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

(Police Accident Case Analysis)

Project ID:
TMP-23-224

Project Proposal Report

Dharsan.R - IT20003982

Supervisor: Ms. Hansika Mahaadikara

Co-Supervisor: Ms. Sanjeevi Chandrasiri

BSc (Hons) in Information Technology

Specializing in Software Engineering

Department of Computer Science & Software Engineering (CSSE)

Sri Lankan Institute of Information Technology
Sri Lanka

March 2023

Declaration of The Candidate & Supervisor

We declare that this is our own work and this proposal 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 our 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.

Name	Student ID	Signature
Dharsan.R	IT20003982	

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor
(Ms. Hansika Mahaadikara)

Date 17/09/2023

ABSTRACT

Crime prevention is a critical aspect of public safety, and police departments are always looking for new ways to prevent crimes from occurring. The importance of crime prevention in public safety cannot be overemphasized. Over the years, police investigators have relied on traditional policing methods to prevent crimes. However, with the advancements in technology, there is a need for new approaches such as machine learning in police case analysis. Machine learning can identify patterns and correlations in data that might be difficult for humans to detect. Hence, this research proposal seeks to explore the use of machine learning in police accident case analysis to predict and prevent accidents. The research proposal aims to use machine learning in police case analysis to predict and prevent crimes, primarily focusing on analyzing and predicting accident locations to improve road safety. Based on the research, this proposal report will have a background section discussing the importance of crime prevention in public safety and the need for new approaches such as machine learning in police case analysis. The introduction section will provide a literature review, identify the research gap, and state the research problem. The main objective and sub-objectives of the research will also be outlined. The methodology section will explain how the research will be conducted, including the data collection and analysis methods to be used. There will also be a requirements section that outlines the functional and non-functional requirements necessary to achieve the research objectives. Additionally, the proposal will discuss the commercial values of the research, indicating how it will contribute to the development of more effective crime prevention strategies. Overall, the research proposal aims to use machine learning in police case analysis to predict and prevent crimes, primarily focusing on analyzing and predicting accident case analysis to improve road safety.

TABLE OF CONTENT

LIST OF TABLES

.

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to take this opportunity to express my heartfelt gratitude to my project supervisor Ms. Hansika Mahaadikara and Co-Supervisor Ms. Sanjeevi Chandrasiri for their invaluable guidance, inspiration, motivation, and constructive feedback that have been instrumental in shaping my research project proposal. I am also thankful to Mr. Jayantha Amararachchi, our project coordinator, for providing me with this opportunity and for his continued support throughout the project. Furthermore, I extend my sincere appreciation to all the educators, students, and parents who have provided me with valuable insights and suggestions, which have helped me improve my solution.

1. Introduction

The rise of crime in recent years has posed significant challenges for law enforcement agencies worldwide. The police are continuously seeking innovative ways to prevent and reduce crime, particularly in high-risk areas. Traditional methods of crime prevention and analysis are no longer sufficient, and there is a growing need for advanced technologies and tools that can process and analyze vast amounts of data to identify crime patterns, trends, and future criminal activities.

To address these challenges, we propose a project titled "The Future of Crime Prevention: Police Case Analysis using Machine Learning," which aims to develop a comprehensive system to help law enforcement agencies predict, prevent, and solve crimes using machine learning techniques. The system consists of four components:

- Analyze, predict accident locations and finding connections and patterns between police accident cases [2].
- Analyzing and grouping commonalities among criminal cases and predicting the future crimes in terms of the nature of the crime
- Automated document classification - Analyzing and classifying similar case documents and predict category [4]
- Clustering crimes against women and crime forecasting prediction using machine learning to create an effective solution to reduce and prevent such crimes [5].

In this introduction, the background of the project, the research gap and problem that our system aim to address will be discussed.

1.1 Background

The rise of crime has become an increasing challenge for law enforcement agencies worldwide. Traditional policing methods are no longer sufficient to deal with the complex nature of crimes and accidents. To address this issue, advanced technology is needed for crime prevention and analysis. The field of police investigation has undergone significant changes in recent years with the increasing reliance on technology and data analysis to solve crimes. Machine learning is a powerful tool that can process vast amounts of data to identify patterns and trends for predicting and preventing accidents. The below chart depicts the number of road accidents and fatalities that have been reported by the Sri Lankan police from 2013 to 2021. The information presented in the graph illustrates the scale of road accidents and the resultant loss of life in Sri Lanka during the specified time period.

