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

(Clustering crimes against women and crime forecasting prediction)

Project Id

TMP-23-224

Project Proposal Report

Anubama Logaratnam - IT20068196

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
Anubama Logaratnam	IT20068196	

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 ..... Date: 17/04/2023.

(Ms. Hansika Mahaadikara)

ABSTRACT

The increasing economic crisis in Sri Lanka has led to a drastic increase in criminal cases. However, most of these cases are still analyzed manually, leading to longer investigation times and potential inaccuracies. At times some of the cases are not even taken to count from the police stations, considering minor and common cases. And, as the cases are handwritten and classified using human at times the accuracy level can be also doubtful, and the time taken for investigation can be lengthy. Manual processes for data collection, recording, and analysis are subject to human error, which can lead to inaccuracies in the information used to make decisions and allocate resources. Moreover, there is not any automated police system to analyze crime and to make future predictions based on this analysis and to identify connection between the crimes. Furtherly, police officers may unconsciously bring biases and discriminatory attitudes to their work, which can affect their ability to make impartial and fair judgements when analyzing criminal cases. Considering the importance of crime forecasting and crime-based analysis, this system provides a feature to provide timely and useful information on patterns, trends, and future crime predictions of crimes. This paper proposes a solution to reduce and prevent crimes against women using machine learning techniques. The approach is to cluster crimes against women based on their characteristics, such as location, type of crime, year of crime, etc. The clustering algorithm will help in identifying patterns and trends in the data, which will be used to develop a crime forecasting model. The crime forecasting model will use historical data to predict the likelihood of a crime occurring in a particular area and time. This will enable law enforcement agencies to take preventive measures to reduce crimes against women. The proposed solution aims to create an effective approach to address crimes against women and ensure the safety of women in society.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I am grateful to our project supervisor Ms. Hansika Mahaadikara and Co-Supervisor - Ms. Sanjeevi Chandrasiri for their kind guidance, inspiration, motivation, and constructive suggestion that was helpful for me in my research project idea and preparation of this proposal. I also wish to express my gratitude to the project coordinator Mr. Jayantha Amararachchi for this support and opportunity. Also, I'd want to express my gratitude to all the educators, students, and parents who have been in touch with me for their insightful remarks and suggestions on how I may enhance the solution.

1. Introduction

1.1 Background

The future of crime prevention is rapidly evolving with advancements in technology and data analytics. While traditional policing methods will continue to play a critical role in law enforcement, the use of data-driven approaches is becoming increasingly important in identifying and preventing criminal activity.

In addition to the general advancements in technology and data analytics, there are specific areas within crime prevention that are being targeted for improvement. These include crime against women, accident-related issues, crime clustering, and risk assessment. THE FUTURE OF CRIME PREVENTION system provides police investigators with a powerful tool to effectively analyze pastcriminal cases and make accurate predictions about future crimes. The system will use vast amounts of crime, accident, crimes against women and case file datasets to train machine learning algorithms. The resulting predictions will help police investigators make informed decisions and allocate resources effectively to prevent crimes and protect the public. This system will provide a comprehensive and sophisticated solution to the challenges faced by police in reducing crime and ensuring public safety. This study will make use of supervised and unsupervised learning, decision trees, random forests and neutral networks as well as knowledge of the training, validating, and testing procedures.

Crimes against women are a significant problem in many societies worldwide. Such crimes include sexual harassment, assault, domestic violence, human trafficking, and other forms of violence that specifically target women. These crimes have a severe impact on women's physical, emotional, and mental health and often go unreported due to stigma, fear, and lack of trust in the legal system.

To address this problem, various initiatives have been taken by governments and non-governmental organizations worldwide. However, prevention and reduction of such crimes remain a challenge due to the complexity of the problem and the limited resources available to law enforcement agencies.

Complaints received from January 2020 to May 2021 to the 1938 helpline

Complaint category	No of cases
Domestic violence	2442
Family dispute	1119
Cyber crime	404
Hurt (physical)	283
Sexual harassment	79
Rape	16
Maintenance & divorce	137
Land dispute	28
Child abuse	13
Employment promotion	33
Guardianship	4
Miscellaneous	1002
Violence of human rights	1
Total	5561

Figure 1.1: Summary of complaints received to women helpline.
https://www.dailymirror.lk/news-features/Pandemic-related-Domestic-Violence%3A-Victims-don%E2%80%99t-often-get-the-desired-help/131-213849

As indicated in Figure 1.1, the data gathered by the survey indicates that Although violence against women cases have been categorized into many, there are similar categories in the list that relate to domestic violence.

As shown in Figure 1.2, one in five (20.4%) ever-partnered women have experienced physical and/or sexual violence by an intimate partner in their lifetime and 6.0% have experienced it in the last 12 months. Comparing the separate rates for physical violence (experienced by 18.9% of women), and sexual violence (6.8%), the combined rate shows a strong overlap, meaning most women who experience sexual violence have also been subject to physical violence. Looking across the five forms of violence by a partner, lifetime prevalence is highest for controlling behaviors (19.1%), physical violence (18.9%) and economic violence (18.1%) – all in similar rates. Most recent violence, in the last 12 months, was notable highest for controlling behavior (9.9%) followed by economic (7.7%) emotional (6.2%) and physical (4.8%) violence. Sexual violence was least often reported during lifetime (6.8%) and in the last 12 months (2.2%).

