THE FUTURE OF CRIME PREVENTION: POLICE CASE ANALYSIS USING MACHINE LEARNING.

(Criminal Case Analysis: Analyzing and grouping commonalities among criminal casesand predicting the future crimes in terms of the nature of the crime.)

Project Id

TMP-23-224

Project Proposal Report

Krishanthini.M- IT19980928

Supervisor – Ms. Hansika Mahaadikara

Co-Supervisor - Ms. Sanjeevi Chandrasiri

B.Sc. (Hons) Degree in Information Technology

Specializing in Software Engineering

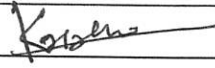Department of Computer Science & Software Engineering (CSSE)

Sri Lanka 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
Krishanthini.M	IT19980928	*(signature)*

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: *(signature)*

(Ms. Hansika Mahaadikara)

Date: 17/04/2023

ABSTRACT

The increasing amount of data being generated by police departments necessitate the adoption of innovative methods for data analysis. This study focuses on four sub-objectives: predicting accident locations and identifying patterns in police accident cases to reduce the frequency of accidents, analyzing, and grouping commonalities among criminal cases in terms of the nature of future crimes, classifying similar case documents to increase the efficiency of case management, and clustering crimes against women and predicting future crimes using machine learning techniques to prevent crimes against women. The proposed research will involve collecting and analyzing data from police case files and applying various machine learning techniques to develop predictive models for crime prevention. Overall study can contribute to the field of crime prevention and offer insights into the use of machine learning in public safety. And this component mainly aims to use machine learning algorithms to analyze criminal records and identify patterns and commonalities in crimes committed in terms of the nature of the crime. By doing so, the study seeks to help law enforcement agencies allocate their resources more effectively by identifying the areas where crimes are most likely to occur and can help to develop more effective crime prevention strategies to reduce crime rates and enhance public safety. The use of machine learning algorithms such as decision tree and clustering will allow to process large amounts of data quickly and accurately. The results of the analysis will be presented in a user-friendly format to enable law enforcement officials to understand and utilize the findings effectively. Overall, the proposed research project has the potential to contribute significantly to the field of criminal justice and public safety. It can improve the efficiency of police stations, reduce crime rates, and offer new insights into the factors that drive criminal behavior.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

1. Introduction

Crime prevention is a critical aspect of public safety, and law enforcement agencies are constantly seeking new ways to prevent crimes from occurring. In recent years, the use of machine learning algorithms in police case analysis has gained significant attention due to its potential to revolutionize the way crimes are prevented and investigated. The proposed research paper, titled "The Future of Crime Prevention: Police Case Analysis using Machine Learning," aims to investigate the effectiveness of machine learning algorithms in analyzing past criminal cases and predicting future crimes. The objective is to develop a powerful tool that can be used by police investigators to make predictions about future crimes and allocate resources effectively to prevent them. By leveraging vast amounts of crime and accident data, crimes against women, and case file datasets to train machine learning algorithms, the proposed system aims to provide a comprehensive and sophisticated solution to the challenges faced by police in reducing crime and ensuring public safety.

1.1 Background

Criminal activities and incidents are widespread today, making it crucial for law enforcement agencies to have a comprehensive understanding of the patterns and trends associated with criminal activities. However, it can be challenging for investigators to comb through vast amounts of data manually to identify patterns and connections that could help solve a case. The utilization of machine learning in criminal case analysis allows for the efficient and effective processing of large volumes of data sets collected from past criminal incidents. The technology works by using algorithms and statistical models to identify patterns and commonalities that exist within the data. These patterns can be used to identify potential suspects, predict future criminal behavior, and provide insights into criminal activities.

