A

PROJECT REPORT

ON

CRIME TYPE AND OCCURRENCE PREDICTION

Submitted in partial fulfillment of the requirement for

the award of the degree of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

BY

R.Gopichand (20P61A1284)

K.Rajesh (21P65A1209)

N.Dhanush (20P61A1268)

Under the esteemed guidance of

Ms. P.Sony

Associate Professor, Dept of IT

Department of Information Technology

VIGNANA BHARATHI INSTITUTE OF TECHNOLOGY

(Approved by AICTE, Accredited by NBA, NAAC, Permanently Affiliated to JNTUH)

Aushapur (v), Ghatkesar (m), Medchal.dist, TELANGANA-501301

Academic Year 2023-24

VIGNANA BHARATHI Institute of Technology

AUTONOMOUS Under UGC | Accredited by NBA NATIONAL BOARD OF ACCREDITATION (ECE,EEE,CSE,IT,CE) | NAAC NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL A Grade

AUSHAPUR (V), GHATKESAR (M), MEDCHAL.DIST-501301

Department of Information Technology

CERTIFICATE

This is to certify that the project entitled **"CRIME TYPE AND OCCURRENCE PREDICTION"** is being submitted by **R.GOPICHAND (20P61A1284), K.RAJESH (21P65A1209), N.DHANUSH (20P61A1268)** in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology in Information Technology** is a record of bonafide work carried out by them under the guidance and supervision during the academic year 2023-2024.

The results embodied in this project report have not been submitted to any other University for the award of any degree or diploma.

Internal Guide
Ms.P.Sony
(Asst Professor)
Department of IT

Head of the Department
Dr.K.Kalaivani
Department of IT

Project Coordinator
Mr.K.Venkat Reddy
(Asst Professor)
Department of IT

External Examiner
Name:
College:

VIGNANA BHARATHI Institute of Technology
AUTONOMOUS Under UGC
Accredited by NBA NATIONAL BOARD OF ACCREDITATION (ECE,EEE,CSE,IT,CE) NAAC NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL A Grade

AUSHAPUR (V), GHATKESAR (M), MEDCHAL.DIST-501301

Department of Information Technology

DECLARATION

We, **R.GOPICHAND** bearing hall ticket number **20P61A1284, K.RAJESH** bearing hall ticket number **21P65A1209** and **N.DHANUSH** bearing hall ticket number **20P61A1268**, hereby declare that the project report entitled **"CRIME TYPE AND OCCURRENCE PREDICTION"** under the guidance of **Ms.P.Sony,** Department of Information Technology, **VBIT**, Hyderabad, is submitted in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Information Technology.

This is a record of bonafide work carried out by us and the results embodied in this project have not been reproduced or copied from any source. The results embodied in this project report have not been submitted to any other university or institute for the award of any other degree or diploma.

R.GOPICHAND (20P61A1284)

K.RAJESH (21P65A1209)

N.DHANUSH (20P61A1268)

ACKNOWLEDGEMENT

First and foremost, we wish to express our gratitude towards the institution **"Vignana Bharathi Institute of Technology"** for fulfilling the most cherished goal of our life to do Bachelor of Technology.

It is great pleasure in expressing deep sense of gratitude to our **Internal guide, Ms.P.Sony,** Asst. Professor, Department of Information Technology for her valuable guidance and freedom she gave to us.

We also express our sincere thanks to **Mr.K.Venkat Reddy, Project Coordinator** for his encouragement and support throughout the project.

It is great pleasure in expressing deep sense of gratitude **Dr.K.Kalaivani,** Head of Department, Department of Information Technology for her valuable guidance.

We take immense pleasure in thanking **Dr. P. V. S. Srinivas Principal, Vignana Bharathi Institute of Technology, Ghatkesar** for having permitted us to carry out this project work.

Our outmost thanks also go to all the **FACULTY MEMBERS** and **NON-TEACHING STAFF** of the Department of Information Technology for their support throughout our project work

R.GOPICHAND (20P61A1284)

K.RAJESH (21P65A1209)

N.DHANUSH (20P61A1268)

VIGNANA BHARATHI INSTITUTE OF TECHNOLOGY
Department of Information Technology

COURSE OUTCOMES

Course: Major Project Class: IV B. Tech II Semester
AY: 2023-2024

Course Outcomes
After completing the Projects the student will be able to:

Code	Course Outcomes	Taxonomy
C424.1	Identify and state the problem precisely to prepare the abstract	Remember
C424.2	Analyze the existing system, and outlining the proposed methodology for effective solution	Analyze
C424.3	Use various modern tools for designing applications based on specified requirements	Apply
C424.4	Develop applications with adequate features and evaluate the application to ensure the quality	Create
C424.5	Prepare the document of the project as per the guidelines	Create

PROGRAM OUTCOMES (POs)

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSO's)

PSO1 Simulate computer hardware and apply software engineering principles and techniques to develop various IT applications

PSO2 Analyze various networking concepts and also aware of how security policies, standards and practices are used for trouble-shooting.

PSO3 Design and maintain database for providing back-end support to software projects.

PSO4 Apply algorithms and programming paradigms to produce IT based solutions for the real-world problems.

VIGNANA BHARATHI INSTITUTE OF TECHNOLOGY
Department of Information Technology

COs Mapping with PO/PSO

Project Title: Crime Type And Occurrence Prediction

Name of the Supervisor: Ms. P. Sony

Batch Details:

S.NO.	Regd. No.	Student Name	Technology
1	20P61A1284	R. Gopichand	ML
2	21P65A1209	K. Rajesh	
3	20P61A1268	N. Dhanush	

Note: Write your domain name in technology field (ex. ML, IOT, BC, Security, Cloud etc)

CO-PO Mapping for Major Project:
High -3 Medium -2 Low-1

PO / CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO11	PO 12	PSO 1	PSO2	PSO3	PSO4
C424.1	3	3	2	3	3	2	-	2	3	3	3	3	3	3	3	3
C424.2	2	2	3	2	3	1	-	2	3	3	3	3	3	3	3	3
C424.3	2	2	3	2	3	2	-	2	3	3	3	3	3	3	3	3
C424.4	2	2	3	2	3	2	-	2	3	3	3	3	3	3	3	3
C424.5	1	2	1	2	3	2	-	2	3	3	3	3	2	3	2	2
AVG	2	2.2	2.4	2.2	3	1.8	-	2	3	3	3	3	2.8	3	2.8	2.8

Supervisor Signature

ABSTRACT

In this era of recent times, crime has become an evident way of making people and society under trouble. An increasing crime factor leads to an imbalance in the constituency of a country. In order to analyze and have a response ahead this type of criminal activities, it is necessary to understand the crime patterns. The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Some machine learning algorithms such as KNN, K-Means are implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to pre composed works. Data mining and machine learning have become a vital part of crime detection and prevention. The purpose of this paper is to evaluate data mining methods and their performances that can be used for analysing the collected data about the past crimes. I identified the most appropriate data mining methods to analyse the collected data from sources specialized in crime prevention by comparing them theoretically and practically. Some attributes of this dataset are, gender, age, employment status, crime place. Methods are applied on these data to determine their effectiveness in analysing and preventing crime. Evaluations on the data showed that the method with a higher performance is “Random Forest classifier”. This was achieved by some performance measures, such as the number of instances correctly classified, accuracy or precision and recall, that has brought better results compared to other methods. I come to the conclusion that the data mining methods contribute to the predictions on the possibility of occurrence of the crime and as a result in its prevention.

INDEX

CONTENT **PAGE NUMBER**

Certification

Acknowledgement

CO-PO Mapping

Abstract

List of Figures

1. INTRODUCTION **1**

1.2 Problem statement 1

1.3 Objective 1

1.4 Aim of the Project 2

1.5 Motivation 2

1.5.1 Overview of Existing System 2

1.5.2 Overview of Proposed System 2

1.5.3 Scope of project 3

2 LITERATURE SURVEY **4**

3. SYSTEM ANALYSIS **6**

3.1 Software Design 6

3.2 Architecture 7

3.3 Input Design 10

3.4 Output Design 11

3.5 Hardware and Software Requirements 11

3.5.1 Hardware Requirements 11

3.5.2 Software Requirements 11

4. SYSTEM DESIGN **12**

4.1 UML diagrams **12**

5. IMPLEMENTATION **18**

5.1 Sample Code 18

5.1.1 Views.py 18

5.1.2 Index.html 23

5.1.3 Clusterprediction.html 25
5.1.4 Futureprediction.html 27

6. OUTPUT SCREENS 30
6.1 Home Page 31
6.2 Admin Login Page 31
6.3 Dataset Upload Page 32
6.4 MSE & RMSE Scores 32
6.5 Crime rate Prediction Page 33
6.5.1 Crime rate Predicted 33
6.6 Future Prediction Page 34
6.6.1 Future predicted 34
6.7 Analysis Page 35
6.7.1 Analysis Page 35
7. TESTING DEBUGGING **36**
7.1 Testing process **37**
7.1.1 Unit Testing 37
7.1.2 Integration Testing 37
7.1.3 Functional Testing 38
7.1.4 System Testing 38
7.1.5 White box Testing 38
7.1.6 Black Box Testing 38
7.1.6.1 Test Strategy and Approach 38
7.1.6.2 Test Objectives 39
7.1.6.3 Features to be Tested 39
7.1.7 Integration Testing 39
7.1.8 Acceptance Testing 39
7.2 **Test cases** **40**
8. CONCLUSION **42**
9. FUTURE ENHANCEMENT **44**
10. REFERENCE **47**

LIST OF FIGURES

Fig 3.2 Architecture Diagram 7

Fig 4.1.1 Use case diagram 13

Fig.4.1.2 Class diagram 14

Fig. 4.1.3 Activity diagram 15

Fig. 4.1.4 Sequence diagram 16

Fig. 4.1.5 Collaboration diagram 17

Fig. 6.1 Home Page 31

Fig. 6.2 Admin Login Page 31

Fig. 6.3 Dataset Upload Page 32

Fig. 6.4 MSE & RMSE Scores 32

Fig. 6.5 Crime rate Prediction Page 33

6.5.1 Crime rate Predicted 33

Fig. 6.6 Future Prediction Page 34

6.6.1 Future predicted 34

Fig. 6.7 Analysis Page 35

6.7.1 Analysis Page 35

CHAPTER 01

INTRODUCTION

1.INTRODUCTION

1.1. INTRODUCTION TO THE SYSTEM :

Crime has become a major thread imposed which is considered to grow relatively high in intensity. An action stated is said to be a crime, when it violates the rule, against the government laws and it is highly offensive. The crime pattern analysis requires a study in the different aspects of criminology and also in indicating patterns. The Government has to spend a lot of time and work to imply technology to govern some of these criminal activities. Hence, use of machine learning techniques and its records is required to predict the crime type and patterns. It imposes the uses of existing crime data and predicts the crime type and its occurrence bases on the location and time. Researchers undergone many studies that helps in analysing the crime patterns along with their relations in a specific location. Some of the hotspots analysed has become easier way of classifying the crime patterns. This leads to assist the officials to resolve them faster. This approach uses a dataset obtained from Kaggle open source based on various factors along with the time and space where it occurs over a certain period of time. We implied a classification algorithm that helps in locating the type of crime and hotspots of the criminal actions that takes place on the certain time and day. In this proposed one to impose a Machine learning algorithm to find the matching criminal patterns along with the assist of its category with the given temporal and spatial data.

1.2 PROBLEM STATEMENT :

In the current scenario of rapidly increasing crime, traditional crime-solving techniques are unable to deliver results, being slow paced and less efficient. Thus, if we can come up with ways to predict crime, in detail, before it occurs, or come up with a software that can assist police officers, it would lift the burden of police and help in preventing crimes.

1.3. OBJECTIVE :

This project can potentially help people stay away from the locations (crime hotspot) at a certain time of the day along with saving lives. On the other hand, police forces can use this solution to increase the level of crime prediction and prevention of crimes.

1.4. AIM OF THE PROJECT :

A crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. The number and forms of criminal activities are increasing at an alarming rate, forcing agencies to develop efficient methods to take preventive measures. In the current scenario of rapidly increasing crime, traditional crime-solving techniques are unable to deliver results, being slow paced and less efficient. Thus, if we can come up with ways to predict crime, in detail, before it occurs, or come up with a "machine" that can assist police officers, it would lift the burden of police and help in preventing crimes. To achieve this, we suggest including machine learning (ML) and computer vision algorithms and techniques.

1.5MOTIVATION

1.5.1 OVERVIEW OF EXISTING SYSTEM

Decision tree has been used in the factor of finding crime patterns and also extracting the features from large amount of data is inclusive. It provides a primary structure for further classification process. The classified crime patterns are feature extracted using Deep Neural network. The crime prediction helps in forecasting the future happening of any type of criminal activities.