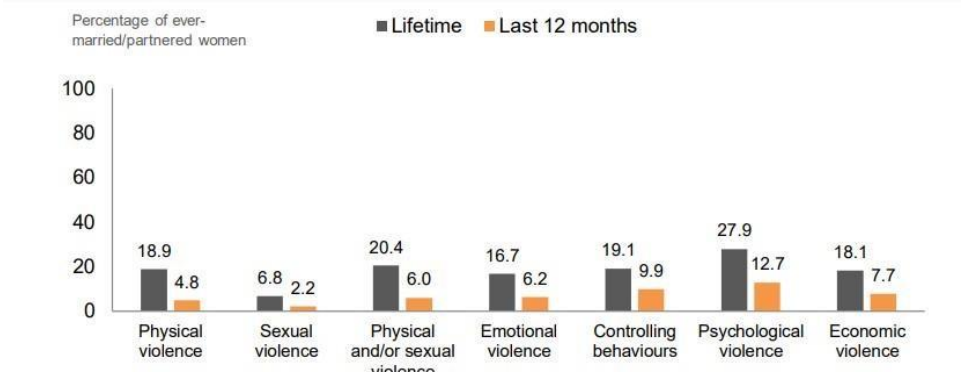

Figure 1.2: physical, sexual, physical and/or sexual, emotional violence, controlling behaviors, psychological violence and economic violence by any partner during lifetime and the last 12 months among ever-partnered women, Sri Lanka (2022)https://srilanka.unfpa.org/sites/default/files/pub-pdf/womens_wellbeing_survey.pdf

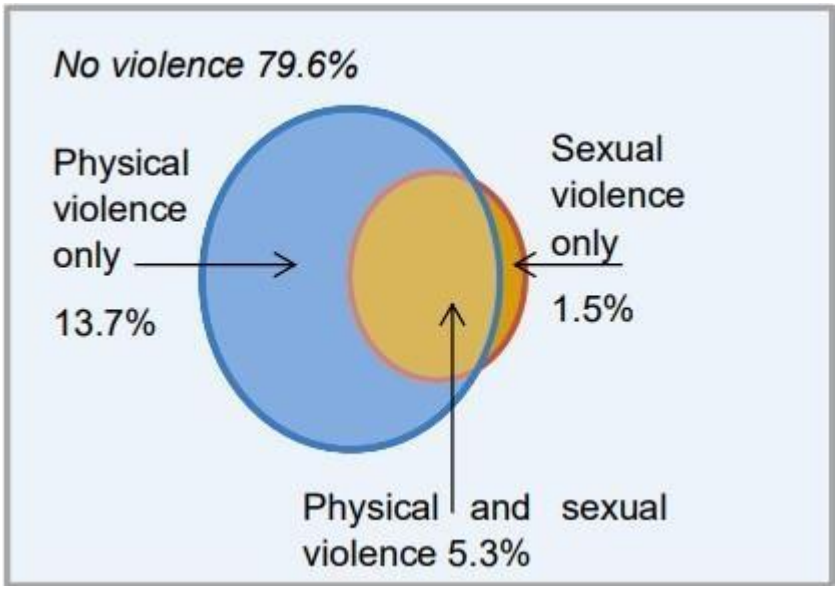

Figure 1.3: Overlap of the prevalence of physical and/or sexual violence by a partner in lifetime, Sri Lanka.https://srilanka.unfpa.org/sites/default/files/pub-pdf/womens_wellbeing_survey.pdf

In analyzing women's coping strategies when living with violence by a partner, the study found that nearly half (49.3%) of the women who experienced sexual violence by a partner did not seek formal help anywhere due to reasons such as shame, embarrassment and fear of being blamed or not being believed, and/or thinking the violence was normal or not serious enough to seek help.

Through this research, this system proposes an application to clustering crimes against women and crime forecasting prediction system where Machine learning techniques offer a promising approach to address this challenge. Clustering algorithms can help identify patterns and trends in the data related to crimes against women. These patterns can be used to develop a crime forecasting model that can predict the likelihood of a crime occurring in a particular area and time. The crime forecasting model can enable law enforcement agencies to take proactive measures to prevent such crimes, including increasing patrols, installing surveillance cameras, and deploying officers in high-risk areas.

Overall, clustering crimes against women and using machine learning for crime forecasting prediction can provide an effective solution to prevent and reduce crimes against women, promoting gender equality and women's safety in society.