The site *macrotrends* depicts that the rate of the crimes in Sri Lanka as visibly increased in the recent years. And, according to their statics[13]:

Sri Lanka crime rate & statistics for 2019 was 3.48, a 43.72% increase from 2018.
Sri Lanka crime rate & statistics for 2018 was 2.42, a 5.26% increase from 2017.
Sri Lanka crime rate & statistics for 2017 was 2.30, a 8.77% decline from 2016.
Sri Lanka crime rate & statistics for 2016 was 2.52, a 8.25% increase from 2015.
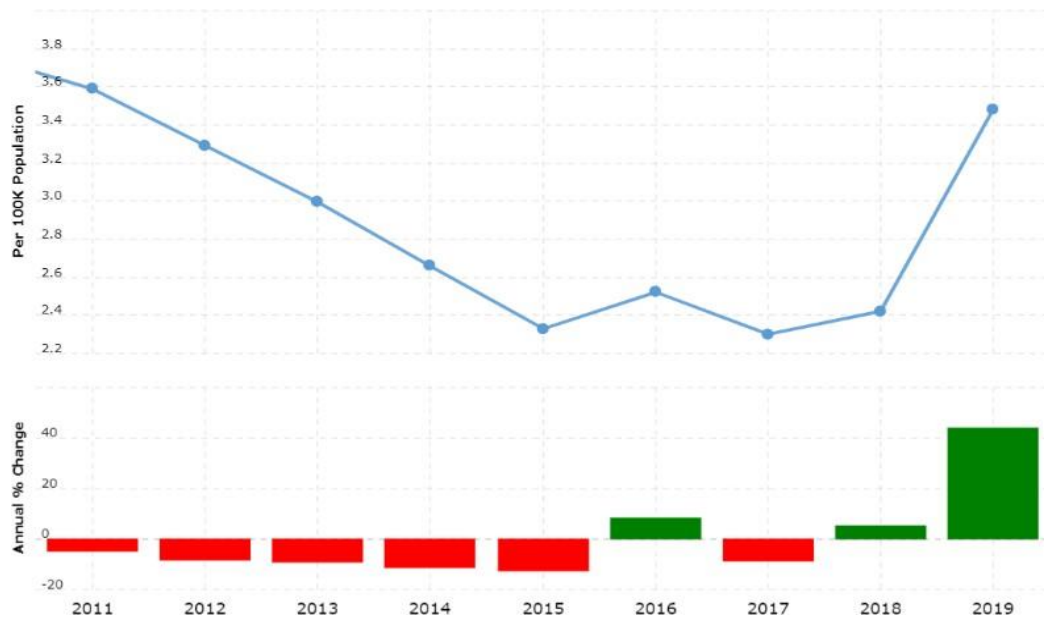
Figure 1.1: Crime rate in Sri Lanka depicted by macrotrends[13]

The collection of past criminal datasets is a critical aspect of this process. The data sets are typically gathered from various sources, including police reports, court records, and other public records. These datasets are then subjected to various machine learning algorithms such as decision tree and clustering which can analyze the data and identify patterns that might not be easily noticeable to human investigators.

The analysis of past criminal datasets can also help identify areas that require more attention from law enforcement agencies. For example, if there is an increase in a particular type of crime in a particular area, it can help police departments allocate resources more efficiently and prevent future incidents. And the patterns of targeted victims and the crime objects can be also identified. Furthermore, the use of machine learning algorithms can help law enforcement agencies identify trends and patterns associated with specific types of criminal activities. This can lead to the development of more effective prevention and intervention strategies to battle criminal activities.

This research proposes an application which gives output of the analysis in a user-friendly format, such as graphs and tables, allowing for easy interpretation and decision-making. This will help police departments optimize their operations and deploy resources more efficiently.

Learning is an emerging field that has the potential to revolutionize law enforcement. The technology can provide valuable insights into criminal activities and help police departments improve their efficiency, leading to better outcomes and safer communities.

1.2 Literature Review

This section focuses on previous research that has been conducted related to this research. The use of machine learning algorithms in crime analysis has become an increasingly popular research topic in recent years. Many studies have demonstrated that the use of different algorithms can enhance the efficiency and accuracy of crime detection, prevention, and investigation in terms of the nature of the crime.

One of the significant studies in [2] this field is conducted by Neil Shah, Nandish Bhagat & Manan Shah, which focuses on crime forecasting using a machine learning and computer vision approach. The authors highlight the importance of using data such as facial recognition, number plate recognition, augmented and mixed realities, location determination, and object identification in predicting and preventing crimes. They propose using the motive as a judgment for the nature of the crime and categorizing every aspect of a crime for efficient analysis.

