The Future of Crime Prevention: Police Case Analysis Using Machine Learning

(Accident Case Analysis)

Project ID: 2023-224

Dharsan. R – IT20003982

Final (Draft) Report

BSc (Hons) in Information Technology

Specializing in Software Engineering

Department of Computer Science & Software Engineering

Sri Lankan Institute of Information Technology

Sri Lanka

September 2023

The Future of Crime Prevention: Police Case Analysis Using Machine Learning

(Accident Case Analysis)

Project ID: 2023-224

Dharsan. R – IT20003982

Final (Draft) Report

BSc (Hons) in Information Technology

Specializing in Software Engineering

Department of Computer Science & Software Engineering

Sri Lankan Institute of Information Technology

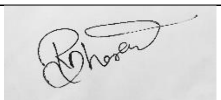
Sri Lanka

September 2023

DECLARATION

I declare that this is my own work, and this Thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my Thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Signature
Dharsan. R	IT20003982	

The above candidate has carried out this research thesis for the Degree of Bachelor of Science (honors) Information Technology (Specializing in Software Engineering) under my supervision.

Signature of the supervisor

(Ms. Hansika Mahaadikara) Date

Signature of co-supervisor

(Ms. Sanjeevi Chandrasiri) Date

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to our module coordinator, Mr. Jayantha Amararachchi, for his unwavering support, motivation, and valuable insights that propelled our project forward with enthusiasm. We would also like to express our appreciation to Ms. Hansika Mahaadikara and our co-supervisor, Ms. Sanjeevi Chandrasiri, for their invaluable guidance and support from the project's inception to its completion. Their wealth of ideas greatly enriched our project's development, and their patience and unwavering commitment helped us overcome challenges. Additionally, we acknowledge the contributions of our lecturers, assistant lecturers, instructors, fellow group members, and the dedicated academic and non-academic staff at SLIIT, who provided unwavering support and assistance throughout the project. Finally, our heartfelt thanks go out to our beloved families and friends, who stood by us as pillars of strength and provided invaluable moral support during challenging moments in the project's journey.

ABSTRACT

This research thesis embarks on an exploration of the future landscape of crime prevention, propelled by innovative machine learning applications. Comprising four interconnected components – accident case analysis, crime analysis, case document classification, and crimes against women analysis – the study presents a holistic approach to enhancing public safety. This individual report is dedicated to delving into the intricacies of police accident case analysis, an integral facet within the overarching machine learning-driven crime prevention system. Its primary goal revolves around predicting and dissecting accident attributes, with a specific emphasis on the Sri Lankan context. Drawing from a rich reservoir of historical accident data, this component harnesses the prowess of advanced statistical models, including the ARIMA model and K-means clustering. These models not only facilitate the forecasting of future accident percentages but also delve deeper to unearth the underlying causal factors. Beyond mere prediction, the system extends its capabilities by actively formulating tailored accident prevention strategies. It is through this synergy of prediction and prevention that the report underscores the potential of technology, advanced analytics, and machine learning to reshape the landscape of accident analysis and prevention. Moreover, the report highlights the practical application of these insights, empowering law enforcement agencies to proactively enhance road safety. By enabling the practical implementation of prevention strategies, this system equips police departments to take targeted actions aimed at reducing accident rates. Ultimately, the findings underscore the transformative potential of technology, advanced analytics, and machine learning in shaping the future of accident analysis and prevention. The individual report serves as an in-depth exploration of the police accident case analysis component, offering valuable insights into its methodologies, outcomes, and its overarching role within the broader context of data-driven crime prevention.

Table of Content

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table **Page**

LIST OF ABBREVIATIONS

Abbreviations	Description
ML	Machine Learning
ARIMA	Auto Regressive Integrated Moving Average
IO	Input/Output
CSV	Comma-Separated Values
SDLC	Software Development Life Cycle
GUI	Graphical User Interface
DIG	Deputy Inspector General
UAT	User Acceptance Testing
API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
SNN	Spiking Neural Networks
CNN	Convolutional Neural Networks
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets

1. INTRODUCTION

1.1 Background Study and Literature

1.1.1 Introduction

Police departments and law enforcement agencies in Sri Lanka face more challenges in effectively addressing road accidents and ensuring public safety. Conventional methods of accident analysis and prevention have proven to be labour-intensive, error-prone, and lacking in predictive capabilities. These traditional approaches often result in delayed responses, with the ever-increasing volume of accident data overwhelming human analytical capacities. The absence of proactive predictive tools hampers law enforcement's ability to anticipate and prevent accidents efficiently. In light of these challenges, there is a solution need for a more advanced, data-driven solution to enhance road safety and optimize police department and law enforcement efforts.

This research project, titled "The Future of Crime Prevention: Police Accident Case Analysis Using Machine Learning," tries to reconsider how law enforcement organizations and police departments address the complex issue of road accidents. This integrated system is made up of four main components, with the "Accident Case Analysis" component being the central focus of this report. Road accidents are inherently unpredictable events with potentially severe consequences, necessitating a deeper understanding of their patterns and causative factors. The "Accident Case Analysis" component aims to empower law enforcement agencies by predicting accident percentages, fatal percentages, and accident causes for the next three years within specified divisions. This innovative approach serves as a solution to the time-consuming and labour-intensive nature of traditional accident case analysis. Leveraging state-of-the-art machine learning techniques, this component forecasts future accident trends, identifies causative factors, and facilitates in-depth analyses. It empowers law enforcement with actionable insights to proactively address safety concerns and allocate resources effectively, even predicting peak accident occurrence times. By harnessing the capabilities of machine learning and data-driven analysis, this research aims to revolutionize how road accidents are analyzed and prevented, ultimately contributing to safer roadways and communities in Sri Lanka. In the following sections of this report, we will delve into the methodology, data analysis, and findings specific to the Accident Case Analysis component, shedding light on its potential to transform and enhance road safety.

1.1.2 Literature Review

Numerous studies have extensively explored approaches in police accident case analysis, crime analysis, text classification, and crimes against women. Li, Wu, and Peng [1] developed a traffic accident prediction model using spiking neural networks (SNNs) and convolutional neural networks (CNNs). This research addresses the critical need for accurate post-impact prediction in traffic accident management, utilizing SNNs to capture spatial and temporal features effectively, potentially improving prediction accuracy over traditional methods. V. Prasannakumar, H. Vijith, R. Charutha, and N. Geetha [2] conducted research on road accident analysis using geo-information technology to identify accident hotspots and employed spatial statistics, including Moran's I and Getis-Ord Gi* statistics, within a GIS-based framework to assess spatial clustering, aiding traffic management and safety decisions. It aimed to assess the spatial and temporal patterns of accidents in the Thiruvananthapuram city area, providing insights into accident hotspots and cold spots. This research contributed to understanding the distribution and variations of road accidents, which can inform traffic management strategies and accident reduction efforts, emphasizing the significance of geospatial and temporal data analysis in accident research and prevention.

In the study conducted by Manzoor et al. [3], the focus was on predicting the severity of road accidents, a critical concern given the numerous factors involved. The authors employed an ensemble model called RFCNN and compared its performance with various base learner models, including tree-based ensemble models (RF, AC, ETC, GBM) and an ensemble of regression algorithms (Voting classifier, LR+SGD). Notably, Random Forest (RF) identified 20 significant features from the dataset, and these features were used for the experiments. The results demonstrated that RFCNN outperformed other models in terms of accuracy, achieving an impressive accuracy rate of 99.1%, while also excelling in precision, recall, and F-score. The study emphasized the importance of identifying significant features and highlighted the potential for improving accident severity prediction while reducing data collection costs. Senanayake and Joshi [4] developed the Road Accident Pattern Miner (RAP miner), an innovative system using a hybrid learning algorithm and the Self-Expiring Association (SEA) algorithm for association rule mining. This system achieved an impressive accuracy rate of 92.75%, surpassing other algorithms' maximum accuracy of 92.25%. Integrating SEA with Case-Based Reasoning (CBR) significantly enhanced the RAP miner's performance, reducing processing time by a remarkable 67%. This advancement holds promise for real-time accident prediction based on location and severity, offering significant contributions to road safety. Yang, Han, and Chen’s study [5], the prediction of traffic accident severity was achieved using a Random Forest algorithm. Their model exhibited a remarkable accuracy rate of 80%, outperforming other machine learning models. Key accident characteristics, such as location, collision pattern, road information, and speed limits, were employed to enhance prediction accuracy, offering valuable insights for traffic safety management and accident prevention.

In an influential study conducted by M. I. Sameen and B. Pradhan [6] a Recurrent Neural Network (RNN) was employed to predict traffic accident injury severity. Analyzing 1130 accident records from Malaysia's North-South Expressway over six years, they highlighted RNNs' aptness for handling sequential data and capturing temporal correlations within accident records. After an extensive grid search, they identified an optimal network structure, including a Long-Short Term Memory (LSTM) layer, two fully-connected layers, and a SoftMax layer. Through sensitivity analysis and comparison with Multilayer Perceptron (MLP) and Bayesian Logistic Regression (BLR) models, the RNN demonstrated superior performance, achieving a validation accuracy of 71.77%. This study underscores the promise of RNNs in deep learning frameworks for accurate traffic accident severity prediction.

In a study by S. K. Singh [7], the focus is on analyzing road accidents at national and state levels in India. The research highlights the distribution of accident-related fatalities and injuries across different parameters, including age, gender, month, and time of occurrence. Notably, the study identifies the age group of 30-59 years as the most vulnerable, with males experiencing a higher rate of accidents than females. It also emphasizes the influence of weather conditions and working hours on accident occurrences. Furthermore, the research reveals significant variations in fatality risks among Indian states and cities. Despite lower accident burdens in metropolitan areas, nearly half of them face higher fatality risks compared to their rural counterparts. The study concludes by emphasizing the worsening road safety situation in India and the urgent need for effective interventions to curb the escalating number of road traffic deaths, projecting a grim scenario if no substantial actions are taken.

A noteworthy study by T. Beshah and S. Hil [8] delve into the pressing issue of road traffic accidents (RTAs) within Ethiopia, a formidable public health challenge. They draw attention to the scarcity of studies exploring the intricate relationship between road characteristics and the severity of RTAs specific to the Ethiopian context. Employing data mining techniques, the study successfully connects road attributes with accident severity and develops predictive models, achieving an impressive accuracy rate of 95%. These findings not only furnish valuable insights for policymakers and the implementation of effective traffic safety measures but also underscore the immense potential of data mining in advancing research on road safety. The authors further envision future research endeavours that encompass driver-related factors for a more comprehensive analysis of road safety dynamics.

The study by M. K. Islam, I. Reza, U. Gazder, et al. [9] explores road traffic accident (RTA) severity prediction in Al-Ahsa, Saudi Arabia, using machine learning techniques. Random Forest (RF) achieved the highest prediction accuracy (94%) among the models. Faulty tires, not giving way, sudden turning, and running over were identified as key factors contributing to severe crashes. Spatial analysis revealed clustering of RTAs, providing valuable insights for traffic safety management.

Further another study by J. Fattah, L. Ezzine, Z. Aman, et al. [10] focuses on demand forecasting in a food company using the ARIMA model, a time series approach. It explores how historical demand data can inform future demand predictions and influence supply chain decisions. The chosen ARIMA (1, 0, 1) model proves effective in forecasting, aiding production planning and cost reduction. Demand forecasting is recognized as a crucial aspect of supply chain management, and the study aims to provide valuable insights for managerial decision-making and future improvement of forecasting techniques, potentially including neural networks and other approaches.

In conclusion, the studies highlighted in this literature review contribute significantly to the field of accident analysis, prediction, and road safety management. Various methodologies, including machine learning, data mining, spatial analysis, and neural networks, were employed to address different aspects of traffic accidents, from severity prediction to hotspot identification and nationwide analysis. The importance of harnessing advanced techniques and data-driven approaches for enhancing road safety, reducing accident-related costs, and informing evidence-based policy decisions is underscored by these research endeavours. Moving forward, further interdisciplinary collaboration and innovation in data analysis will play a pivotal role in the ongoing efforts to mitigate the devastating impact of road traffic accidents worldwide

1.2 Research Gap

In the realm of traffic accident prediction and analysis, a significant research gap looms large. This gap becomes especially apparent when considering the amalgamation of accident percentage predictions for future years, causative factors, and comprehensive accident case analysis into a unified and cohesive system. Although previous studies have made commendable progress in individual aspects such as predicting accident locations and identifying causes, the holistic approach proposed in this research remains relatively uncharted territory. The core novelty of this research lies in the seamless integration of these critical facets, creating a unified system that promises to provide invaluable insights for long-term traffic safety planning, targeted prevention strategies, and a profound understanding of accident dynamics. It's important to recognize that this unified approach represents a substantial research gap, as there currently exists no comprehensive system within the domain of traffic safety management that successfully combines accident prediction, causative factors analysis, and in-depth case analysis under a single overarching framework.

The prevailing research landscape in traffic safety management primarily consists of fragmented studies and efforts that often address specific aspects of the problem in isolation. Some research may focus on predicting accident locations using advanced spatial analysis techniques, while others delve into the causative factors behind accidents, attempting to identify patterns or contributing variables. In parallel, there are endeavors dedicated to in-depth accident case analysis, which delve into the fine details of specific accidents, aiming to derive lessons for prevention. However, these individual strands of research often remain disconnected, failing to harness the potential synergy that could emerge from their integration. The research gap becomes evident when one considers the profound benefits that could arise from a unified system capable of seamlessly integrating accident prediction, causative factor identification, and detailed case analysis.

