EMPOWERING COMMUNITIES WITH CRIME PREDICTION SYSTEM AND ITS WEB ACCESSIBILITY

A PROJECT REPORT

Submitted by

HARITHA D [REGISTER NO :211420104093]

JAYALANI SHIRIN I [REGISTER NO :211420104110]

KALAIYARASI S [REGISTER NO :211420104118]

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE ENGINEERING

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

MARCH 2024

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **“EMPOWERING COMMUNITIES WITH CRIME PREDICTION SYSTEM AND ITS WEB ACCESSIBILITY”** is the bonafide work of “**HARITHA D (211420104093), JAYALANI SHIRIN I (211420104110), KALAIYARASI S (211420104118)”** who carried out the project work under my supervision.

Signature of the HOD with date

Dr.L.JABASHEELA, M.E., Ph.D.,
PROFESSOR AND HEAD,

Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai - 123

Signature of the Supervisor with date

Dr.M.MAHESWARI, M.TECH., Ph.D.,
SUPERVISOR,
ASSISTANT PROFESSOR,

Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai - 123

Certified that the above candidate(s) was examined in the End Semester Project Viva-Voce Examination held on ....25/03/2024..........................

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION BY THE STUDENT

We **HARITHA D (211420104093), JAYALANI SHIRIN I (211420104110), KALAIYARASI S (211420104118)"** hereby declare that this project report titled " **EMPOWERING COMMUNITIES WITH CRIME PREDICTION SYSTEM AND ITS WEB ACCESSIBILITY**", under the guidance of **Dr.MAHESWARI M, M.TECH.**, **Ph.D.,** is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

HARITHA D

JAYALANI SHIRIN I

KALAIYARASI S

ACKNOWLEDGEMENT

Our profound gratitude is directed towards our esteemed Secretary and Correspondent, **Dr. P. CHINNADURAI, M.A., Ph.D**., for his fervent encouragement. His inspirational support proved instrumental in galvanizing our efforts, ultimately contributing significantly to the successful completion of this project

We want to express our deep gratitude to our Directors, **Tmt. C. VIJAYARAJESWARI, Dr. C. SAKTHI KUMAR, M.E., Ph.D., and Dr. SARANYASREE SAKTHI KUMAR, B.E., M.B.A., Ph.D.,** for graciously affording us the essential resources and facilities for undertaking of this project.

Our gratitude is also extended to our Principal, **Dr. K. MANI, M.E., Ph.D.,** whose facilitation proved pivotal in the successful completion of this project.

We express our heartfelt thanks to **Dr. L. JABASHEELA, M.E., Ph.D.,** Head of the Department of Computer Science and Engineering, for granting the necessary facilities that contributed to the timely and successful completion of project.

We would like to express our sincere thanks to **Project Coordinator**, **Dr. KAVITHA SUBRAMANI, M.E., Ph.D.,** and **Project Guide, Dr. MAHESWARI M, M.TECH.**, **Ph.D.,** and all the faculty members of the Department of CSE for their unwavering support for the successful completion of the project.

HARITHA D
JAYALANI SHIRIN I
KALAIYARASI

ABSTRACT

A crime is an intentional act that, depending on its seriousness, may result in punishment from the state or another authority, as well as material damage or loss, psychological harm, or both. The worrisome pace of growth in both the quantity and types of criminal activity is pressuring authorities to devise effective strategies for taking preventive action. Because they are inefficient and slow-moving, traditional crime-solving approaches cannot keep up with the present fast-rising crime rate. This work aims to provide a succinct overview of the use of these algorithms in crime prediction by surveying and introducing a variety of criminal analysis and prediction methods utilizing a number of machine learning techniques. These methods are based on the percentage of an accuracy measure. This review study aims to support future research in developing these techniques for crime analysis by presenting some challenges related to crime identification, risk rate, and classifications with a comparative study. And it also helps to classify whether an area is safe or not. The system attains an impressive accuracy of 84% with a precision of 0.84. It is anticipated that this will be useful in introducing these techniques to crime researchers. The use of past crime data trends helps us to correlate factors that might help in understanding the risk probability of a particular area.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

In recent years, the rise in criminal activities has underscored the need for advanced tools and methodologies to enhance public safety and law enforcement efforts. As societies become more complex, traditional methods of crime prevention and detection are proving insufficient to address the dynamic nature of criminal behaviors. Crime is a pervasive issue affecting communities globally, with law enforcement agencies continuously seeking innovative approaches to stay ahead of evolving criminal tactics. The conventional reactive approach to law enforcement, which focuses on investigating and solving crimes after they occur, has limitations. The Crime Prediction System seeks to shift the paradigm by adopting a proactive strategy that anticipates potential criminal activities based on historical data, patterns, and trends. In response to this challenge, the Crime Prediction System aims to leverage cutting-edge technologies and data analytics to predict and prevent criminal activities, thereby contributing to the creation of safer communities.

Constructing a Crime Prediction System lowers the crime rate and expedites the process of solving crimes. The primary objectives of the Crime Prediction System include various factors. Predictive Analytics, Utilizing advanced data analytics and machine learning algorithms to analyse historical crime data and identify patterns that can aid in predicting future criminal activities. Resource Optimization, which enhances law enforcement efficiency by strategically deploying resources to areas with higher predicted crime rates, thus optimizing the allocation of personnel and equipment. Community Engagement is based on fostering collaboration between law enforcement agencies and the community to create a symbiotic relationship that encourages information sharing and proactive crime prevention initiatives.

Real-time Monitoring involves implementing a system capable of real-time monitoring and analysis of data streams to promptly identify emerging crime trends and respond swiftly to potential threats. Our various approaches are based on previously published data that has been time- and location-stamped. The Crime Prediction System makes use of data that has been recorded, analyses it using a variety of methods, and then employs any of the methods listed below to forecast patterns and trends in crime. Systems for predicting crime are intended to aid law enforcement in preventing and solving crimes more successfully. These systems examine crime trends and identify possible suspects by utilizing cutting-edge technology like artificial intelligence (AI), machine learning (ML), and data analytics and forecast potential criminal scenes. Crime prediction systems can aid law enforcement authorities in more effective resource allocation by analyzing past crime data to find patterns and trends. Law enforcement might potentially dissuade criminal activity by increasing patrols and monitoring high-risk locations before crimes occur.

Our idea will focus on the development of a robust and scalable platform capable of processing large volumes of historical and realtime data. The system will employ machine learning models to analyze patterns, identify risk factors, and generate predictive insights. The scope encompasses collaboration with law enforcement agencies, community stakeholders, and technology experts to create a comprehensive solution. A system that effectively integrates technology and uses resources efficiently can help to improve public safety.

1.2 PROBLEM DEFINITION

The Crime prediction system utilizes historical crime data, socio-economic factors, demographic information, and other relevant variables to predict future criminal activities. Used to Improve law enforcement efficiency by enabling proactive resource allocation based on predicted crime hotspots. The need for a predictive system arises to identify patterns, trends, and potential crime hotspots before incidents happen.

Various machine learning algorithms are used to build a successful model with a higher accuracy rate. With the help of the predictive analysis of crimes, developing a Flask web application that acts as a user-friendly interface for interacting with the system in order to take necessary preventive methods to overcome the crime. To Design and implement a robust crime prediction system that leverages historical crime data and various contextual factors to forecast the likelihood, location, and timing of future criminal Incidents.

The system should employ machine learning techniques for accurate predictions, considering temporal and spatial patterns, while addressing ethical considerations, and transparency, and providing a user-friendly interface for law enforcement agencies to enhance proactive policing strategies and community safety. Address the ethical concerns related to data privacy, bias in predictions, and transparency in the decision-making process to ensure the fairness and accountability of the system. For that various clustering techniques like boosting technologies are employed to classify the crimes. To Develop a user-friendly interface using the flask framework which is tailored for law enforcement agencies, facilitating effective interaction with the system visualization of predictions and informed decision-making.

In the Flask web application, the users can check for the future crime location and the type of the crime and some preventive measures to be taken either by the government side or by the individual in order to get rid of social criminal activities, Based on the predictive analysis future predictions will be calculated and necessary precaution methods needed to be taken will be specified. For the future prediction user can enter the date, month, year, state, address, latitude, and longitude of the place to be searched, to know the future predicted result with some preventive measures. This idea is useful for law enforcement agencies to proactively allocate resources, plan interventions, and prioritize areas based on the predicted crime risks. By Contributing to overall community safety by creating a system that aids in preventing and mitigating criminal activities through strategic and data-driven approaches.