Another significant contribution in the field is the research conducted by Steven Walczak, which outlines the use of neural network models to predict specific types of crime using location and time information [4]. The author demonstrates that the neural network models can predict the type of the crime 16.4% of the time for 27 different types of crime or 27.1% of the time when similar crimes are grouped into seven categories of crime. This study highlights the effectiveness of using neural network models in police decision-making for crime prevention.

[1]Karabo Jenga, Cagatay Catal & Gorkem Kar, in their research on machine learning in crime prediction, focus on using data mining and machine learning to perform tasks related to crime prevention and detection. They emphasize the versatility of machine learning and data mining in predicting and preventing crime.

Moreover, [5]Suhong Kim, Param Joshi, Paraminder Singh Kalsi, and Pooya Taheri have contributed to the field by demonstrating the importance of using decision trees in analyzing crimes. They used more than 560,000 records to predict crimes in Vancouver and obtained a crime prediction accuracy of 39% to 44% based on the category and time of the crime. This study emphasizes the importance of analyzing crime data using decision trees to improve crime prediction accuracy.

In conclusion, all these research areas emphasize the use of analyzing criminal cases in relation to the nature of the crime to the society and the efficiency and accuracy by using Machine Learning algorithms. The previous research in this field has contributed significantly to the understanding of how these technologies can enhance the efficiency and accuracy of crime prevention and investigation. The proposed research aims to build upon these contributions by analyzing past criminal data and identifying patterns and commonalities in the nature of the crime using machine learning algorithms to enhance police station efficiency.

1.3 Research Gap

Although there are different ways and ideas in implementing criminal case analysis using machine learning use of decision tree and clustering algorithms are limited. While these algorithms have shown promise in other areas of machine learning, such as marketing and finance, their potential use in criminal case analysis has not been fully explored. Decision trees have the advantage of being easy to interpret and visualize, making them an ideal tool for law enforcement officials who may not have a background in data science. They can be used to identify patterns and factors that contribute to specific types of criminal behavior, helping law enforcement officials to develop more effective crime prevention strategies. Clustering algorithms have the advantage of being able to group together similar data points, even in cases where there are no pre-defined categories or labels. In criminal case analysis, clustering algorithms can be used to identify different groups of offenders or to group together similar types of crimes. This can help law enforcement officials to better understand the nature of criminal activity in each area and to develop more targeted responses.

RESEARCH	FUTURE PREDICTION	FINDING THE SIMILARITIES	GROUPING THE CRIMINAL ACTIVITIES	LOCATION DETECTION	CRIME PREVENTION
A MACHINE LEARNING AND COMPUTER VISIONAPPROACH TO CRIME PREDICTION AND PREVENTION [1]	✓	✗	✗	✗	✓
PREDICTING THE TYPES OF CRIMES COMMITTED IN THE CITY OF CHICAGO USING NEURAL NETWORKS. NEURAL COMPUTING AND APPLICATIONS [3]	✓	✗	✗	✗	✗
CRIME PREDICTION USING DECISION TREES. [5]	✓	✓	✗	✗	✗
PREDICTIVE MAPPING OF CRIME BY PROMAP: ACCURACY, UNITS OF ANALYSIS, AND THE ENVIRONMENTAL BACKCLOTH[7]	✓	✗	✗	✓	✓
CRIMINAL CASE ANALYSIS AND FUTURE PREDICTION	✓	✓	✓	✓	✓

Table 1.1: Comparison of other available research papers

1.4 Research Problem

Criminal cases are one of the major parts in the police cases. Almost 45% reported as the police case are criminal activities. With criminal activity rates increasing rapidly, there is a growing need for efficient and accurate analysis of criminal cases to assist law enforcement agencies in solving crimes and apprehending criminals. Although there are many research papers which deals with different aspects of using Machine learning for criminal case analysis, grouping the cases in terms of the patterns, trends, victims and objects is not proposed yet. Current manual methods of analysis are inefficient and often lead to inaccuracies, making it difficult for police officers to efficiently investigate cases. By leveraging advances in machine learning technologies, it may be possible to develop a system that can automatically analyze criminal cases, thereby reducing the workload on police officers and increasing the accuracy and efficiency of criminal investigations. However, such a system would need to be carefully designed and thoroughly tested to ensure that it is effective and reliable, as well as privacy- compliant and ethical.