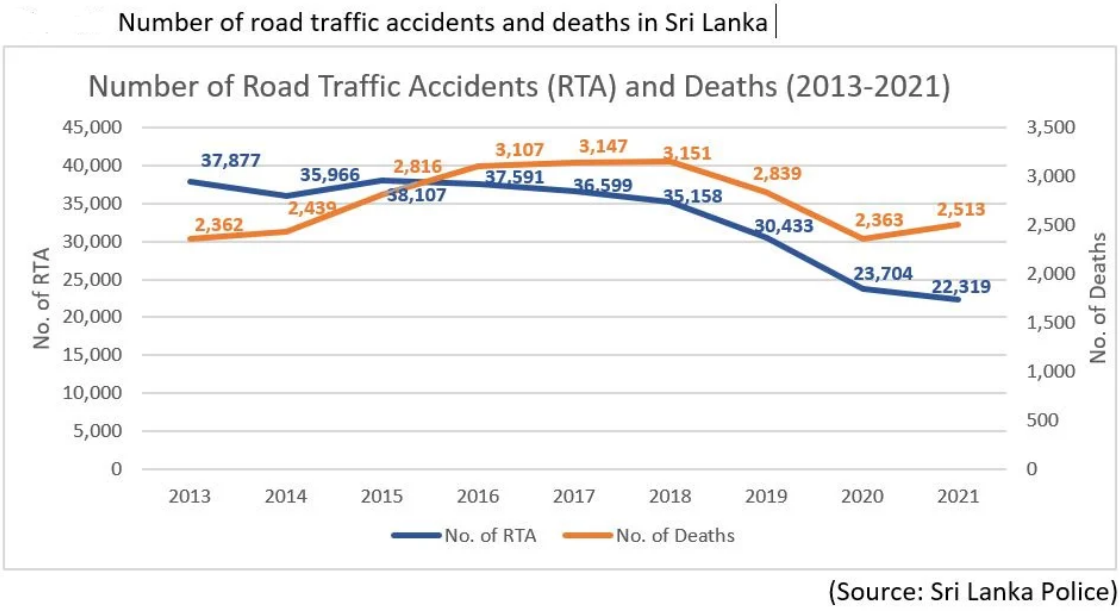

Figure 1.1: Sri Lanka Road Accidents and Fatalities (2013-2021)

To address these challenges, the research project "The Future of Crime Prevention: Police Case Analysis using Machine Learning" aims to leverage the power of machine learning to enhance public safety and prevent crime. The project will consist of four components: police accident case analysis, criminal case analysis, automated document classification, and clustering crimes against women. Each component will be developed to address specific research problems related to crime prevention and analysis.

The police accident case analysis component aims to develop a machine learning system that can analyze and predict accident locations, as well as identify connections and patterns between police accident cases. This will inform targeted safety measures and accident prevention strategies. The criminal case analysis component aims to classify police case documents, automating the time-consuming and tedious task of manual classification. The automated document classification component will save valuable time for investigators and ensure that cases are prioritized appropriately. Finally, the clustering crimes against women component aims to identify patterns and trends in crimes against women, helping law enforcement agencies to prevent such crimes from occurring in the future.

Overall, the project's goal is to provide accurate and reliable analysis, prediction, and classification of accident and crime-related data. By doing so, 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.

1.2 Literature Review

Accidents on roads are a significant cause of death and injury worldwide, making road safety a significant concern for governments and law enforcement agencies [8]. Machine learning techniques have become increasingly popular in the field of road safety as they can identify patterns and correlations in data that may be difficult for humans to detect [9]. To address this problem, a system has been proposed that can analyze accident data and identify common patterns, similarities, causes, or contributing factors. This analysis is done by preprocessing and feature engineering data, conducting exploratory and clustering analysis, developing machine learning models, evaluating their performance, and visualizing outputs. By doing so, the system can identify high-risk areas for accidents, predict future accident locations, and develop targeted safety measures [3].

Previous studies have shown that machine learning algorithms can be used to predict road accidents [10]. One study used clustering algorithms to identify hotspots of accidents and identified contributing factors such as driver behavior and road conditions [11]. Another study used decision trees to predict the severity of accidents based on factors such as the time of day and weather conditions [12]. These studies demonstrate the potential of machine learning algorithms in identifying causes and correlations in accident data.