1.2 Literature Review

This section briefly reviews previous research on violence against women and how to predict it using various machine learning techniques. A mobile application designed for rape victims by Mahmud et al. [14] provides basic self-defense courses, location monitoring, alarm messages, and legal support for rape survivors. In a different study, A Comprehensive Study on crime predictive systems [9] where it explains about the Crime predictive techniques, refer to the use of data analysis and machine learning algorithms to forecast criminal activities. A comprehensive study of these techniques involves examining their effectiveness, ethical consi derations, and potential biases. A mobile app called "Joy 109" was also released by the Information and Communication Technology (ICT) division of the Bangladeshi government to provide urgent aid in the struggle against violence against women and children. The app enables users to send real-time distress SMS to the National Helpline Center, three FNF numbers, the Police Super, and Deputy Metropolitan Police Commissioner (109). Moreover, it has the capability of sharing a GPS location, an audio recording, and even a photo of the immediate area without notifying the offender [17]. Analysis and crime prediction [10] by clustering involves using clustering algorithms to identify patterns in crime data and forecast criminal activities in specific areas. The research aims to develop a more efficient and accurate crime prediction model by analyzing and grouping similar crime incidents.

Nonetheless, Islam et al. [18] created the wearable "SafeBand"[1] for women's safety, which was linked to two mobile applications utilized by the victim assistance program and the police, respectively. The victim's location will be determined using the GPS of that device by pressing an emergency button, and messages will be sent to surrounding police stations as well as pre-saved cell contacts. Like this, Hossain et al. [19] created another mobile application for women's safety that has an SOS button and may call trusted contacts with voice instructions or by shaking the device.

1.3 Research Gap

For immediate victim support, some studies created wearable technology, mobile and online applications, and others worked on the analysis and prediction of women's and children's violence. Only a few studies suggested preventing violences from happening and taking safety measures regarding the violences against women. There is no standard method or approach for clustering crimes against women, and as a result, different studies use different methods and algorithms. This makes it difficult to compare the results obtained from different studies. To help the government and other law enforcement groups take preventive action, our system provides and helps the government and other law enforcement organizations to take preventive steps by providing the forecast of the violences and locations. Nonetheless, the system's crime prediction module can predict crimes against women and provide a report for taking precautions, minimizing the crime rate. And showed comparisons of different types of crime statistics based on location. Moreover, during this research we'll be trying to cover predict data analysis of different crimes state-wise and year-wise as well.

Table 1.1: Shows a tabularized format of the explanation.

RESEARCH	FUTURE PREDICTION OF CRIMES	FOCUS SPECIFIC CHARACTERISTICS OF WOMEN CRIMES	POLICE RELATED	CRIME PREVENTION	LOWER COST
A COMPREHENSIVE STUDY ON CRIME PREDICTION TECHNIQUES[9]	✓	✗	✗	✗	✓
CLUSTERING-BASED CRIME PREDICTION WITH APPLICATION TO HUMAN TRAFFICKING[6]	✓	✗	✗	✗	✓
"JOY 109"	✗	✗	✓	✗	✗
SAFE BAND[1]	✗	✓	✓	✗	✗
ANALYSIS AND CRIME PREDICTION BY CLUSTERING [10]	✓	✗	✗	✗	✓
CRIME CLUSTERING AND FORECASTING PREDICTION SYSTEM FOR CRIME AGAINST WOMEN	✓	✓	✓	✓	✓

Table 1.1: Comparison of other available applications

1.4 **Research Problem**

Crimes against women, such as domestic violence, sexual assault, and human trafficking, continue to be a major social issue in many parts of the world. Despite efforts to combat these crimes, there is still a need for effective crime prevention measures. Crime forecasting is a technique that can help to predict future criminal activities by analyzing historical data and identifying patterns and trends.

There are existing solutions in the market for crime prediction but the problem with current crime forecasting models is that they tend to focus on overall crime rates and do not consider the specific characteristics of crimes against women. This makes it difficult for law enforcement agencies to take preventive measures targeted at such crimes. The problem is exacerbated by the fact that it is often difficult to predict and prevent these crimes. Traditional policing methods are often reactive, meaning that law enforcement agencies wait for crimes to occur before responding. However, clustering and prediction models offer the potential to identify areas with high rates of crimes against women and predict future criminal activity. This would allow law enforcement agencies to allocate resources more effectively and prevent crimes from occurring.

The proposed research aims to develop an Effective Clustering-based Crime Forecasting Model for Crimes Against Women to Enable Preventive Step, which can predict crimes against women and provide information on the locations and types of crimes. This will enable law enforcement agencies to take preventive measures and allocate resources more effectively. The research will contribute to the development of effective crime prevention strategies, particularly for crimes against women.

2. Objectives

2.1 Main Objectives

The main objective of this component is Clustering crimes against women and crime forecasting prediction to take preventive steps by providing the forecast of the violences and locations.

2.2 Sub Objectives

- Identify the key characteristics of crimes against women that can be used to develop a clustering-based crime forecasting model.
- Analyze historical data on crimes against women to identify patterns and trends that can help in forecasting future crimes.
- Evaluate different clustering algorithms for grouping crimes against women into meaningful clusters and select the most appropriate algorithm.
- Develop a clustering-based crime forecasting model that can predict future crimes against women and provide information on the locations and types of crimes.
- Evaluate the effectiveness of existing crime prevention programs and interventions targeted at crimes against women and identify areas where additional resources or strategies may be needed.
- Validate the accuracy of the clustering-based crime forecasting model using statistical tests.
- Evaluate the effectiveness of the clustering-based crime forecasting model in preventing crimes against women.