2. Objectives

2.1 Main Objectives

The main objective of this component is to improve criminal case analysis for security agencies by identifying patterns and trends in crimes. By analyzing past criminal cases, the system will find commonalities, trends, predict future crime and their rate, presented through interactive visualizations. The research aims to enhance public safety and prevent criminal activity by providing a tool for security agencies to forecast potential future criminal activities.

2.2 Specific Objectives

- Develop a system that can analyze large amounts of criminal case data to identify patterns, trends, and similarities in crimes.

- Use the insights gained from the analysis of past criminal cases to predict future crimes and the rate of crimes, providing security agencies with a tool to forecast potential future criminal activities.

- Present the output of the analysis in interactive visualizations such as word clouds, network graphs, and dashboards, making it easily understandable for security agencies.

- Evaluate the accuracy and effectiveness of the system by training the model, selecting appropriate models, and identifying patterns and trends that may indicate the likelihood of a particular type of crime occurring in a specific location or time frame.

3. Methodology

This research component involves developing a system that can automatically analyze a large amount of criminal case data to identify patterns, trends, and similarities in the nature of crimes, predict future crimes, and present the output in interactive visualizations. The methodology involves data gathering, preparation, analysis, model selection, training, and evaluation, along with visualization of the output. The system is then tested on real-world scenarios, deployed in security agencies, and monitored to ensure its accuracy and effectiveness. This methodology aims to provide security agencies with a tool to enhance public safety and prevent criminal activity by predicting potential future criminal activities.
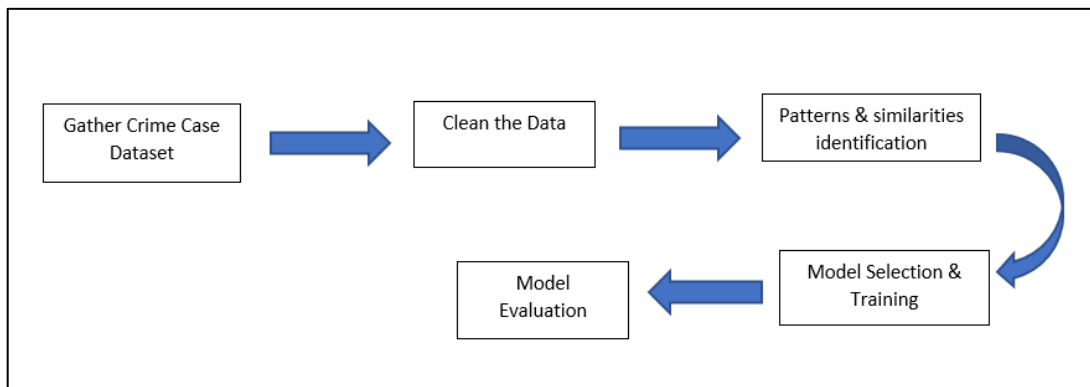

Figure 3: Methodology

3.1 Procedures

Data Gathering
- Collect a large amount of criminal case data from various sources, such as police departments and courts, ensuring that the data is comprehensive and reliable.

Data Preparation
- Clean the data and ensure it is in a consistent format, removing any duplicates, and converting any unstructured data into structured data.

Data Analysis
- Analyze the prepared data to identify patterns, trends, and similarities in crimes using statistical techniques and machine learning algorithms.

Model Selection and Training
- Choose an appropriate model, such as a decision tree or neural network, and train the model using the prepared data to predict future crimes and their rate.

Model Evaluation

- Evaluate the accuracy and effectiveness of the model by testing it on a subset of the data.

Visualization

- Visualize the output of the analysis in interactive visualizations such as word clouds, network graphs, and dashboards to make it easily understandable for security agencies.

Testing

- Test the system on real-world scenarios, evaluate its effectiveness and adjust if necessary. And also, system will be compared with the manually system will is used currently.

Deployment

- Deploy the system in security agencies and provide them with the necessary training to utilize the system efficiently.

Monitoring

- Continuously monitor the system's performance, regularly updating the model and data as necessary to ensure it remains accurate and effective.