Furthermore, recent studies have shown the potential of machine learning algorithms in identifying reasons and causes between accident cases [6]. For example, one study used association rule mining to identify patterns in accident data and found that certain driver behaviors were associated with specific types of accidents [2]. Another study used network analysis to identify the connections between different types of accidents in some areas [13].

In conclusion, machine learning techniques have great potential in the field of road safety, and the proposed system can help to identify high-risk areas for accidents, predict future accident locations, and develop targeted safety measures. The system's ability to identify connections and patterns between accident cases can also aid in developing effective prevention strategies. Further research in this area could help to refine the proposed system and improve road safety outcomes.

1.3 Research Gap

The existing literature reveals research gaps in the field of police-related accident analysis. Previous studies, such as [10], have implemented prediction techniques for identifying high-risk areas without addressing the need for finding connections and patterns, providing prevention strategies, or considering police-related accidents. The research study [20], on the other hand, uses cameras to detect vehicle accidents, which significantly increases the cost. In contrast, this proposed study aims to address these research gaps by developing a

comprehensive system for police-related accident analysis that integrates predictive modeling, connection identification, and prevention strategy planning, while ensuring a lower cost by avoiding external devices.

Furthermore, research study [15] identifies connections between accident cases but fails to distinguish between planned and unplanned accidents. In this proposed study, the focus is on identifying planned accidents and investigating their connections, which is a novel aspect that has not been addressed in previous studies. Moreover, the proposed system aims to develop prevention strategies for future accidents prediction and causes, which is another research gap in the existing literature.

In summary, the proposed study aims to fill the research gaps in the existing literature by developing a comprehensive system for police-related accident analysis that integrates predictive modeling, connection identification, prevention strategy planning, and distinguishing between planned and unplanned accidents. By addressing these research gaps, the proposed study will make a significant contribution to the field of police accident case analysis. The following table presents a summary of the research gaps identified in the proposed study compared to existing literature in the field of police accident case analysis and prediction.

RESEARCH	FUTURE PREDICTION	CONNECTIONS AND PATTERNS	PREVENTION STRATEGIES	POLICE RELATED	LOWER COST
ROAD TRAFFIC ACCIDENT PREDICTION USING MACHINE LEARNING [10]	✓	✗	✗	✗	✓
PREDICTING FUTURE DRIVING RISK OF CRASH-INVOLVED DRIVERS BASED ON A SYSTEMATIC MACHINE LEARNING FRAMEWORK. [20]	✓	✗	✗	✗	✗
ROAD TRAFFIC ACCIDENTS ANALYSIS USING DATA MINING TECHNIQUES[15]	✗	✓	✗	✗	✗
PREDICTION OF TRAFFIC ACCIDENT SEVERITY BASED ON RANDOM FOREST[22]	✓	✗	✗	✗	✓
POLICE ACCIDENT CASE ANALYSIS:	✓	✓	✓	✓	✓

Table 1.1: Comparison of other available applications

1.4 Research Problem

The main problem addressed by this research is the high incidence of road accidents and the need for effective accident prevention strategies. The existing methods of analyzing accident data and identifying high-risk areas for accidents are time-consuming and inefficient. Manual analysis of accident data requires a significant amount of time and resources, and there is a high probability of human error.

To overcome these limitations, the proposed system aims to develop a machine learning model that can analyze and predict accident locations and identify connections and patterns between police accident cases. By leveraging labeled and unlabeled datasets, the system can accurately predict future accident locations and identify high-risk areas, thereby helping to inform targeted safety measures and accident prevention strategies.

The system will use various machine learning techniques such as data preprocessing, feature engineering, exploratory and clustering analysis, and model development to achieve its objectives. The outputs generated by the system will be displayed in graphical formats, providing users with easy-to-understand insights that can help them identify high-risk areas, predict future accidents, and develop targeted safety measures.

By developing a machine learning model that can accurately predict future accident locations and identify connections and patterns between police accident cases, this research aims to provide a valuable tool for police staff and administrators, helping to reduce the number of accidents and improve road safety.

2. Objectives

2.1 Main Objectives

The main objective of my component is to develop a machine learning system that can analyze and predict accident locations, as well as identify connections and patterns between police accident cases, in order to inform targeted safety measures and accident prevention strategies.

2.2 Specific Objectives

- To preprocess and engineer the accident data to prepare it for machine learning analysis.
- To conduct exploratory analysis to gain insights into the factors that contribute to accidents.
- To develop a machine learning model that can accurately predict accident locations based on relevant features.
- To evaluate the performance of the machine learning model using appropriate metrics and techniques.
- To visualize the outputs of the machine learning model in graphical formats to facilitate interpretation and decision-making.
- To use clustering analysis to identify connections and patterns between police accident cases.
- To develop targeted safety measures based on the insights gained from the machine learning analysis to prevent accidents and improve road safety.