3. Methodology

This section briefly discusses the study's methodology which includes two stages: developing the application and evaluating the web application.

Figure [3.1] illustrates the High-level architecture diagram of our proposed system from the beginning of the data collection.

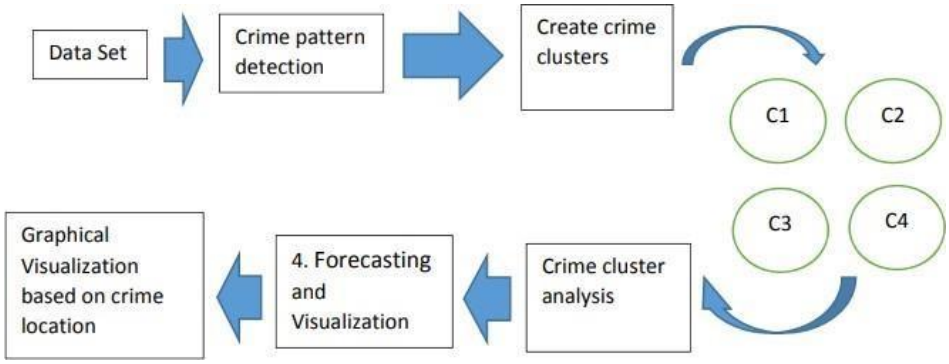

Figure 3.1 High level architecture diagram

Clustering crimes against women and predicting crime forecasts are two distinct tasks in data analysis. However, they can be linked by using clustering as a preprocessing step for crime forecasting.

This can be divided into several steps:

Data collection

The first step is to collect the relevant data on crimes against women. This data can be obtained from online, official crime records, police reports, or other reliable sources. The data should include information on the type of crime, location, and year.

Data preprocessing

Once the data has been collected, it must be preprocessed to remove any irrelevant or redundant information. This step involves cleaning, transforming, and formatting the data in a way that can be easily analyzed.

Clustering analysis

After preprocessing the data, the next step is to perform clustering analysis to group similar crimes together. This can be done using various clustering techniques such as K-means, hierarchical clustering, or density-based clustering. The clustering algorithm should be chosen based on the characteristics of the data and the specific research questions being addressed.

Crime trend analysis

Once the crimes have been clustered, the next step is to analyze crime trends over time. This can be done using various statistical techniques such as time-series analysis or regression analysis. The analysis should focus on identifying patterns and trends in the data, such as increasing or decreasing crime rates over time.

Crime forecasting

Finally, the last step is to use the results of the clustering and trend analysis to forecast future crime trends. This can be done using various forecasting techniques such as ARIMA, exponential smoothing, or neural networks. The forecasting model should be chosen based on the characteristics of the data.

Overall, the methodology for clustering crimes against women and predicting crime trends involves a combination of data collection, preprocessing, clustering analysis, trend analysis, and forecasting. The results of this analysis can be used to develop effective crime prevention and intervention strategies that target specific types of crimes in specific locations.

3.1 Procedures

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

3.2 Testing

In the testing part of the research, testing will assess the effectiveness of the suggested system by contrasting the results produced by the system with the results produced manually. This evaluation will be carried out by contrasting the precision and dependability of the results produced by the system with the manual results obtained from historical crimes against women cases and criminal activity data. This comparison will enable to evaluate the usefulness and efficiency of the system in precisely evaluating, forecasting, and categorizing data connected to crimes against women. The review will allow to pinpoint any areas where the system might need to be improved and then hone it so that it can produce results that are as precise and trustworthy as possible. In the end, this will assist us in creating a system that can provide valuable insights into crimes against women 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)