Such a comprehensive system would not only provide invaluable tools for long-term traffic safety planning, but it would also allow for the development of targeted prevention strategies that are grounded in both predictive analytics and a deep understanding of accident causation. Moreover, it would pave the way for a more nuanced and sophisticated approach to accident dynamics, enabling stakeholders in traffic safety management to make informed decisions that have a real impact on reducing accidents and their associated costs. In essence, the research gap in the domain of traffic accident analysis can be summed up as the absence of an integrated and holistic system that brings together diverse components of accident prediction, causative factor analysis, and detailed case examination. This research endeavour aims to bridge this gap comprehensively, offering a novel and integrated approach that holds the potential to revolutionize safety management and accident prevention.

As the research progresses and this unified system takes shape, it is expected that its contributions to the field of traffic safety management will be substantial. By addressing this critical research gap, the potential benefits are not only academic but also extend to practical applications that can save lives, reduce injuries, and mitigate the social and economic burdens associated with traffic accidents. Ultimately, the pursuit of a unified approach to traffic accident analysis represents a promising avenue for advancing the field and making our roads safer for all.

RESEARCH	FUTURE YEAR PREDICTION	CAUSE PREDICTION	PREVENTION STRATEGIES	POLICE RELATED	LOWER COST
ROAD TRAFFIC ACCIDENT PREDICTION USING MACHINE LEARNING	✓	✗	✗	✗	✓
PREDICTING FUTURE DRIVING RISK OF CRASH-INVOLVED DRIVERS BASED ON A SYSTEMATIC MACHINE LEARNING FRAMEWORK.	✓	✗	✗	✗	✗
ROAD TRAFFIC ACCIDENTS ANALYSIS USING DATA MINING TECHNIQUES	✗	✓	✗	✗	✗
PREDICTION OF TRAFFIC ACCIDENT SEVERITY BASED ON RANDOM FOREST	✓	✗	✗	✗	✓
POLICE ACCIDENT CASE ANALYSIS	✓	✓	✓	✓	✓

***Figure 1.2.1**:* Research gap for accident case analysis

1.3 Research Problem

The research problem at hand encompasses multiple domains, each with its unique challenges and gaps. In the field of traffic accident prediction and analysis, there is a pressing issue: the absence of a unified system combining accident percentage prediction for future years, causative factors identification, and in-depth accident case analysis. Within the realm of police accident case analysis, a complex and multifaceted research problem comes to the fore. This challenge revolves around the absence of a comprehensive and cohesive system that can effectively unite various essential components within this domain, leading to the emergence of a fragmented landscape. The central research problem revolves around the imperative need for an integrated framework capable of harmonizing accident case analysis, the prediction of accident percentages for future years, and the identification of causative factors. While individual studies and methodologies have made commendable strides in addressing specific aspects of police accident case analysis, the core issue lies in bridging the gaps that exist between these disparate elements. The research problem becomes pronounced when considering the practical implications. The fragmented nature of existing research hinders the development of a unified system that can significantly enhance accident investigations, enable more accurate predictions of accident trends, and facilitate the formulation of targeted safety measures. It underscores the crucial need for synergy and coordination among these diverse components to glean a more profound understanding of accident dynamics and improve safety outcomes on our roadways. In essence, the research problem in police accident case analysis can be distilled into the absence of a comprehensive and integrated system that seamlessly combines accident prediction, causative factor analysis, and in-depth case examination under a single overarching framework. Addressing this research problem is paramount for advancing the field and empowering stakeholders in traffic safety management to make informed decisions that contribute to reducing accidents and their associated costs. As research progresses in this direction, the potential contributions extend beyond academia to practical applications that can save lives, reduce injuries, and alleviate the societal and economic burdens linked to traffic accidents. Thus, the pursuit of a unified approach to police accident case analysis signifies a promising avenue for driving advancements in the field and enhancing safety on our roadways.

1.4 Research Objectives

1.4.1 Main Objective

The main objective of this project is to develop a robust machine learning system capable of comprehensive analysis and prediction within the domain of accidents. This system aims to achieve a multifaceted set of goals, including forecasting accident percentages for specific locations, identifying the root causes of accidents, and conducting in-depth analyses to formulate precise and targeted safety measures and accident prevention strategies. By amalgamating predictive analytics, causative factor identification, and thorough accident case analysis, this research endeavour seeks to create a unified application that not only enhances our understanding of accident dynamics but also empowers police department and law enforcement agencies in traffic safety to proactively mitigate accidents and reduce accidents.

1.4.2 Specific Objectives

- Develop a location-specific accident prediction system leveraging machine learning techniques to forecast accident percentages accurately, aiding in targeted safety measures.
- Create predictive models to estimate accident percentages for different causative factors, enhancing the ability to prioritize interventions and allocate resources effectively.
- Utilize advanced statistical analyses to uncover intricate accident patterns, enabling data-driven decision-making and proactive safety planning.
- Formulate comprehensive accident prevention strategies based on data insights, fostering a safer environment through evidence-based interventions.
- Visualize the outcome in graph format.
- Providing simple and user-friendly interface to user

2. METHODOLOGY

2.1 Introduction

In this section, we outline the methodological approach utilized to tackle the functions and elements of the envisioned system. Our research adheres to a well-structured methodology grounded in the software lifecycle model, guaranteeing a methodical and orderly path towards system implementation. We draw upon extensive prior research within the same domain, leveraging valuable insights and expertise to achieve both the primary and secondary objectives of this study. The wealth of information and discoveries from preceding studies forms a robust foundational basis, providing direction for the development and enactment of the proposed system.

2.2 System Overview

The integrated system comprises four primary components, each designed to fulfill specific purposes: police accident case analysis, crime analysis, case document classification via text analysis, and crimes against women analysis. All these components contribute to the predictive and analytical capabilities of the system.

The police accident case analysis component plays a pivotal role in improving road safety planning and optimizing resource allocation. It consists of several stages, each contributing to a comprehensive understanding of accident predictions, patterns, and their root causes. Initially, historical data is analyzed using the autoregressive integrated moving average (ARIMA) model to predict future accident percentages, taking into account variables like division name and year. This predictive analysis enables proactive measures for enhancing road safety. Next, the K-means clustering technique is applied to categorize accidents based on factors such as light conditions, weather, drunken driving, and traffic. This clustering process identifies patterns and commonalities within the dataset, allowing the system to categorize accidents according to their underlying causes. The ARIMA model is then employed to forecast future accident trends within each cluster, leveraging statistical methods to enhance prediction accuracy. Furthermore, the component utilizes various statistical analysis techniques to uncover meaningful patterns, trends, and relationships within accident cases. This analysis aims to comprehend the distribution and trends in accident data concerning temporal factors like day, month, day of the week, and hour of the day. With this information, the system can derive tailored prevention strategies to address specific accident causes. This comprehensive approach empowers authorities to implement proactive measures and enhance road safety. Users of this component can input historical and current accident data, and the system dynamically predicts accident percentages and fatal percentages for future years, incorporating updated datasets. The outcomes are presented through percentages and graphical representations, enabling law enforcement agencies to make informed decisions, formulate prevention strategies, and take actions to prevent accidents, ultimately contributing to enhanced public safety.

Understanding the underlying causes of accidents is critical for effective safety measures. The system utilizes the K-means clustering technique to categorize accidents based on a wide range of factors, including light conditions, weather, drunken driving, and traffic conditions. These clusters help identify patterns and similarities within the dataset, shedding light on the root causes of accidents. By discerning the primary causes, law enforcement agencies and safety planners can implement targeted prevention strategies.

In-depth analysis of accidents is pivotal in deriving meaningful insights for prevention. Statistical analysis techniques are employed to uncover patterns, trends, and relationships within accident cases. This analysis extends to temporal factors such as day, month, day of the week, and hour of the day. By dissecting the distribution and trends in accident data, the system provides actionable information that can guide the formulation of prevention strategies. This comprehensive approach ensures that the strategies are tailored to address specific accident causes, making them more effective in reducing accidents. Armed with the predictions and in-depth analysis, the system offers prevention strategies. These strategies are derived from a profound understanding of accident patterns, causative factors, and historical data. By combining predictive analytics and empirical insights, the system provides law enforcement agencies and safety planners with the tools needed to proactively address potential accident risks. These strategies can range from targeted law enforcement efforts to infrastructure improvements, all aimed at enhancing road safety.

In the background processes of achieving these system functionalities, a structured methodology rooted in the software lifecycle model is followed. Extensive prior research in the domain of traffic safety management serves as a foundational resource, providing valuable insights and knowledge. The collected historical accident data undergoes preprocessing to ensure its suitability for analysis. Machine learning algorithms, such as ARIMA for prediction and K-means clustering for cause identification, are trained and fitted to the data to generate expected outcomes.

The system's outputs are presented in a user-friendly application, making the insights easily accessible to law enforcement agencies and safety planners. By seamlessly integrating accident prediction, causative factor analysis, and detailed case examination, the system empowers police departments to make informed decisions, formulate effective prevention strategies, and take actions aimed at preventing accidents, ultimately contributing to enhanced public safety.

The proposed integrated web application combines the power of data analysis and prediction to provide a comprehensive solution for police departments and investigators. By harnessing diverse algorithms and analytical techniques, this system empowers users to gain deeper insights into accident predictions,cause predictions and prevention strategies.