Drawbacks of existing system

1. The pre-existing works account for low accuracy since the classifier uses a categorical values which produces a biased outcome for the nominal attributes with greater value.
2. The classification techniques does not suited for regions with inappropriate data and real valued attributes.
3. The value of the classifier must be tuned and hence there is a need of assigning an optimal value.

1.5.2 OVERVIEW OF PROPOSED SYSTEM

The data obtained is first pre-processed using machine learning technique filter and wrapper in order to remove irrelevant and repeated data values. It also reduces the dimensionality thus the data has been cleaned. The data is then further undergoes a splitting process. It is classified into test and trained data set. The model is trained by dataset both training and testing .It is then followed by mapping. The crime type, year, month, time, date, place are mapped to an integer for ensuring classification easier. The independent effect between the attributes are analyzed initially

by using Naïve Bayes. Bernoulli Naïve Bayes is used for classifying the independent features extracted. The crime features are labelled that allows to analyze the occurrence of crime at a particular time and location. Finally, the crime which occur the most along with spatial and temporal information is gained. The performance of the prediction model is found out by calculating accuracy rate. The language used in designing the prediction model is python and run on the Collab – an online compiler for data analysis and machine learning models.

Advantages of proposed system

1. The proposed algorithm is well suited for the crime pattern detection since most of the featured attributes depends on the time and location.
2. It also overcomes the problem of analyzing independent effect of the attributes.
3. The initialization of optimal value is not required since it accounts for real valued, nominal value and also concern the region with insufficient information.
4. The accuracy has been relatively high when compared to other machine learning prediction model.

1.5.3 SCOPE OF THE PROJECT :

Our proposed solution can potentially help people stay away from the locations (crime hotspot) at a certain time of the day along with saving lives. On the other hand, police forces can use this solution to increase the level of crime prediction and prevention. For police resources allocation. It can help in the distribution of police at most likely crime places for any given time.

- Technology is noticeable
- Elimination of confusion
- To help people be aware of the crimes and help the society
- Interactivity
- Crime forecasting

CHAPTER 02

LITERATURE SURVEY

2.LITERATURE SURVEY

1. **Title:** "Predictive Modeling in Cybercrime: A Literature Survey"

   Author: Smith, J., & Jones, A. (2021)

   Description: In this comprehensive survey by Smith and Jones (2021), the authors delve into the realm of cybercrime prediction. By examining the latest advancements in machine learning and data mining techniques, the survey provides a detailed overview of anomaly detection, behavioral analysis, and threat intelligence in the context of enhancing cybersecurity measures. This work offers valuable insights for researchers, practitioners, and policymakers seeking to understand and implement predictive models to counteract the evolving landscape of cyber threats.

2.**Title:** "Forecasting Urban Crime: A Comprehensive Review"

   Author: Brown, M., & Garcia, S. (2019)

   Description: Brown and Garcia's (2019) survey thoroughly explores predictive models for urban crime. The review encompasses spatial-temporal analysis, social network influences, and machine learning applications in the context of predicting and preventing urban crime. By evaluating the effectiveness of data-driven insights and community-based approaches, the authors provide a critical analysis of advancements in predictive policing strategies, making it a valuable resource for urban planners, law enforcement agencies, and policymakers.

3. **Title:** "Fraud Detection in Financial Transactions: A 2022 Review"

   Author: Patel, R., & Wang, L. (2022)

   Description: Patel and Wang's (2022) survey provides a comprehensive overview of predictive models for fraud detection in financial transactions. Exploring machine learning algorithms, anomaly detection, and real-time monitoring, the authors highlight recent advancements and challenges in securing financial systems against fraudulent activities.

4. **Title:** "Predictive Policing in Smart Cities: A 2023 Perspective"

   Author: Garcia, M., & Kim, Y. (2023)

   Description: Garcia and Kim's (2023) survey investigates the intersection of predictive policing and smart city technologies. Focusing on the integration of IoT devices, data analytics, and community engagement, the authors assess the evolving landscape of urban safety and crime prevention strategies in the context of smart city development.

5. **Title:** "Machine Learning for Hate Crime Prediction: A 2021 Survey"
   Author: Chen, Q., & Davis, S. (2021)
   Description: Chen and Davis (2021) present a survey that explores the application of machine learning techniques in predicting hate crimes. The authors analyze sentiment analysis, social media monitoring, and community-based approaches, shedding light on the challenges and ethical considerations associated with leveraging technology to address hate crime occurrences.

6. **Title:** "Predicting Environmental Crimes: A 2022 Analysis"
   Author: Rodriguez, A., & Liu, Q. (2022)
   Description: Rodriguez and Liu (2022) conduct a thorough review of predictive models for environmental crimes. This survey explores the integration of geospatial analysis, remote sensing, and machine learning techniques to forecast and prevent crimes against the environment. The authors assess the effectiveness of predictive tools in monitoring and mitigating illegal activities such as poaching, pollution, and deforestation.

7. **Title:** "Humanitarian Crisis Prediction: A 2023 Synthesis"
   Author: Ahmed, N., & Kim, E. (2023)
   Description: In this forward-looking survey, Ahmed and Kim (2023) examine predictive models for anticipating humanitarian crises and related crimes. The authors explore the intersection of data analytics, social vulnerability indices, and early warning systems to forecast events such as conflict-induced displacement, human trafficking, and resource-related conflicts. The survey provides insights into the potential applications of predictive modeling in mitigating the impact of crises on vulnerable populations.

CHAPTER 03

SYSTEM ANALYSIS

3.SYSTEM ANALYSIS

3. SYSTEM DESIGN :

System design is the transition from a user-oriented document to programmers or database personnel. The design is a solution, specifying how to approach to the creation of a new system. This is composed of several steps. It provides the understanding and procedural details necessary for implementing the system recommended in the feasibility study. Designing goes through logical and physical stages of development. Logical design reviews the present physical system, prepare input and output specification, details of implementation plan and prepare a logical design walkthrough. The database tables are designed by analysing functions involved in the system and format of the fields is also designed. The fields in the database tables should define their role in the system. The unnecessary fields should be avoided because it affects the storage areas of the system. Then, in the input and output screen design, the design should be made user friendly. The menu should be precise and compact.

3.1. SOFTWARE DESIGN :

In designing the software, the following principles are followed:

• Modularity and partitioning: software is designed in such a way that; each system should consist of hierarchy of modules and serve to partition into separate function.

• Coupling: modules should have little dependency on the other modules of a system.

• Cohesion: modules should carry out the operations in a single processing function.

• Shared use: avoid duplication by allowing a single module which is called by other, that needs the function it provides.

What is Python: -

Below are some facts about Python. Python is currently the most widely used multi-purpose, high level programming language. Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java. Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

Modules Used in Project: -

• TensorFlow

• NumPy

• Pandas

• Matplotlib

• Scikit – learn.

3.2. ARCHITECTURE :

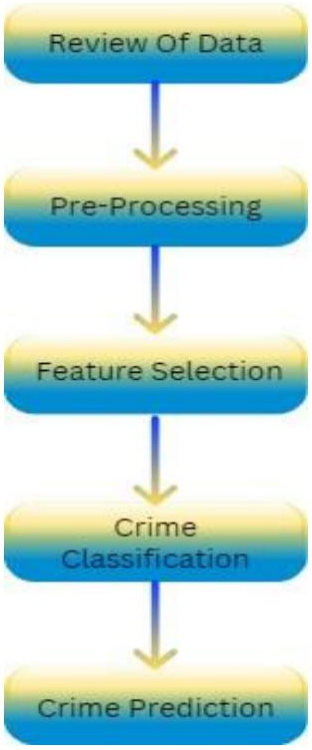

Fig3.2- Architecture Diagram of The System.

Our project involves analyzing crime data to understand crime patterns, predict future crime occurrences, and enhance crime prevention strategies.

Here's a comprehensive overview of the process involved:

1. Data Collection:

- Gather crime-related data from various sources, including law enforcement agencies, government databases, and research institutions.
- Collect information such as the type of crime, date and time of occurrence, location (latitude and longitude), demographic data of offenders and victims (gender, age, employment status), and other relevant attributes.

2. Data Preprocessing:

- Clean the collected data to remove inconsistencies, missing values, and outliers.
- Perform data transformation and normalization to ensure uniformity and compatibility across different attributes.
- Encode categorical variables into numerical format using techniques like one-hot encoding.
- Split the dataset into training and testing sets for model evaluation.

3. Exploratory Data Analysis (EDA):

- Conduct exploratory data analysis to gain insights into the characteristics and patterns present in the crime dataset.
- Visualize crime trends over time, geographical distribution of crimes, correlations between different attributes, and demographic patterns associated with crime occurrences.
- Identify any notable patterns or anomalies in the data that could inform subsequent modeling efforts.

4. Feature Selection:

- Identify the most relevant features (attributes) that have a significant impact on crime prediction.
- Use techniques such as correlation analysis, feature importance from machine learning models, and domain knowledge to select informative features.
- Remove redundant or irrelevant features to improve model efficiency and performance.

5. Model Building:

- Choose appropriate machine learning algorithms for crime prediction, such as Random Forest, Support Vector Machines (SVM), or Gradient Boosting Machines (GBM).
- Train multiple models using the training dataset and evaluate their performance using appropriate metrics such as accuracy, precision, recall, and F1-score.
- Fine-tune model hyperparameters using techniques like grid search or random search to optimize model performance.

6. Model Evaluation:

- Assess the predictive performance of trained models using the testing dataset.
- Evaluate models based on their ability to accurately predict crime occurrences, detect patterns, and generalize to unseen data.
- Compare the performance of different models and select the one with the highest predictive accuracy and reliability.

7. Crime Prediction:

- Deploy the trained model to predict future crime occurrences based on input data such as time, location, and demographic attributes.
- Generate crime predictions for specific time periods and geographical areas to identify potential hotspots or high-risk areas.
- Provide actionable insights to law enforcement agencies and policymakers for proactive crime prevention and resource allocation.

8. Monitoring and Refinement:

- Continuously monitor the performance of the crime prediction model in real-world scenarios.
- Gather feedback from stakeholders and incorporate new data to refine the model and improve its predictive accuracy over time.
- Update the model periodically to adapt to changing crime trends, demographics, and environmental factors.

By following these steps, our project aims to leverage data-driven insights and machine learning techniques to analyze crime patterns, predict future crime occurrences, and enhance crime prevention efforts for the benefit of society and public safety.

ALGORITHM:

The Random Forest algorithm is a powerful ensemble learning method commonly used for classification and regression tasks in machine learning. It is particularly suitable for our project, which involves analyzing crime data and predicting crime patterns. Here's an explanation of the Random Forest algorithm and its relevance to our project:

1. Random Forest Algorithm:

- **Ensemble Learning:** Random Forest belongs to the ensemble learning category, which combines multiple individual models to produce a more robust and accurate prediction.
- **Decision Trees:** At the core of the Random Forest algorithm are decision trees, which are hierarchical structures used for classification and regression tasks.
- **Bagging Technique:** Random Forest builds multiple decision trees by randomly selecting subsets of the training data (with replacement) and features. This process is known as bagging (bootstrap aggregating).
- **Voting Mechanism:** During prediction, each decision tree in the Random Forest independently makes a prediction. The final prediction is determined by aggregating the predictions of all decision trees through a voting mechanism (for classification) or averaging (for regression).
- **Reduced Overfitting:** By building multiple decision trees with randomness, Random Forest reduces overfitting compared to individual decision trees, resulting in better generalization performance on unseen data.
- **Feature Importance:** Random Forest provides a measure of feature importance, indicating the relative contribution of each feature to the prediction task.

2. Relevance to Our Project:

- **Crime Prediction:** In your project, you aim to predict crime patterns based on various attributes such as gender, age, employment status, and crime location. Random Forest is well-suited for this task as it can handle both categorical and numerical data effectively.
- **Handling Complex Relationships:** Crime prediction involves analyzing complex relationships between different attributes and crime occurrences. Random Forest excels in capturing nonlinear and interactive effects among features, making it suitable for modeling intricate crime patterns.
- **Robustness:** Random Forest is robust to noisy data and outliers, which are common in real-world crime datasets. It can handle missing values and maintain performance even with incomplete or imperfect data.
- **Interpretability:** While Random Forest is an ensemble of decision trees, it still provides insights into feature importance, allowing you to understand which attributes contribute most to crime prediction. This interpretability is valuable for understanding crime patterns and informing decision-making processes.

Overall, the Random Forest algorithm serves as a reliable and effective tool for crime prediction in our project, offering robustness, interpretability, and high predictive performance. By leveraging Random Forest, you can uncover valuable insights from crime data and enhance crime detection and prevention strategies.