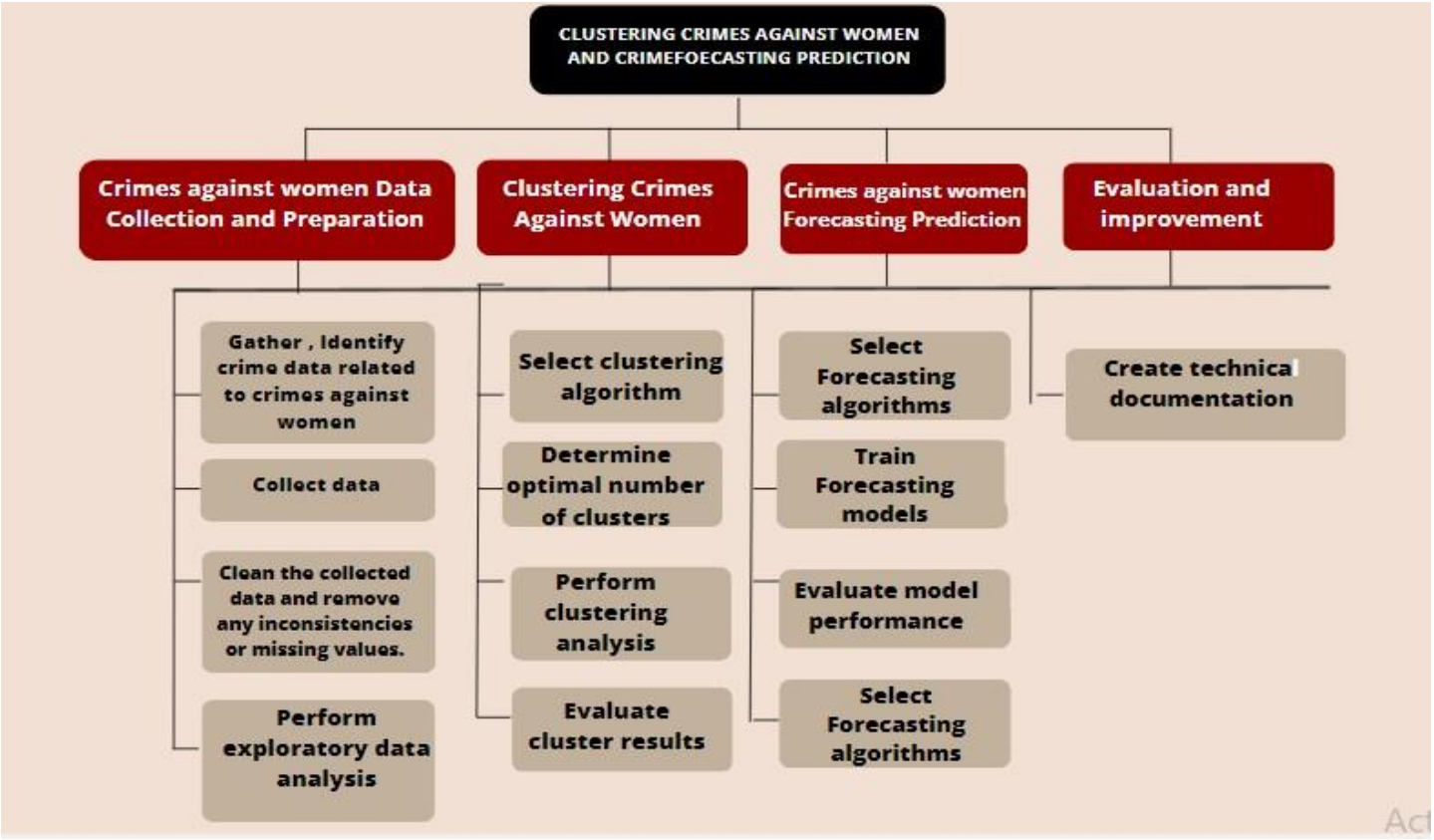

Figure 4.1: WBS diagram

4.2 Gantt Chart

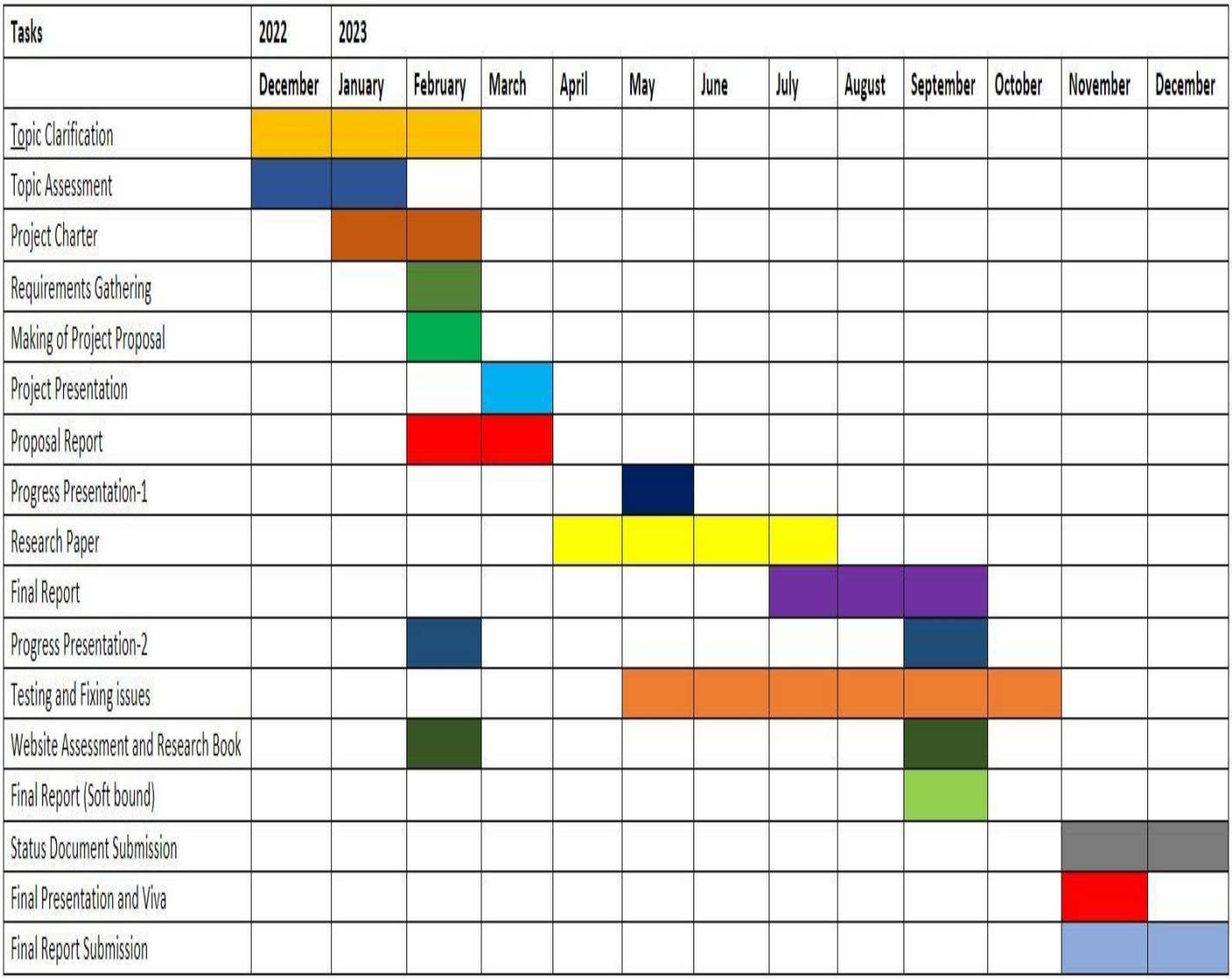

Figure 4.2: Gannt Chart