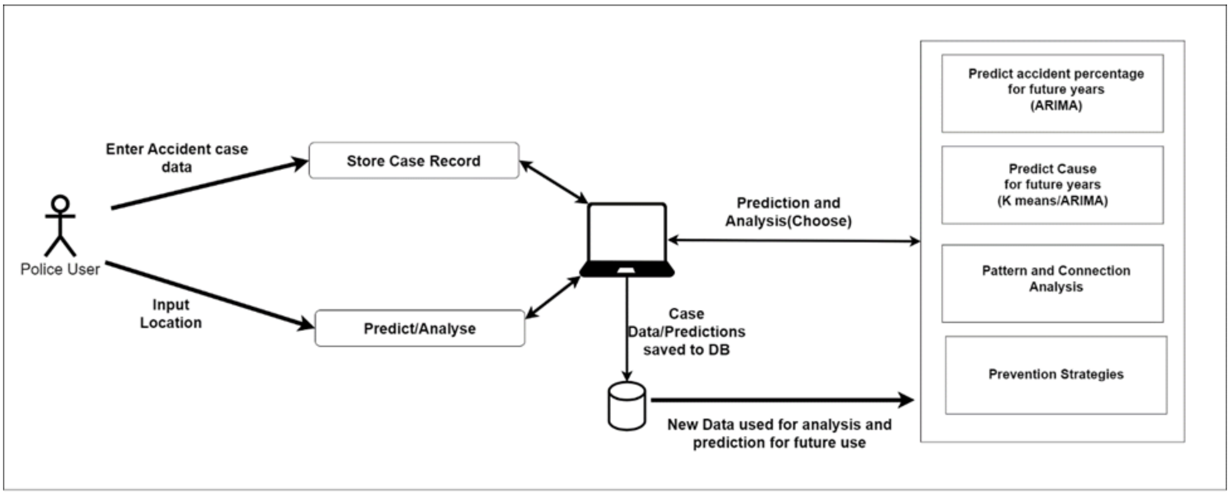


Figure 2.2.1: - High Level Architectural Diagram of police case analysis

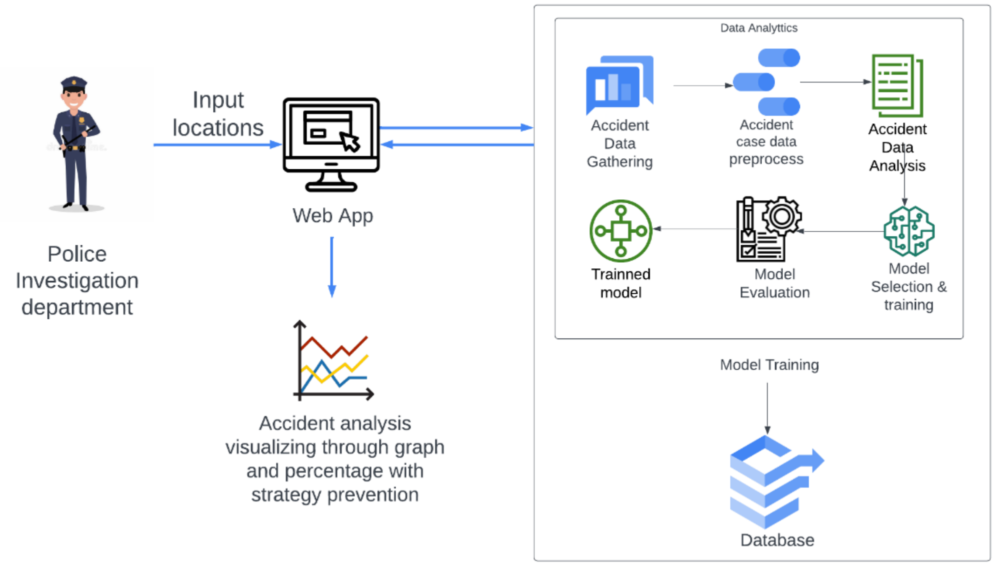


Figure 2.2.2: - More detailed diagram about model training

2.3 Diagrams

2.3.1 Work Breakdown Structures

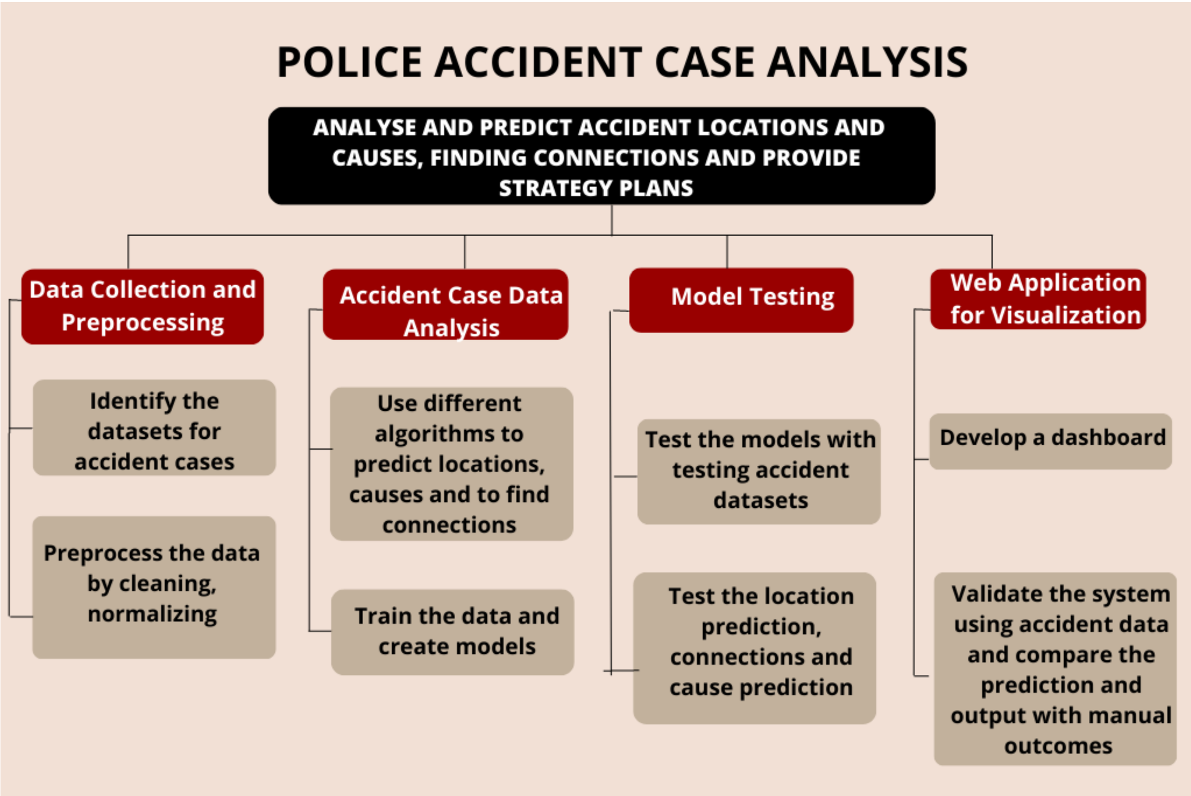


Figure 2.3.1.1: - WBS of accident case analysis

2.3.2 Flow Chart Diagram

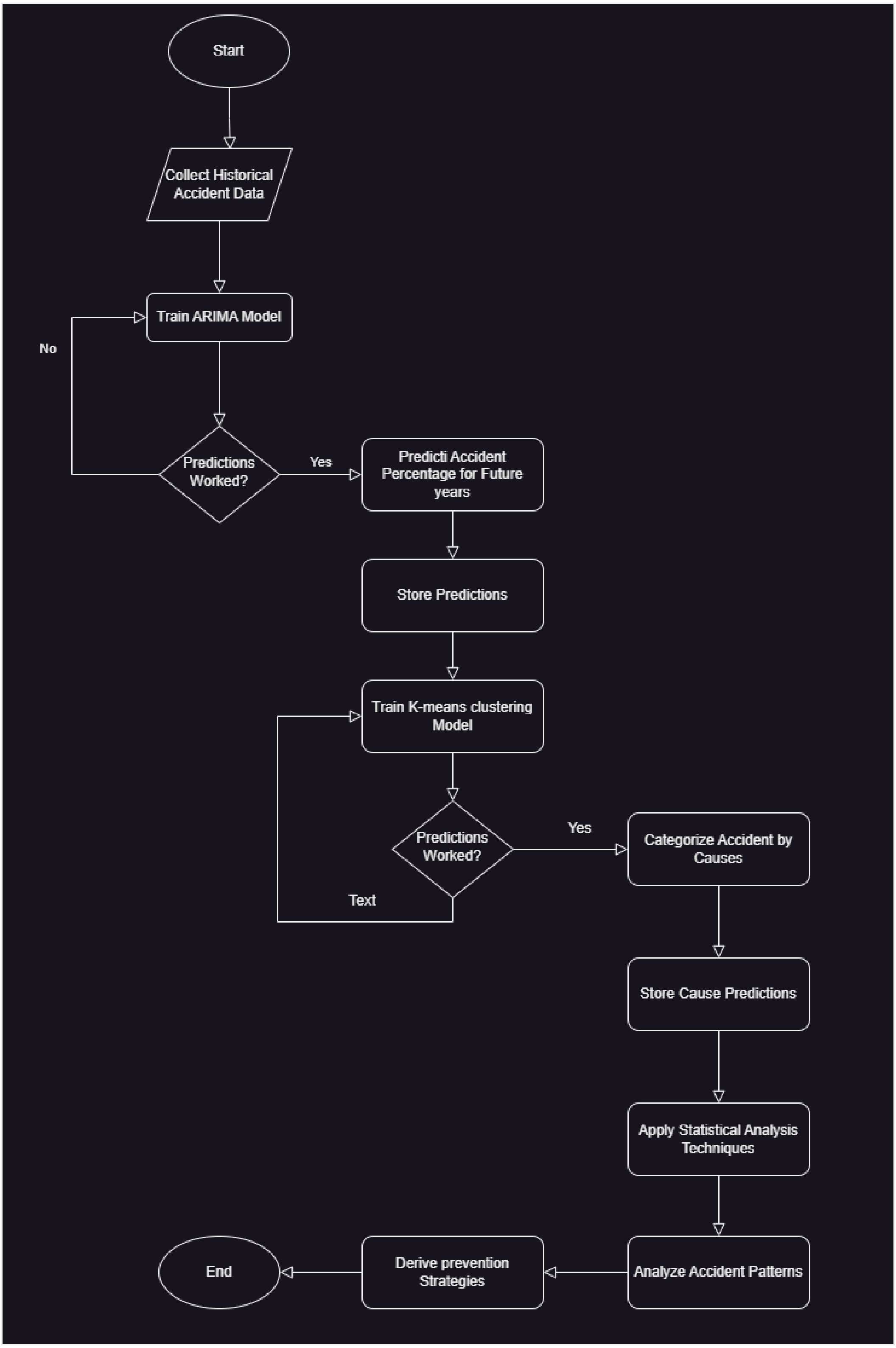


Figure 2.3.2.1: - Flow chart diagram of accident case analysis

2.4 Development Process

The development process chosen for our system follows a Waterfall model. This model is selected because it offers a linear and systematic approach, making it well-suited for our project's needs and clearly defined specifications. The Waterfall model involves distinct stages, each with tasks that can be scheduled and completed within specific timeframes. These stages and the associated strategy proceed sequentially without overlapping. Following the guidelines in this section, we will address and investigate the issues raised, provide a figurative characterization of the system's functions aimed at resolving these issues, examine the expected outcomes, and emphasize the importance of effective time management, which has been crucial throughout our year-long research endeavour. The system's requirements have been systematically divided into functional phases, as illustrated in Figure 2.4.1, which include Requirement Analysis, Design, Development, Testing, and Maintenance.

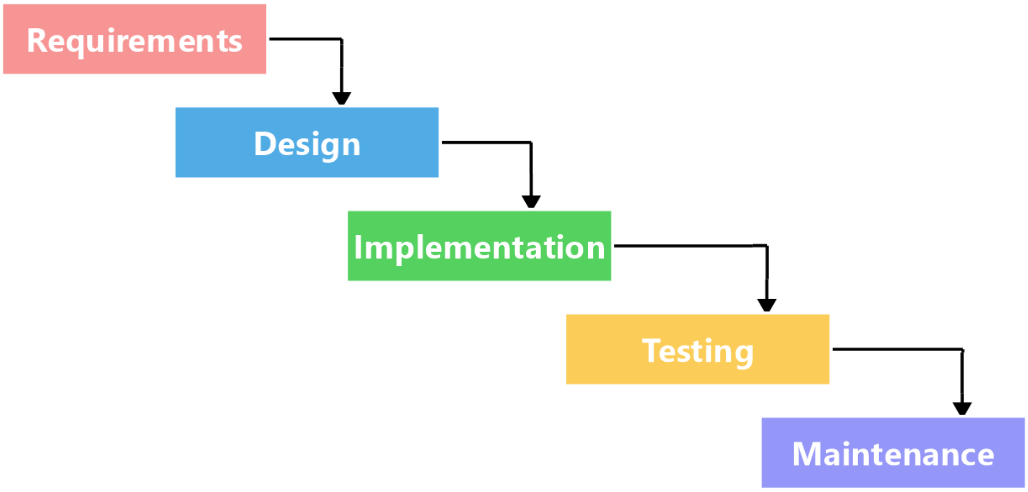


Figure 2.4.1: Development process of the system

2.4.1 Project Management

Project Management in the context of our Police Case Analysis project involves a distinctive approach due to the unique software development life cycle. Unlike traditional project management, software development entails multiple iterations, encompassing testing, updates, and continuous user feedback. To align with the dynamic nature of our project and adapt to evolving requirements from law enforcement agencies and stakeholders, we have embraced an agile methodology. Within our project teams, we have used a collaborative tool such as Microsoft teams for scheduling meetings and formulating project plans. These teams are responsible for both group and individual tasks, ensuring that every aspect of the project is systematically addressed and executed. The visual representations below provide insights into our structured approach to project management and task allocation

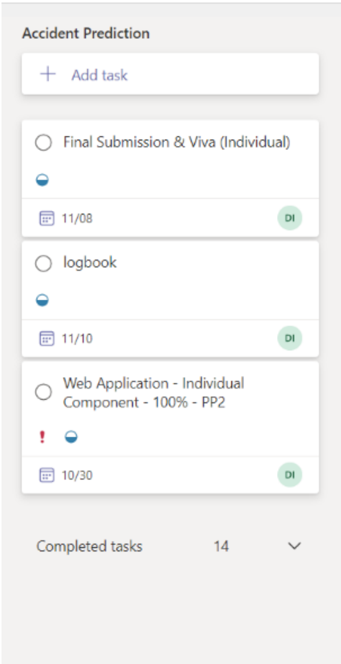


Figure 2.4.1.1:(Individual tasks) Project Management

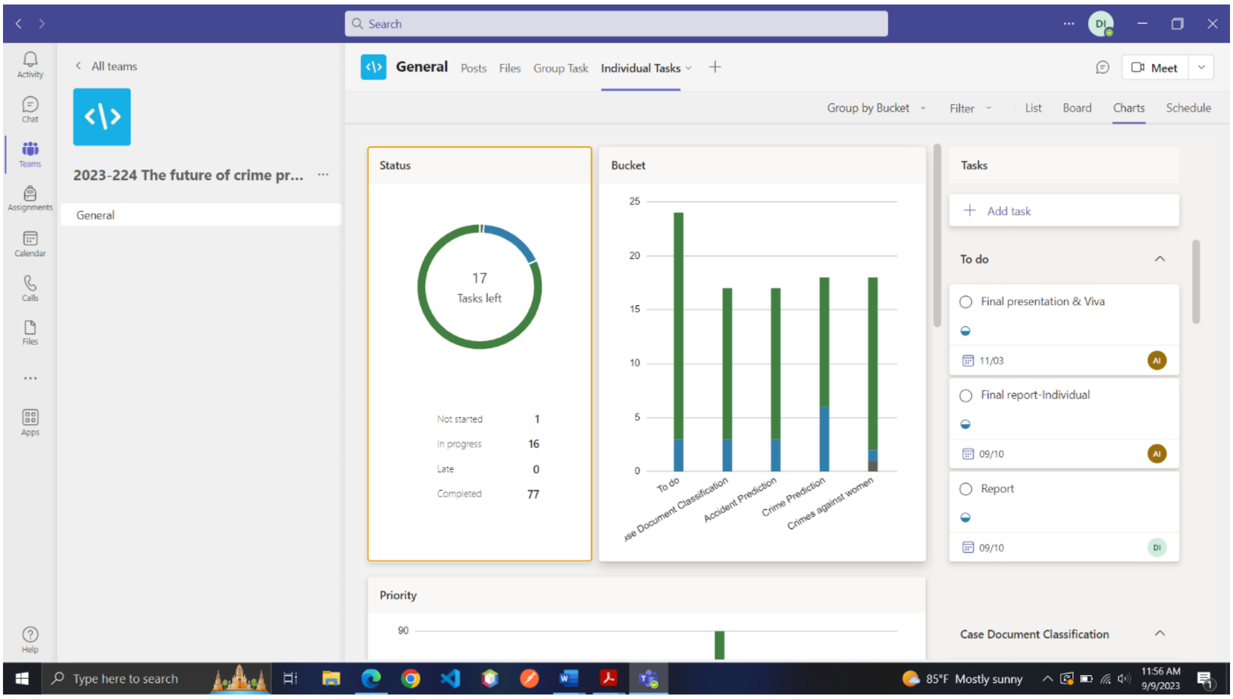


Figure 2.4.1.2: Project management of the system through MS teams (Individual tasks)

