Sophie to be temporarily incapable of forming a criminal intent.
84	a person who, at the time of doing it, by reason of unsoundness of mind	Experiences a bout of mental incapacity, rendering James unable to comprehend the nature of his actions.
84	a person who, at the time of doing it, by reason of unsoundness of mind	Sarah, diagnosed with severe mental illness, enters a shopping mall convinced a common misconception.
84	a person who, at the time of doing it, by reason of unsoundness of mind	Consider a situation where an individual, due to a severe and temporary episode of mental illness, unintentionally damages a neighbor's property.
85	Act of a person incapable of judgment by reason of intoxication caused by another person	Mark unknowingly consumes a drink spiked with a potent substance, causing him to be incapable of forming a criminal intent.
85	Act of a person incapable of judgment by reason of intoxication caused by another person	Sophie attends a party where someone intentionally slips a substance into her beverage, rendering her incapable of forming a criminal intent.
85	Act of a person incapable of judgment by reason of intoxication caused by another person	At a bar, James unintentionally consumes a drink with a substance added without his knowledge, causing him to be incapable of forming a criminal intent.

BERT MODEL



BERT MODEL

- BERT, which stands for Bidirectional Encoder Representations from Transformers, is a pre-trained natural language processing (NLP) model
- the BERT base-uncased model serves as a powerful feature extractor and context-aware representation learner.
- It has already been pre-trained on a diverse and extensive corpus of text, allowing it to capture contextual relationships between words.
- When fine-tuned for the specific task of classifying FIR descriptions and offenses into IPC sections, the model leverages its pre-learned knowledge to extract relevant features and make accurate predictions.
- The choice of the "bert-base-uncased" variant specifically indicates the use of a model that does not differentiate between uppercase and lowercase letters, which is suitable for many text classification tasks.

TRAINING:

```
# Train the model
model.fit(preprocess_text(x_train['Description'], x_train['Offense']),
          y_train,
          epochs=60,
          batch_size=16,
          validation_data=(preprocess_text(x_test['Description'], x_test['offense']), y_test))
```

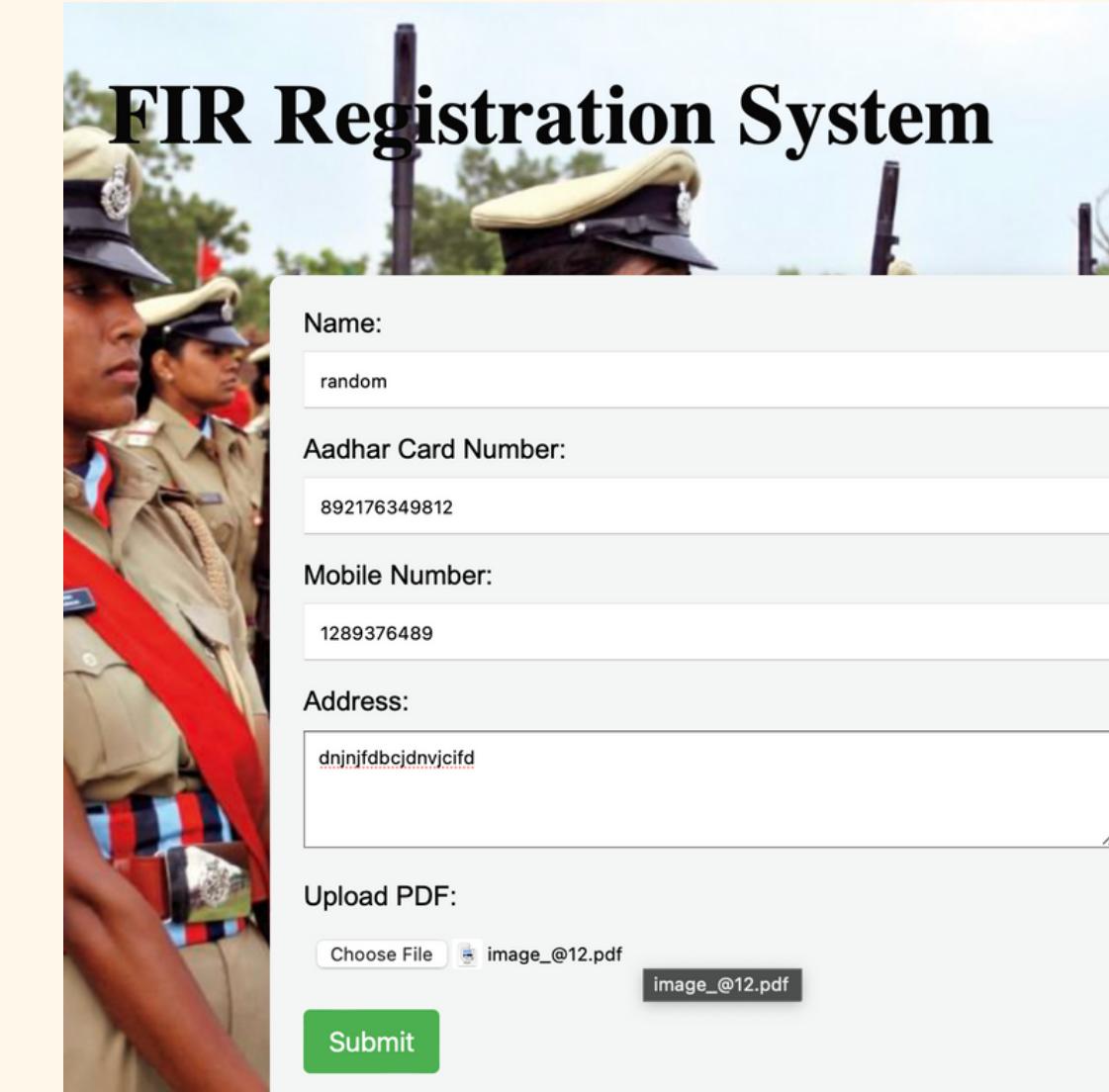
- this code snippet trains the BERT-based model on the provided FIR dataset. The model is updated based on the optimization process using the training data, and its performance is assessed on the validation set to monitor generalization.
- The model is compiled with the AdamW optimizer, sparse categorical crossentropy loss, and accuracy metric.
- The code evaluates the final trained model on both the training and testing datasets, providing insights into its overall performance.
- The trained model, its configuration, and weights are saved for future use or deployment.