By achieving these sub-objectives, the project aims to provide a valuable tool for police staff and administrators to reduce the number of accidents and improve road safety through the use of machine learning.

3. Methodology

This research proposes the development of a machine learning system to analyze and predict accident locations, as well as identify connections and patterns between police accident cases, in order to inform targeted safety measures and accident prevention strategies. The proposed methodology will involve the following steps:

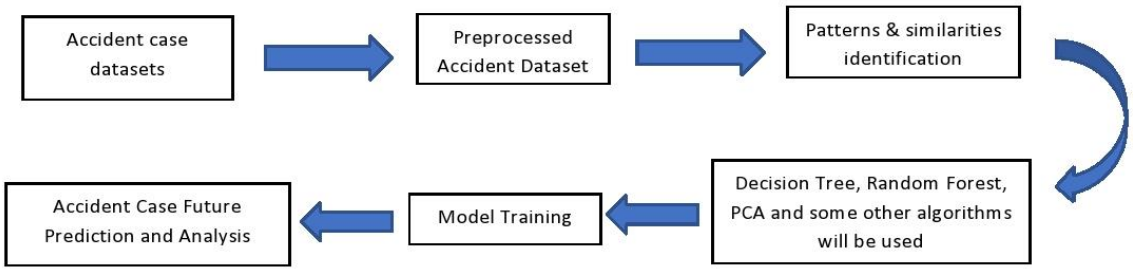

Figure 3.1: Flow diagram for methodology

3.1 Procedures

Data Collection: The research will use a combination of internet and local Sri Lankan datasets, collected through meetings with police personnel and other relevant authorities. The internet dataset will be obtained from publicly available sources such as accident databases, news articles, and social media platforms, while the local Sri Lankan dataset will be collected through interviews and discussions with police officers, traffic authorities, and other experts in the field.

Data Preprocessing: The collected data related to police accident cases will undergo preprocessing to remove any irrelevant or redundant information and to ensure that it is in a format that can be used for analysis. This will involve techniques such as data cleaning, data transformation, and data normalization to ensure that the data is consistent and accurate [14].

Feature Extraction: The relevant features will be extracted from the preprocessed data. This will include location, time, road conditions, vehicle type, and driver details.

Algorithm Selection: The appropriate machine learning algorithms will be selected to analyze and predict police accident locations and identify patterns between accident cases. The selected algorithms will include decision trees, random forests, DBSCAN, PCA and neural networks, which have been shown to be effective in predicting accident locations and identifying patterns in accident data [16, 17].

Model Training and Evaluation: The selected algorithms will be trained on the preprocessed and feature-extracted data. The models will be evaluated based on various performance metrics such as accuracy, precision, recall, and F1-score. The best performing model will be selected for predicting future accident locations, identifying the causes of the accident, and providing prevention strategies. Additionally, the trained model will be used to identify patterns and

connections between accident cases, allowing for a more comprehensive understanding of the factors that contribute to police accidents.

System Development: The selected machine learning models will be integrated into a user-friendly system with a simple and organized user interface. The system will provide reliable reports on predicted accident locations and patterns, along with the causes of the accidents and prevention strategies. The system will also identify connections between accident cases, including whether they were planned accident or not.

3.1.1 Data Collection

For this police accident case analysis, the dataset will be collected from the Sri Lankan Police Department, which contains detailed information about past accidents including the location, time, cause of the accident, vehicle type, driver behavior, and any injuries or fatalities. The dataset will be preprocessed to remove any irrelevant or redundant information and to address missing values or outliers. Additionally, internet datasets from other countries will be used to supplement the analysis and provide a broader understanding of accident patterns and causes. The collected datasets will be carefully screened and cleaned to ensure accuracy and completeness using techniques such as outlier detection and missing value imputation [14]. The datasets will then be preprocessed to extract relevant features using techniques such as dimensionality reduction and feature selection [17]. This will enable the identification of patterns and relationships between accident cases, and improve the accuracy of the machine learning models