3.3. INPUT DESIGN :

Considering the requirements, procedures are adopted to collect the necessary input data in most efficiently designed format. The input design has to be done keeping in view that, the interaction of the user with the system should be in the most effective and simplified way. Also, the necessary measures are taken for the following :

- Controlling the amount of input
- Avoid unauthorized access to the users
- Eliminating the extra steps
- Keeping the process simple
- At this stage the input forms and screens are designed.

3.4. OUTPUT DESIGN :

All the screens of the system are designed with a view to provide the user with easy operations in a simpler and efficient way, with minimum key strokes possible. Important information is emphasized on the screen. Almost every screen is provided with no error and important messages and option selection facilitates. Emphasis is given for faster processing and speedy transactions between the screens. Each screen assigned to make it as much user friendly as possible by using interactive procedures. In other words, we can say that the user can operate the system without much help from the operating manual.

3.5.HARDWARE AND SOFTWARE REQUIREMENTS

3.5.1. HARDWARE REQUIREMENTS :

- Processor: Intel i3 or above
- RAM: 4GB or above
- Hard Disk: 20gb or above
- Keyboard: Standard Windows Keyboard
- Mouse: Two or Three Button Mouse
- Monitor: SVGA

3.5.2. SOFTWARE REQUIREMENTS :

- Operating system- Windows 7 or above
- Coding Language- Python.
- Front-end - Python
- Back-end – Django-ORM
- Designing – Html, CSS, JavaScript.
- Data Base- MySQL (WAMP Server).

CHAPTER 04

SYSTEM DESIGN

4.SYSTEM DESIGN

4.1. UML DIAGRAMS :

Unified Modelling Language The Unified Modelling Language (UML) is a standard language for specifying, visualizing, constructing, and documenting the artefacts of software systems, as well as for business modelling and other non-software systems. The UML represents a collection of best engineering practices that have proven to be successful in the modelling of large and complex systems. The UML is a very important part of developing object-oriented software and the software development process. UML mostly uses graphical notations to express the design of software projects. Using the UML helps the project teams to communicate, explore potential designs and validate the architectural design of the software. The Unified Modelling Language (UML) is a standard language for writing software blue prints. The UML is a language for :

• Visualizing

• Specifying

• Constructing

• Documenting the artifacts of a software system.

UML is a language which provides vocabulary and the rules for combining words in that vocabulary for the purpose of communication. A modelling language is a language whose vocabulary and the rules focus on the conceptual and physical representation of a system.

Contents:

• Use case Diagram

• Class Diagram

• Activity Diagram

• Sequence Diagram

• Collaboration Diagram

4.1.1. USE CASE DIAGRAM :

• Use case Diagram consists of use case and actors.

• The main purpose is to show the interaction between the use cases and the actor.

• It intends to represent the system requirements from user's perspective.

• The use cases are the functions that are to be performed in the module.

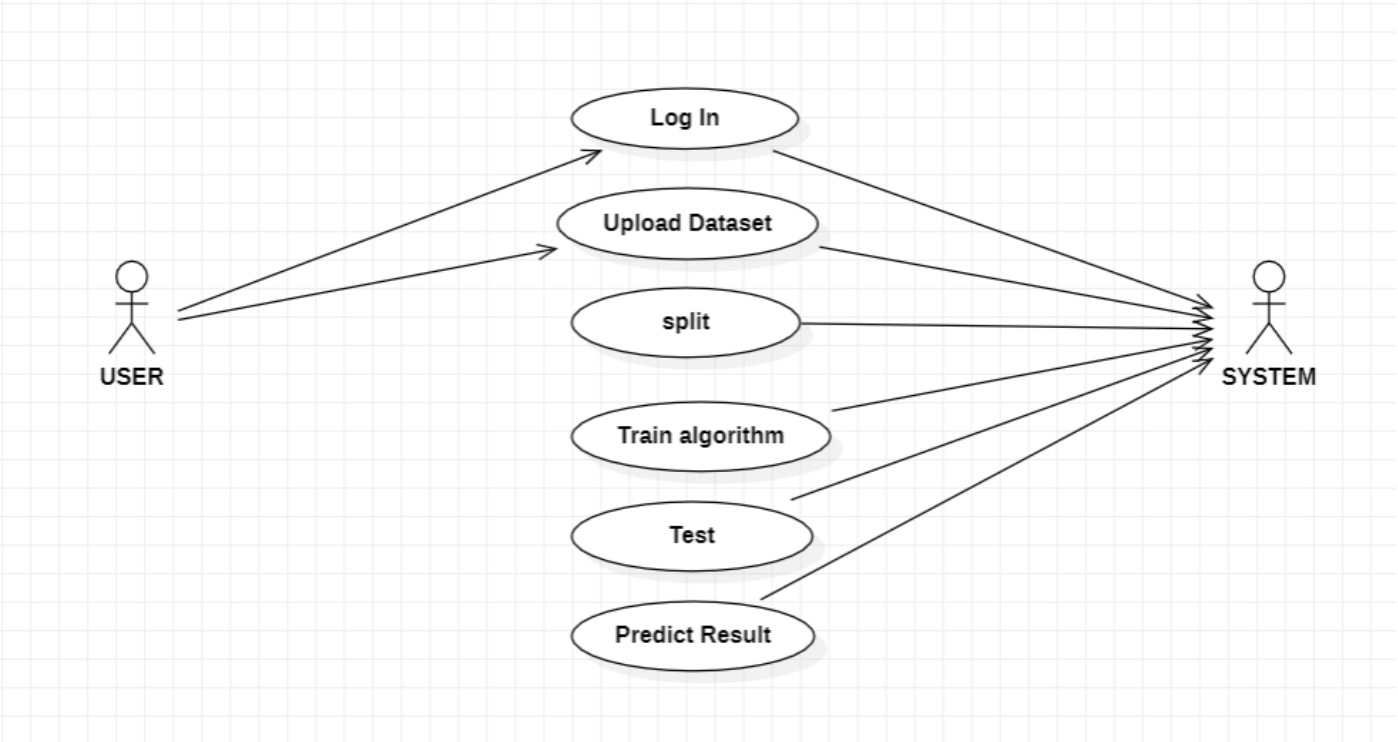

fig 4.1.1- Use Case diagram showing interaction between user and system

4.1.2. CLASS DIAGRAM :

• It contains the classes involved and shows the connections between the various classes.

• Class diagram includes classes, which further has a class label or name, attributes of the class and the operations of functions performed by the class.

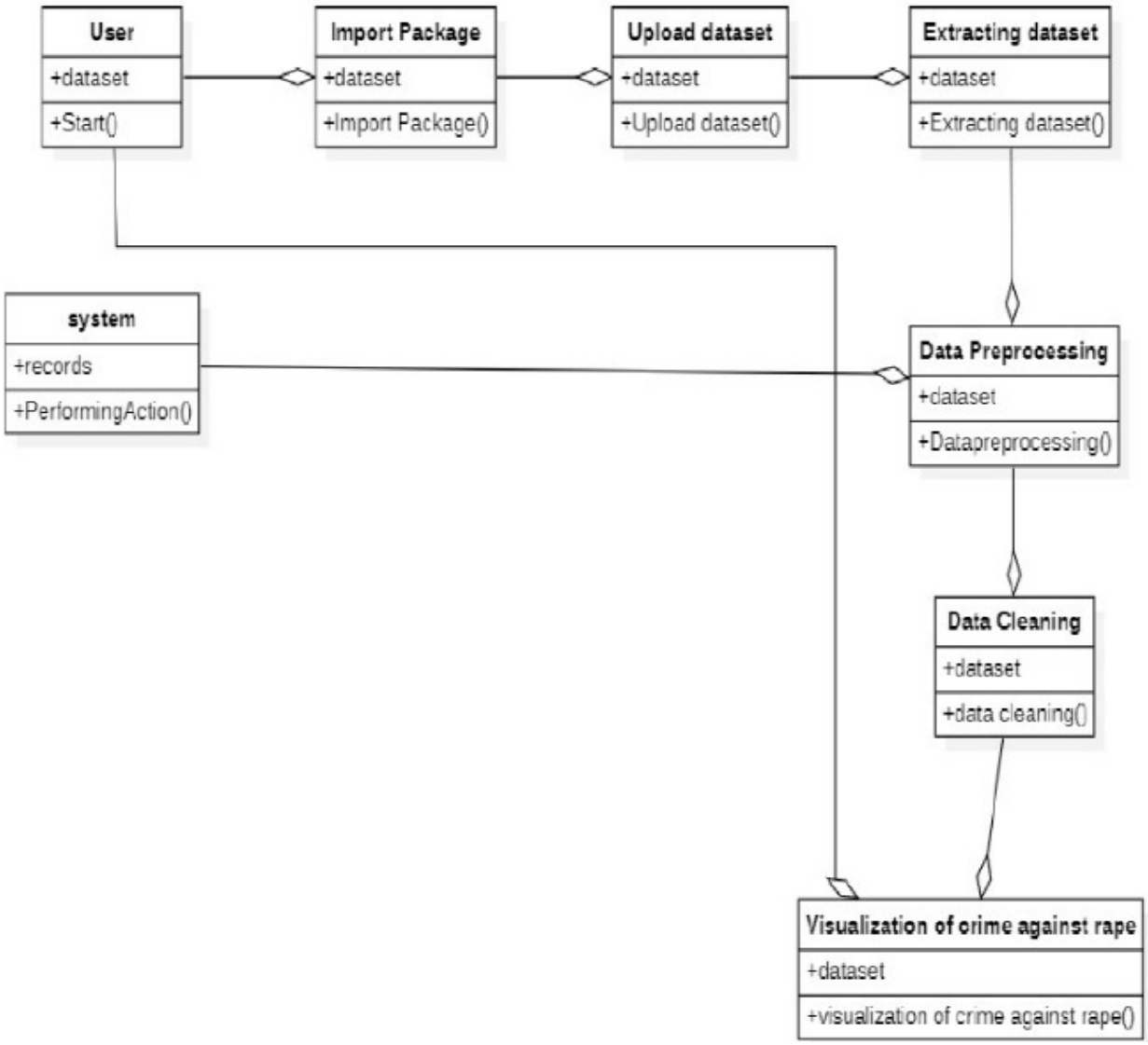

Fig 4.1.2- Class Diagram showing connections between different classes.

4.1.3. ACTIVITY DIAGRAM :

• It shows the flow of the various activities that are undergone from the beginning till the end.

• It consists of the activities that are held and carried out throughout the session from starting till the ending stage.

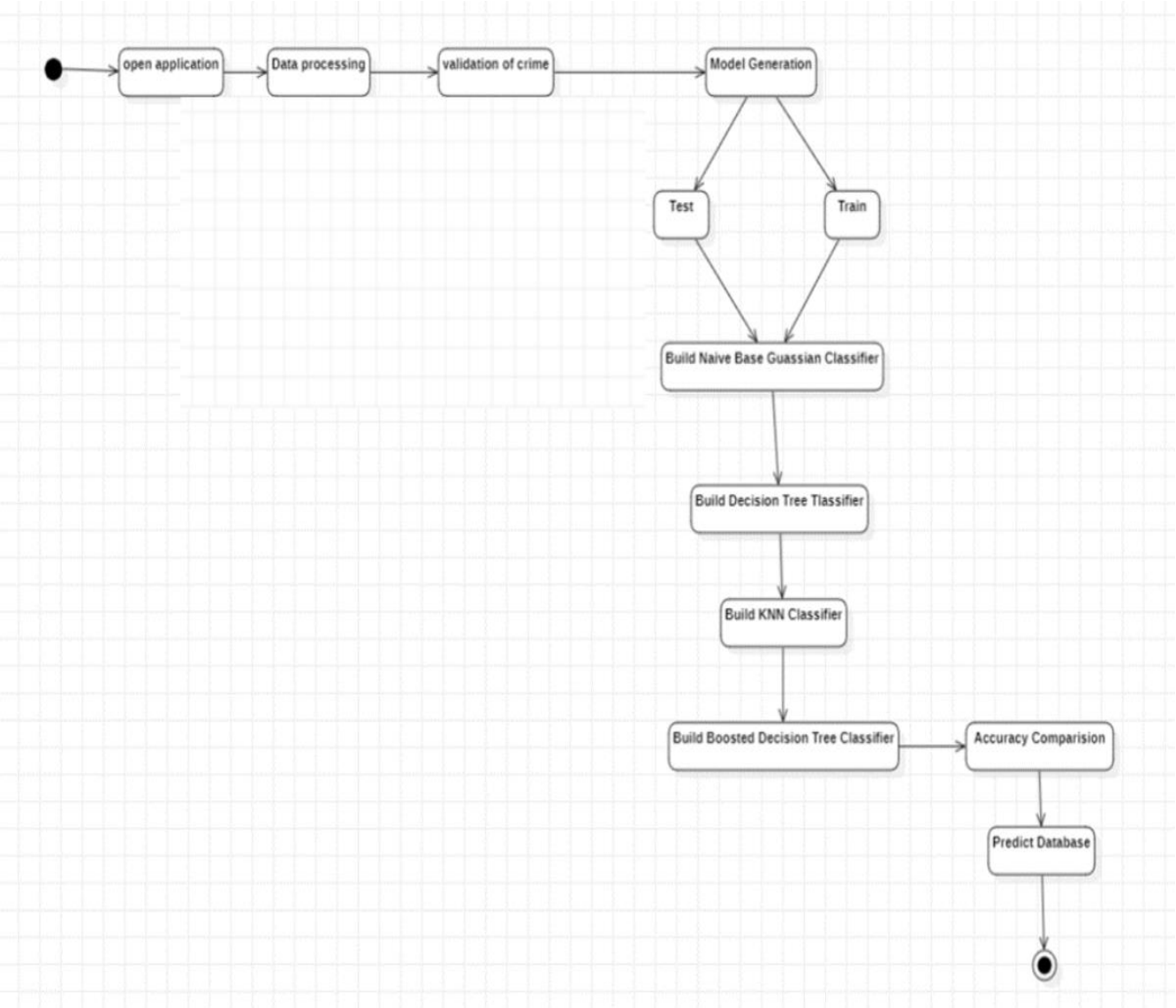

Fig 4.1.3- Activity Diagram all steps from user input to result

4.1.4.SEQUENCE DIAGRAM :

• It shows the sequence of the steps that are carried out throughout the process of execution.