3.2 System Architecture
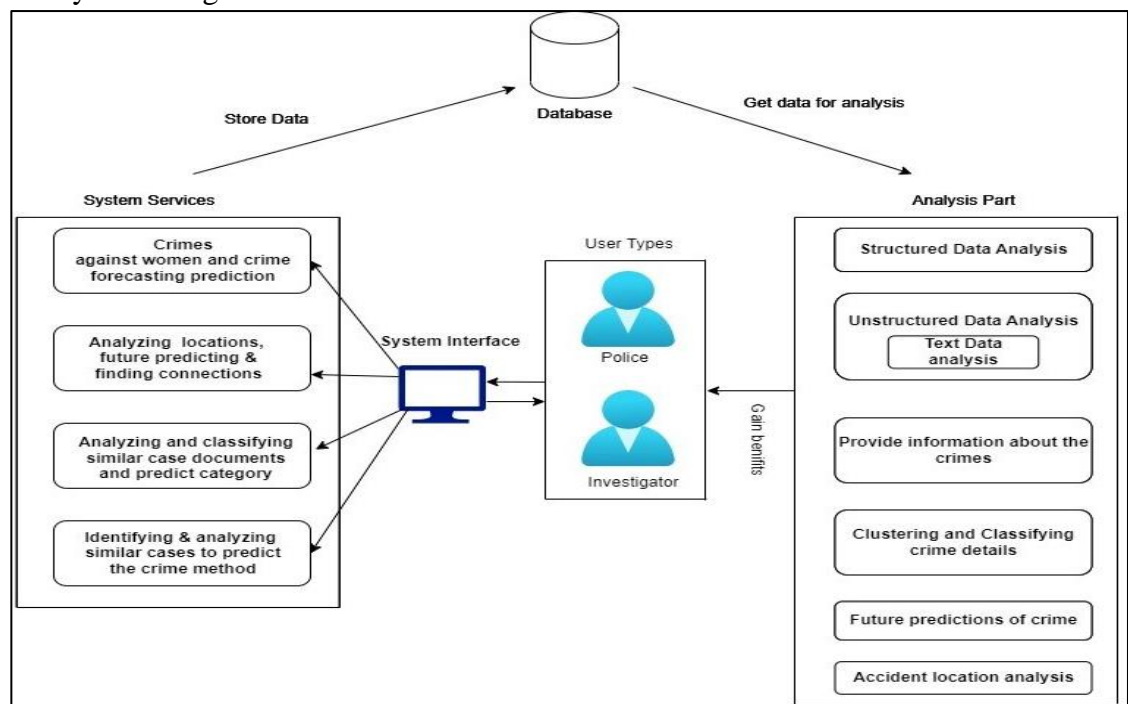
3.2.1 System Diagram of overall Research:

Figure 3.2.1: System diagram of overall system

3.2.2 System Diagram of individual Research Component:
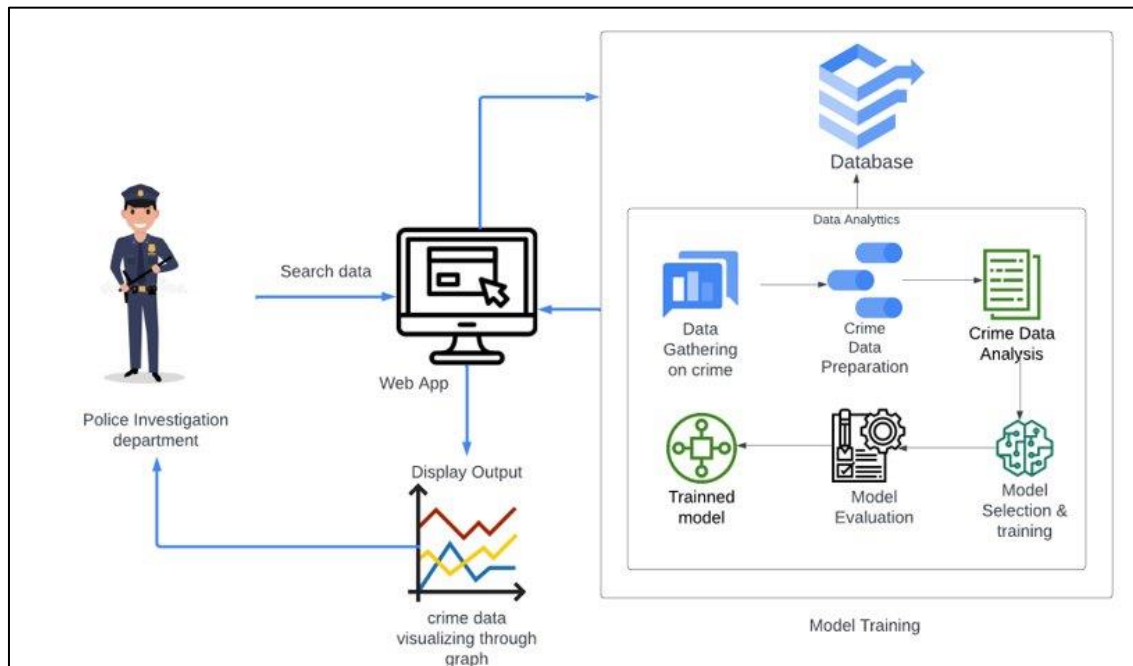
Figure 3.2.2: System diagram of individual component