2.4.1.1 Project Code Management

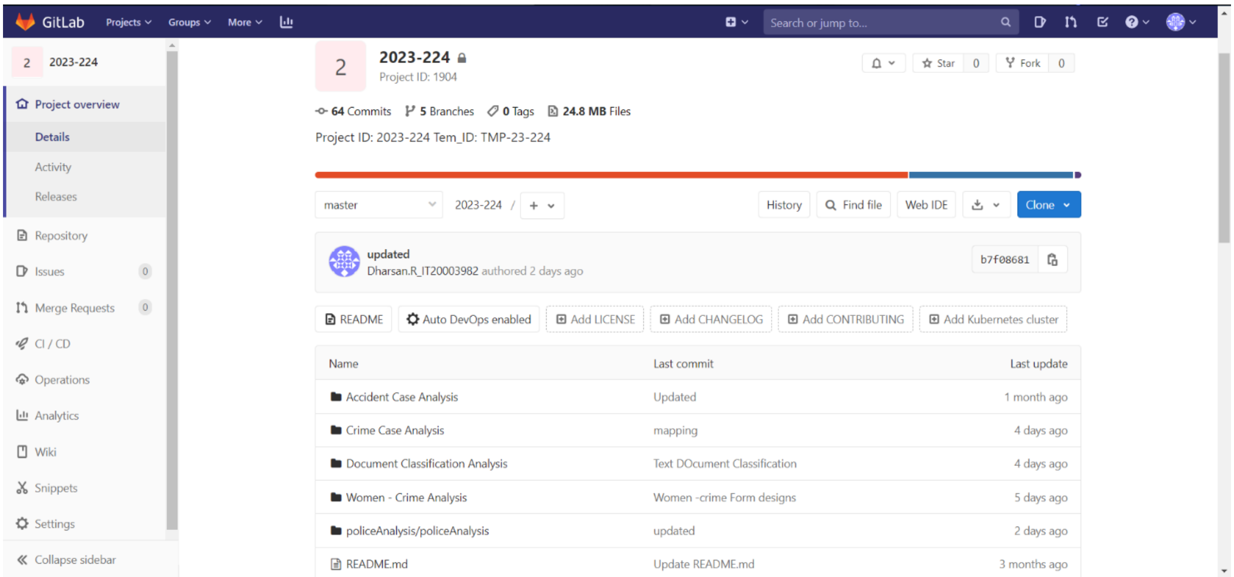


Figure 2.4.1.1.1: Code management in GitLab

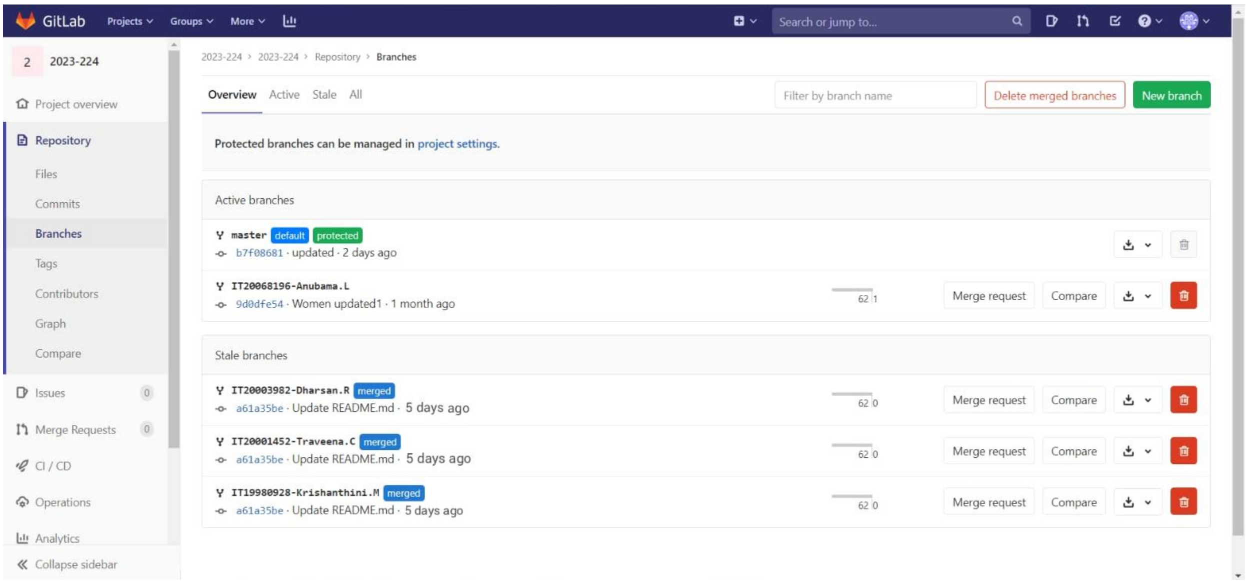


Figure 2.4.1.1.2: Merge branches

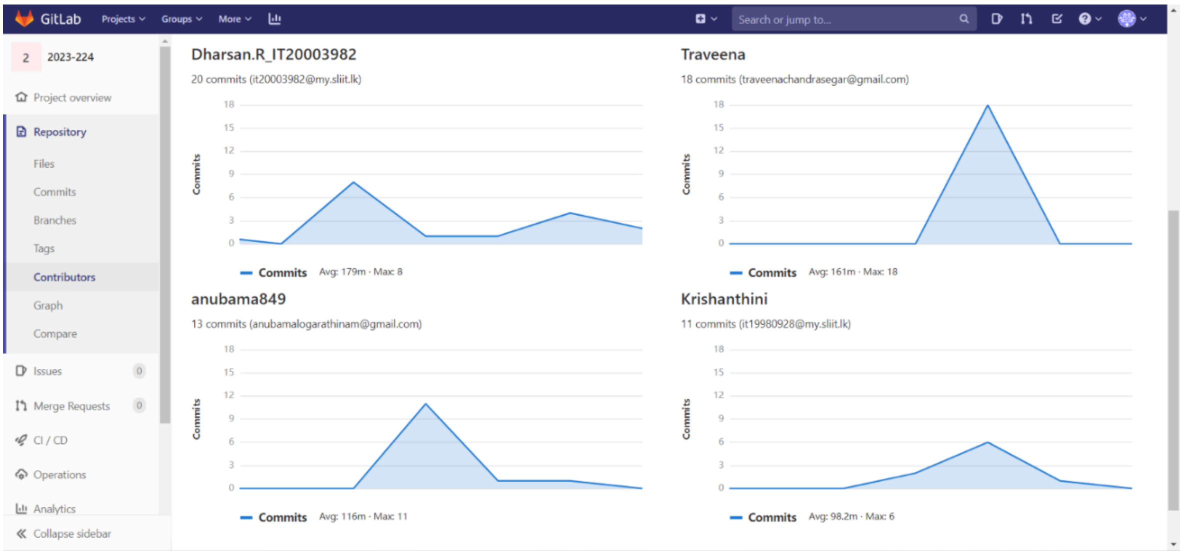


Figure 2.4.1.1.3: Individual commits

2.5 Requirement Gathering

The central objective of this system is to cater to the specific requirements of the police force, with no commercial objectives involved. In the endeavour to gather the requisite data for each component of the system, collaborative efforts were undertaken in coordination with key stakeholders. Discussions were initiated with the Deputy Inspector General (DIG) in Kurunegala, resulting in the procurement of essential datasets. Furthermore, the expertise of legal professionals and senior lecturers from the Faculty of Law in Colombo was sought, and they graciously provided valuable data contributions. The datasets meticulously compiled encompassed information related to accidents and crimes, as well as case documents. This comprehensive dataset compilation exceeded five thousand records for each component. Importantly, the data collected pertains exclusively to case details and does not involve any personally identifiable information, ensuring the utmost privacy and security.

This extensive dataset serves as the foundational bedrock of our machine learning algorithms, enabling the derivation of accurate predictions and the conduct of comprehensive analyses. Users of our system will have the opportunity to visualize these outcomes through graphical representations. It is noteworthy that the process of dataset acquisition presented significant challenges, necessitating a year-long effort and close collaboration with law enforcement authorities, ultimately resulting in the successful completion of our project.

Our requirements encompass both functional and non-functional aspects, which we will elaborate on in the sections below.

2.5.1 Functional Requirements:

- Accident Percentage Prediction: The system should be able to predict accident percentages for specified divisions and future years based on historical data and relevant variables.

- Accident Forecasting: It should have the capability to forecast accident percentages and causes for future years, enabling proactive safety planning and resource allocation.

- Accident Pattern Analysis: The system must conduct an in-depth accident pattern analysis to identify high-risk days, hours, and months, providing insights into temporal factors contributing to accidents.

- Prevention Strategies: Based on the analysis results, the system should offer targeted accident prevention strategies, aiding law enforcement agencies and authorities in implementing proactive measures.

2.5.2 Non-functional Requirements:

- Performance
- Reliability
- Security
- Usability
- Scalability

2.6 Resources Used

2.6.1.1 Software Boundaries

Backend - Python Language

Python plays a fundamental role as the foundation of the application, responsible for crafting server-side logic to manage HTTP requests and responses. Within this context, Python takes charge of user input processing, machine learning-based predictions, and data manipulation tasks. The utilization of libraries such as Pandas, NumPy, Scikit-Learn, and Django further enhances its capabilities in data management, computation, and machine learning model implementation.

Visual Studio Code Editor

Visual Studio Code (VS Code) serves as the integrated development environment (IDE) for the creation and development of Python scripts. Offering capabilities such as code autocompletion, debugging tools, and Python extensions, it emerges as a robust and efficient platform for the coding and maintenance of Python scripts.

Frontend – HTML, CSS, JavaScript

HTML, CSS, JavaScript: The frontend of the web application is built using HTML, CSS, and JavaScript. HTML is used for creating the structure and content of web pages. CSS is used for styling the user interface, making it visually appealing. JavaScript is employed for adding interactivity to the web application, such as handling user inputs and triggering requests to the backend.

Framework – Django

Django, a renowned high-level Python web framework recognized for its capacity to expedite the creation of secure and robust websites, has been employed within the software boundaries. Developed by experienced professionals, Django simplifies many of the intricacies inherent in web development, allowing for the construction of our application without necessitating the reinvention of fundamental elements. It is crucial to emphasize that all four components function seamlessly within the Django framework, ensuring system-wide cohesion and consistency.

Web Application

The web application serves as the frontend interface, allowing users to input data and trigger predictions via web forms. It communicates with the Python backend, exchanging HTTP requests and receiving prediction results. This collaborative use of Python and VS Code for backend operations and machine learning, along with HTML, CSS, and JavaScript for frontend design, results in an interactive and visually appealing user-friendly web application for crime prediction and prevention.

Libraries Used

1. **Pandas** - Pandas is a versatile library that provides data structures like DataFrame and Series. It's essential for data manipulation tasks such as data cleaning, transformation, and analysis. It simplifies working with structured data and supports various data sources.

2. **Matplotlib and Seaborn -** Matplotlib is a widely-used library for creating static, animated, or interactive visualizations in Python. Seaborn is built on top of Matplotlib and specializes in creating informative and attractive statistical graphics. Together, they facilitate data exploration and presentation.

3. **Scikit-Learn (sklearn) -** Scikit-Learn is a comprehensive machine learning library offering tools for classification, regression, clustering, and more. The imported classifiers, such as LinearSVC and MultinomialNB, are used for building machine learning models. It also includes modules for feature extraction, selection, and model evaluation.

4. **Django** - Django is a robust web framework that simplifies web application development. It follows the Model-View-Controller (MVC) architectural pattern and provides built-in tools for handling HTTP requests, managing databases, and rendering views. Django is particularly well-suited for building data-driven web applications.

5. **Numpy -** Numpy is the fundamental library for numerical computing in Python. It offers efficient array operations and mathematical functions, making it indispensable for scientific and mathematical applications.

6. **String -** The `string` library provides constants and utilities for working with strings, including character sets like ASCII letters and punctuation. It's handy for text manipulation tasks.

7. **Tempfile and Os -** Tempfile and Os are essential for managing temporary files and handling file paths. They ensure efficient and safe file operations within the application.

8. **Urllib and Base64 -** Urllib is used for making HTTP requests and handling URLs. Base64 encodes and decodes binary data, which can be valuable for encoding and decoding data for transmission or storage.
9. **IO -** The `io` module provides classes and functions for handling input and output operations. It's often used for working with streams and data buffers.

10. **Pickel -** The `pickle` library is used for serializing (pickling) and deserializing (unpickling) Python objects. It's valuable for storing and retrieving complex data structures.

11. **Base64 (import base64) -** Base64 is used to encode binary image data into a format that can be embedded directly into HTML and displayed as images in the web application.

12. **CSV (import csv) -** The CSV module is used for reading and writing CSV (Comma-Separated Values) files. It's used to add new data entries to the dataset.

13. **Random (import random) -** The Random module is used for generating random values. In this project, it's used to generate random colours for data visualization.

14. **Plotly Express (import plotly.express as px) -** Plotly Express is a library for interactive data visualization. It's used for creating interactive charts and graphs, although it's not used extensively in this code.

15. **statsmodels.tsa. arima.model.ARIMA:** This is a part of the stats models library and is specifically used for fitting ARIMA (AutoRegressive Integrated Moving Average) models. ARIMA is employed for time series forecasting and is crucial for predicting future accident percentages.

16. **sklearn.cluster.KMeans:** This module is part of scikit-learn, a machine learning library in Python. It's used for K-means clustering, which groups data points into clusters based on similarity.

2.7 Commercialization aspects of the product

The proposed system is intended for a service purpose, with the goal of leveraging machine learning techniques to address important societal issues related to public safety and crime prevention. By providing accurate and reliable analysis, prediction, and classification of accident and crime-related data, the system has the potential to help law enforcement agencies, policy makers, and other stakeholders make informed decisions, allocate resources effectively, and take proactive measures to prevent and reduce crime. Furthermore, the system's ability to identify patterns and commonalities across cases can facilitate the development of targeted intervention strategies and support efforts to improve public safety in a more holistic and comprehensive manner.