CHAPTER 2
LITERATURE SURVEY

CHAPTER 2
LITERATURE SURVEY

A few techniques for forecasting crime statistics are described in the recent literatures. Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Direction by Varun Mandalapu, Lavanya Elluri, Piyush Vyas, Nirmalya Roy. The thorough summary of studies on machine learning and deep learning techniques to crime prediction that is provided in this publication is a useful resource for academics working in this area. Law enforcement organisations can create more effective preventative and response plans for criminal activity by investigating crime prediction techniques in greater detail [1]. And [2] Sapna Singh Kshatri, Deepak Singh, Bhavana Narain, Surbhi Bhatia, Mohammed Tabrez Quasim, G. R. Sinha's, An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach gives an effective authentic approach for selecting acceptable crime predictions through the implementation of learning-based methods utilizing MATLAB is proposed: the assemble-stacking based crime prediction method (SBCPM), which is based on SVM algorithms.

Ayisheshim Almaw, Kalyani Kadam presented [5] Survey Paper on Crime Prediction using Ensemble Approach.This review study provides a fair examination of data mining techniques and ensemble classification approaches for the detection and prediction of future crimes. It also summarises the methods and techniques used in crime data analysis and prediction. Forecasting districts with a high probability of crimes occurring on a given day are those where Naïve Bayes predicts more accurately when it comes to crime data analysis and prediction and the Apriori Algorithm is utilized to identify recurring trends in criminal activity [9] is presented by Shiju Sathyadevan, M.S. Devan, S. Surya Gangadharan in their Crime analysis and prediction using data mining in IEEE 2014.

Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention by Neil Shah, Nandish Bhagat, Manan Shah seeks to have an impact by predicting the type of crime utilising ML algorithms and computer vision techniques [3]. Combining machine learning (ML) methods to function as a database for all crimes that have been reported according to categories with computer vision techniques to provide visual information of the surrounding area, enables prediction of crime to occur.

TABLE I – LITERATURE SURVEY

Sno	**TITLE**	**AUTHORS & JOURNAL**	**ALGORITHMS**	**RESULTS**
1	Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions	Varun Mandalapu, Lavanya Elluri, Piyush Vyas, Nirmalya Roy. (IEEE)	Application of ARIMAX – TFM with a single input, XGBoost and applied multi-class classification models like OVR-XGBoost and OVO-XGBoost	Presents a graph-based ensemble classification approach for predicting crime reports better than traditional classifiers.
2	An Empirical Analysis of Machine Learning Algorithms for Crime	Sapna Singh kshatri, Deepak Singh, Bhavana Narain,	Stacking, Naïve Bayes, Sequential minimal	The stacking method is more reliable than the Ensemble method.

	Prediction Using Stacked Generalization: An Ensemble Approach	Surbhi Bhatia, Mohammed Tarbez Quasim, G.R.Sinha (IEEE)	optimization (SMO).	
3	Crime Prediction and Analysis	Pratibha Kumari, Akanksha Gahalot, Uprant, Suraina Dhiman, Lokesh Chouhan.	Extra Tree Classifier, K-Neighbour Classifier, Support Vector Machine (SVM), Decision Tree Classifier and Artificial Neural Network (ANN)	Decision tree, k-neighbour, support vector machine are working best with optimal training and good accuracy.
4	Survey Paper on Crime Prediction using Ensemble Approach	Ayisheshim Almaw, Kalyani Kadam.	Naïve Bayes, Decision tree, Artificial Neutral network.	The decision tree performs well compared to the other two algorithms.
5	Crime analysis and prediction using data mining	Shiju Sathyadevan, M.S. Devan, S. Surya Gangadharan.	Naïve Bayes Decision Tree Apriori Algorithm Mongo DB GraphDB	Naive Bayes algorithms provides higher accuracy rate.

CHAPTER 3
THEORETICAL BACKGROUND

CHAPTER 3
THEORETICAL BACKGROUND

3.1 EXISTING SYSTEM

The existing system relies on traditional crime risk estimation methods which detect crime hotspots based solely on historical crime data. It utilizes SVM algorithms for crime prediction, but these methods exhibit low accuracy and performance, especially for large-scale prediction work. The applicability of the existing system is limited to small-level prediction tasks.

Three phases are included in the schematic establishment and definition of the study process, computational measurements, and analytic methodologies.

(1) The creation of a combined crime prediction model based on machine learning stack.

(2) To recognize and investigate the types of crimes Computer-based empirical research investigation and construction of crime prediction models through the application of several basic, simple, and aggregate machine learning techniques to estimate crime rates and pre-identified crime factors;

(3) Implementation of multiple ensemble machine learning (ML) models, such as bagging, boosting, and stacking;

(4) Comparison and validation of the results by evaluating the model's performance.

Drawbacks: -

• Previously used the SVM which resulted in low accuracy score

• Does not integrated into application .

3.2 PROPOSED METHODOLOGY

3.2.1 IMPLEMENTATION ENVIRONMENT

SOFTWARE AND HARDWARE REQUIREMENTS

- **Hardware**

- OS – Windows 7, 8 and 10 (32 and 64 bit)
- RAM – 4GB

- **Software**
- Python Navigator
- Anaconda

3.2.2 PROPOSED SYSTEM

The Crime prediction system utilizes historical crime data, socio-economic factors, demographic information, and other relevant variables to predict future criminal activities. The need for a predictive system arises to identify patterns, trends, and potential crime hotspots before incidents happen. Our model consists of two phases one phase for identifying the risk probability of a particular area, and another phase comprised of predicting the crime occurred at a particular area. For Predicting the crimes, we use the following machine-learning algorithms,

- Random forest deals with the construction of more number of decision trees which will give the accurate execution, moreover it is quite less explainable than decision trees.
- KNN is used since it is a simple algorithm that specifies by making use of the majority vote if its nearest neighbors.
- Adaboost is used since here all weak classifiers were treated in a well manner, in our system we introduced a trace of reassigning value in the algorithm and the classification report for both original and modified were displayed as the results obtained.
- A Decision tree is a supervised machine learning algorithm used for both classification and regression tasks. It works by recursively partitioning the dataset into subsets based on the values of input features.

For risk evaluation phase we are using,

- K-means is one of the most popular clustering algorithms in machine learning. It's an unsupervised learning algorithm used to partition a dataset into a set of K clusters. The objective of the algorithm is to minimize the variance within clusters and maximize the variance between clusters.

After the model evaluation, with the help of the predictive analysis of crimes, by developing a Flask web application that acts as a user-friendly interface for interacting with the system to take necessary preventive methods to overcome the crime by identifying the occurrences of crime in any particular district. Based upon the risk level of a particular area government can take necessary precaution measures.

3.2.3 SYSTEM ARCHITECTURE

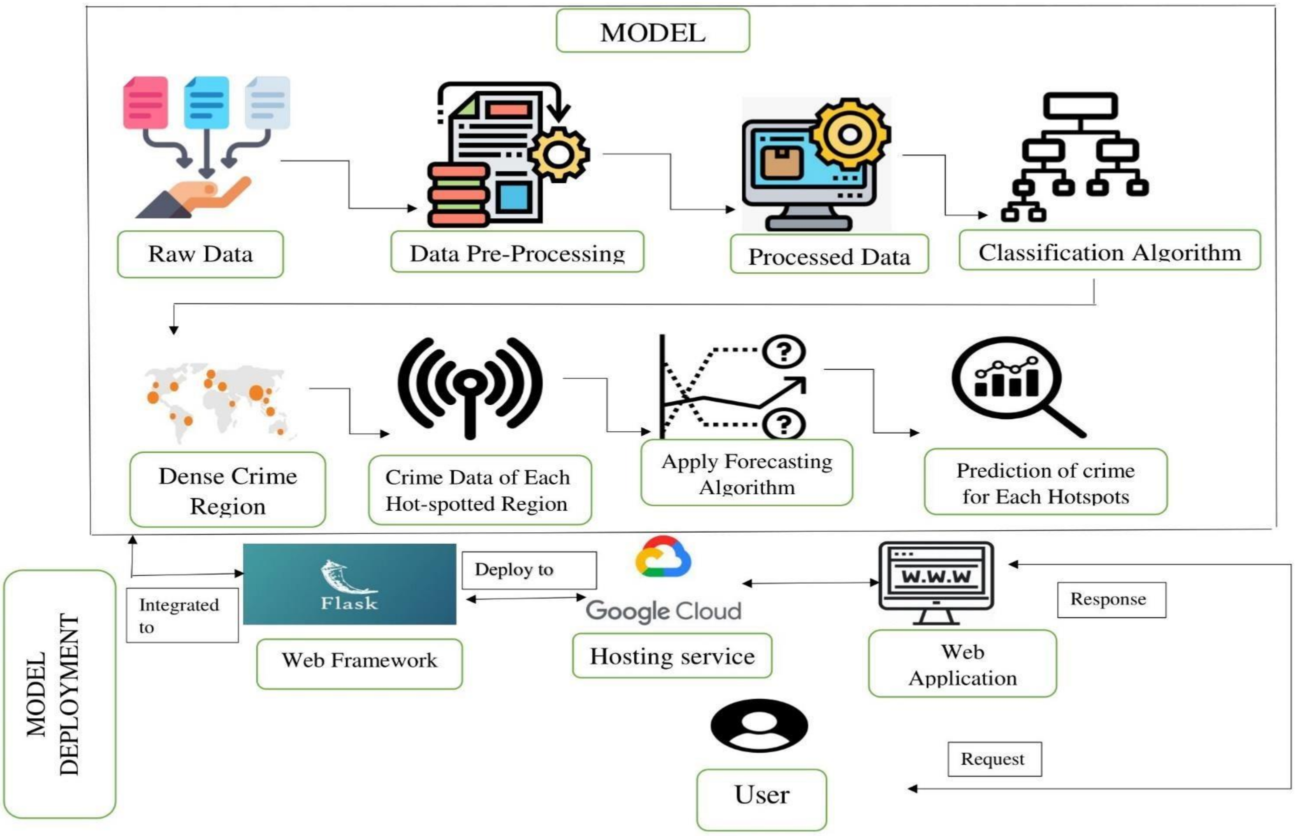


Fig 3.2.3.1 Architecture diagram

This displays the whole design of our working model including the components and the states occurring during the execution of the process. It displays the very initial process of Extracting the dataset and followed by pre-processing of the dataset. Then we have to extract the feature model and we have to train and test the model using various algorithms and performance analysis based on the acquired accuracy score rating.