3.3 Tools and Technology Selection

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

3.4 Data Collection

The online data set will be taken from sites Google Scholar, Kaggle etc. Both foreign and Sri Lankan dataset will be used. To improve the accuracy and effectiveness of the machine learning algorithms, the proposed research will use both foreign and Sri Lankan datasets for training. Additionally, further testing will be conducted by collecting additional datasets from local police stations. This will enable the system to learn from a wide range of criminal cases and improve its ability to predict and prevent future crimes.

3.5 Testing

In the testing part of the research, the effectiveness of the suggested system will be assessed by contrasting the results produced by the proposed system with the results produced manually. This evaluation will be carried out by contrasting the precision and dependability of the results produced by the proposed system with the manual results obtained from historical accident and criminal activity data. This comparison will lead to evaluate the usefulness and efficiency of the system in evaluating, forecasting, and categorizing data connected to crimes. The review helps to highlight any areas where the system needs to be improved, so that it can produce results that are as precise and trustworthy as possible. In the end, this will assist in creating a system that can provide valuable insights into crime-related data and support efforts to improve public safety and crime prevention.

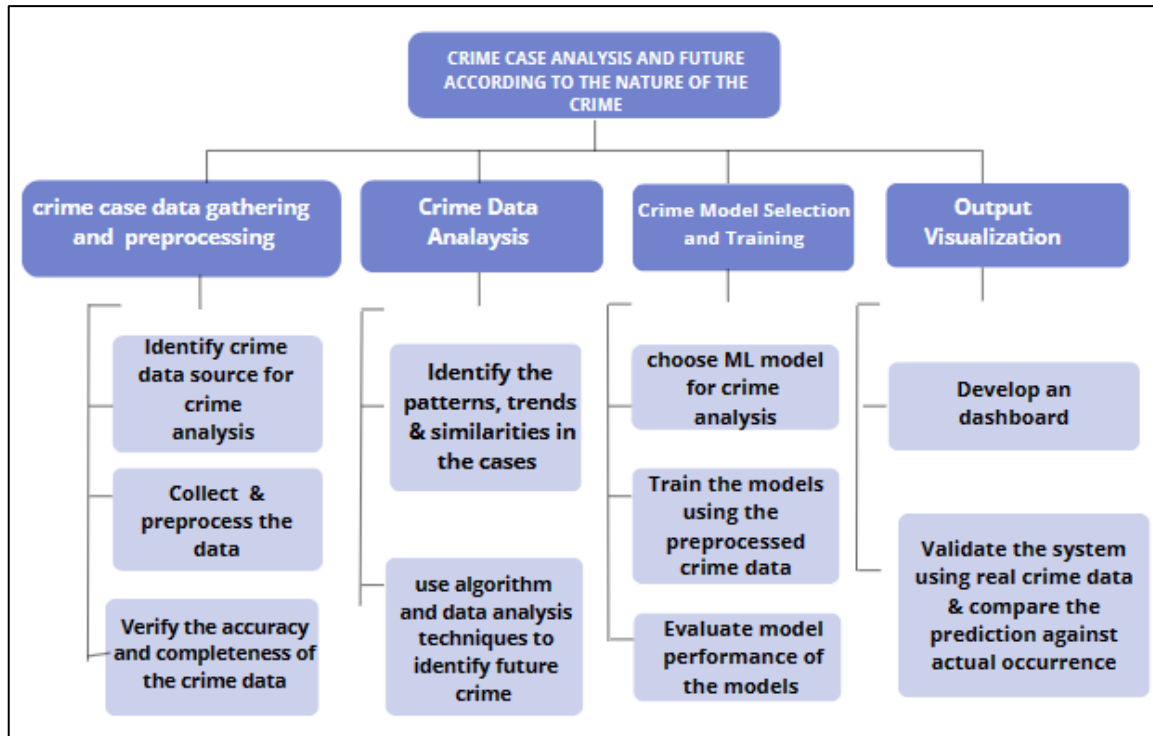# 4. Description of Personal and Facilities