4.3 System Diagram

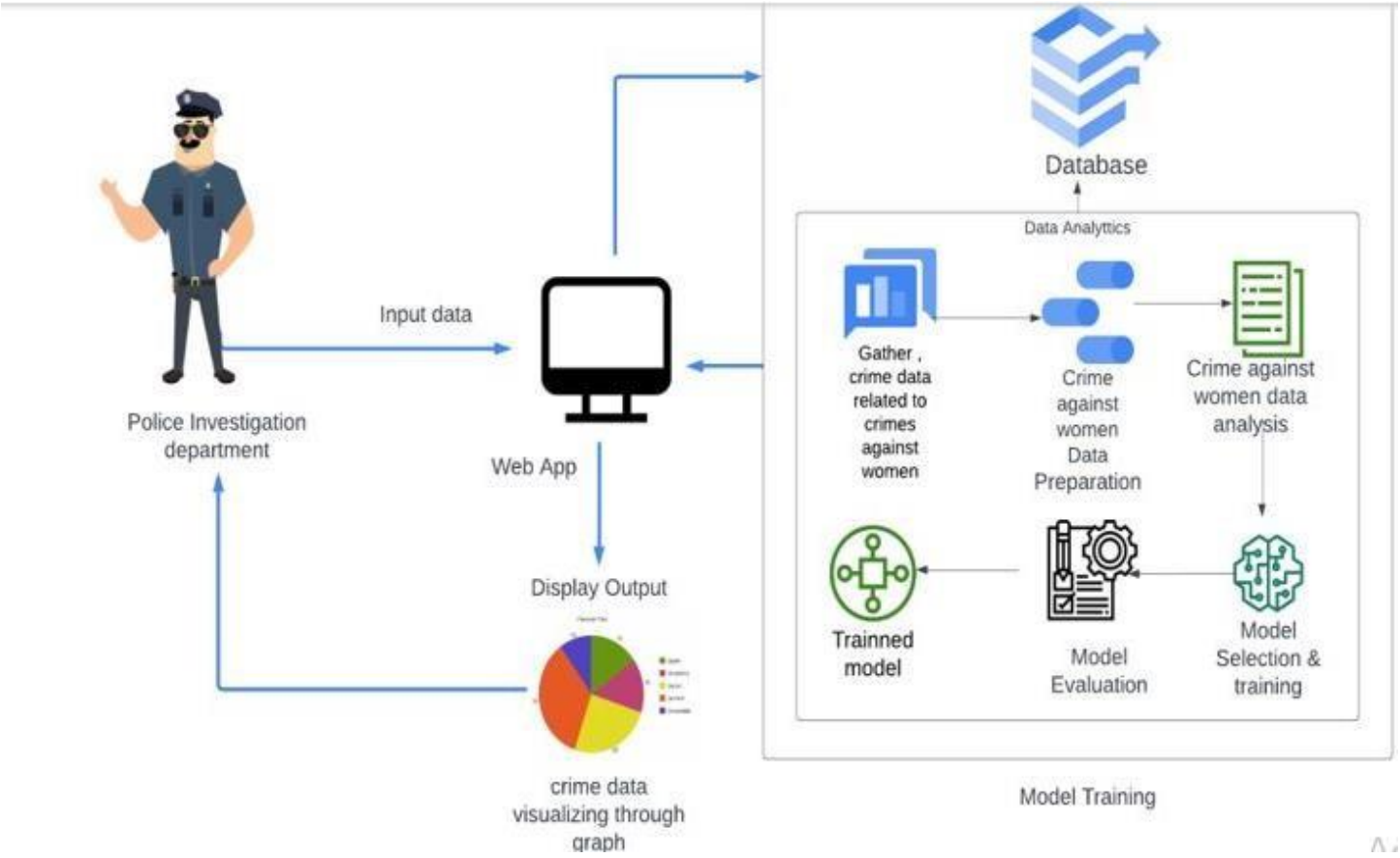

Figure 4.3.1: System diagram for clustering crimes against women and future crime forecasting prediction.