3.2.4 DATASET DESCRIPTION

- The dataset has been collected from the government website called data.gov.in, which consists of the following attributes with values,

▪	Timestamp	**act379**	**Act313**	**act279**	**act323**	**act363**	**act302**	**latitude**	**longitude**
0	28-02-2023 21:00	1	0	0	0	0	0	22.737260	75.875987
1	28-02-2023 21:15	1	0	0	0	0	0	22.720992	75.876083
2	28-02-2023 10:15	0	0	1	0	0	0	22.736676	75.883168
3	28-02-2023 10:15	0	0	1	0	0	0	22.746527	75.887139
4	28-02-2023 10:30	0	0	1	0	0	0	22.769531	75.888772
5	28-02-2023 14:15	0	0	0	1	0	0	22.735218	75.913366
6	28-02-2023 08:00	0	0	0	0	1	0	22.736766	75.908579

3.2.5 UML DIAGRAMS

3.2.5.1 MODEL WORKFLOW

The diagram specifies the workflow of our system, like how we collect datasets and their processing to build a successful model.

WORKFLOW DIAGRAM

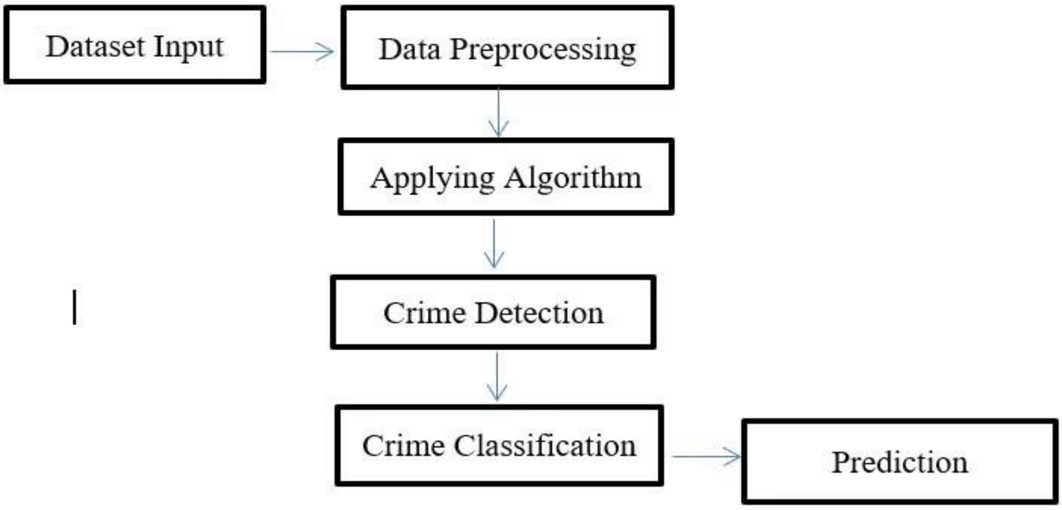


Fig 3.2.5.1 Workflow Diagram

3.2.5.2 USE CASE DIAGRAM

Use case diagrams give a graphic overview of the actors involved in a system, the different functions needed by those actors, and how these functions interact.

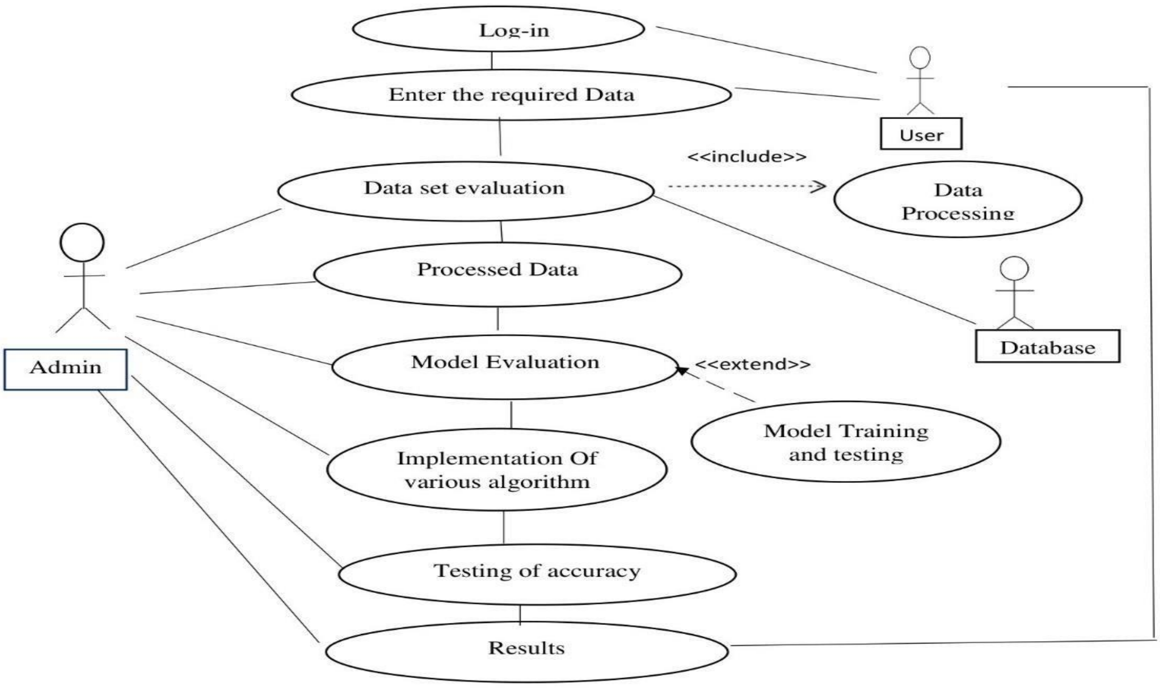


Fig 3.2.5.2 Usecase Diagram

3.2.5.3 ACTIVITY DIAGRAM

Activity diagrams graphically represent workflows. They can be used to describe the business workflow or the operational workflow of any component in a system or the dynamic aspects of the system.