• It involves lifelines or life time of a process that shows the duration for which the processis alive while the steps are taking place in the sequential manner.

• Sequence diagram specifies the order in which the various steps are executed.

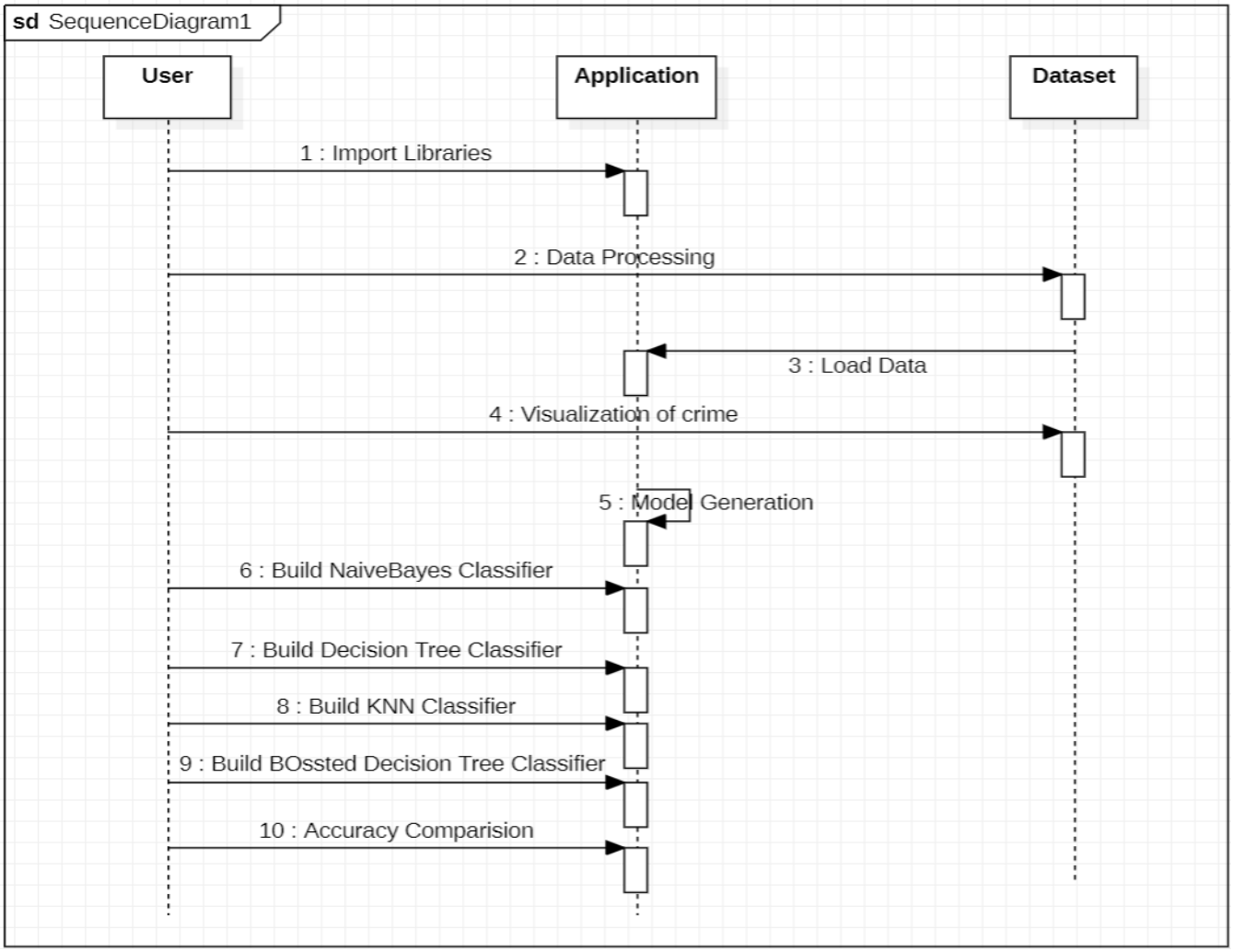

fig 4.1.4- Sequence Diagram showing all steps of executing the process.

4.1.5.Collaboration Diagram :

• A collaboration diagram groups together the interactions between different objects.

• The interactions are listed as numbered interactions that help to trace the sequence of the interactions.

• The collaboration diagram helps to identify all the possible interactions that each object has with other objects.

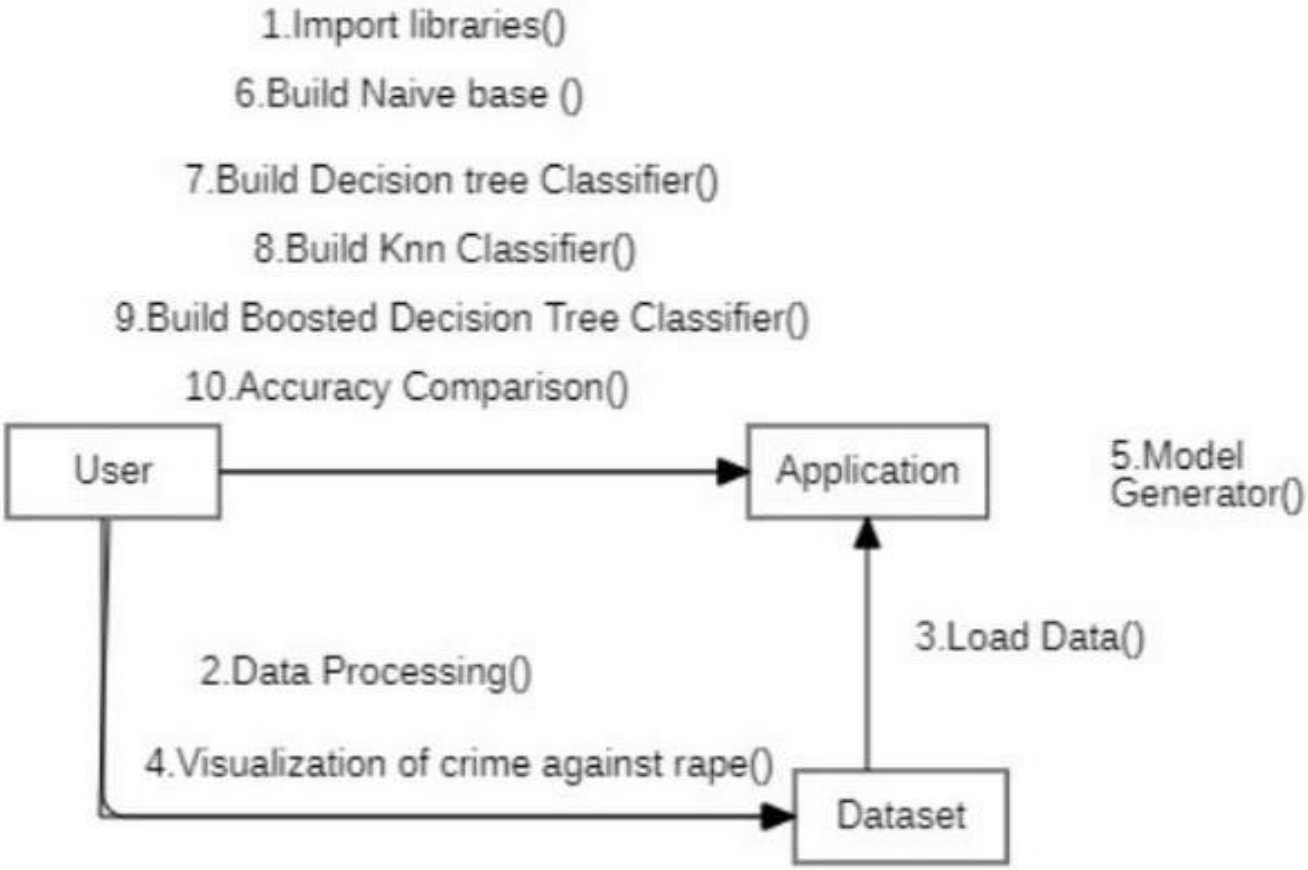

fig 4.1.5-Collaboration Diagram.