3.2 Tools and Technology Selection

- Programming language: Python
- Integrated Development Environment (IDE): PyCharm, Jupyter Notebook
- Machine Learning Libraries: Scikit-Learn, Tensor flow, pyTorch
- Data Analysis Libraries: Pandas, NumPy
- Data Visualization Libraries: Matplotlib, or Octave
- Database Management System: MongoDB
- Data Reporting: Power BI
- Web Frameworks: Django
- Cloud Computing Platforms: Microsoft Azure
- Version Control System: Git
- Collaboration Tool: Gitlab, GitHub

3.3 Testing

During the testing phase of our research, the performance of our proposed system will be evaluated by comparing the outcomes generated by our system with the manual outcomes. This evaluation will be conducted by comparing the accuracy and reliability of the outcomes generated by our system with the manual outcomes derived from past accident and crime-related data. By conducting this comparison, effectiveness and efficiency of our system in accurately analyzing, predicting, and classifying accident and crime-related data will be assessed. The evaluation will enable us to identify any areas where our system may need improvement and refine our system to ensure that it provides the most accurate and reliable outcomes possible. Ultimately, this will help to develop a system that can provide valuable insights into accident and crime-related data and support efforts to improve public safety and crime prevention.

4. Description of Personal and Facilities

4.1 Work Breakdown Structure (WBS)

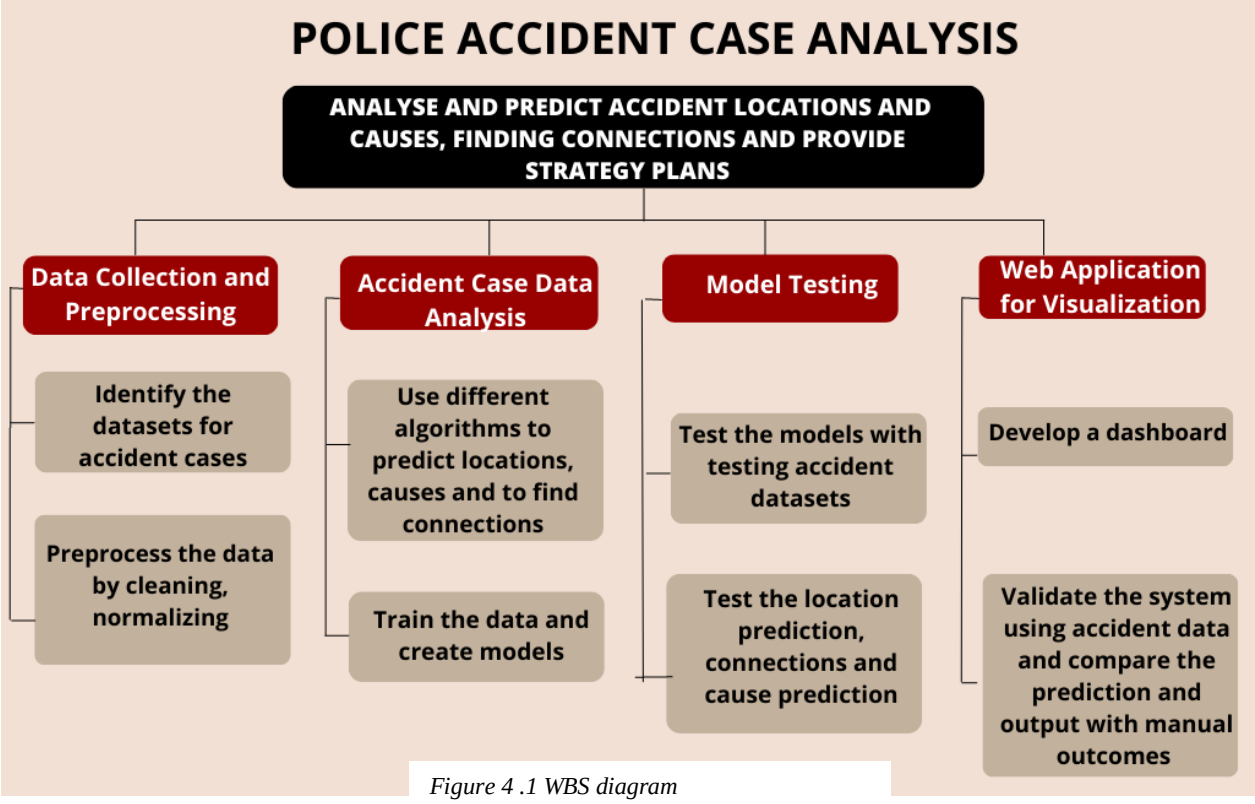

Figure 4 .1 WBS diagram

4.1 System Architecture

System Diagram for police accident case analysis component:

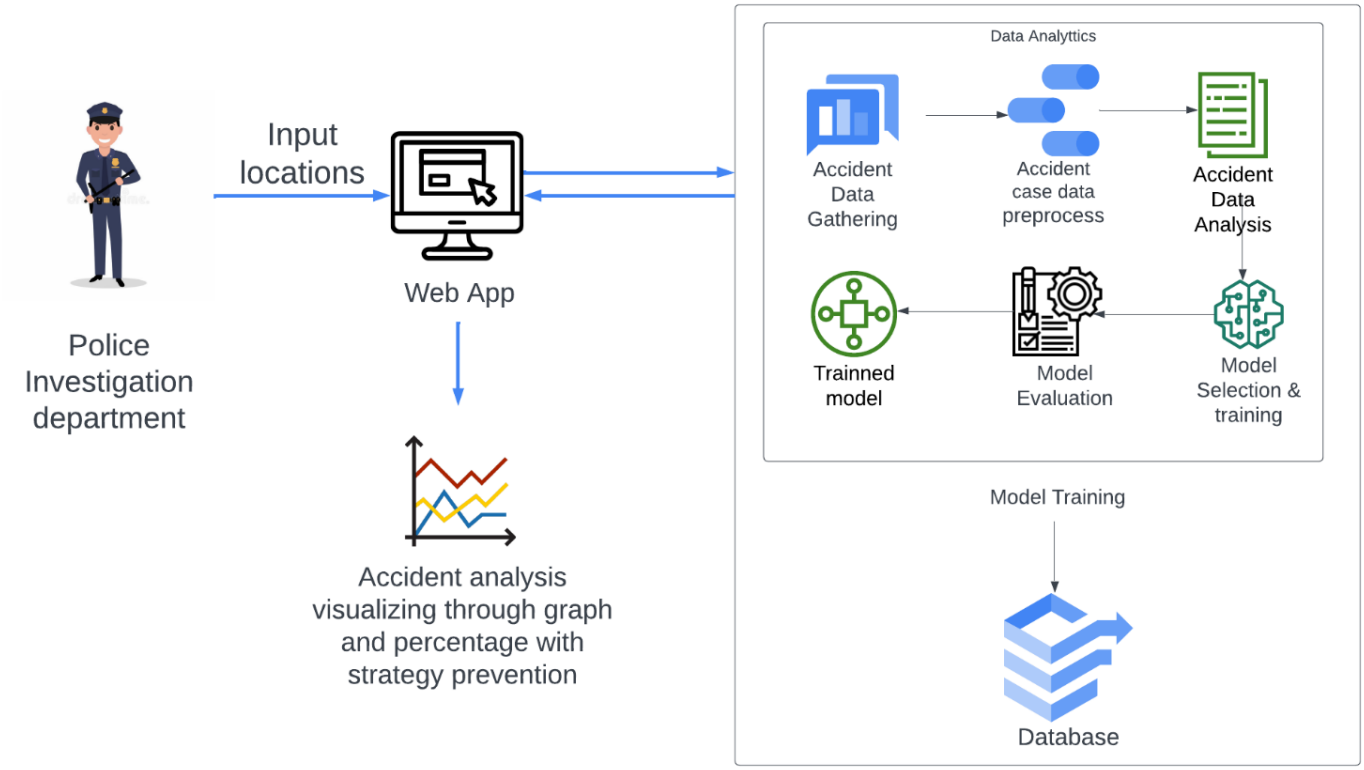

Figure 4.2 System diagram for police accident analysis component

System diagram for overall project:

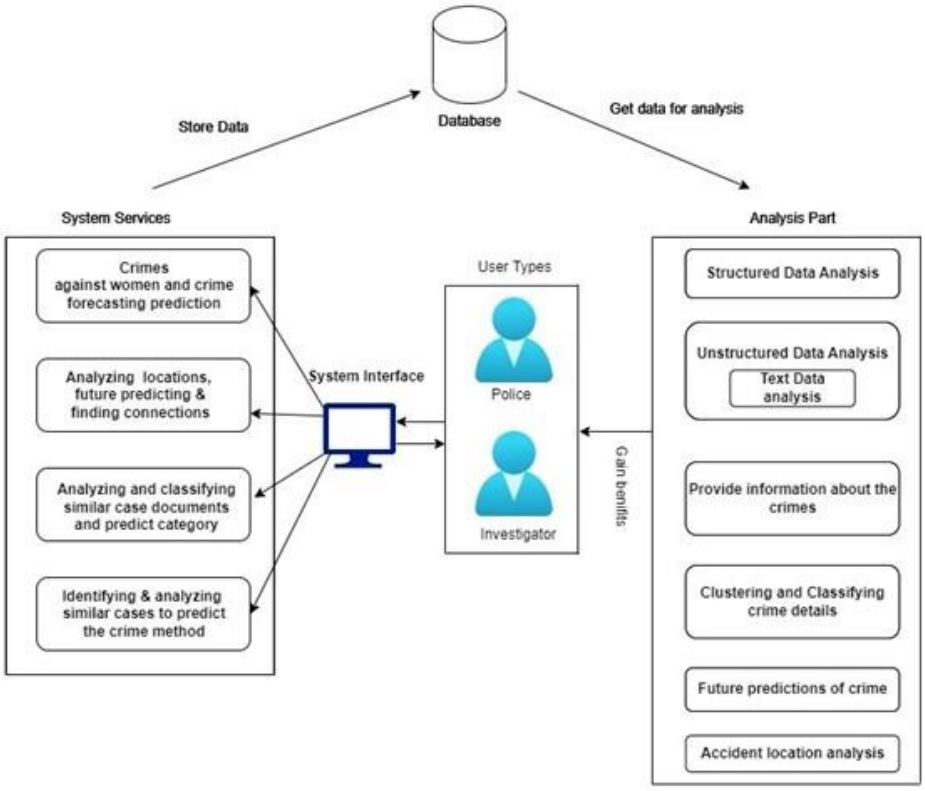

Figure 4.3 System diagram for overall project

4.3 Gantt Chart

Tasks	2022	2023											
	December	January	February	March	April	May	June	July	August	September	October	November	December
Topic Clarification													
Topic Assessment													
Project Charter													
Requirements Gathering													
Making of Project Proposal													
Project Presentation													
Proposal Report													
Progress Presentation-1													
Research Paper													
Final Report													
Progress Presentation-2													
Testing and Fixing issues													
Website Assessment and Research Book													
Final Report (Soft bound)													
Status Document Submission													
Final Presentation and Viva													
Final Report Submission													

Figure 4.4 Gantt Chart

5. Requirements

5.1 Functional Requirements

- The system should have a login page for the user to access the components.
- The system should consist of four components and the user should be able to redirect to the desired component.
- The system should provide a graph and percentage of past data and future prediction details for accident locations.
- The system should allow the user to enter one or multiple locations to view the predicted outcome of accident cases.
- The system should display the future prediction of accident cases based on year-wise analysis for the entered location/s.
- The system should display all the locations which have more than 60% of accident police cases for the entered year/s.
- The system should be able to analyze the causes of the predicted accident for the inputted location
- The system should identify the connections and patterns between accident cases and display the percentage of such cases.
- System should provide prevention strategy plans for the predicted cause of the accident.
- The output should consist of a graph, percentage, and prediction for the entered input.
- System will generate a report according to user's input
- If there is a positive outcome for the entered input, the system should display the result.
- If there is no relevant output to display, the system should show the message accordingly.

5.2 Non-Functional Requirements

- Performance:
    - The system's predictive modeling and pattern analysis components should be designed to efficiently analyze accident data and provide accurate predictions and patterns within a reasonable amount of time. This will ensure that users receive prompt and efficient responses and can take appropriate actions to prevent future accidents.

- Reliability:
    - The system should be highly reliable and stable, with a robust algorithm and error handling mechanisms. This will ensure that it can consistently and accurately analyze accident data, provide accurate predictions and patterns, and operate without interruptions.

- Security:
    - The system should maintain the confidentiality and integrity of all accident-related data, including locations, causes, and patterns. Robust security

mechanisms should be in place to prevent unauthorized access, protect against data breaches, and ensure the system remains secure.

- Usability:
    - The system should be designed with a user-friendly and intuitive interface, with a focus on ease of use and accessibility for users with different levels of technical expertise. The system should enable easy navigation, data retrieval, and analysis, and provide users with actionable insights to prevent future accidents.

- Scalability:
    - The system should be designed to handle large amounts of data and accommodate future increases in data volume. This will ensure that the system can maintain high levels of performance and accuracy as more accident data becomes available.