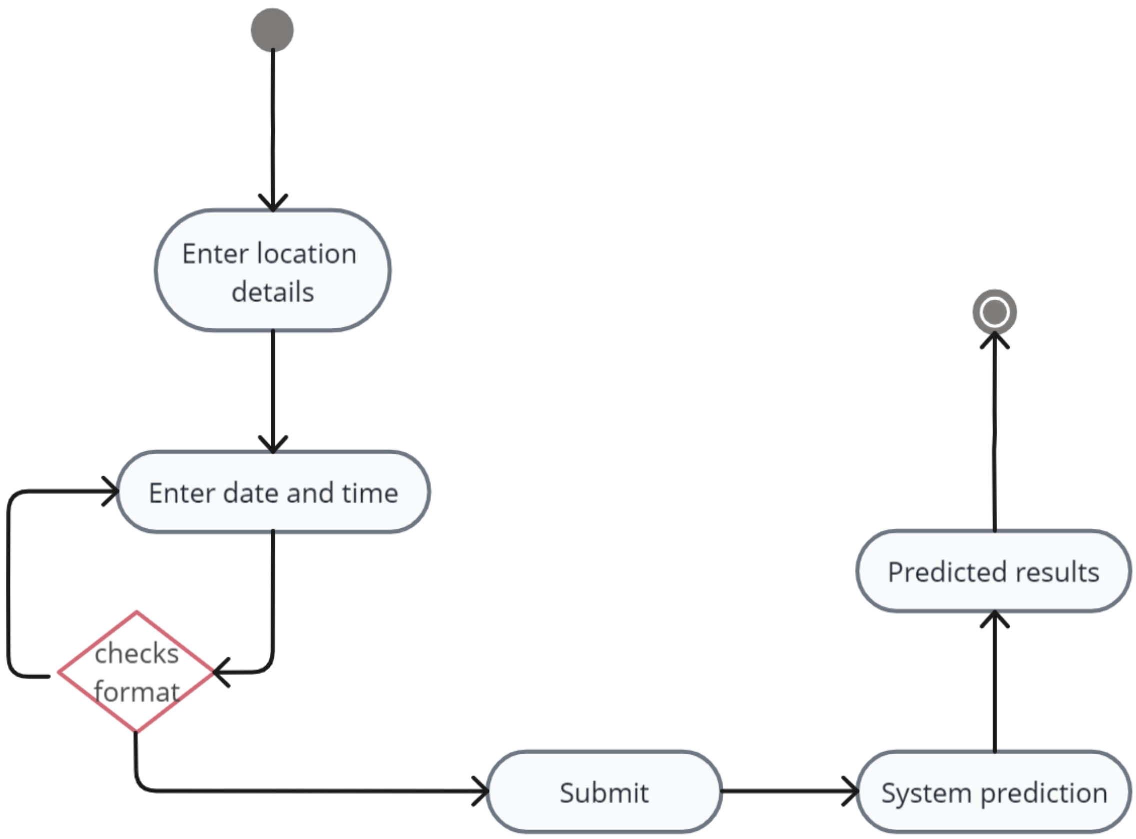


Fig 3.2.5.3 Activity Diagram

The above diagram specifies various activities that are involved in our crime prediction system, after the prediction activity it specifies the activity that the user performs to get use of our model, crime prediction system. User has to enter the location details like latitude and longitude of the area they are going to predict, and then they enter the date and time if the format is correct then the system proceeds with prediction otherwise it asks for reentering data in the correct format, and it finally displays the results.

3.2.5.4 SEQUENCE DIAGRAM

Sequence diagrams in UML show how objects interact with each other and the order those interactions occur. It’s important to note that they show the interactions for a particular scenario.

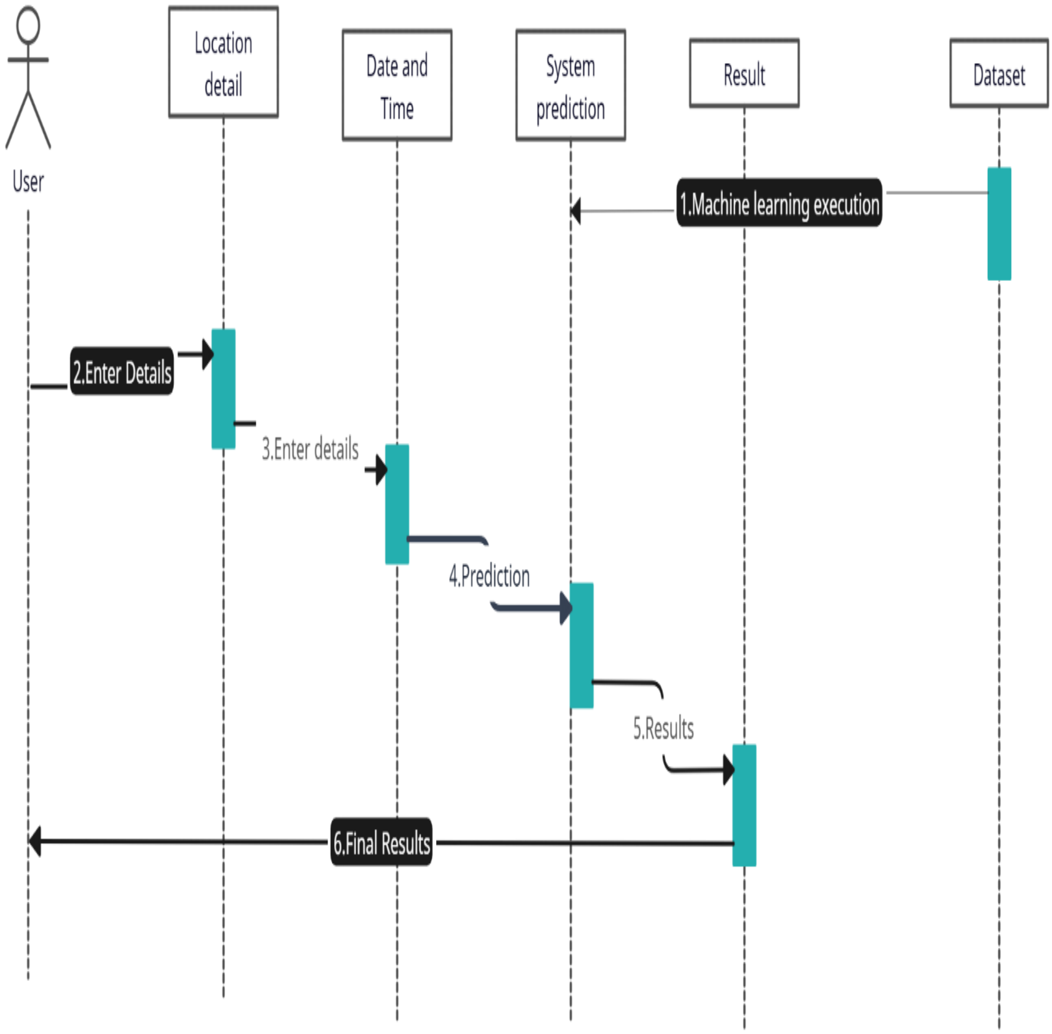


Fig 3.2.5.4 Sequence Diagram

3.2.5.5 DATAFLOW DIAGRAM

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination.

DATA FLOW DIAGRAMS: -

LEVEL 0:

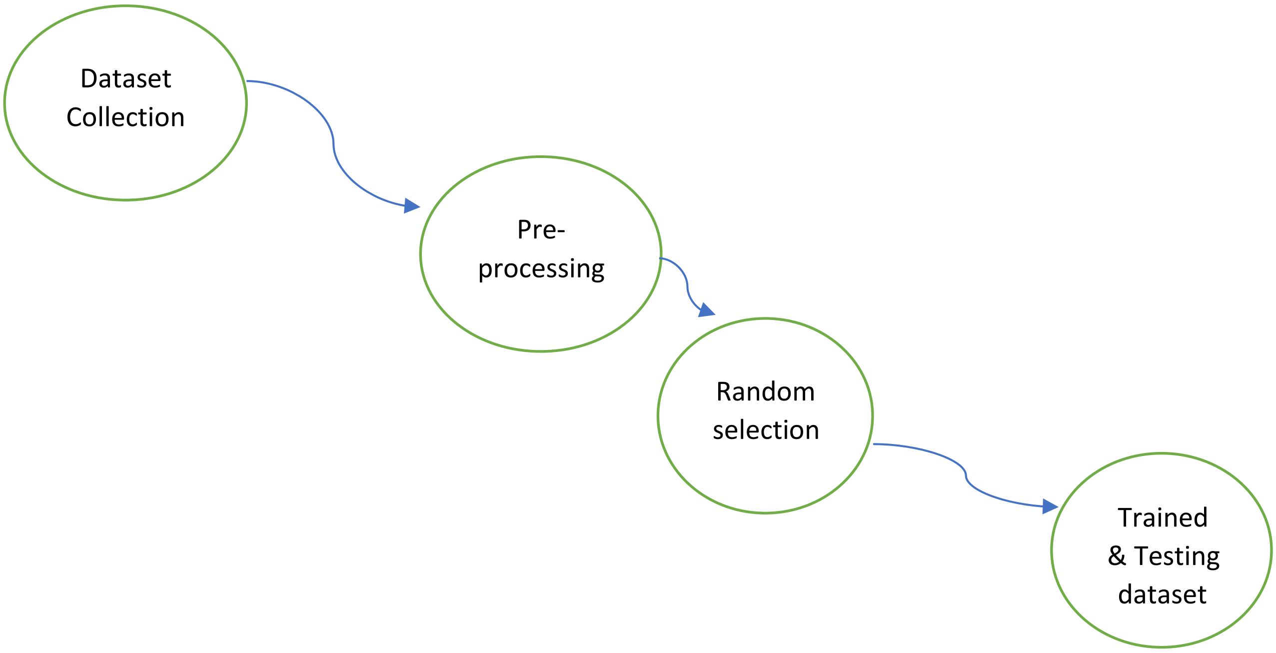


Fig 3.2.5.5 DFD Level 0

LEVEL 1:

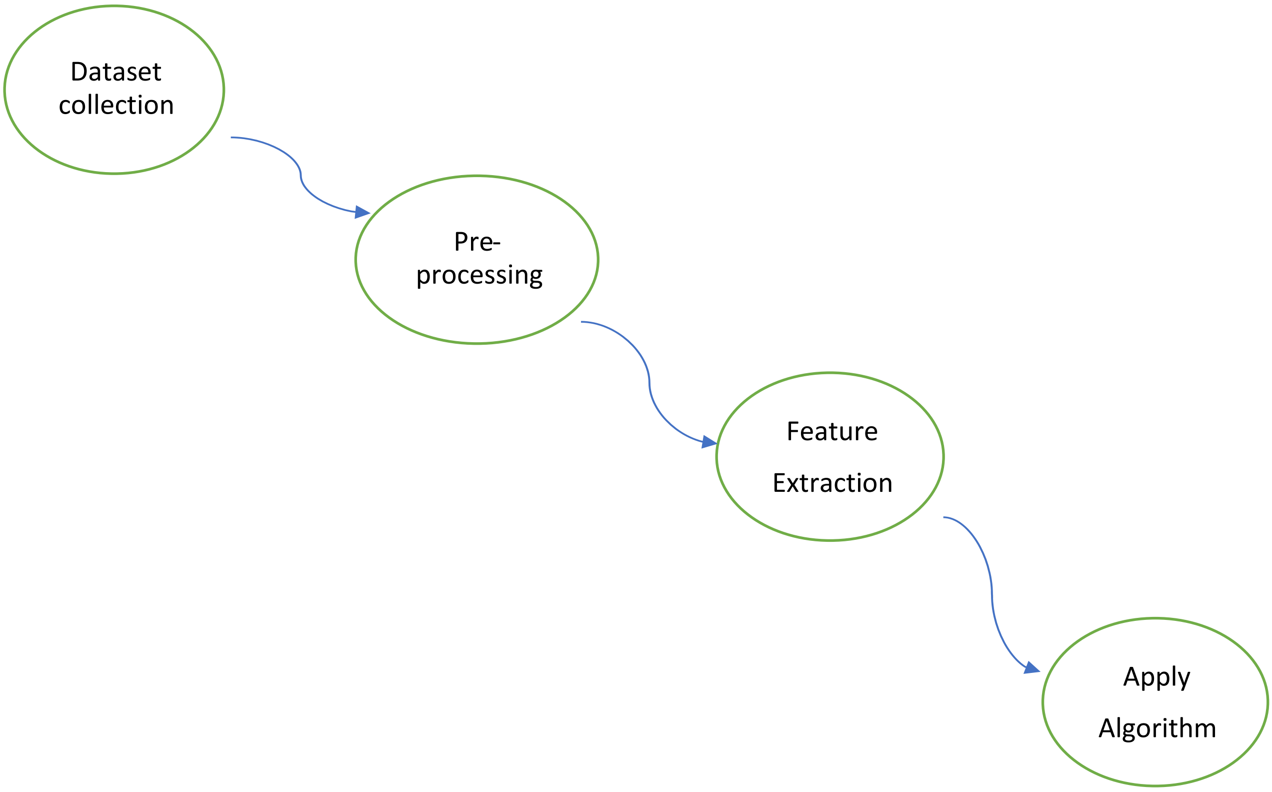


Fig 3.2.5.6 DFD Level 1

LEVEL 2:

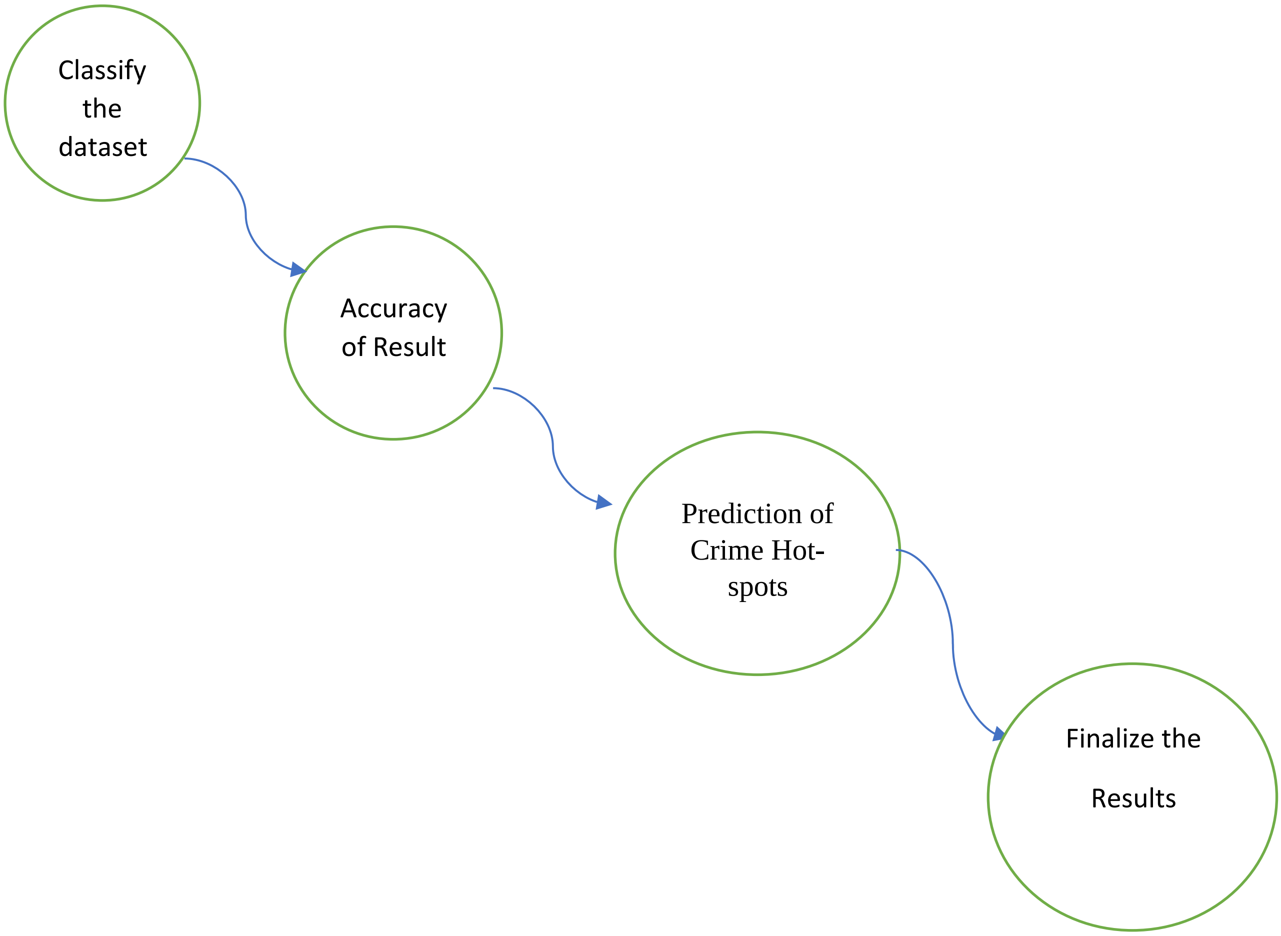


Fig 3.2.5.7 DFD Level 2

CHAPTER 4
SYSTEM IMPLEMENTATION

CHAPTER 4

SYSTEM IMPLEMENTATION

4.1 DATA COLLECTION

- Data used in this paper is a set of product reviews collected from credit card transaction records.
- This step is concerned with selecting the subset of all available data that you will be working with.
- ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer

4.2 DATA PREPROCESSING

Pre-processing: It is the process of three important and common steps as follows:

- **Formatting**: It is the process of putting the data in a legitimate way that it would be suitable to work with. Format of the data files should be formatted according to the need. The most recommended format is .csv files.
- **Cleaning**: Data cleaning is a very important procedure in the path of data science as it constitutes the major part of the work. It includes removing missing data and complexity with naming categories and so on. For most data scientists, Data Cleaning continues for 80% of work.
- **Sampling**: This is the technique of analyzing the subsets from whole large datasets, which could provide a better result and help in understanding the behavior and pattern of data in an integrated way.

4.3 FEATURE EXTRACTION

Feature extraction involves extracting relevant features from the data that can contribute to crime prediction, considering factors like time, day, week, and some other geographical features.

4.4 MODEL EVALUATION

Model Evaluation is an essential part of the model development process. It helps to find the best model that represents our data and how well the selected model will work in the future. By choosing the best deep learning method to be suitable for crime prediction by splitting the dataset into training and testing sets. Training the model using training data by adjusting hyperparameters as needed and evaluating the trained model's performance using metrics like accuracy, precision, recall, and F1 score.

4.5 MODEL DEPLOYMENT

Develop a Flask web application that can take input data and make predictions using the trained model. To design a user-friendly interface for interacting with the system.

4.6 DATA VISUALIZATION

In a pairplot, a grid of scatterplots is created for each pair of features in a dataset. Along the diagonal, histograms or kernel density plots are usually shown for each individual feature. This visualization is helpful for quickly identifying patterns, correlations, and distributions within the data.

Using the same way we can get pairplot for all of the crimes that we used in our project.