CHAPTER 05

IMPLEMENTATION

5.IMPLEMENTATION

5.1 Sample Code

5.1.1 Views.py

```
from django.shortcuts import render
from django.template import RequestContext
from django.contrib import messages
from django.http import HttpResponse
import os
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.cluster import KMeans
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
from math import sqrt

global dataset, kmeans_cluster, theft_cls, rape_cls, murder_cls

sc = MinMaxScaler(feature_range = (0, 1))
le1 = LabelEncoder()
le2 = LabelEncoder()

le3 = LabelEncoder()
le4 = LabelEncoder()

global mse, rmse

def calculateError(alg, X_test, y_test):
    predict = alg.predict(X_test)
    #predict = predict.reshape(predict.shape[0],1)
    #predict = sc.inverse_transform(predict)
    predict = predict.ravel()
    #labels = sc.inverse_transform(y_test)
    labels = y_test.ravel()
    mse_error = mean_squared_error(labels,predict)
    rmse_error = sqrt(mse_error)
    mse.append(mse_error/1000)
    rmse.append(rmse_error)

def UploadDatasetAction(request):
    if request.method == 'POST':
```

```
    global dataset, kmeans_cluster, theft_cls, rape_cls, murder_cls, mse, rmse
    mse = []
    rmse = []
    myfile = request.FILES['t1']
    dataset = pd.read_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Murder', 'Rape',
'Theft', 'Dowry_Deaths', 'Year'])
    dataset.fillna(0, inplace = True)
    cols = ['States/UTs', 'District']
    dataset[cols[0]] = pd.Series(le1.fit_transform(dataset[cols[0]].astype(str)))
    dataset[cols[1]] = pd.Series(le2.fit_transform(dataset[cols[1]].astype(str)))
    X = dataset.values
    X = sc.fit_transform(X)
    kmeans_cluster = KMeans(n_clusters=2, n_init=1200)
    kmeans_cluster.fit(X)

    dataset = pd.read_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Year', 'Theft',
'Murder', 'Rape'])
    dataset.fillna(0, inplace = True)
    print(dataset)
    cols = ['States/UTs', 'District']
    dataset[cols[0]] = pd.Series(le3.fit_transform(dataset[cols[0]].astype(str)))
    dataset[cols[1]] = pd.Series(le4.fit_transform(dataset[cols[1]].astype(str)))
    theft_Y = dataset['Theft'].values
    murder_Y = dataset['Murder'].values
    rape_Y = dataset['Rape'].values
    dataset.drop(['Theft'], axis = 1,inplace=True)
    dataset.drop(['Murder'], axis = 1,inplace=True)
    dataset.drop(['Rape'], axis = 1,inplace=True)
    X = dataset.values

    X_train1, X_test1, y_train1, y_test1 = train_test_split(X, theft_Y, test_size = 0.2)
    X_train2, X_test2, y_train2, y_test2 = train_test_split(X, rape_Y, test_size = 0.2)
    X_train3, X_test3, y_train3, y_test3 = train_test_split(X, murder_Y, test_size = 0.2)

    theft_cls = RandomForestRegressor()
    theft_cls.fit(X, theft_Y)
    calculateError(theft_cls, X_test1, y_test1)

    rape_cls = RandomForestRegressor()
    rape_cls.fit(X, rape_Y)
    calculateError(rape_cls, X_test2, y_test2)

    murder_cls = RandomForestRegressor()
    murder_cls.fit(X, murder_Y)
    calculateError(murder_cls, X_test3, y_test3)

    dataset = pd.read_csv("Dataset/Dataset.csv")
    dataset.fillna(0, inplace = True)
    columns = list(dataset.columns)
    strdata = '<table border=1 align=center width=100%><tr><th><font size=""
```

```
color="black">'+columns[0]+'</th>'
        for i in range(1,len(columns)):
            strdata+='<th><font size="" color="black">'+columns[i]+'</th>'
        strdata += "</tr>"
        dataset = dataset.values
        for i in range(len(dataset)):
            strdata += "<tr>"
            for j in range(len(dataset[i])):
                strdata+='<td><font size="" color="black">'+str(dataset[i,j])+'</td>'
            strdata += "</tr>"
        context= {'data':strdata}
        return render(request, 'ViewDataset.html', context)

def AdminLogin(request):
    if request.method == 'POST':
        user = request.POST.get('t1', False)
        password = request.POST.get('t2', False)
        if user == 'admin' and password == 'admin':
            context= {'data':user}
            return render(request, 'AdminScreen.html', context)
        else:
            context= {'data':"Invalid login details"}
            return render(request, 'Admin.html', context)

def index(request):
    if request.method == 'GET':
       return render(request, 'index.html', {})

def Admin(request):
    if request.method == 'GET':
       return render(request, 'Admin.html', {})

def UploadDataset(request):
    if request.method == 'GET':
       return render(request, 'UploadDataset.html', {})

def MSEGraph(request):
    if request.method == 'GET':
        strdata = '<table border=1 align=center width=100%><tr><th><font size=""
color="black">Algorithm Name</th><th><font size="" color="black">RMSE</th>'
        strdata+='<th><font size="" color="black">MSE</th></tr>'
        strdata+='<tr><td><font size="" color="black">Random Forest Theft Prediction</td><td><font
size="" color="black">'+str(rmse[0])+'</td>'
        strdata+='<td><font size="" color="black">'+str(mse[0])+'</td>'
        strdata+='<tr><td><font size="" color="black">Random Forest Rape Prediction</td><td><font
size="" color="black">'+str(rmse[1])+'</td>'
        strdata+='<td><font size="" color="black">'+str(mse[1])+'</td>'
        strdata+='<tr><td><font size="" color="black">Random Forest Murder
Prediction</td><td><font size="" color="black">'+str(rmse[2])+'</td>'
        strdata+='<td><font size="" color="black">'+str(mse[2])+'</td></tr></table>'
```

```
        df = pd.DataFrame([['Random Forest Theft Prediction','RMSE',rmse[0]],['Random Forest Theft Prediction','MSE',mse[0]],
                           ['Random Forest Rape Prediction','RMSE',rmse[1]],['Random Forest Rape Prediction','MSE',mse[1]],
                           ['Random Forest Murder Prediction','RMSE',rmse[2]],['Random Forest Murder Prediction','MSE',mse[2]],
                           ],columns=['Parameters','Algorithms','Value'])
        df.pivot("Parameters", "Algorithms", "Value").plot(kind='bar')
        plt.show()
        context= {'data':strdata}
        return render(request, 'Comparison.html', context)



def ClusterPrediction(request):
    if request.method == 'GET':
        dataset = pd.read_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Year'])
        dataset.fillna(0, inplace = True)
        states = np.unique(dataset['States/UTs'].values)
        output = '<tr><td><font size="" color="black">States</b></td><td><select name="t1">'
        for i in range(len(states)):
            output += '<option value="'+states[i]+'">'+states[i]+'</option>'
        output += "</select></td></tr>"
        output += '<tr><td><font size="" color="black">District</b></td><td><select name="t2">'
        for i in range(len(states)):
            district = dataset[dataset['States/UTs'] == states[i]]['District']
            district = district.values
            output += '<option value="'+states[i]+'"><b>--'+states[i]+'--</b></option>'
            for j in range(len(district)):
                output += '<option value="'+district[j]+'">'+district[j]+'</option>'
        output += "</select></td></tr>"
        output += '<tr><td><font size="" color="black">Year</b></td><td><select name="t3">'
        year = np.unique(dataset['Year'].values)
        for i in range(len(year)):
            output += '<option value="'+str(year[i])+'">'+str(year[i])+'</option>'
        output += "</select></td></tr>"
        context= {'states':output}
        return render(request, 'ClusterPrediction.html', context)

def ClusterPredictionAction(request):
    if request.method == 'POST':
        state = request.POST.get('t1', False)
        district = request.POST.get('t2', False)
        year = request.POST.get('t3', False)
        murder = request.POST.get('t4', False)
        rape = request.POST.get('t5', False)
        theft = request.POST.get('t6', False)
        dowry = request.POST.get('t7', False)
        temp = []
```

```
        temp.append([state, district, murder, rape, theft, dowry, year])
        test = pd.DataFrame(temp, columns=['States/UTs','District', 'Murder', 'Rape', 'Theft',
'Dowry_Deaths', 'Year'])
        test.fillna(0, inplace = True)
        test['States/UTs'] = pd.Series(le1.transform(test['States/UTs'].astype(str)))
        test['District'] = pd.Series(le2.transform(test['District'].astype(str)))
        test = test.values
        test = sc.transform(test)
        predict = kmeans_cluster.predict(test)
        print(predict)
        output = district+" Low Crime Rate Area"
        if predict == 1:
            output = district+" High Crime Rate Area"
        context= {'data':output}
        return render(request, 'index.html', context)



def FuturePrediction(request):
    if request.method == 'GET':
        dataset = pd.read_csv("Dataset/Dataset.csv", usecols=['States/UTs','District', 'Year'])
        dataset.fillna(0, inplace = True)
        states = np.unique(dataset['States/UTs'].values)
        output = '<tr><td><font size="" color="black">States</b></td><td><select name="t1">'
        for i in range(len(states)):
            output += '<option value="'+states[i]+'">'+states[i]+'</option>'
        output += "</select></td></tr>"
        output += '<tr><td><font size="" color="black">District</b></td><td><select name="t2">'
        for i in range(len(states)):
            district = dataset[dataset['States/UTs'] == states[i]]['District']
            district = district.values
            output += '<option value="'+states[i]+'"><b>--'+states[i]+'--</b></option>'
            for j in range(len(district)):
                output += '<option value="'+district[j]+'">'+district[j]+'</option>'
        output += "</select></td></tr>"
        output += '<tr><td><font size="" color="black">Year</b></td><td><select name="t3">'
        year = np.unique(dataset['Year'].values)
        for i in range(len(year)):
            output += '<option value="'+str(year[i])+'">'+str(year[i])+'</option>'
        output += '<option value="2022">2022</option>'
        output += '<option value=2023>2023</option>'
        output += "</select></td></tr>"
        context= {'states':output}
        return render(request, 'FuturePrediction.html', context)


def FuturePredictionAction(request):
    if request.method == 'POST':
        state = request.POST.get('t1', False)
        district = request.POST.get('t2', False)
```

```
        year = request.POST.get('t3', False)
        classify_type = request.POST.get('t4', False)
        temp = []
        temp.append([state, district, year])
        test = pd.DataFrame(temp, columns=['States/UTs','District', 'Year'])
        test.fillna(0, inplace = True)
        test['States/UTs'] = pd.Series(le1.transform(test['States/UTs'].astype(str)))
        test['District'] = pd.Series(le2.transform(test['District'].astype(str)))
        test = test.values
        predict = 0
        if classify_type == "Theft":
            predict = "Future Predicted Thefts = "+str(int(theft_cls.predict(test)[0]))
        if classify_type == "Murder":
            predict = "Future Predicted Murders = "+str(int(murder_cls.predict(test)[0]))
        if classify_type == "Rape":
            predict = "Future Predicted Rapes = "+str(int(rape_cls.predict(test)[0]))
        context= {'data':predict}
        return render(request, 'index.html', context)

def Analysis(request):
    if request.method == 'GET':
        return render(request, 'Analysis.html', {})

def AnalysisAction(request):
    if request.method == 'POST':
        classify_type = request.POST.get('t1', False)
        strdata = '<table border=1 align=center width=100%>'
        if classify_type == "Theft":
            strdata += '<tr><td><img src="static/analysis/theft_bar.png" height="300"
width="500"/></td></tr>'
            strdata += '<tr><td><img src="static/analysis/theft_pie.png" height="300"
width="500"/></td></tr>'
        if classify_type == "Murder":
            strdata += '<tr><td><img src="static/analysis/murder_bar.png" height="300"
width="500"/></td></tr>'
            strdata += '<tr><td><img src="static/analysis/murder_pie.png" height="300"
width="500"/></td></tr>'
        if classify_type == "Rape":
            strdata += '<tr><td><img src="static/analysis/rape_bar.png" height="300"
width="500"/></td></tr>'
            strdata += '<tr><td><img src="static/analysis/rape_pie.png" height="300"
width="500"/></td></tr>'
        context= {'data':strdata}
        return render(request, 'ViewGraphs.html', context)
```

5.1.2 Index.html

```
{% load static %}
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta name="description" content="" />
<meta name="keywords" content="" />
<title>Crime Prediction & Analysis</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link rel="stylesheet" type="text/css" href="{% static 'style.css' %}"/>
</head>
<body>
<div id="wrapper">
	<div id="header">
		<div id="logo">
			<center><font size="4" color="yellow">Crime Rate Prediction &
Analysis </font></center>
		</div>
		<div id="slogan">

		</div>
	</div>
	<div id="menu">


		<ul><center>
			<li class="first current_page_item"><a href="{% url 'index'
%}"><font size="" color="yellow">Home Page</font></a></li>
			<li class="first current_page_item"><a href="{% url 'Admin'
%}"><font size="" color="yellow">Admin Login</font></a></li>
			<li class="first current_page_item"><a href="{% url
'ClusterPrediction' %}"><font size="" color="yellow">Cluster Prediction</font></a></li>
			<li class="first current_page_item"><a href="{% url 'FuturePrediction'
%}"><font size="" color="yellow">Future Prediction</font></a></li>
			<li class="first current_page_item"><a href="{% url 'Analysis'
%}"><font size="" color="yellow">Analysis</font></a></li>

		</center></ul>
		<br class="clearfix" />
	</div>
	<div id="splash">
		<img class="pic" src="{% static 'images/investor.jpg' %}" width="870"
		height="230" alt="" />
	</div>
	<br/>
		<p align="justify"><font size="3" color="black"  style="font-family: Comic
	Sans MS">
	Abstract:-Crime Rate Prediction & Analysis</p>

	<font size="3" color="blue"  style="font-family: Comic Sans
	MS"><center>{{data|safe}}</center>
	<table>
```

```
</table>
<br/><br/><br/><br/><br/><br/><br/><br/><br/>
</body>
</html>
```

5.1.3 ClusterPrediction.html

```
{% load static %}
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta name="description" content="" />
<meta name="keywords" content="" />
<title>Crime Prediction & Analysis</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link rel="stylesheet" type="text/css" href="{% static 'style.css' %}" />
<script language="javascript">
        function validate(formObj)
        {
        if(formObj.t4.value.length==0)
        {
        alert("Please Enter Murder value");
        formObj.t4.focus();
        return false;
        }
        if(formObj.t5.value.length==0)
        {
        alert("Please Enter Rape value");
        formObj.t5.focus();
        return false;
        }
        if(formObj.t6.value.length==0)
        {
        alert("Please Enter Theft value");
        formObj.t6.focus();
        return false;
        }
        if(formObj.t7.value.length==0)
        {
        alert("Please Enter Dowry Deaths value");
        formObj.t7.focus();
        return false;
        }
        return true;
        }
        </script>
</head>
<body>
```

```
<div id="wrapper">
	<div id="header">
		<div id="logo">
			<center><font size="4" color="yellow">Crime Rate Prediction &
			Analysis</font></center>
		</div>
		<div id="slogan">

		</div>
	</div>
	<div id="menu">
		<ul><center>
			<li class="first current_page_item"><a href="{% url 'index'
			%}"><font size="" color="yellow">Home Page</font></a></li>
			<li class="first current_page_item"><a href="{% url 'Admin'
			%}"><font size="" color="yellow">Admin Login</font></a></li>
			<li class="first current_page_item"><a href="{% url
			'ClusterPrediction' %}"><font size="" color="yellow">Cluster
			Prediction</font></a></li>
			<li class="first current_page_item"><a href="{% url 'FuturePrediction'
			%}"><font size="" color="yellow">Future Prediction</font></a></li>
			<li class="first current_page_item"><a href="{% url 'Analysis'
			%}"><font size="" color="yellow">Analysis</font></a></li>

		</center></ul>		<br class="clearfix" />
				</div>

	<div id="splash">
		<img class="pic" src="{% static 'images/investor.jpg' %}" width="870"
		height="230" alt="" />
	</div>
			<center>
<form name="f1" method="post" action={% url 'ClusterPredictionAction' %}
onsubmit="return validate(this);">
<br/>{% csrf_token %}<br/>
  <h5><b>Crime Rate Prediction Screen Using Clustering</b></h5>

  <font size="" color="red"><center>{{ data|safe }}</center></font>


					<table align="center" width="40" >
					{{ states|safe }}
		 <tr><td><font size="" color="black">Murder</b></td><td><input
		type="text" name="t4" style="font-family: Comic Sans MS"
		size="20"/></td></tr>
	<tr><td><font size="" color="black">Rape</b></td><td><input type="text"
	name="t5" style="font-family: Comic Sans MS" size="20"/></td></tr>
			<tr><td><font size="" color="black">Theft</b></td><td><input
	type="text" name="t6" style="font-family: Comic Sans MS" size="20"/></td></tr>
			<tr><td><font size=""
```

```
color="black">Dowry Deaths</b></td><td><input type="text" name="t7"
style="font-family: Comic Sans MS" size="20"/></td></tr>

                        <tr><td></td><td><input type="submit" value="Submit">

                        </td>
                        </tr>
                        </table>
                                </div>

                                </div>
                                <br/><br/>
        </body>
</html>
```