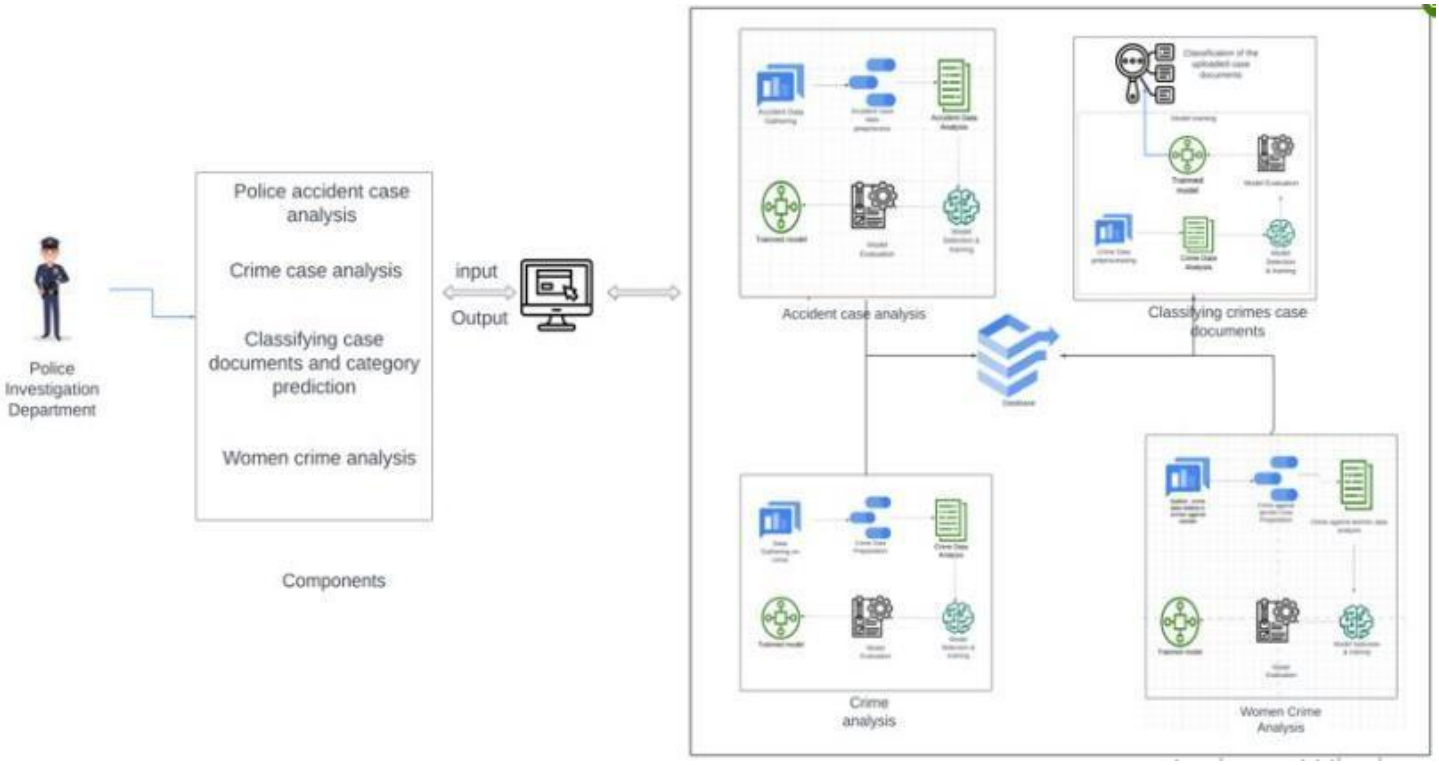

Figure 4.3.2: System diagram for overall project

5. Requirements

5.1 Functional Requirements

1. The system clusters crimes against women based on different factors such as location, type of crime, time of day, etc.
2. The system identifies patterns and trends in crime data related to crimes againstwomen.
3. The system forecasts the likelihood of future crimes against women based on historical crime data.
4. The system provides insights and recommendations to law enforcement agencies and policymakers to prevent and mitigate crimes against women.
5. The system handles large datasets and process them efficiently.

5.2 Non - Functional Requirement

1. Usability

The system has user-friendly interface that is easy to use and navigate.

2. Security

The system secures and protects sensitive data related to crimes against women.

3. Availability

The system is reliable and available 24/7 to ensure that law enforcement agencies and policymakers can access crime data and insights at any time.

4. Scalability

The system is scalable and able to handle an increasing amount of crime data as more data is collected over time.

5. Accuracy

The system provides accurate and timely crime predictions and insights to ensure that law enforcement agencies and policymakers can take proactive measures to prevent crimes against women.