2.8 Testing and Implementation

2.8.1 Testing

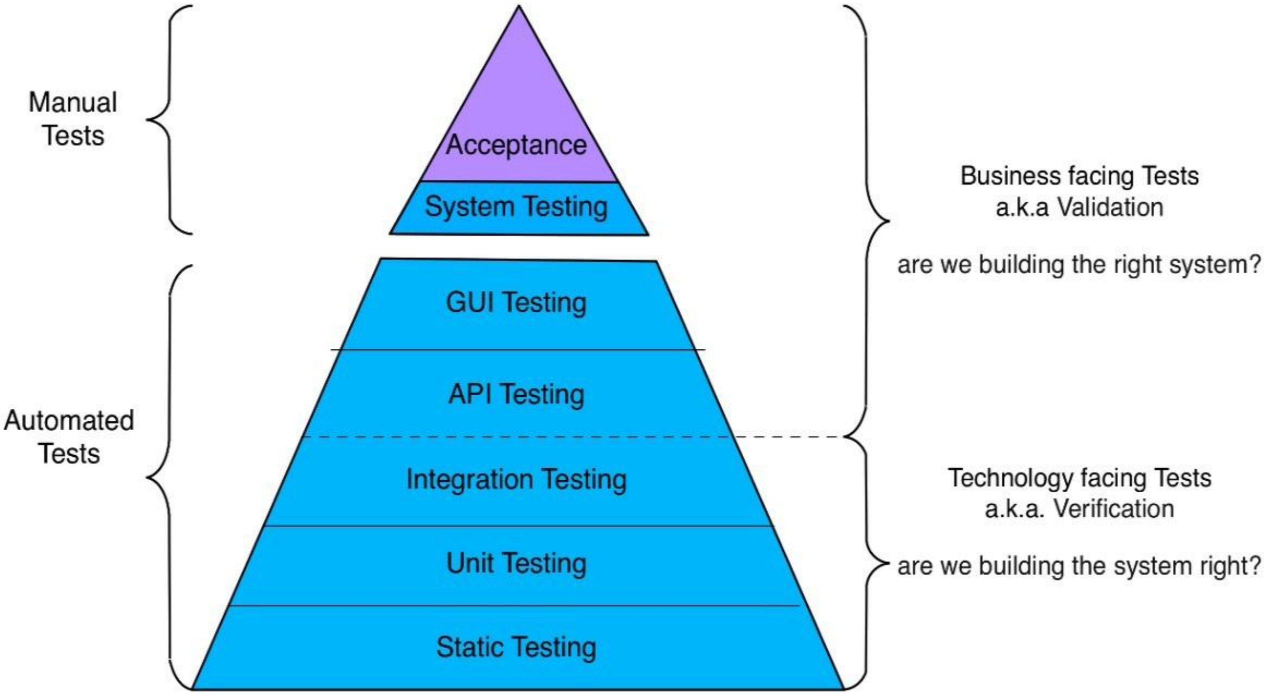


Figure 2.8.1.1: Testing Phase.

The diagram provided above outlines a systematic and rigorous approach to software testing that extends across all stages of development, encompassing applications and defect resolution. This comprehensive testing methodology consistently elevates the quality of applications, ensuring they meet the highest standards. Testing stands as an integral element within the software development life cycle, serving as a means to detect and rectify issues that might have silently accumulated throughout various development phases. Moreover, it is crucial to validate the overall quality of the product or software application. Regardless of the specific component within our SDLC, which includes Accident Case Analysis, Crime Analysis, Case Document Classification, and Crimes Against Women Analysis, each stage necessitates rigorous testing to affirm its functionality and reliability.

1. **Unit Testing**

Unit testing is pivotal in our project's development, validating component functionality and precision individually. For the Accident Case Analysis, tests verify data pre-processing, ARIMA predictions, and visualization tools, ensuring accurate accident forecasts. They also validate data processing and transformation functions crucial for precise analysis in all components.

2. **Module Testing**

In module testing, each of the four key components of our system, namely Accident Case Analysis, Crime Analysis, Case Document Classification, and Crimes Against Women Analysis, undergoes individual evaluation. This phase verifies that each component functions correctly in isolation, addressing specific functionalities and predictions. By examining and validating the accuracy of these individual modules, we ensure their reliability and effectiveness within the larger system Integration Testing.

3. **Integration Testing**

Integration testing involves a rigorous examination of the interactions between these components. This testing evaluates the effectiveness of the modules working together, verifying the smooth flow of data between them and their ability to generate accurate and coherent results as a collective. This critical step serves to detect and address any integration-related issues, thus ensuring the seamless functioning of the integrated web application.

4. **System Testing**

System testing encompasses a comprehensive evaluation of the entire integrated web application. It verifies that the system functions as a cohesive whole, accurately predicting accident percentages, crime patterns, and document classifications. This phase validates the system's ability to provide valuable insights and user-friendly visualizations. System testing is essential to ensure the reliability and robustness of the entire platform.

5. **User Acceptance Testing**

User Acceptance Testing (UAT) is a critical phase where the system is put to the test by actual users, such as police departments and investigators. During UAT, users interact with the system, input parameters, and assess its performance, usability, and the accuracy of predictions. Feedback and observations from users play a pivotal role in refining and fine-tuning the system to meet their needs effectively.

6. **Maintenance**

The maintenance phase is an ongoing commitment to ensure the system's long-term reliability and relevance. This includes addressing real-world issues, updating datasets, and incorporating user feedback for continuous improvement. Maintenance aims to keep the system valuable and dependable for law enforcement agencies and investigators.

Our integrated system, comprising the Accident Case Analysis, Crime Analysis, Case Document Classification, and Crimes Against Women Analysis components, has successfully passed all testing phases without encountering errors. We systematically tested each component independently to ensure rigorous examination and guarantee reliability and accuracy in delivering insights and predictions to law enforcement agencies and investigators

Test cases that are done for each testing method is shown below.

Test Case No	Test Case 01
Description	Predict Accident Percentage for specified division name and future year
Test Steps	1. Login to the system. 2. Navigate to accident 3. Navigate to accident prediction section 4. Type or select the division name 5. Select the future year 6. Click predict button
Test Data	String
Expected Result	Successfully predicting the accident percentage and fatal percentage for future years
Actual Result	Pass
User Role	Police

Table 2.8.1.1 Predict Accident Percentage Test Case 01

Test Case No	Test Case 02
Description	Predict Accident and fatal percentage for future years
Test Steps	1. Login to the system. 2. Navigate to accident 3. Navigate to accident prediction section 4. Enter or select division name

	5. Click analysis button
Test Data	String
Expected Result	Predicted percentages for accident and fatal for future years displayed and graph also displayed.
Actual Result	Pass
User Role	Police

Table 2.8.1.2 Predict accident and fatal percentage for future years - Test Case 02

Test Case No	Test Case 03
Description	Predict cause of the accident and the percentage for each cause
Test Steps	1. Login to the system. 2. Navigate to accident 3. Navigate to cause prediction section 4. Enter or select future year 5. Click submit
Test Data	String
Expected Result	Predicted percentage for each cause should be displayed and the graph for those values should be displayed
Actual Result	Pass
User Role	Police

Table 2.8.1.3 Cause Prediction - Test Case 03

Test Case No	Test Case 04
Description	Provide prevention strategies
Test Steps	1. Login to the system. 2. Navigate to accident 3. Navigate to cause prediction section 4. Click prevention strategy button 5. Select cause 6. Click submit
Test Data	String
Expected Result	All prevention strategies displayed
Actual Result	Pass
User Role	Police

Table 2.8.1.4 Prevention Strategies Test Case 04

Test Case No	Test Case 05
Description	Select the type of analysis in the pattern analysis section
Test Steps	1. Login to the system. 2. Navigate to accident. 3. Navigate to pattern analysis section. 4. Select one option from (Day, Hour, Month).
Test Data	String
Expected Result	Predicted and statistical graph for the selected option and the most accident

	occurring day/hour/month should be displayed
Actual Result	Pass
User Role	Police

Table 2.8.1.5 Pattern Analysis Test Case 05

2.8.2 Implementation

During the implementation phase, we capitalized on the extensive datasets gathered throughout the research and development stages of the project. These datasets, comprising accident and crime-related data, case documents, and demographic information, formed the bedrock upon which our integrated web application was constructed. Employing the Django framework in conjunction with Python, we seamlessly translated our research findings into a fully functional and user-friendly system. This implementation not only facilitates the prediction of accident percentages for future years and the forecasting of causes but also encompasses accident case analysis and pattern analysis. Through the utilization of machine learning algorithms, statistical models, and data visualization tools, we have successfully crafted a potent platform that empowers law enforcement agencies and investigators with the insights necessary to enhance safety planning, crime prevention, and decision-making. Our meticulous implementation process ensures that our system is well-prepared to tackle the intricate challenges in the realm of accident case analysis, ultimately contributing to the enhancement of public safety and the effectiveness of law enforcement efforts.

Code

Accident case analysis implementation

```python
from django.shortcuts import render, redirect
from .accident_Prediction_RP_FINAL import perform_prediction
from .forms import YearForm,ClusterNameForm
import matplotlib.pyplot as plt
from io import BytesIO
import base64
import pandas as pd
import matplotlib
import io
from sklearn.cluster import KMeans
from statsmodels.tsa.arima.model import ARIMA
from django.http import HttpResponse, HttpResponseNotFound
import csv
```

Figure 2.8.2.1: Import libraries for accident prediction

he imported code snippets within our component serve as essential building blocks for the functionality of our integrated web application. We rely on Django for robust web framework support, enabling smooth user interactions. The 'accident_Prediction_RP_FINAL' module is responsible for incorporating predictive analysis using the ARIMA model to forecast accident percentages and fatal percentages for future years and divisions. To collect user inputs for customized queries, we utilize forms like 'YearForm' and 'ClusterNameForm.' 'Matplotlib' aids in data visualization, allowing us to generate graphical representations for user insights. Data manipulation and analysis are efficiently carried out with the assistance of 'pandas,' while 'io' is instrumental in data handling. For K-means clustering, we rely on the 'KMeans' module from 'scikit-learn,' and 'statsmodels' comes into play for time series analysis using ARIMA. This code framework ensures a seamless user experience and delivers accurate predictions, ultimately enhancing the overall functionality of our system.

accident_Prediction_RP_FINAL.py ×

policeAnalysis > accidentAnalysis > accident_Prediction_RP_FINAL.py > perform_prediction

```python
# Modify accident_Prediction_RP_FINAL.py
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA
import warnings
import joblib
import numpy as np

def perform_prediction(division_name, selected_year=None):
    # Disable the warnings
    warnings.filterwarnings("ignore")

    # Load the dataset
    # data = pd.read_csv('C:\\Users\\HP\\DatasetAccident.csv')
    data = pd.read_csv('DatasetAccident.csv')


    # Select relevant columns
    data = data[['DIVISION_NAME', 'YEAR', 'VE_TOTAL', 'FATALS']]

    # Convert YEAR column to datetime
    data['YEAR'] = pd.to_datetime(data['YEAR'], format='%Y')
```

Figure 2.8.2.2: Read dataset and use necessary columns

The presented code snippet is encapsulated within a Python script named 'accident_Prediction_RP_FINAL.py,' serving as a pivotal element within the accident case analysis component of our system. This script plays a fundamental role in conducting predictive analysis, utilizing the ARIMA (AutoRegressive Integrated Moving Average) model for the purpose of forecasting accident percentages and fatal percentages with a specific focus on divisions and years. To ensure smooth execution, the initial step involves the suppression of warnings. Subsequently, the script proceeds to load the accident dataset and meticulously selects pertinent columns that contain essential information, including division names, years, total accidents (VE_TOTAL), and fatalities (FATALS). Furthermore, it takes a proactive approach by converting the 'YEAR' column into a datetime format, thereby facilitating time-based analysis

```python
# Group by DIVISION_NAME and YEAR, calculate total accidents and fatalities
grouped_data = data.groupby(['DIVISION_NAME', 'YEAR']).agg({'VE_TOTAL': 'sum', 'FATALS': 'sum'}).reset_index()

# Define a function to fit ARIMA model and make predictions
def predict_arima(series):
    model = ARIMA(series, order=(1, 0, 0))
    model_fit = model.fit()
    forecast = model_fit.forecast(steps=3)
    return forecast

# Input the future years
future_years = [2024, 2025, 2026]

# Create an empty DataFrame to store the predictions
predictions = pd.DataFrame(columns=['DIVISION_NAME', 'YEAR', 'Accident_Percentage', 'Fatal_Percentage'])
```

Figure 2.8.2.3: fit ARIMA model

In this section of the code, we embark on a crucial data preparation phase, which forms the bedrock for subsequent analysis. Initially, the accident dataset is meticulously organized by grouping the data based on 'DIVISION_NAME' and 'YEAR.' This grouping mechanism sets the stage for structured and systematic examination. Within this organized framework, the data is aggregated, enabling the calculation of total accident occurrences (VE_TOTAL) and fatalities (FATALS) for each division-year combination. This aggregation process results in a well-structured dataset, providing a comprehensive view of accident statistics. Furthermore, the code introduces a pivotal function christened 'predict_arima,' designed to implement the ARIMA (AutoRegressive Integrated Moving Average) model for predictive analytics. The 'predict_arima' function takes as its input a time series data series, embarks on the fitting of the ARIMA model, employing specific parameter settings (order= (1, 0, 0)), and embarks on forecasting accident trends. These forecasts are extended to future years, with a forward-looking scope of three years (specifically, 2024, 2025, 2026). To capture and organize these forecasts systematically, the code initializes an empty DataFrame named 'predictions.' This DataFrame serves as a structured repository for storing forecasted accident percentages and fatal percentages, meticulously categorized by division and year. This intricate code segment lays a robust foundation for predicting key accident-related metrics and generating invaluable insights within the framework of our integrated web application.

```
# Iterate over each district
for district in data['DIVISION_NAME'].unique():
    # Get the data for the current district
    district_data = grouped_data[grouped_data['DIVISION_NAME'] == district]

    # Extract the relevant columns
    years = district_data['YEAR']
    accident_totals = district_data['VE_TOTAL']
    fatal_totals = district_data['FATALS']

    # Fit ARIMA model and make predictions for accidents
    accident_predictions = predict_arima(accident_totals)

    # Fit ARIMA model and make predictions for fatalities
    fatal_predictions = predict_arima(fatal_totals)

    # Create a DataFrame for the predictions
    district_predictions = pd.DataFrame({
        'DIVISION_NAME': [district] * len(future_years),
        'YEAR': future_years,
        'Accident_Percentage': accident_predictions / accident_totals.sum(),
        'Fatal_Percentage': fatal_predictions / fatal_totals.sum()
    })

    # Append the district predictions to the overall predictions DataFrame
    predictions = pd.concat([predictions, district_predictions])
```