5.1.4 FuturePrediction.html

```
{% load static %}
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta name="description" content="" />
<meta name="keywords" content="" />
<title>Crime Prediction & Analysis</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link rel="stylesheet" type="text/css" href="{% static 'style.css' %}" />
<script language="javascript">
function validate(formObj)
{
if(formObj.t4.value.length==0)
{
alert("Please Enter Murder value");
formObj.t4.focus();
return false;
}
if(formObj.t5.value.length==0)
{
alert("Please Enter Rape value");
formObj.t5.focus();
return false;
}
if(formObj.t6.value.length==0)
{
alert("Please Enter Theft value");
formObj.t6.focus();
return false;
```

```
}
if(formObj.t7.value.length==0)
{
alert("Please Enter Dowry Deaths value");
	formObj.t7.focus();
	return false;
	}
	return true;
	}
	</script>
</head>
<body>
<div id="wrapper">
	<div id="header">
		<div id="logo">
			<center><font size="4" color="yellow">Crime Rate Prediction & Analysis using K-Means Clustering Algorithm</font></center>
		</div>
		<div id="slogan">

		</div>
	</div>
	<div id="menu">
		<ul><center>
			<li class="first current_page_item"><a href="{% url 'index' %}"><font size="" color="yellow">Home Page</font></a></li>
			<li class="first current_page_item"><a href="{% url 'Admin' %}"><font size="" color="yellow">Admin Login</font></a></li>
			<li class="first current_page_item"><a href="{% url 'ClusterPrediction' %}"><font size="" color="yellow">Cluster Prediction</font></a></li>
			<li class="first current_page_item"><a href="{% url 'FuturePrediction' %}"><font size="" color="yellow">Future Prediction</font></a></li>
			<li class="first current_page_item"><a href="{% url 'Analysis' %}"><font size="" color="yellow">Analysis</font></a></li>

		</center></ul>		<br class="clearfix" />
				</div>

	<div id="splash">
		<img class="pic" src="{% static 'images/investor.jpg' %}" width="870" height="230" alt="" />
	</div>
			<center>
<form name="f1" method="post" action={% url 'FuturePredictionAction' %} onsubmit="return validate(this);">
<br/>{% csrf_token %}<br/>
  <h5><b>Crime Rate Prediction Screen Using Future Prediction</b></h5>
```

```
  <font size="" color="red"><center>{{ data|safe }}</center></font>


                                        <table          align="center"
width="40" >
                                        {{ states|safe }}

                    <tr><td><font                                   size=""
color="black">Prediction Type</b></td><td><select name="t4">
                    <option value="Theft">Theft</option>
                    <option value="Murder">Murder</option>
                    <option value="Rape">Rape</option>
                    </td></tr>

                    <tr><td></td><td><input          type="submit"
value="Submit">

                    </td>
                    </tr>
                    </table>
                        </div>

                        </div>
                        <br/><br/>
        </body>
</html>
```

CHAPTER 06

OUTPUT SCREENS

6.OUTPUT SCREENS

6.1 Home Page :

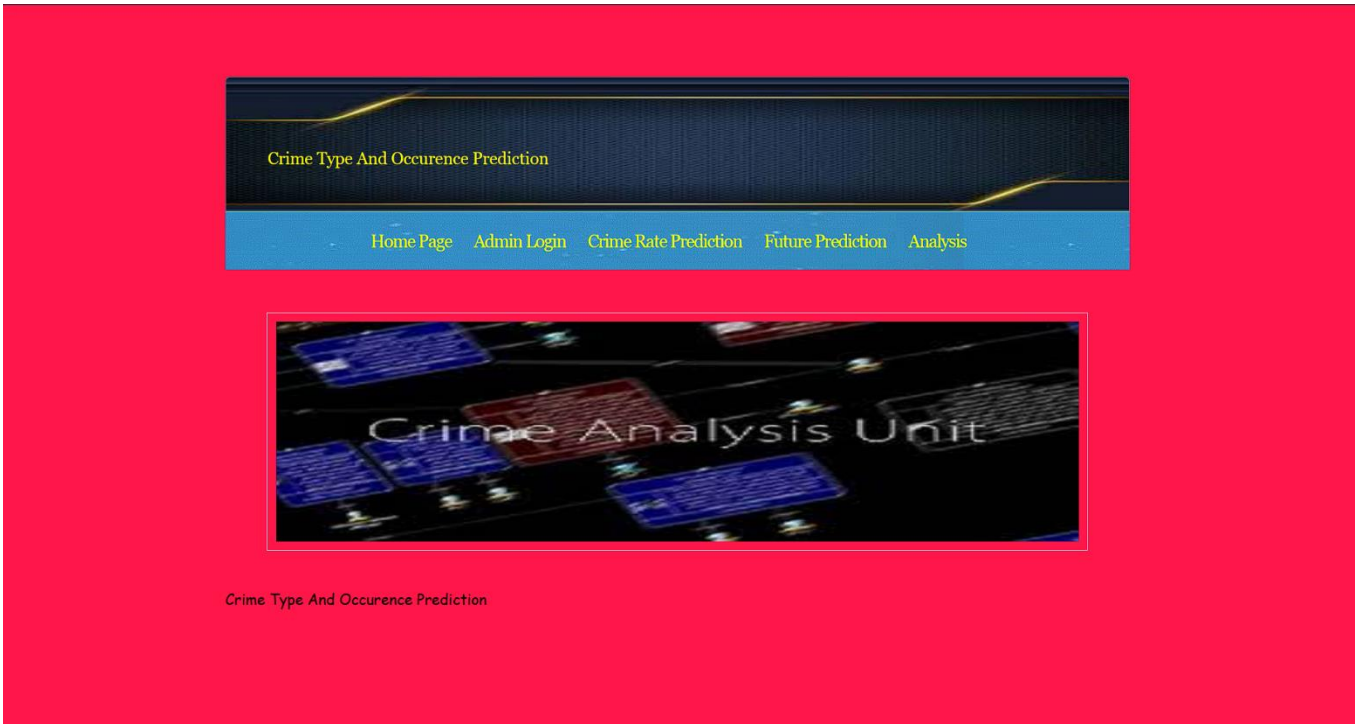

Fig 6.1 Home Page

6.2 Admin Login Page :

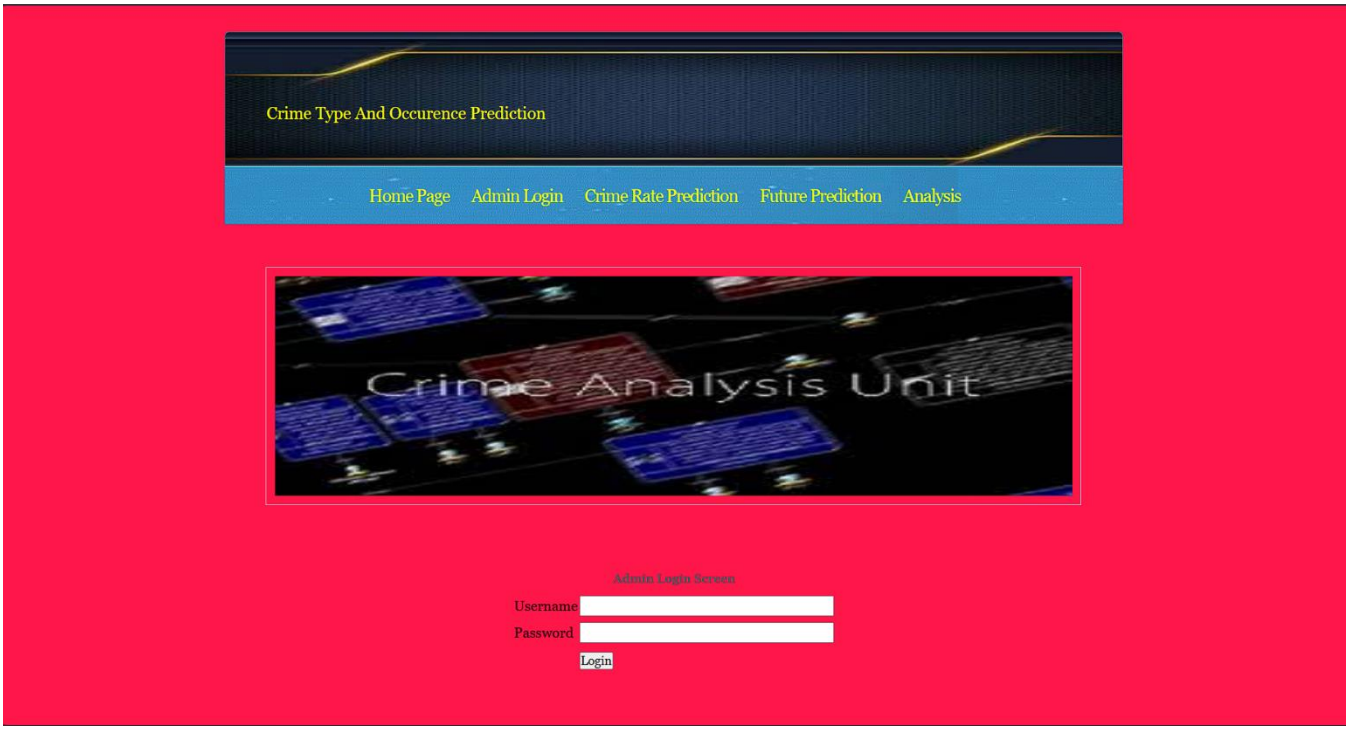

Fig 6.2 Admin Login Page

6.3 Dataset Upload Page :

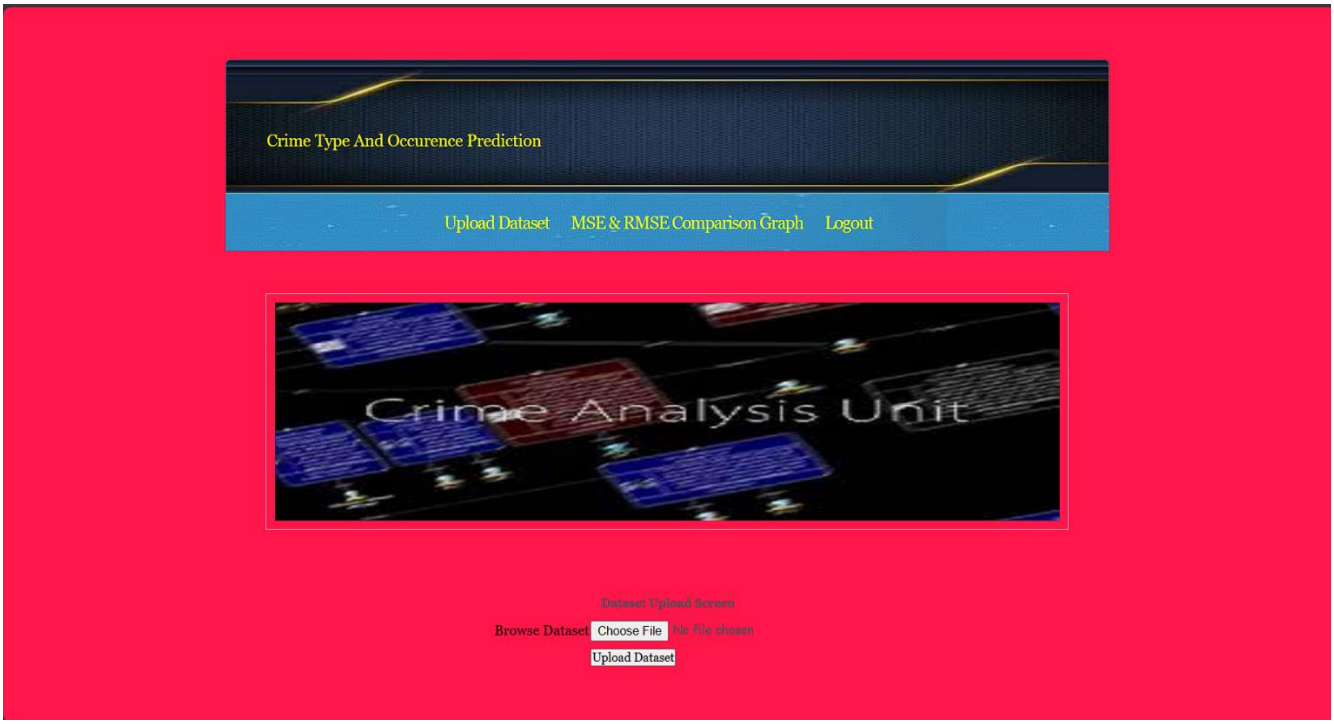

Fig 6.3 Dataset Upload Page

6.4 MSE & RMSE Scores :

Algorithm Name	RMSE	MSE
Random Forest Theft Prediction	1211.6525274687579	1468.101847321429
Random Forest Rape Prediction	98.22899630091776	9.648935714285715
Random Forest Murder Prediction	53.7108146199893	2.8848516071428567