4.1 Work Breakdown Structure (WBS)

Figure 4.1 Work Breakdown Structure

4.2 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 4.2 Gantt Chant

5. Requirements

5.1 Functional Requirements

- The system should be able to collect and ingest large amounts of criminal case data from various sources, ensuring that the data is comprehensive and reliable.

- The system should be able lean the data, convert any unstructured into structured data and ensure that it is in a consistent format. And also, system should remove any duplicates and reduce redundancy.

- The system should be able to identify the similarities in the crime cases and group accordingly.

- The system should be able to identify and predict the locations where the crime rates are high.

- The system should be able to analyze the prepared data using statistical techniques and machine learning algorithms to identify patterns and trends.

- The system should be able to visualize the output of the analysis in interactive visualizations such as word clouds, network graphs, and dashboards to make it easily understandablefor security agencies.

5.2 Non-Functional Requirements

Performance

The system should be able to handle large amounts of data and provide quick analysis results to the users. It should have the capability to process a high volumeof data and provide near real-time analysis, ensuring that the security agencies can react promptly to prevent potential criminal activities.

Scalability

The system should be able to handle many users and multiple concurrent requests. The component should be designed in a way that allows it to scale up as thevolume of data and the number of users grows, ensuring that it remains responsive and efficient.

Security

The system should comply with the security standards and policies of the security agencies, including data protection, access control, and privacy regulations. The component should ensure that the data remains secure and confidential and that only authorized personnel can access the data.

Usability

The system should be user-friendly and easy to use, even for non-technical personnel. The component should be designed with an intuitive user interface that allows users to navigate and interact with the data easily and efficiently.

Reliability

The system should be highly available and reliable, with minimum downtime or interruptions. The component should be designed with fault tolerance and backup and recovery mechanisms to ensure that data and analysis results remain available in case of failures or disasters.

6. Commercial Value

- Commercial Value of the Overall System:
  - The commercial value of this research lies in the potential for the developed system to provide law enforcement agencies with a highly efficient and accurate tool for crime prevention and public safety. By leveraging machine learning algorithms and vast amounts of data, the system can help investigators make informed decisions and allocate resources effectively. This can result in reduced crime rates, improved public safety, and a more efficient allocation of law enforcement resources. Moreover, the developed system can potentially be marketed to law enforcement agencies worldwide as a cutting-edge solution to crime prevention and public safety challenges. The research can also attract funding from government and private organizations, as it offers a promising solution to one of the most pressing societal issues. Overall, this research has significant commercial potential in the law enforcement and public safety industries.

- Commercial Value of the individual Component:
  - The research component described above has significant commercial value as it provides a powerful tool for security agencies to analyze and predict criminal activities. By using machine learning algorithms and data analysis techniques, the component can identify patterns, trends, and group according to the similarities among criminal cases, which can help law enforcement agencies to prevent future crimes and enhance public safety. This capability can be highly valuable to government agencies, and other organizations involved in crime prevention and public safety. Additionally, the interactive visualization and user-friendly interface of the component make it easy to use, providing an efficient and effective solution for analyzing and predicting criminal activities. Overall, the research component help the police and other law enforcement agencies to take decision more efficiently and accurately enhances the commercial value.

References

[1] Karabo Jenga, Cagatay Catal , and Gorkem Kar , "Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 10, pp. 4329-4345, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s12652-023-04530-y.