6. Commercial Value

Clustering crimes against women and developing a crime forecasting prediction system can have significant commercial value for a variety of stakeholders and make it stand out among the other products. Here are some potential ways in which such a system could be monetized:

1. Law enforcement agencies: Police departments and other law enforcement agencies can use the system to better understand crime patterns and allocate resources more effectively. This could lead to increased efficiency and effectiveness in preventing and solving crimes against women, which would in turn enhance public safety and improve community relations.

2. Private security firms: Private security firms can use the system to provide enhanced protection services to women, particularly in high-risk areas. This could include providing real-time alerts and emergency response services to women who are at risk of being victimized.

3. Government agencies: Government agencies can use the system to track crime patterns and develop policies and initiatives aimed at reducing crime against women. This could include funding programs that address the root causes of violence against women, such as poverty, lack of education, and social inequality.

In summary, clustering crimes against women and developing a crime forecasting prediction system has the potential to provide significant commercial value to a wide range of stakeholders, including law enforcement agencies, private security firms, insurance companies, government agencies, and research institutions.

References

[1] “Developing a Machine Learning Based Support System for Mitigating the Suppression Against Women and Children” available on https://www.researchgate.net/publication/357724993_Developing_a_Machine_Learning_Based_Support_System_for_Mitigating_the_Suppression_Against_Women_and_Children

[2]“Who women violence,” available onhttps://www.who.int/news-room/fact-sheets/detail/violence-against-women [Online, Accessed: 2021-07-15].

[3] P. Patil, "K Means Clustering : Identifying F.R.I.E.N.D.S in the World of Strangers," 20 May 2018. [Online]. Available: https://towardsdatascience.com/k-means-clustering- identifying-f-r-i-e-n-d-s-in-the-world-of-strangers-695537505d

[4] “Violence against women in the United States,” available on https://centerforfamilyjustice.org/community-education/statistics/?fbclid=IwAR1L48Js6P079ZRfLRsslS0zyA8bzLelP0J36la0YiSHgzoCAMxP5kBwMAM ,[Online, Accessed: 2021-07-15].

[5] “UNICEF south Asia,” available onhttps://www.unicef.org/rosa/topics/gender-based-violence , [Online, Accessed: 2021-07-15]

[6] "Using Clustering and Classification Techniques for Predicting Crime Hotspots For Human trafficking" by R. Subramanian and S. Padmavathi. This research paper discusses the use of clustering techniques to identify crime hotspots, and machine learning algorithms for prediction.

[7] "Forecasting Crime using Machine Learning and Social Media" by S. Adhikari et al. This research paper discusses the use of machine learning algorithms for crime prediction using data from social media.

[8] "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations" by U. Bhatti and N. Ullah. This book chapter discusses the use of crime forecasting in law enforcement operations, including the use of clustering techniques and machine learning algorithms.

[9]A Comprehensive Study on crime predictive systems https://www.researchgate.net/publication/226799979_Review_of_Current_Crime_Prediction_Techniques

[10] ”Analysis and Prediction of Crimes by Clustering and Classificaion”https://www.researchgate.net/publication/281478329_Analysis_and_Prediction_of_Crimes_by_Clustering_and_Classification

[11] Editors. (2019). Merriam Webster. Retrieved 12 15, 2019, fromhttps://journals.plos.org/plosone/article?id=10.1371/journal.pone.0217015

[12] Guruge, S., Illesinghe, V. J., Gunawardhana, N., & Perera, J. (2015). Intimate Partner Violence in Sri Lanka;A scoping review. Ceylon Medical Journal , 60 (4), 6.

[13] Helpguide.org. (2019). Helpguide.org. Retrieved 12 16, 2019, from https://www.helpguide.org/articles/abuse/domestic-violence-and-abuse.htm

[14] Ibe, C. (2017). Harvard Political Review. Retrieved 12 16, 2019, from https://harvardpolitics.com/world/the-nordic-paradox-gender-equity-and-sexual-assault/
Samules, F., Jones, N., & Gupta, T. (2017). Tackling Intimate Partner Violence in South Asia. ODI.

[15] S. R. Mahmud, S. N. Tumpa, A. B. Islam, C. N. Ferdous, N. Paul, and T. T. Anannya, "Bonitaa: A smart approach to support the female rapevictims," in 2017 IEEE Region 10 Humanitarian Technology Conference(R10-HTC). IEEE, 2017, pp. 730–733.

[16]S. I. Ahmed, S. J. Jackson, N. Ahmed, H. S. Ferdous, M. R. Rifat,A. Rizvi, S. Ahmed, and R. S. Mansur, "Protibadi: A platform for fighting sexual harassment in urban bangladesh," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014,pp. 2695–2704.

[17]M. N. Islam, N. T. Promi, J. M. Shaila, M. A. Toma, M. A. Pushpo, F. B.Alam, S. N. Khaledur, T. T. Anannya, and M. F. Rabbi, "Safeband: A wearable device for the safety of women in bangladesh," in Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia, 2018, pp. 76–83.