Figure 2.8.2.4: Predict accident and fatal percentage

Within this code section, an iterative process unfolds as a loop cycle through the unique division names found within the accident dataset. For each distinct division, a precise data extraction operation is initiated, retrieving the corresponding data from the meticulously grouped and aggregated dataset. The focus narrows down to essential columns, encompassing 'YEAR,' 'VE_TOTAL' (representing total accidents), and 'FATALS' (indicating total fatalities), setting the stage for in-depth analysis. The crux of the analysis hinges on the application of two distinct ARIMA models. These models are meticulously tailored to forecast trends related to accidents and fatalities within the current division. These predictions are anchored in historical data, offering a glimpse into potential future outcomes. To methodically capture and organize these predictive insights, a fresh DataFrame named 'district_predictions' is summoned into existence. This DataFrame acts as a structured repository for consolidating the anticipated accident and fatality percentages. These percentages are expressly attributed to the specified forthcoming years (namely, 2024, 2025, 2026), carefully linked to the current division under examination. The comprehensive code execution extends further details, comprehensively chronicled in the accompanying appendix.

```python
form = ClusterNameForm()

if request.method == 'POST':
    form = YearForm(request.POST)
    if form.is_valid():
        selected_year = form.cleaned_data['selected_year']

        # Load the forecast data from the CSV file
        forecasts_data = pd.read_csv(r'C:\Users\HP\forecasts_data.csv')

        forecast_outcome = f"<p>Prediction Outcome for {selected_year}:</p><br>"
        cluster_info = []
        # clusters = []  # Initialize clusters list

        for index, row in forecasts_data.iterrows():
            cluster_name = row['Cluster Name']
            abbreviated_name = cluster_shortcuts[cluster_name]
            forecast_amount = row[f'Forecast {selected_year}']
            forecast_outcome += f"<p>Prediction for {cluster_name} ({abbreviated_name}):<br>"
            forecast_outcome += f"Amount: {forecast_amount:.2f}</p>"
            cluster_info.append(f"{cluster_name}: {abbreviated_name}")



        # Create a histogram graph for the selected year
        forecast_values = forecasts_data[f'Forecast {selected_year}']
        matplotlib.use('Agg')

        plt.figure(figsize=(10, 6))
        bars = plt.bar(range(len(forecasts_data['Cluster Name'])), forecast_values, color='#002863', edgecolor='black')
        plt.xlabel('Causes')
        plt.ylabel('Number of Accidents')
        plt.title(f'Forecasted Number of Accidents for {selected_year}')
```

Figure 2.8.2.5: cause prediction code

3. RESULTS & DISCUSSION

3.1 Results

Within the integrated web application, the accident case analysis component stands as the cornerstone of our comprehensive system. This particular segment provides users with a user-friendly interface where they can input precise details, including the division of interest and future years. Subsequently, leveraging the predictive power of our machine learning algorithms, the system generates tailored forecasts for accident and fatality percentages specific to the user's chosen parameters. These predictions play a pivotal role in empowering law enforcement agencies with critical insights to shape their strategies and interventions effectively. By proactively addressing potential accident scenarios in the upcoming years, law enforcement can contribute significantly to the overarching goal of enhancing road safety and reducing accident-related harm. To visualize these predictions and insights, users can refer to Figure 3.1.1, which graphically represents the results, enhancing their interpretability and utility.

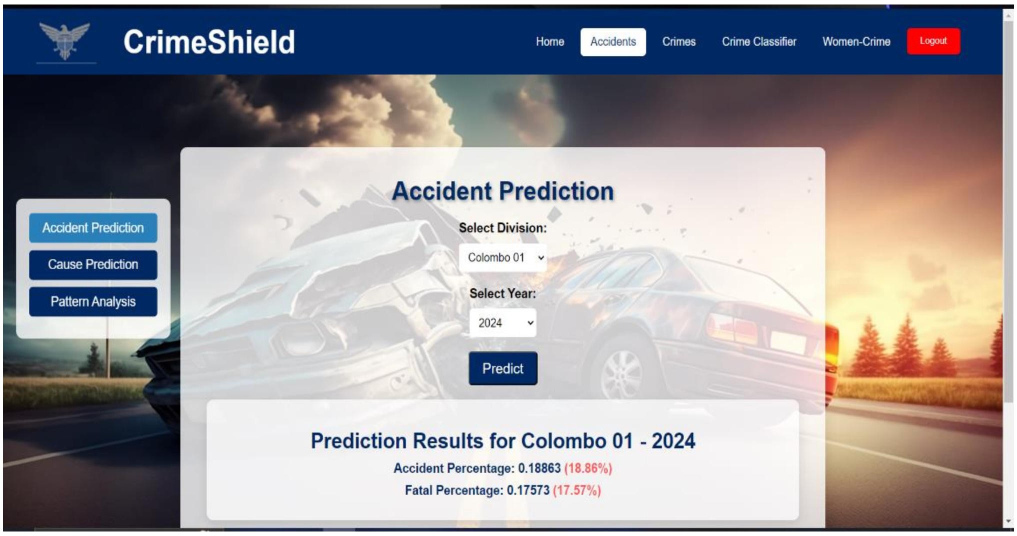


Figure 3.1.1: - Prediction results output

To accomplish this, we employed the ARIMA model for forecasting future years. The results are further illustrated through the accompanying graph in the subsequent image**.**

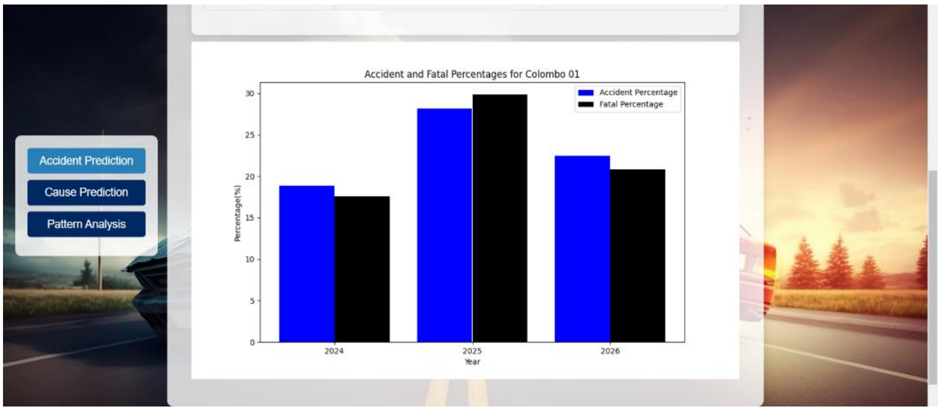


Figure 3.1.2: - Accident prediction graph

Moreover, the predictive analysis of accident causes was carried out utilizing the K-means clustering technique, a data-driven method that segments causes into distinct groups based on similarity. The resulting outcomes reveal five distinct cause clusters, each accompanied by its corresponding forecasted values for the specified future year, as elaborated in the subsequent section presenting the results. Additionally, for enhanced user interaction and accessibility, user interfaces pertaining to various components and functionalities of the system have been thoughtfully incorporated into the appendix section of this report, ensuring a comprehensive and user-friendly experience for system users.

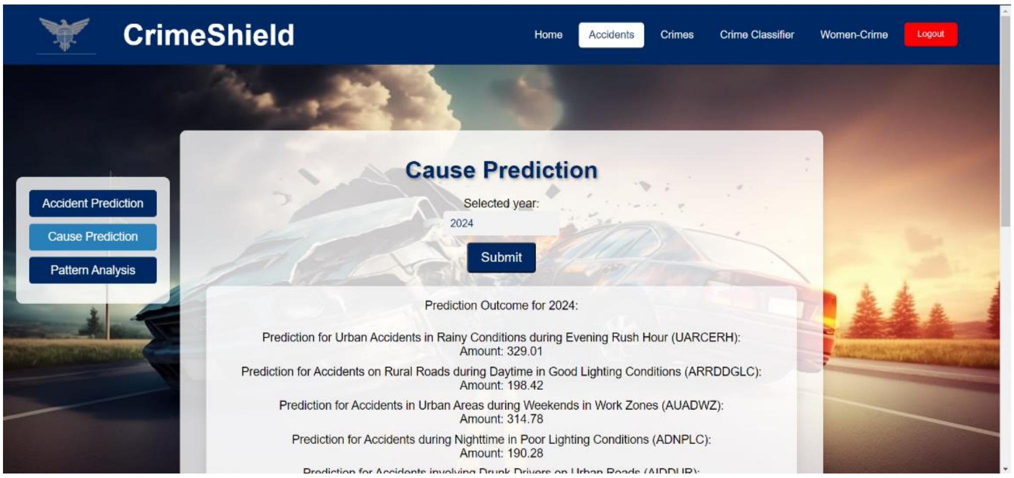


Figure 3.1.3: Cause Prediction outcome

Each of the four components generates its unique predictions and analyses, equipping law enforcement agencies with valuable insights to formulate proactive strategies and plans aimed at enhancing public safety and preventing incidents.

3.2 Research Finding

The culmination of our research efforts across diverse system components has yielded a ground-breaking approach to bolstering public safety. In the domain of Accident Case Analysis, our system effectively harnesses the ARIMA model's predictive capabilities to anticipate accident percentages, making it invaluable for law enforcement agencies in their road safety planning endeavours. Concurrently, the implementation of K-means clustering within our system enables the identification of underlying causes of accidents, facilitating the formulation of highly targeted and tailored prevention strategies. This multifaceted approach equips police departments with a comprehensive toolkit to not only comprehend the intricacies of accidents but also to take proactive measures in reducing their occurrence. In sum, our research findings represent a robust and innovative resource for comprehending, mitigating, and proactively addressing the myriad facets of crime and accidents, ultimately culminating in the enhancement of public safety.

3.3 Discussion

The police accident case analysis component within this comprehensive project serves as a linchpin in the broader mission of bolstering road safety. By harnessing sophisticated techniques, notably the ARIMA model and K-means clustering, the system excels in its capacity to forecast accident percentages for forthcoming years and to categorize accidents based on a multitude of influential factors. These predictive capabilities furnish law enforcement agencies with a wealth of actionable insights, empowering them to proactively design and implement prevention strategies tailored to the unique characteristics of each scenario. Beyond prediction, the component's forte lies in its aptitude for discerning intricate patterns concealed within accident data. This analytical prowess facilitates the formulation of precisely targeted safety measures, optimizing resource allocation and intervention strategies. The introduction of an innovative web application, the first of its kind in Sri Lanka, represents a milestone achievement, enabling users to access bespoke predictions and conduct in-depth analyses tailored to specific divisions and timeframes. This holistic approach, encapsulating accurate predictions, root cause identification, meticulous case analysis, and well-informed prevention strategies, equips law enforcement authorities with a multifaceted toolkit. This toolkit, in turn, enhances their ability to mitigate accidents, elevate road safety standards, and ultimately realize the overarching project's mission—to enhance public safety through data-driven insights and proactive measures. In summation, the police accident case analysis component signifies a substantial leap toward accident reduction and the creation of safer road networks.

4. FUTURE WORK

The following can be done to extend our work:

Future advancements in police accident case analysis are poised to build upon the existing foundation, with a primary focus on the expansion and enhancement of datasets. The incorporation of a more extensive dataset will play a pivotal role in refining the accuracy of accident percentage predictions and deepening our understanding of the root causes of accidents. To bolster responsiveness and effectiveness, the integration of real-time data sources and traffic patterns is on the horizon, enabling more immediate accident prevention interventions. Embracing advanced machine learning models and geospatial analysis techniques holds the potential to pinpoint accident hotspots and contributing factors with a heightened degree of precision. Furthermore, we envision the continuous development of user interfaces and visualization tools, ensuring that the insights generated remain easily accessible and actionable for law enforcement agencies and policymakers alike. To fortify the system's resilience, we also plan to devise a comprehensive Contingency Plan, pre-emptively addressing potential risks and vulnerabilities that could affect its performance and security. Through these future endeavours, this component remains committed to advancing road safety measures and fortifying accident prevention efforts effectively.

5. CONCLUSION

In conclusion, the machine learning system developed for police accident case analysis, which harnessed a diverse array of machine learning algorithms, has paved the way for significant advancements in enhancing public safety and road accident prevention. By delving into the analysis of accident locations, predictive modelling, clustering of accidents based on causes, and the generation of prevention strategies, this component has armed law enforcement agencies with data-driven insights derived from past data. The comprehensive toolkit provided encompasses not only the prediction and prevention of accidents but also insightful pattern analysis. These data-driven efforts, empowered by historical data, have fortified our mission to reduce accidents and improve road safety. As we look to the future, the potential for expanding datasets, incorporating real-time information, and refining analytical tools holds the promise of further elevating the effectiveness of our accident case analysis component, ultimately resulting in safer roadways and enhanced public safety.

6. REFERENCE

[1] D. Li, J. Wu, and D. Peng, *"Online Traffic Accident Spatial-Temporal Post-Impact Prediction Model on Highways Based on Spiking Neural Networks,"* Journal of Advanced Transportation, vol.2021, pp. 20, Dec 2021.