Comparison of act13 with act363 and act279

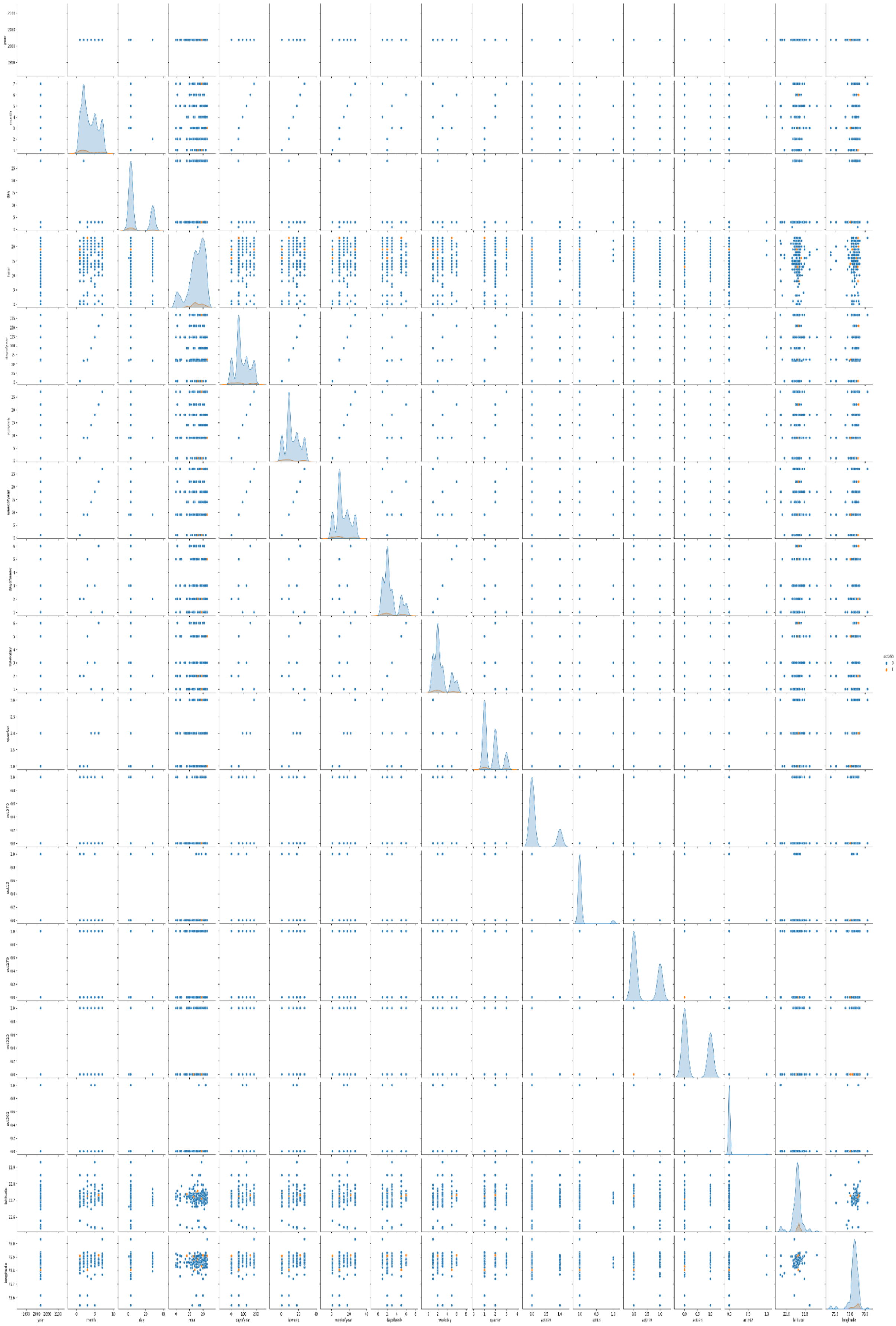


Fig 4.6.1 Pairplot of act363

Comparison of act13 with act363 and act279

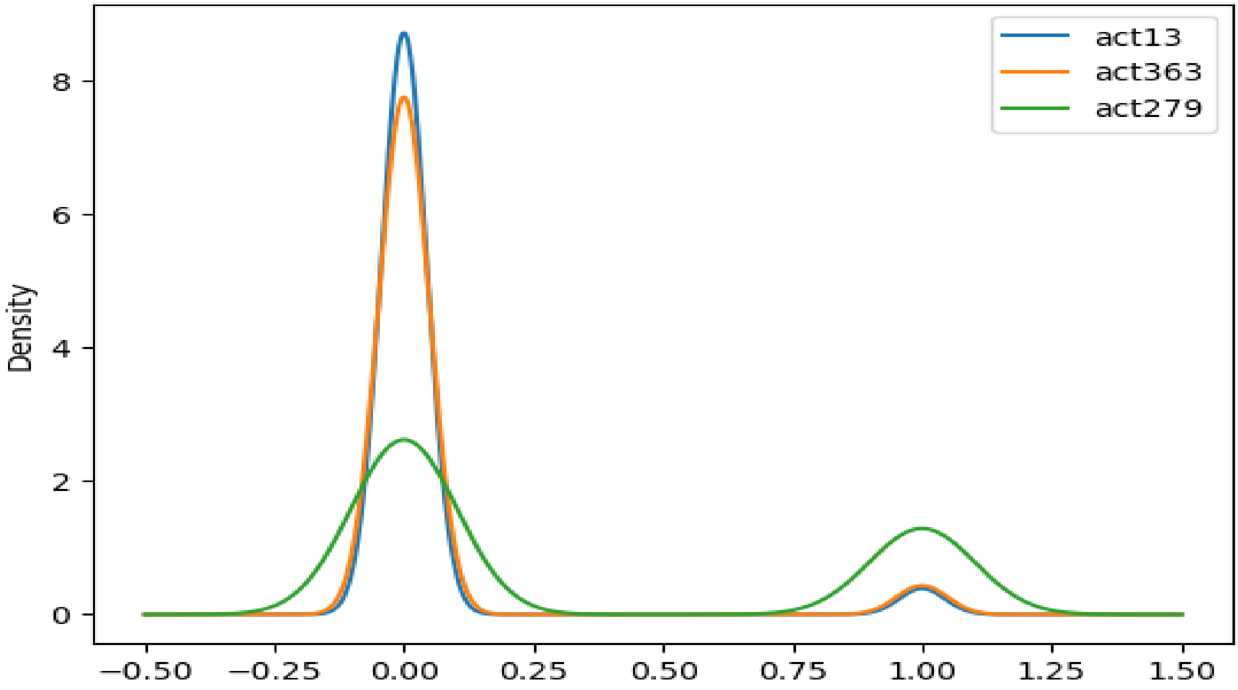


Fig 4.6.2 Comparison of act13 with act363 and act279

Comparison of act13 with act323 and act379

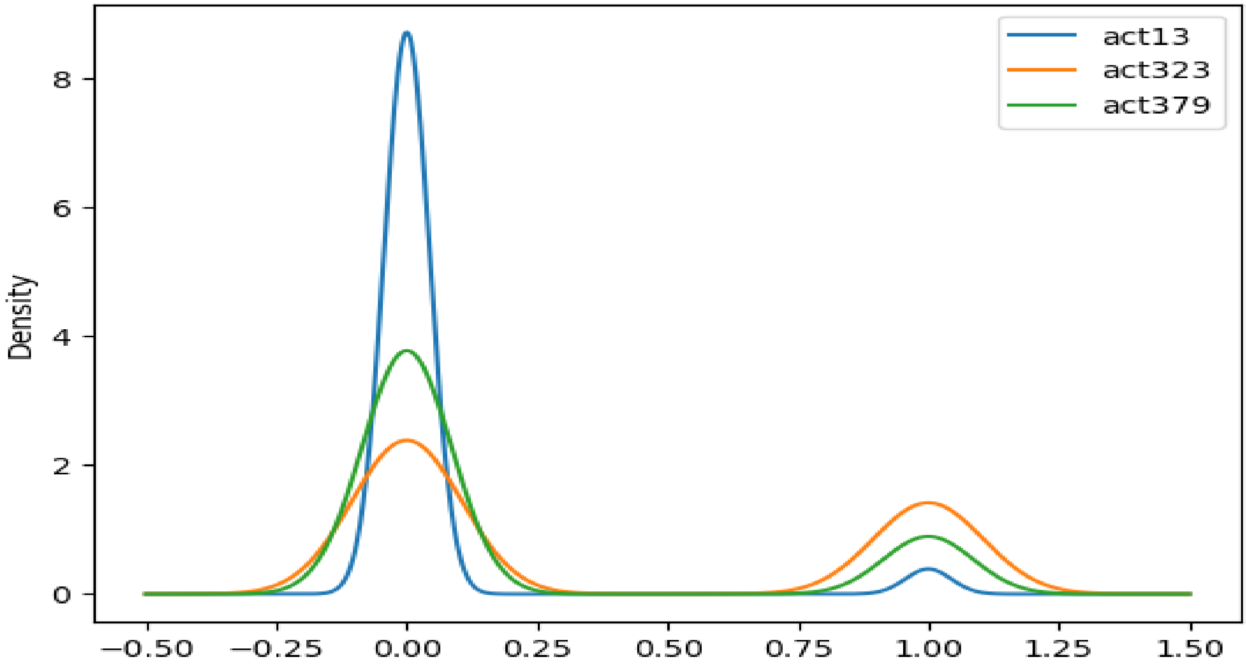


Fig 4.6.3 Comparison of act13 with act323 and act379

Comparison of act13 with act302 and act323

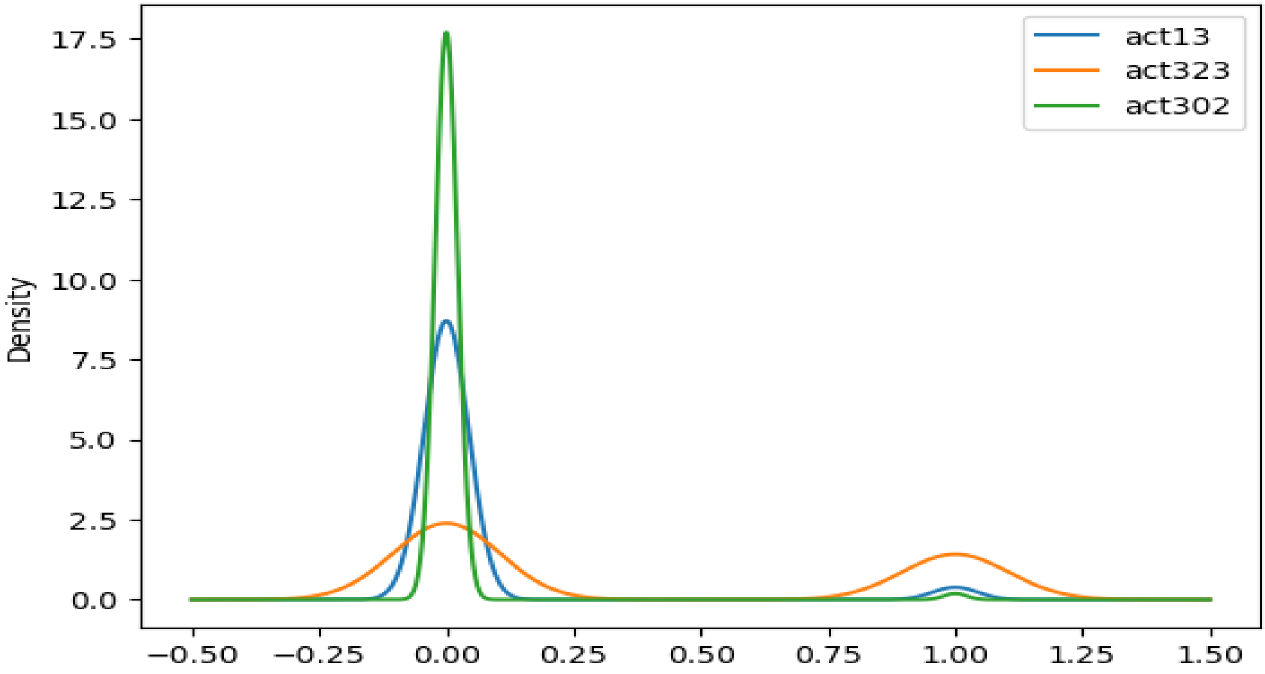


Fig 4.6.4 Comparison of act13 with act333 and act302

By using the same method we can visualize comparison of each crime with other crimes according to our dataset.

Visualization of clusters on a map

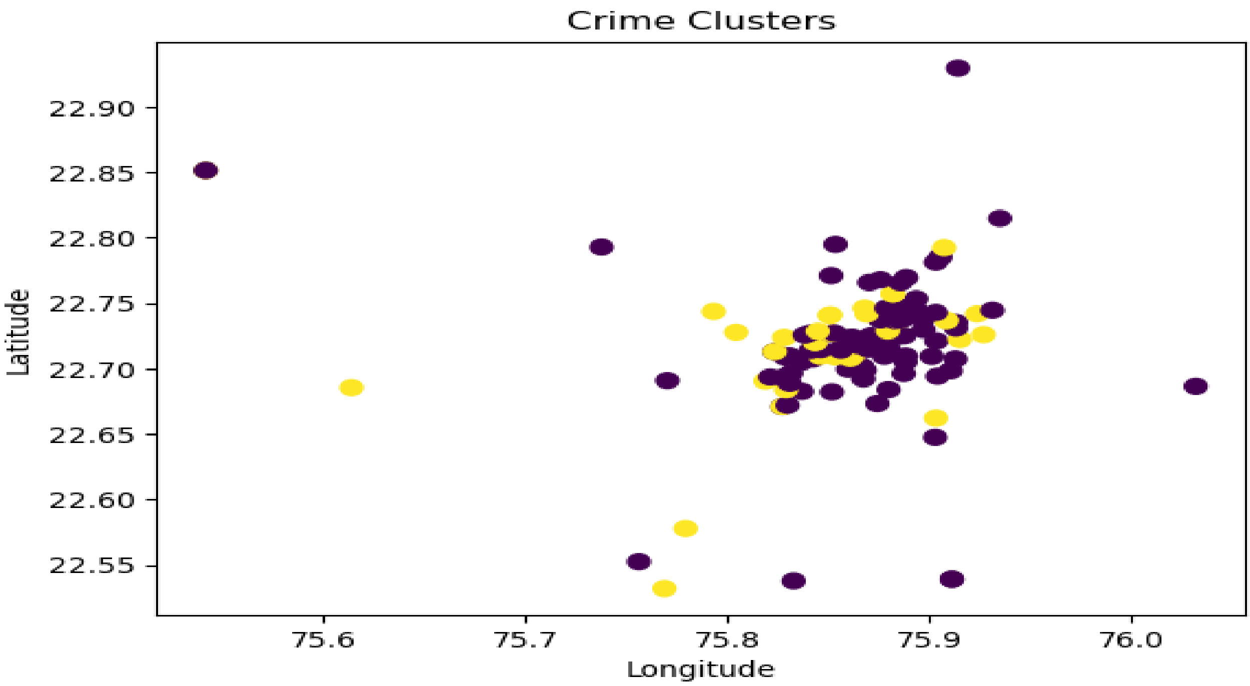


Fig 4.6.5 Visualization of clusters on a map

4.7 PROPOSED ALGORITHMS

RANDOM FOREST

RF is an aggregation style and is regarded as association learning for identifying foundational concepts and regression. To understand uneven patterns, deep woods are utilized. Nevertheless, RF uses a mean of the variance of its figure, which may be decreased using this technique. if equal portions of the training sample are examined by deep trees. A random pattern is chosen and replaced with the training set that fits for those samples by the training statistics(p = p1. pn) with responses(Q = q1,, qn) and bagging(X times). The following list of guidelines is provided as an algorithm, Random Forest.

The Random Forest Algorithm's operation is described in the phases that follow:

Step 1: Choose a Random selection from a specified data collection or training set in step 1.

Step 2 For each training set of data, this algorithm will build a decision tree.

Step 3:Voting will be conducted using an average of the decision tree.

Step 4:Lastly, choose the prediction result that received the most votes as the predicted result.

There are more specified features for using this random forest algorithm, the most specified one is as it deals with working with more number of decision trees which leads to predicting the result accurately.

DECISION TREE