[2] Shah, N., Bhagat, N., & Shah, M. (2020). Crime forecasting using machine learning and computer vision. Journal of Ambient Intelligence and Humanized Computing, 11(9), 3697-3705.

[3] Walczak, S. (2019). Predicting the types of crimes committed in the city of Chicago using neural networks. Neural Computing and Applications, 31(8), 3795-3802.

[4] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100-108.

[5] Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2015). Crime prediction using decision trees. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 1285-1292. DOI: 10.1109/ASONAM.2015.7339281

[6] Wang, X., Bai, Y., & Shi, H. (2019). A novel crime pattern detection algorithm based on machine learning. Future Generation Computer Systems, 99, 139-146.

[7] Johnson, S. D., Bowers, K. J., & Pease, K. (2014). Predictive mapping of crime by ProMap: Accuracy, units of analysis, and the environmental backcloth. Journal of Quantitative Criminology, 30(2), 165-187.

[8] Xu, X., Jiang, C., Cai, Q., & Zhang, H. (2019). Machine learning for crime analysis: a systematic literature review. Journal of Ambient Intelligence and Humanized Computing, 10(5), 1815-1834.

[9] Li, Y., Li, X., Gao, J., Liu, J., & Li, Y. (2021). Crime hotspots detection using machine learning and social media. IEEE Access, 9, 59554-59563.

[10] Porta, J. M., & Crucitti, P. (2017). Urban crime prediction through machine learning techniques. Applied Geography, 88, 61-70.

[11] Zheng, J., Zhao, Z., & Zhang, C. (2019). Crime prediction based on machine learning: A systematic literature review. Computers, Materials & Continua, 59(2), 597-617.

[12] Moustafa, N., & Slay, J. (2019). A novel hybrid crime prediction model using neural networks and decision trees. Expert Systems with Applications, 123, 299-316.

[13] Macrotrends. (n.d.). Sri Lanka Crime Rate 1960-2021. Retrieved March 19, 2023, from https://www.macrotrends.net/countries/LKA/sri-lanka/crime-rate-statistics.

[14] Prithi S, Aravindan S, Anusuya E, Kumar AM (2020) GUI based prediction of crime rate using machine learning approach. Int J Comput Sci Mob Comput 9(3):221–229.

[15] Jha P, Jha R, Sharma A (2019) Behavior analysis and crime prediction using big data and machine learning. Int J Recent Technol Eng 8(1):461–468

[16] Shojaee S, Mustapha A, Sidi F, Jabar MA (2013) A study on classification learning algorithms to predict crime status. Int J Digital Content Technol Appl 7(9):361–369

[17] Alves LGA, Ribeiro HV, Rodrigues FA (2018) Crime prediction through urban metrics and statistical learning. Phys A Stat Mech Appl 505:435–443. https://doi.org/10.1016/j.physa.2018.03.084

[18] Sun CC, Yao CL, Li X, Lee K (2014) Detecting crime types using classification algorithms. J Digit Inf Manag 12(8):321–327. https://doi.org/10.14400/JDC.2 014.12.8.321

[19] Bharati A, Sarvanaguru RAK (2018) Crime prediction and analysis using machine learning. Int Res J Eng Technol 5(9):1037–1042

[20] Obuandike GN, Isah A, Alhasan J (2015) Analytical study of some selected classification algorithms in WEKA using real crime data. Int J Adv Res Artif Intell 4(12):44–48. https://doi.org/10.14569/IJARAI.2015.041207

[21] Ahishakiye E, Taremwa D, Omulo EO, Niyonzima I (2017) Crime prediction using decision tree (J48) classification algorithm. Int J Comput Inf Technol 6(3):188–195

[22] Gorr W, Olligschlaeger AM, Thompson Y (2000) Assessment of crime forecasting accuracy for deployment of police. Int J Forecast 2000:743–754

[23] Jangra M, Kalsi S (2019) Crime analysis for multistate network using naive Bayes classifier. Int J Comput Sci Mob Comput 8(6):134–143

[24] Iqbal R, Murad MAA, Mustapha A, Panahy PHS, Khanahmadliravi N (2013) An experimental study of classification algorithms for crime prediction. Indian J Sci Technol 6(3):4219–4225. https://doi.org/10.17485/ijst/2013/v6i3.6