Fig 6.4 MSE & RMSE Scores

6.5 Crime Rate Prediction Page :

Fig 6.5 Crime Rate Prediction Page

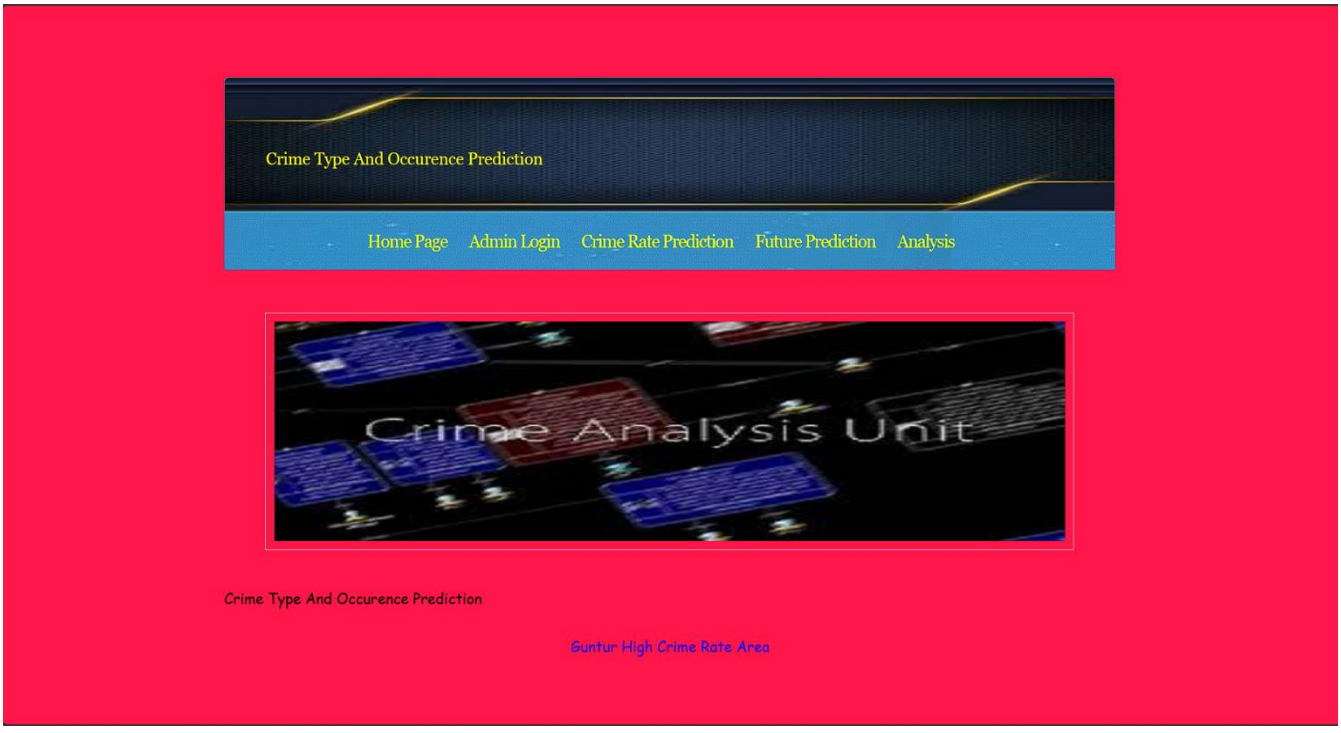

Fig 6.5.1 Crime Rate Prediction Page

6.6 Future Prediction Page :

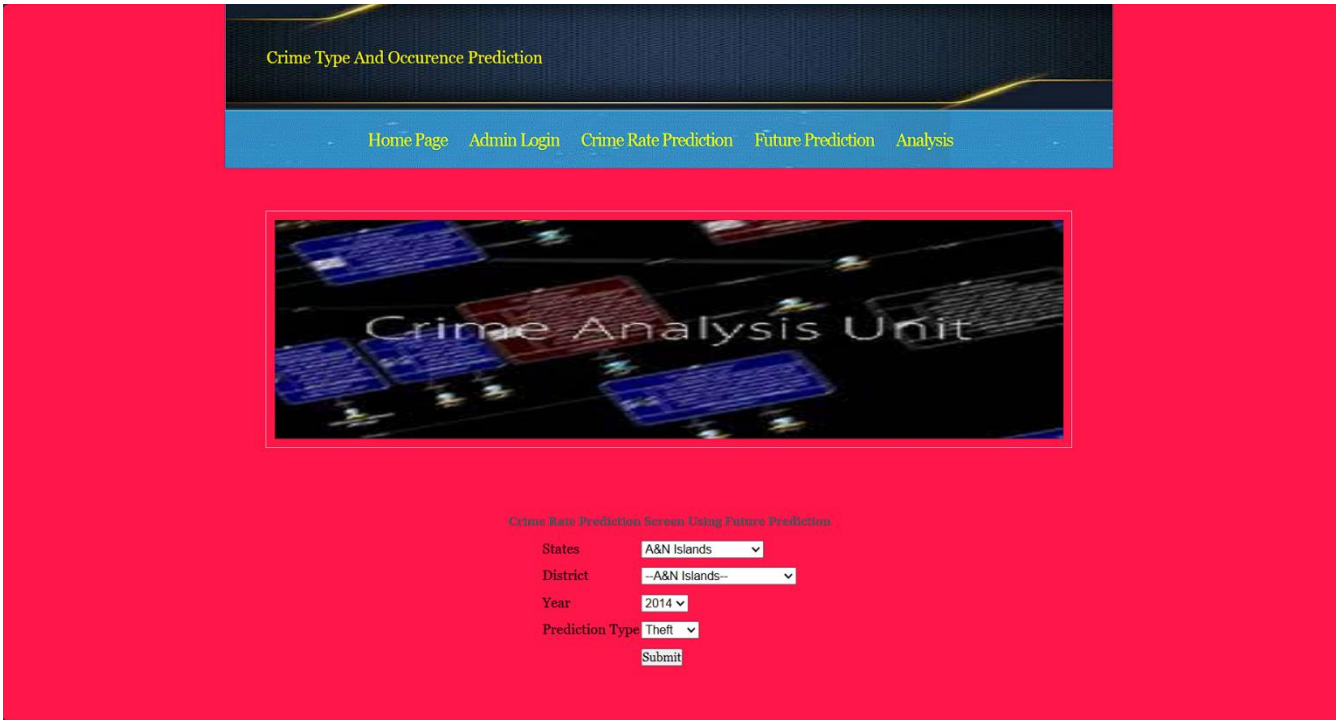

Fig 6.6 Future Prediction Page

Fig 6.6.1 Future Prediction Page

6.7 Analysis Page :

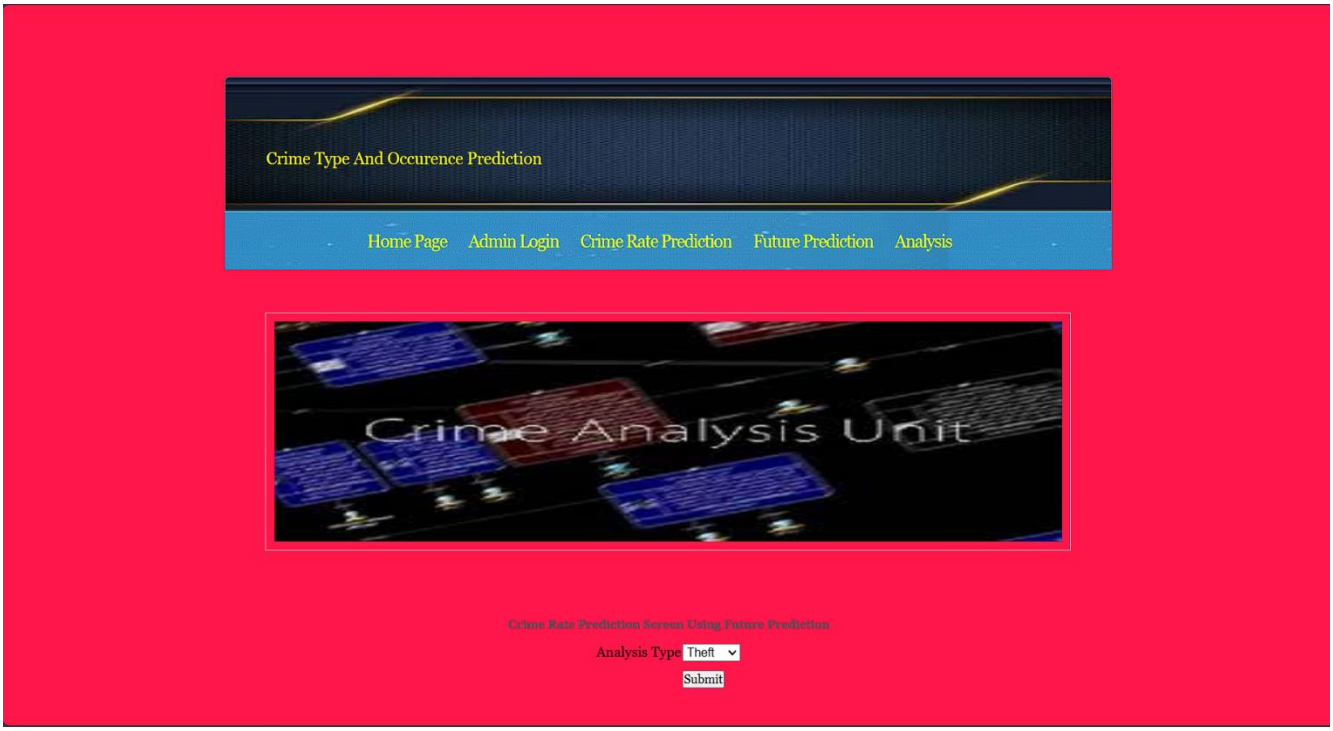

Fig 6.7 Analysis Page

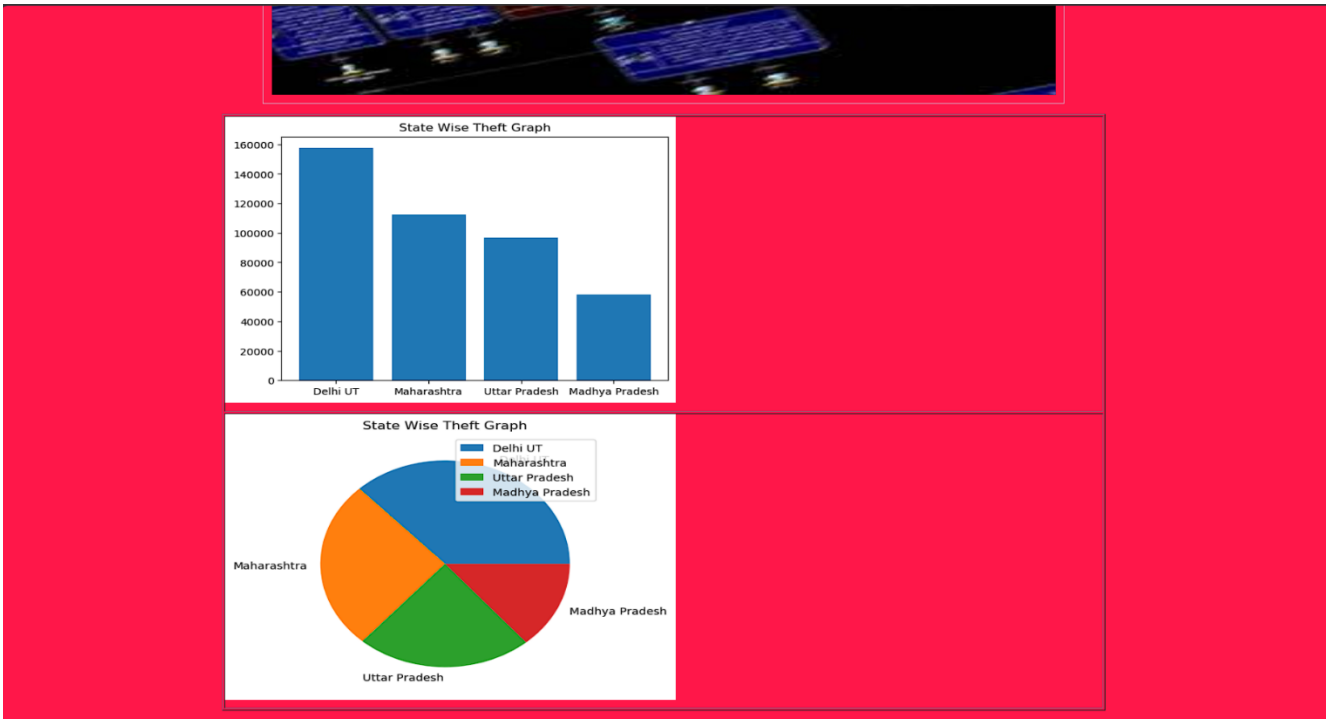

Fig 6.7.1 Analysis Page

CHAPTER 07

TESTING AND DEBUGGING

7.TESTING AND DEBUGGING

7.1 TESTING PROCESS

Firstly, we will start testing with individual module and will perform unit testing on that. So we successfully verify and validate all modules by successively integrating each module. Moreover, checking the work done was very important to reduce risk factor. Checking was being ultimately handled by testing but interim checking was required. So we planned work done by one member was tested by other for some time and again revolved for other level check. This technique proved to be very much helpful as it came out with innovative ideas to reduce error very low level. The objective of this testing phase is to prove that the developed system satisfies the requirements defined earlier. Several types of tests will be conducted in this phase. Testing is an important phase of system development because it can ensure the system matches the specifications. Besides that, testing also ensures that the system functions in the correct and proper manner with the minimum amount of errors. Applications are susceptible to bugs and malfunctions, such as scripts that not run properly. Besides that, there might be compatibility problems because a project may run perfectly on one device but may not display properly on another. Bottom-up testing strategy is used in this system to avoid unnecessary duplication of effort. Individual objects will be tested in isolation using unit testing and gradually integrated for the higher level integration testing and system testing. Failed components will be migrated back to the development phase for rework, and components that work properly will migrate ahead for implementation.

7.1.1 UNIT TESTING

Unit testing reveals syntax and semantic errors from the smallest programming unit. In this thesis, unit testing is used to test each individual page. Errors that are found in a particular page of the application are thoroughly debugged and removed before starting to develop another. Due to the dynamic nature of testing, there is no proper testing documentation created.

7.1.2 LINK /INTEGRATION TESTING

Testing when each application of a particular Section in the System passed the unit testing, integration test was carried out to ensure that pages are linked in the correct flow an integrate properly into the entire website. All the buttons, text boxes and navigation bars were tested.

7.1.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

7.1.4 SYSTEM TESTING

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

7.1.5 WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester has knowledge of theinner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached.

7.1.6 BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the inner workings,structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document,such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

7.1.6.1 TEST STRATEGY AND APPROACH

Field testing will be performed manually and functional tests will be written in detail.

7.1.6.2 TEST OBJECTIVES

• All field entries must work properly.

• Pages must be activated from the identified link.

•The entry screen, messages and responses must not be delayed.

7.1.6.3 FEATURES TO BE TESTED

• Verify that the entries are of the correct format.

• No duplicate entries should be allowed.

• All links should take the user to the correct page.

7.1.7 INTEGRATION TESTING

Software integration testing is the incremental integration testing of two or more integratedsoftware components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or one step up software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

7.1.8 ACCEPTANCE TESTING

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

7.2 TEST CASES

TEST ID	TEST DESCRIPTION	TEST DESIGN	EXPECTED OUTPUT	ACTUAL OUTPUT	STATUS
1	Analyze crime patterns	Use machine learning algorithms (KNN, K-Means)	Accurately classify crime patterns	Accurately classify crime patterns	Pass
2	Evaluate data mining methods	Compare theoretical and practical performance	Identify most appropriate data mining methods	Identify most appropriate data mining methods	Pass
3	Admin Login	Username, password	Admin Login successful	Admin Login Successful.	Pass
4	Analyze dataset attributes	Use methods on attributes (gender, age, etc.)	Determine effectiveness in analyzing crime	Determine effectiveness in analyzing crime	Pass
5	Test robustness against incomplete or missing data	Introduce incomplete or missing data	Handle missing data gracefully	Prediction accuracy decreases due to missing data	Fail
6	Cross-validation assessment with imbalanced dataset	Apply cross-validation techniques on imbalanced dataset	Accurately assess model performance despite imbalance	Model performance metrics affected by data imbalance	Fail
7	Assess performance of crime prediction based on location	Analyze predictions based on crime location	Accurately predict crime occurrences	Accurately predict crime occurrences	Pass
8	Evaluate Random Forest classifier performance	Measure number of instances correctly classified	Achieve higher performance compared to other methods	Achieve higher performance compared to other methods	Pass

9	Integration with real-time data sources	Integrate the system with real-time crime data sources	Seamless integration with real-time data	sourcesIssues encountered during integration with APIs	Fail

Table :7.1 Test cases for system

CHAPTER 08

CONCLUSION

8.CONCLUSION

The purpose of this study is to examine crime analysis through the applicability of data mining methods in the process of crime prediction and prevention. The results of experiments conducted in this research by implementing algorithms of data mining methods have revealed that these methods are applicable in the process of crime prediction. The decision tree as a data mining classification method has classified crime data at an accuracy rate of 76%. This method has shown promising results for the problem of crime prediction as the accuracy rate is high in the experiments performed.

CHAPTER 09

FUTURE ENHANCEMENTS

9. FUTURE ENHANCEMENTS