Although decision trees are a supervised learning technique, they are primarily employed to solve classification problems. However, they can also be used to solve regression problems. The classifier is tree-structured with internal nodes standing in for dataset attributes, branches for decision rules, and leaf nodes for each outcome. The decision node and leaf node are the two nodes that make up the decision tree. While leaf nodes represent the result of decisions and do not have any more branches, decision nodes are used to make any kind of decision and have numerous branches. The characteristics of the provided dataset are used to inform the decisions or the test. It is a graphical tool that shows all potential solutions to an issue or choice. The complete process can be known by the following algorithm,

Step 1 According to S, start the tree at the root node, which has the entire dataset.

Step-2 Use the Attribute Selection Measure(ASM) to determine which attribute in the dataset is the best.

Step 3 Create subgroups inside the S that include potential values for the greatest qualities.

Step 4 Create the decision tree node with the =e best attribute at its core.

Step 5 Using the dataset subsets generated in step 3, recursively develop new decision trees. This process should be continued until the nodes can no longer be classified further, at this point, the final node is referred to as a leaf node.

K NEAREST NEIGHBOUR

The quantity in the information that we require to analyze is referred to as the item or the recognized Variable in the context of supervised learning, which is the literacy that takes into account the volume or result that we desire or expect within the training data (labelled data). Next, we build the interpretation of how the "k- Neighbors Classifier" model operates for the k- Closest Neighbors (KNN) and choose the number "five" as the NN's representative. The number of the "n-neighbors" is called randomly, but it may be determined appreciatively by repeating a number of values, assessed using appropriate methods, and putting the predicted values into the "knn- yhat" variable. The pseudocode for administering the KNN with set of rules from scrape

- cargo the training information.
- put together statistics by means of scaling, missing value remedy, and dimensionality reducing.
- discover the top- quality figure for okay

- anticipate a order value for brand new information o Calculate distance(X, Xi) from i = ,2, three, n. o where X = new data factor, Xi = education records, distance as per your favoured distance metric. o kind those distances in growing order with corresponding train records. o From this taken care of table, pick out the zenith ‘ k ’ rows. o discover the maximum frequent class from those named ‘ k ’ rows

.