[2] V. Prasannakumar, H. Vijith, R. Charutha, and N. Geetha, *"Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment,"* Procedia – Social and Behavioural Sciences, Dec 2011.

[3] M. Manzoor, M. Umer, S. Sadiq, et al., *"RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model,"* in IEEE Access, pp. 02 – 05, 2021.

[4] S. M. N. A. Senanayake, and S. Joshi, *"A road accident pattern miner (RAP miner),"* Journal of Information and Telecommunication, Aug 2021.

[5] J. Yang, S. Han, and, Y. Chen, *"Prediction of Traffic Accident Severity Based on Random Forest,"* Journal of Advanced Transportation, pp. 1-8, Feb 2023.

[6] M. I. Sameen and B. Pradhan, "*Severity Prediction of Traffic Accidents with Recurrent Neural Networks*," Applied Sciences, June 2017.

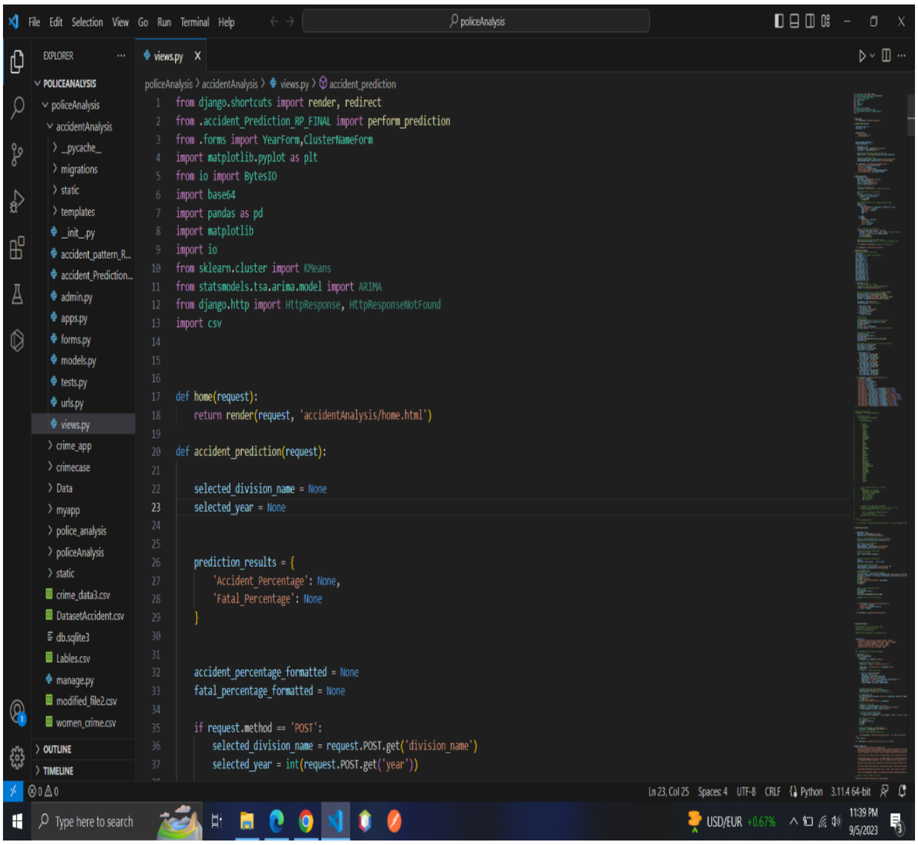
[7] S. K. Singh, *"Road traffic accidents in India: issues and challenges,"* Transportation research procedia, Vol 25, pp. 4708-4719, 2017.

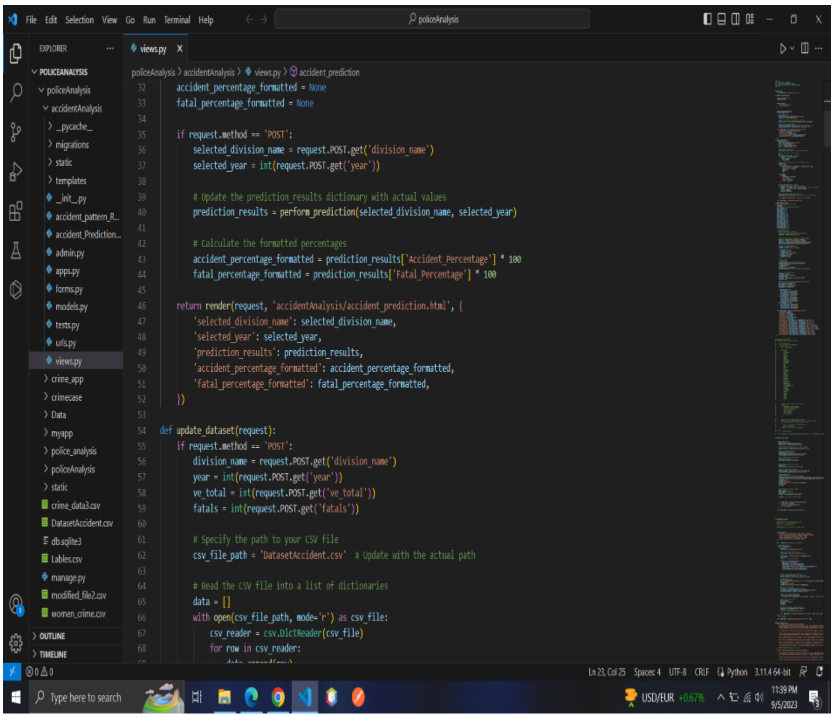
[8] T. Beshah and S. Hill*, "Predicting Road Crash Severity Using Classifier Models and Crash Hotspots,"* In AAAI Spring Symposium: Artificial Intelligence for Development, vol. 24, pp.1173–1181, Jan 2010.

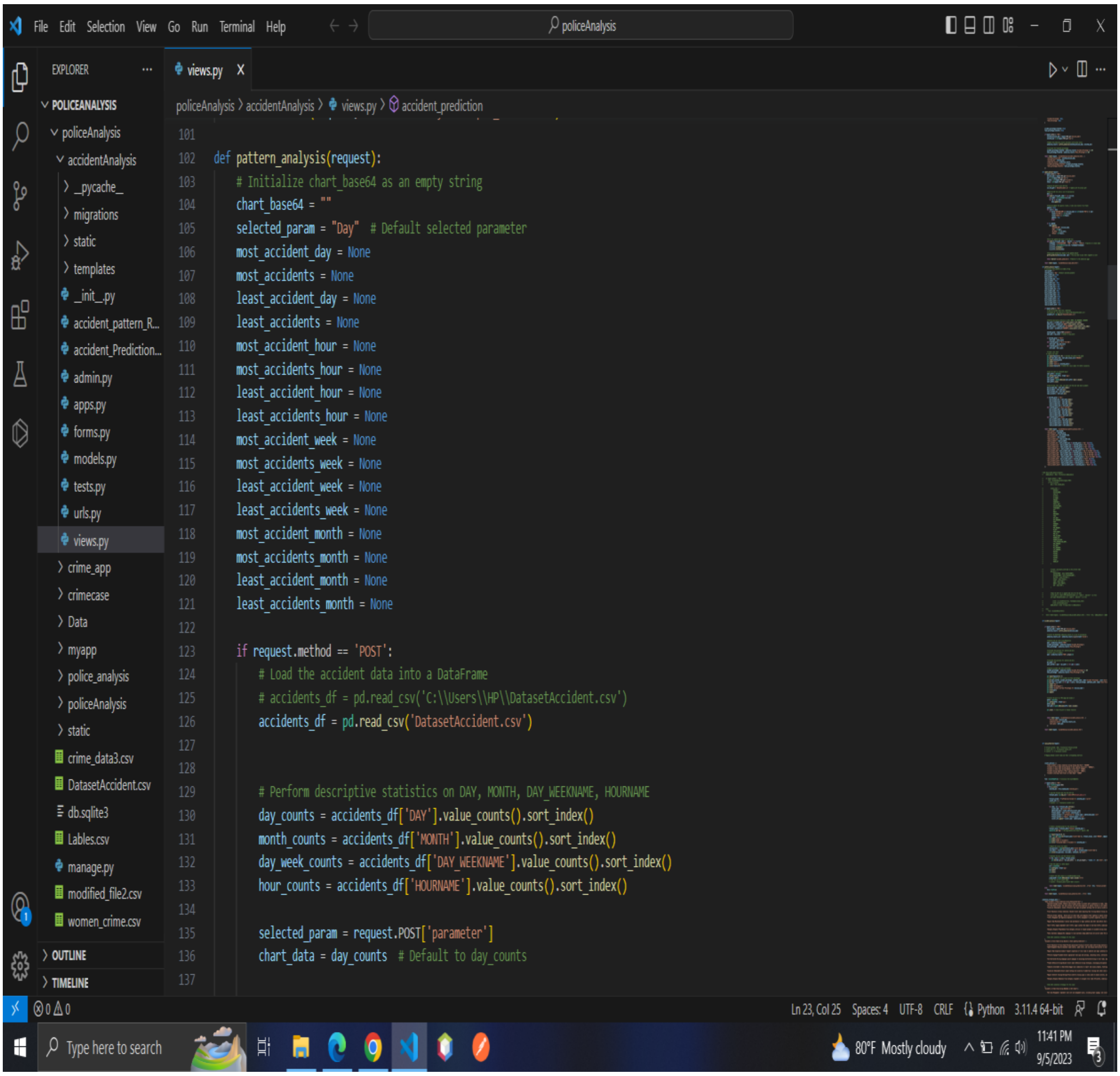
[9] M. K. Islam, I. Reza, U. Gazder, et al., *"Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia,"* Applied Sciences vol. 12, Nov 2022.

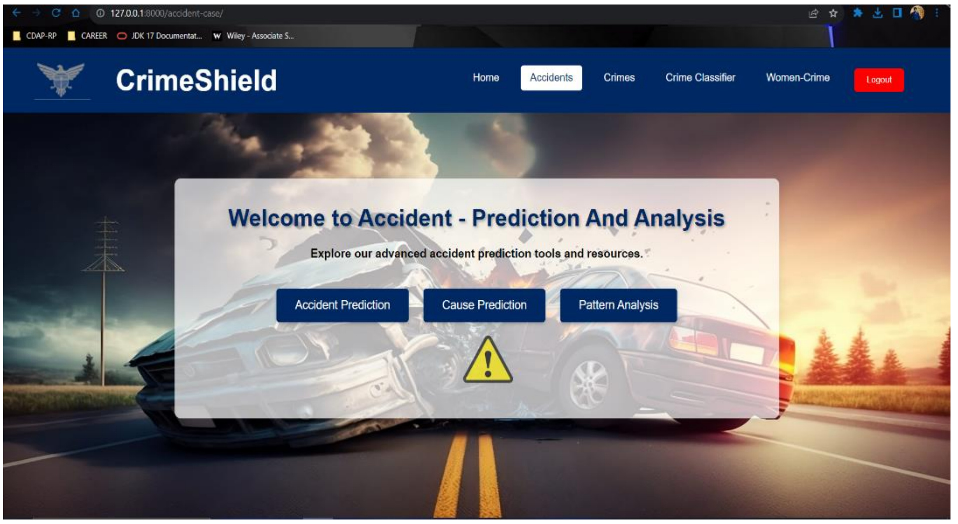
[10] J. Fattah, L. Ezzine, Z. Aman, et al. *"Forecasting of demand using ARIMA model,"* International Journal of Engineering Business Management, vol. 10, Oct 2018.

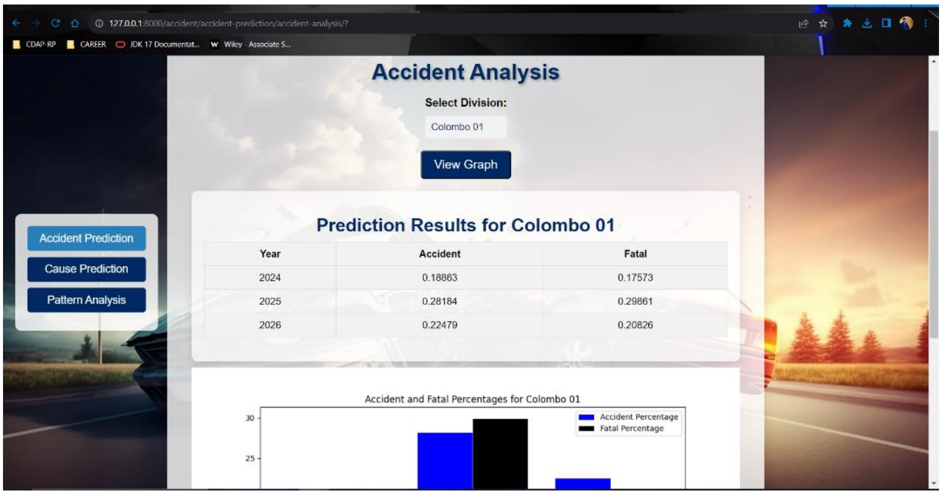
APPENDICES

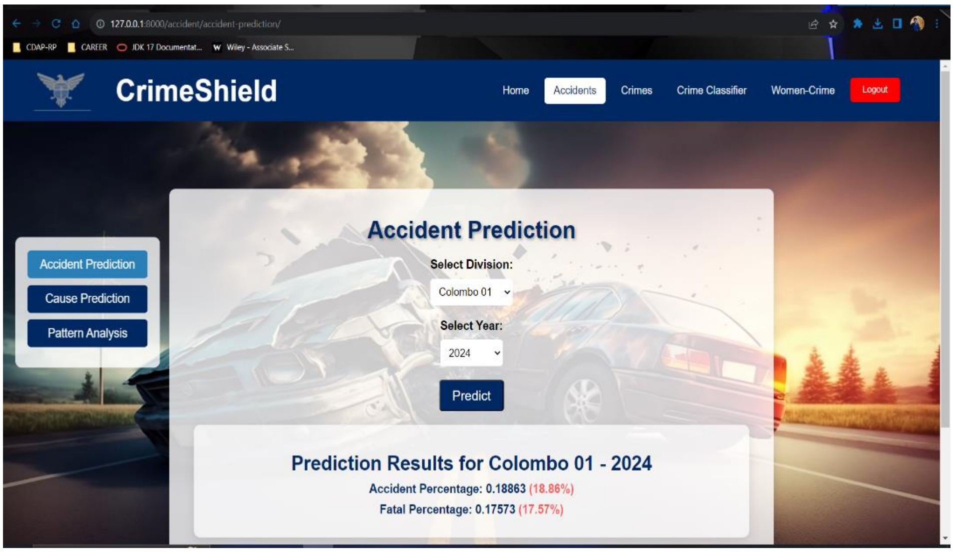

from django.shortcuts import render, redirect
from .accident_Prediction_RP_FINAL import perform_prediction
from .forms import YearForm,ClusterNameForm
import matplotlib.pyplot as plt
from io import BytesIO
import base64
import pandas as pd
import matplotlib
import io
from sklearn.cluster import KMeans
from statsmodels.tsa.arima.model import ARIMA
from django.http import HttpResponse, HttpResponseNotFound
import csv
def home(request):
return render(request, 'accidentAnalysis/home.html')
def accident_prediction(request):
selected_division_name = None
selected_year = None
prediction_results = {
'Accident_Percentage': None,
'Fatal_Percentage': None
}
accident_percentage_formatted = None
fatal_percentage_formatted = None
if request.method == 'POST':
selected_division_name = request.POST.get('division_name')
selected_year = int(request.POST.get('year'))