TESTING:

Example:

उभरते युवा राजनेता ने गुस्से में अपनी पत्नी को गोली मार दी, फिर शव के टुकड़े किए और उसे एक रेस्तरां के तंदूर में जलाने की कोशिश की। दिल्ली प्रदेश युवा कांग्रेस के पूर्व अध्यक्ष सुशील शर्मा को 2003 में अपनी पत्नी नैना साहनी की हत्या के लिए दोषी ठहराया गया था। महिला कांग्रेस की दिल्ली इकाई की पदाधिकारी.

INPUT DATA



FIR Registration System

Name: random

Aadhar Card Number: 892176349812

Mobile Number: 1289376489

Address: dnjnjdabcjdnvjcfid

Upload PDF: Choose File image_@12.pdf

Submit

HTML FORM

Predicted IPC Section:

302: 2.5749075

Description: Description of IPC Section 302 According to section 302 of Indian penal code, Whoever commits murder shall be punished with death, or imprisonment for life, and shall also be liable to fine. IPC 302 in Simple Words Section 302 of the Indian Penal Code states that anyone who commits murder can be punished with the death penalty or life imprisonment, and may also be fined. Cited by Ramdeo Rai Yadav vs State Of Bihar Mohammad Irfan vs State Of Karnataka Prasad Pradhan vs The State Of Chhattisgarh Nandu Singh vs The State Of Madhya Pradesh Kala Singh Gurnam Singh vs The State Of Punjab

Offense: Murder

Punishment: Death or Imprisonment for Life + Fine

Bailable: Non-Bailable

Court: Court of Session

Predicted IPC-SECTION

FEATURES:

- OCR functionality
- States specific rules along with its punishments
- Hindi and english FIRs can be currently analyzed
- Tracking of fake cases(aadhar card)
- Predicts multiple IPCs
- All the details of previous and current users will saved in csv file

Name	Adhar	PhoneNo	ImagePath
Aayushi Personal			/content/drive/MyDrive/user_uploads/uploaded.pdf
Wilson Sir	987654321012	8980103107	/content/drive/MyDrive/user_uploads/uploaded.pdf
Zahera Saiyed	456789321023	8128870155	/content/drive/MyDrive/user_uploads/uploaded.pdf
Dad Nokia	123456789012	7265067921	/content/drive/MyDrive/user_uploads/uploaded.pdf
Wilson Sir	987654321023	8980103107	/content/drive/MyDrive/user_uploads/uploaded.pdf
Real Sandy	342156789123	9558886560	/content/drive/MyDrive/user_uploads/uploaded.pdf
Sangeeta Mausi	123456789012	7903790833	/content/drive/MyDrive/user_uploads/uploaded.pdf
			/content/drive/MyDrive/user_uploads/uploaded.pdf
poooooo	111111111111	1111111111	/content/drive/MyDrive/user_uploads/uploaded.pdf

FIR Registration System



Name:

Aadhar Card Number:

Mobile Number:

Address:

Upload PDF: image_@12.pdf

FUTURE SCOPES:

If you want to adapt our model for future purposes, such as handling input in multiple languages, incorporating voice input, and expanding to include additional legal acts and sections like CRPC (Code of Criminal Procedure) or other relevant legal frameworks.

Multilingual Support:

- Explore models that support multiple languages. Models like mBERT (Multilingual BERT) are designed to understand and generate representations for text in various languages.

Voice Input Integration:

- Incorporate a speech-to-text (STT) system to convert voice input into text. Popular libraries like Google Speech-to-Text or Mozilla DeepSpeech can be used for this purpose.

Inclusion of CRPC and Other Acts:

- Extend your dataset to include FIRs labeled with sections from CRPC or other legal acts. Ensure that the dataset is balanced and representative of the variety of cases we want the model to handle.

multi language
input in terms of voice
include crpc and other act and section