In the future extension of this study some models will be created for predicting the crime hot spots that would help the deployment of police to places of crimes. Algorithms' behaviour changes will be looked at when more data is added. I also plan to look into developing social link networks of criminals, suspects and gangs. I also intend to implement this study to an integrated enterprise software that will be created.

1. **Real-time data integration:** Implementing mechanisms to ingest and process real-time data from various sources such as social media, surveillance cameras, and IoT devices to provide timely insights and predictions.

2. **Enhanced predictive models:** Continuously improving machine learning algorithms and predictive models by incorporating more advanced techniques such as deep learning, ensemble learning, and anomaly detection to better identify patterns and trends in crime data.

3. **Geographic profiling:** Developing geographic profiling algorithms to identify hotspots and areas with high crime rates, enabling law enforcement agencies to allocate resources more effectively and implement targeted interventions.

4. **Predictive analytics for specific crime types:** Developing specialized models for predicting specific types of crimes such as burglary, theft, or cybercrime, allowing for tailored prevention strategies and interventions.

5. **Integration with law enforcement systems:** Integrating the crime prediction system with existing law enforcement systems and databases to facilitate seamless data sharing and collaboration among agencies, enabling faster response times and more efficient resource allocation.

6. **Community engagement and citizen involvement:** Implementing features that allow citizens to report suspicious activities or incidents through mobile apps or

online platforms, enabling proactive community policing initiatives and fostering collaboration between law enforcement and the public.

7. **Automated risk assessment and decision support:** Developing automated risk assessment tools and decision support systems that assist law enforcement agencies in prioritizing cases, allocating resources, and making informed decisions based on the predicted likelihood of criminal activity.

8. **Enhanced visualization and reporting capabilities:** Improving data visualization techniques and reporting functionalities to present insights and predictions in a more intuitive and actionable manner, enabling stakeholders to easily interpret and act on the information provided by the system.

9. **Ethical and privacy considerations:** Implementing robust privacy safeguards and ethical guidelines to ensure responsible use of data and mitigate potential biases or unintended consequences associated with predictive policing technologies.

10. **Continuous evaluation and feedback:** Establishing mechanisms for ongoing evaluation and feedback from users, stakeholders, and experts to iteratively improve the system's performance, usability, and effectiveness over time.

CHAPTER 10
REFERENCES

10. REFERENCES

[1] K. Zakir Hussain, M. Durairaj and G. R. J. Farzana, "Criminal behavior analysis by using data mining techniques," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, 2012, pp. 656-658.

[2] Keyvanpour, Mohammad & Javideh, Mostafa & Ebrahimi, Mohammadreza. (2011). Detecting and investigating crime by means of data mining: A general crime matching framework. Procedia CS. 3.872-880. 10.1016/j.procs.2010.12.143.

[3] Ioannis Kavakiotis OlgaTsave Athanasios Salifoglou Nicos Maglaveras Ioannis Vlahavas Ioanna Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, Computational and Structural Biotechnology Journal Volume 15, 2017, Pages 104-116

[4] Frank, Eibe & Hall, Mark & Holmes, Geoffrey & Kirkby, Richard & Pfahringer, Bernhard & Witten, Ian & Trigg, Len. (2010). Weka-A Machine Learning Workbench for Data Mining. 10.1007/978-0-387- 09823-4_66.

[5] Pang-Ning Tan; Michael Steinbach; Anuj Karpatne; Vipin Kuma Introduction to Data Mining 2nd ed, Publisher: Pearson, 2019, Print ISBN: 9780133128901, 0133128903 eText ISBN: 9780134080284,013408028

[6] M. Kantardzic, Data Mining Concepts, Models, Methods, and Algorithms, 2nd ed, John Wiley & Sons, Inc., Hoboken, New Jersey 2011, ISBN 978-0-470-89045-5 , oBook ISBN: 978- 1-118-02914-5,ePDF ISBN: 978-1-118-02912-1, ePub ISBN: 978-1-118-02913-8

[7] Ahishakiye, Emmanuel & Opiyo, Elisha & Wario, Ruth & Niyonzima, Ivan. (2017). A Performance Analysis of Business Intelligence Techniques on Crime Prediction. International Journal of Computer and Information Technology. 06. 84 - 90.

[8] Marlina, Leni & Muslim, Muslim & Siahaan, Andysah Putera Utama. (2016). Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms). International Journal of Emerging Trends & Technology in Computer Science. 38. 380-383.10.14445/22315381/IJETTV38P268.

[9] Himani Bhavsar, Mahesh H. Panchal, (2012). A Review on Support Vector Machine for Data Classification, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 1, Issue 10, December 2012, ISSN: 2278 – 1323.

[10] Wang, Y., Li, Z., & Guo, Q. (2019). Application of Machine Learning in Crime Prediction. In 2019 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 154-157). IEEE.

[11] Chawla, S., & Kashyap, V. (2020). Crime Analysis and Prediction using Machine Learning Techniques. International Journal of Computer Applications, 174(9), 29-33.

[12] Javed, S., & Shahbaz, M. (2018). Crime Prediction using Machine Learning Techniques: A Survey. In 2018 3rd International Conference on Communication and Information Processing (ICCIP) (pp. 227-231). IEEE.

[13] Lu, C. T., & Wu, Y. H. (2017). A Study on Crime Prediction Using Machine Learning Techniques. International Journal of Distributed Sensor Networks, 13(3), 1550147717694458.

[14] Singh, R., & Singh, S. (2016). Crime Detection and Prediction using Machine Learning Algorithms. International Journal of Advanced Research in Computer Science, 7(4), 126-131.

[15] Mohanty, M. S., & Chandra, S. (2015). Crime Prediction using Data Mining and Machine Learning Techniques. International Journal of Computer Applications, 117(22), 29-34.

[16] Sharma, A., & Soni, V. (2014). Predictive Policing using Data Mining and Machine Learning. In 2014 International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 490-493). IEEE.

[17] Bhatt, R., & Gupta, R. (2013). Crime Prediction and Analysis using Data Mining. International Journal of Computer Applications, 82(12), 25-29.

[18] Verma, S., & Shukla, A. K. (2012). Crime Data Analysis and Prediction Using Machine Learning. International Journal of Computer Science Issues, 9(5), 218-224.

[19] Patel, S., & Sharma, R. (2011). Crime Prediction System using Data Mining. In 2011 Second International Conference on Emerging Applications of Information Technology (EAIT) (pp. 273-276). IEEE.