- Maintainability:
    - The system should be easy to maintain and update, with clear documentation and coding standards. This will enable quick and seamless modifications to be made to the system without impacting its performance or accuracy.

- Compatibility:
    - The system should be compatible with different devices and operating systems, to ensure that users can access the system using a range of devices and platforms. Compatibility testing should be conducted regularly to ensure the system works seamlessly across all platforms.

6. Commercial Value

- Commercial Value of the Overall System:

    - 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.

- Commercial value for police accident case analysis component:

    - Enhanced public safety through proactive accident prevention measures.
    - More informed decision-making for police departments.
    - Efficient allocation of resources for accident response and prevention.
    - Targeted intervention strategies for accident reduction and improvement of public safety.
    - System helps to provide efficient service for police, and the police department can provide better service for public.

References

[1] Fisher, B. S., & Lab, S. P. (2010). Encyclopedia of victimology and crime prevention. Sage Publications.

[2] Hoque, N., & Khan, M. A. (2018). Accident analysis and prevention by machine learning approaches: A systematic review. Safety science, 105, 225-244.

[3] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. Journal of Business Research, 70, 263-286.

[4] Choudhury, P., & Kshirsagar, M. (2019). Automated document classification and categorization for crime investigation using machine learning. In Advances in Signal Processing and Intelligent Recognition Systems (pp. 527-538). Springer.

[5] Arora, R., & Pandey, A. K. (2021). A comprehensive review on clustering and forecasting techniques in crimes against women. Pervasive and Mobile Computing, 79, 101607.

[6] Zhang, J., Qin, Z., & Liu, X. (2021). Machine learning techniques for predicting crime: A systematic review. Expert Systems with Applications, 167.

[7] R. Polikar, "Ensemble based systems in decision making," IEEE Circuits and Systems Magazine, vol. 6, no. 3, pp. 21-45, 2006.

[8] World Health Organization, "Road traffic injuries," 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. [Accessed: Mar. 23, 2023].

[9] He, J., & Wu, J. (2019). Using machine learning to predict road accidents: A review of recent studies. Journal of Advanced Transportation, 2019.

[10] Wan, J., Zhang, D., Li, D., & Liu, C. (2019). Road traffic accident prediction using machine learning: A review. Sustainability, 11(20), 5754.

[11] Chai, Y., Li, H., Wang, Z., & Li, K. (2017). Clustering-based hotspot identification for road traffic accidents: A review of methodological progress and existing issues. International Journal of Environmental Research and Public Health, 14(6), 651.

[12] Zhang, X., Qu, X., & Zhang, X. (2019). A decision tree-based approach for predicting road traffic accident severity. IEEE Access, 7, 58608-58616.

[13] Zhang, Y., & Wang, W. (2019). Identifying risk factors for pedestrian crashes using network analysis of big data. Journal of Advanced Transportation, 2019.

[14] Z. Wu, X. Wu, and C. Zhang, "Data cleaning: Problems and current approaches," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 10, pp. 1355-1369, Oct. 2008.

[15] Guru, J., Devi, N.R. (2018). Road Traffic Accidents Analysis Using Data Mining Techniques. JITA - Journal of Information Technology and Applications (Banja Luka) - APEIRON 14(2). DOI: 10.7251/JIT1702084J

[16] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, Oct. 2001.

[17] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798-1828, Aug. 2013.

[18] S. S. S. M. R. Fernando, T. Fernando, and N. De Silva, "A survey of evaluation metrics used in machine learning," in 2019 4th International Conference on Advances in Computing and Technology (ICACT), Colombo, Sri Lanka, Aug. 2019, pp. 1-5.

[19] Shi, X., Li, Q., Qi, Y., Huang, T., & Li, J. (2017, November). An accident prediction approach based on XGBoost. In 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) (pp. 29-34).

[20] Wang, C., Liu, L., Xu, C., & Lv, W. (2021). Predicting Future Driving Risk of Crash-Involved Drivers Based on a Systematic Machine Learning Framework. IEEE Access, 9, 129605-129613

[21] Venkat, A., Gokulnath, M., Vijey, G.K.P., Thomas, I.S., & Ranjani, D. (2020). Machine Learning Based Analysis for Road Accident Prediction. IJETIE, 6(2), 1-7. Retrieved March 7, 2020

[22] Yang, J., Han, S., & Chen, Y. (2023). Prediction of Traffic Accident Severity Based on Random Forest. Journal of Advanced Transportation, 2023(2), 1-8. DOI: 10.1155/2023/7641472