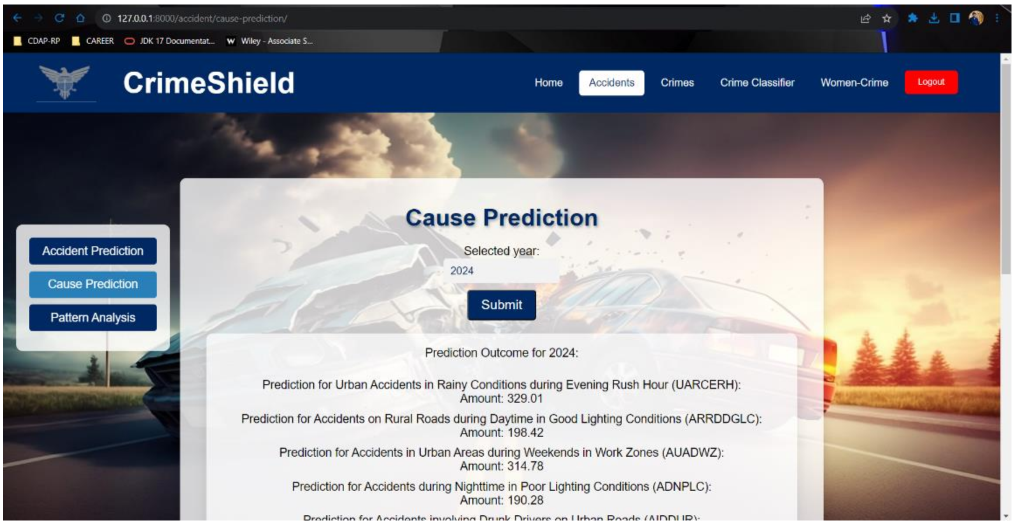

accident_percentage_formatted = None
fatal_percentage_formatted = None
if request.method == 'POST':
selected_division_name = request.POST.get('division_name')
selected_year = int(request.POST.get('year'))
# Update the prediction_results dictionary with actual values
prediction_results = perform_prediction(selected_division_name, selected_year)
# Calculate the formatted percentages
accident_percentage_formatted = prediction_results['Accident_Percentage'] * 100
fatal_percentage_formatted = prediction_results['Fatal_Percentage'] * 100
return render(request, 'accidentAnalysis/accident_prediction.html', {
'selected_division_name': selected_division_name,
'selected_year': selected_year,
'prediction_results': prediction_results,
'accident_percentage_formatted': accident_percentage_formatted,
'fatal_percentage_formatted': fatal_percentage_formatted,
})
def update_dataset(request):
if request.method == 'POST':
division_name = request.POST.get('division_name')
year = int(request.POST.get('year'))
ve_total = int(request.POST.get('ve_total'))
fatals = int(request.POST.get('fatals'))
# Specify the path to your CSV file
csv_file_path = 'DatasetAccident.csv' # Update with the actual path
# Read the CSV file into a list of dictionaries
data = []
with open(csv_file_path, mode='r') as csv_file:
csv_reader = csv.DictReader(csv_file)
for row in csv_reader:

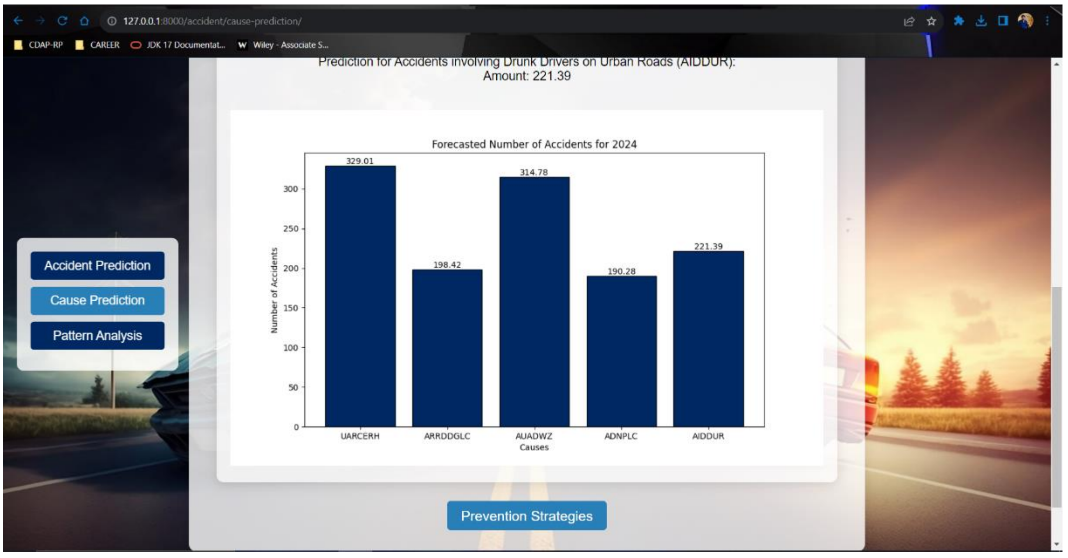

File Edit Selection View Go Run Terminal Help
policeAnalysis
EXPLORER
views.py
policeAnalysis > accidentAnalysis > views.py > accident_prediction
def pattern_analysis(request):
# Initialize chart_base64 as an empty string
chart_base64 = ""
selected_param = "Day" # Default selected parameter
most_accident_day = None
most_accidents = None
least_accident_day = None
least_accidents = None
most_accident_hour = None
most_accidents_hour = None
least_accident_hour = None
least_accidents_hour = None
most_accident_week = None
most_accidents_week = None
least_accident_week = None
least_accidents_week = None
most_accident_month = None
most_accidents_month = None
least_accident_month = None
least_accidents_month = None
if request.method == 'POST':
# Load the accident data into a DataFrame
# accidents_df = pd.read_csv('C:\\Users\\HP\\DatasetAccident.csv')
accidents_df = pd.read_csv('DatasetAccident.csv')
# Perform descriptive statistics on DAY, MONTH, DAY_WEEKNAME, HOURNAME
day_counts = accidents_df['DAY'].value_counts().sort_index()
month_counts = accidents_df['MONTH'].value_counts().sort_index()
day_week_counts = accidents_df['DAY_WEEKNAME'].value_counts().sort_index()
hour_counts = accidents_df['HOURNAME'].value_counts().sort_index()
selected_param = request.POST['parameter']
chart_data = day_counts # Default to day_counts
POLICEANALYSIS
policeAnalysis
accidentAnalysis
pycache
migrations
static
templates
init.py
accident_pattern_R...
accident_Prediction...
admin.py
apps.py
forms.py
models.py
tests.py
urls.py
views.py
crime_app
crimecase
Data
myapp
police_analysis
policeAnalysis
static
crime_data3.csv
DatasetAccident.csv
db.sqlite3
Lables.csv
manage.py
modified_file2.csv
women_crime.csv
OUTLINE
TIMELINE
Ln 23, Col 25 Spaces: 4 UTF-8 CRLF Python 3.11.4 64-bit
Type here to search
80°F Mostly cloudy
11:41 PM
9/5/2023

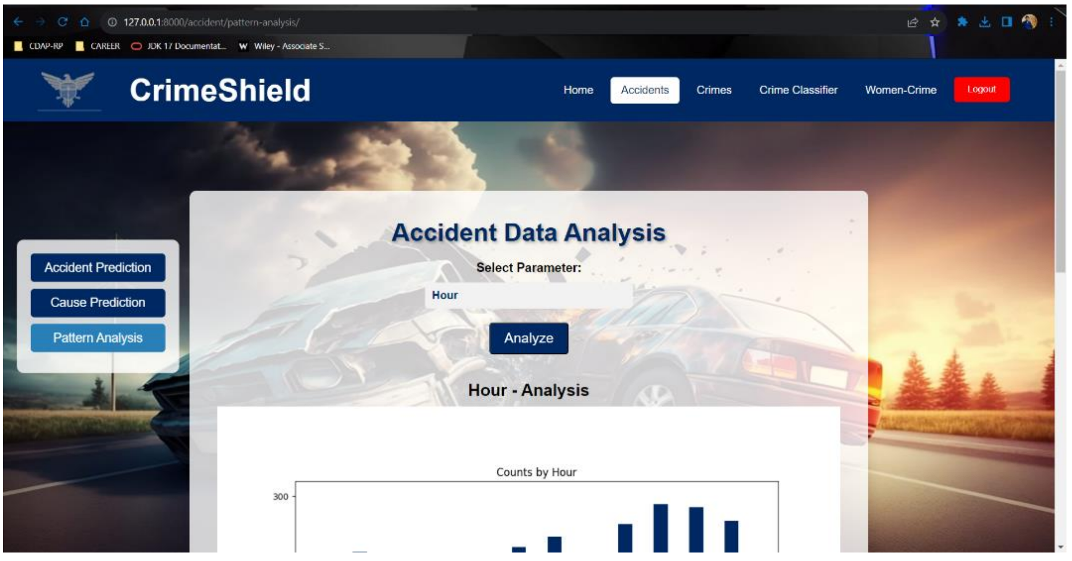

CrimeShield
Home
Accidents
Crimes
Crime Classifier
Women-Crime
Logout
Welcome to Accident - Prediction And Analysis
Explore our advanced accident prediction tools and resources.
Accident Prediction
Cause Prediction
Pattern Analysis


Accident Analysis
Select Division:
Colombo 01
View Graph
Accident Prediction
Cause Prediction
Pattern Analysis
Prediction Results for Colombo 01
Year Accident Fatal
2024 0.18863 0.17573
2025 0.28184 0.29861
2026 0.22479 0.20826
Accident and Fatal Percentages for Colombo 01
Accident Percentage
Fatal Percentage
30
25


CrimeShield
Home
Accidents
Crimes
Crime Classifier
Women-Crime
Logout
Accident Prediction
Cause Prediction
Pattern Analysis
Accident Prediction
Select Division:
Colombo 01
Select Year:
2024
Predict
Prediction Results for Colombo 01 - 2024
Accident Percentage: 0.18863 (18.86%)
Fatal Percentage: 0.17573 (17.57%)


127.0.0.1:8000/accident/cause-prediction/
CrimeShield
Home
Accidents
Crimes
Crime Classifier
Women-Crime
Logout
Accident Prediction
Cause Prediction
Pattern Analysis
Cause Prediction
Selected year:
2024
Submit
Prediction Outcome for 2024:
Prediction for Urban Accidents in Rainy Conditions during Evening Rush Hour (UARCERH):
Amount: 329.01
Prediction for Accidents on Rural Roads during Daytime in Good Lighting Conditions (ARRDDGLC):
Amount: 198.42
Prediction for Accidents in Urban Areas during Weekends in Work Zones (AUADWZ):
Amount: 314.78
Prediction for Accidents during Nighttime in Poor Lighting Conditions (ADNPLC):
Amount: 190.28


127.0.0.1:8000/accident/cause-prediction/
Prediction for Accidents involving Drunk Drivers on Urban Roads (AIDDUR):
Amount: 221.39
Accident Prediction
Cause Prediction
Pattern Analysis
Forecasted Number of Accidents for 2024
Number of Accidents
Causes
UARCERH
ARRDDGLC
AUADWZ
ADNPLC
AIDDUR
329.01
198.42
314.78
190.28
221.39
Prevention Strategies


127.0.0.1:8000/accident/pattern-analysis/
CrimeShield
Home
Accidents
Crimes
Crime Classifier
Women-Crime
Logout
Accident Prediction
Cause Prediction
Pattern Analysis
Accident Data Analysis
Select Parameter:
Hour
Analyze
Hour - Analysis
Counts by Hour

127.0.0.1:8000/accident/pattern-analysis/
Accident Prediction
Cause Prediction
Pattern Analysis
Count
Most Accident Occurring - Hour: 7:00pm-7:59pm(Accidents: 295)
Least Accident Occurring - Hour: Unknown Hours (Accidents: 27)