ADABOOST

AdaBoost, sometimes called Adaptive Boosting, is a method for device learning that is applied like an Ensemble learning. Because of this want wood with best 1 split, goal trees with just tier are the most popular estimator used with AdaBoost. These are also known as "choice Stumps" trees. This approach creates a model while assigning equal weights to each piece of input. Then, it gives points that would have been classified incorrectly a higher weight. The new model now puts more importance on all the elements with improved weights. It will continue using these educational models until and until fewer errors are made.

The procedure is as follows,

Step 1: On the basis of the education data alone and just using the weighted samples, a weak classifier (such as a decision stump) is created. The weights of each pattern indicate how important accurate labelling is in this situation. The samples are first delivered with identical weights for the first stump.

Step 2: to develop a selection strategy for each variable and observe how well each one categorizes samples according to its intended use. For illustration, we examine Age, Junk Food Consumption, and Exercise in the diagram below. For each individual stump, we would examine how many samples were rightly or wrongly identified as meritorious or unsuitable.

Step 3: In order to label the erroneously categorized samples effectively inside the next choice stump, greater weight is given to them. Every classifier is also given a weight based on how accurate it is; hence, a classifier with a high accuracy is given a high weight.

Step 4: Continue as in Step 2, repeating the process until all data points have been effectively classified or the most production stage has been achieved.

FOR IDENTIFYING RISK PROBABILITY

KMEANS ALGORITHM

K-means is one of the most popular clustering algorithms in machine learning. It's an unsupervised learning algorithm used to partition a dataset into a set of K clusters. The objective of the algorithm is to minimize the variance within clusters and maximize the variance between clusters. Following are the steps involved,

- Group by latitude and longitude, count the number of occurrences (crime count).
- Extract latitude and longitude columns.
- Use KMeans clustering on crime counts.
- Calculate mean and standard deviation of crime counts.
- Define thresholds based on standard deviations from the mean.
- Categorize clusters based on thresholds.
- Combine clusters for each risk level.
- Map risk categories to a single cluster for simplicity.
- Check unique values in 'risk_category'.
- Print information about each cluster and its risk category.

CHAPTER 5
RESULTS AND DISCUSSION

CHAPTER 5

RESULTS AND DISCUSSION

The metrics we are used in our project is accuracy score rating.

5.1 ACCURACY SCORE

One often used metric to assess a classification model's performance is accuracy. The ratio of accurately predicted instances to all instances in the dataset is measured. The accuracy formula is:

1)Accuracy

Accuracy= Total Number of Predictions / Number of Correct Predictions

Accuracy=((TN+TP)/(TP+FP+TN+FN))

5.2 CLASSIFICATION REPORT

The classification report specifies all the factors or metrics as one and the evaluation has made for predicting the system performance.

TABLE-II MODEL EVALUATION TABLE II

ALGORIHMS USED	F1 SCORE
Random Forest	0.8429
Decision Tree	0.8333
K-Nearest Neighbour	0.789
ADABoost	0.408

.

5.3 OBSERVATION OF RESULTS

ACCURACY AND CLASSIFICATION REPORT

1) K NEAREST NEIGHBOUR

```
0.7898550724637681
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.76      0.80        86
           1       0.77      0.53      0.62        19
           2       0.98      0.82      0.90       147
           3       0.96      0.83      0.89       140
           4       0.55      0.63      0.59        19
           5       1.00      1.00      1.00         3

   micro avg       0.91      0.79      0.85       414
   macro avg       0.85      0.76      0.80       414
weighted avg       0.92      0.79      0.85       414
 samples avg       0.79      0.79      0.79       414
```

Fig 5.3.1 Accuracy of k nearest neighbour

2) DECISION TREE

```
0.8333333333333334
Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.74      0.82        86
           1       0.70      0.74      0.72        19
           2       1.00      0.81      0.89       147
           3       0.93      0.91      0.92       140
           4       0.64      0.95      0.77        19
           5       1.00      1.00      1.00         3

   micro avg       0.92      0.83      0.87       414
   macro avg       0.86      0.86      0.85       414
weighted avg       0.93      0.83      0.87       414
 samples avg       0.83      0.83      0.83       414
```

Fig 5.3.2 Accuracy of decision tree

3) RANDOMFOREST

```
0.8429951690821256
Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.74      0.82        86
           1       0.70      0.74      0.72        19
           2       1.00      0.81      0.89       147
           3       0.92      0.93      0.92       140
           4       0.63      1.00      0.78        19
           5       1.00      1.00      1.00         3

   micro avg       0.91      0.84      0.87       414
   macro avg       0.86      0.87      0.85       414
weighted avg       0.92      0.84      0.87       414
 samples avg       0.84      0.84      0.84       414
```

Fig 5.3.3 Accuracy of random Forest

4) ADABOOST

```
0.4082125603864734
Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.74      0.82        86
           1       0.70      0.74      0.72        19
           2       1.00      0.81      0.89       147
           3       0.92      0.93      0.92       140
           4       0.63      1.00      0.78        19
           5       1.00      1.00      1.00         3

   micro avg       0.91      0.84      0.87       414
   macro avg       0.86      0.87      0.85       414
weighted avg       0.92      0.84      0.87       414
 samples avg       0.84      0.84      0.84       414
```

Fig 5.3.4 Accuracy of Adaboost

CHAPTER 6
CONCLUSION AND FUTUREWORK

CHAPTER 6

CONCLUSION AND FUTUREWORK

6.1 CONCLUSION

In conclusion, the proposed crime prediction system showcases the versatility and effectiveness of various machine learning algorithms, when integrated with traditional models. The initial calibration using historical crime data sets the foundation, and the subsequent inclusion of built environment features significantly improves predictive accuracy. Through rigorous comparative analysis, the system identifies the most robust model, which is then deployed for crime prediction. In addition to that, the proposed model will help find the probability of risks due to criminal activity in any particular area, to enhance more security measures. The incorporation of continuous monitoring and periodic updates ensures the adaptability of the model to evolving crime patterns. This holistic approach not only enhances the precision of crime prediction but also underscores the importance of leveraging diverse algorithms and environmental data for more robust and reliable predictive modeling in the field of crime prevention.

6.2 Future Scope

The functionalities of this project can be scaled up in the future. These functionalities could be:

- Real-time data analysis of crime data: This could help us obtain crime patterns and forecasts of the future instantly using real-time datasets.
- Data mining of social media to generate datasets, and then reprocess and analyze them to spot trends of the current crime situation in a particular place or region.
- Compare and display the results of all available and applicable forecasting, predicting, and classification models side by side, such that the user can select any of those methods.

APPENDICES

APPENDICES

A.1 SDG GOAL

The Sustainable Development Goal 16 (SDG 16) aims to promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels. People everywhere should be free of fear from all forms of violence and feel safe as they go about their lives whatever their ethnicity, faith or sexual orientation. High levels of armed violence and insecurity have a destructive impact on a country's development, while sexual violence, crime, exploitation and torture are prevalent where there is conflict or no rule of law, and countries must take measures to protect those who are most at risk.

Governments, civil society and communities need to work together to find lasting solutions to conflict and insecurity. Strengthening the rule of law and promoting human rights is key to this process, as is reducing the flow of illicit arms, combating corruption, and ensuring inclusive participation at all times.

Goal 16 aligns with the broader human rights framework by promoting societies that respect and uphold individual rights, as well as the right to privacy, freedom of expression, and access to information. Peace is a fundamental precondition for social and economic development. Without peace, societies are often plagued by conflict, violence, and instability, which can hinder progress and result in the loss of lives and resources. Equal access to justice is essential for protecting the rights of individuals, resolving disputes, and ensuring that vulnerable populations are not marginalized or mistreated. Crimes threatening peaceful societies, including homicides, trafficking and other organized crimes, as well as discriminatory laws or practices, affect all countries.

A.2 SOURCE CODE

With the help of the dataset, we are implementing machine learning algorithms to find out which algorithm works well. This work will be the backend of our project, after implantation of machine learning part we are converting it into a pickle file as a dumb file , after that we are loading the pickle file into our flask library , by executing our python file we get our finalized output as a user-friendly website.

BACKEND IMPLEMENTATION

PREPARING THE DATASET

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

DATA COLLECTION

```
data=pd.read_csv('crime_data.csv')
```

DATA PREPROCESSING

```
data.head(7)
data.shape
data.info()
data['timestamp'] = pd.to_datetime(data['timestamp'], errors='coerce')
data['timestamp'] = pd.to_datetime(data['timestamp'], format = '%d/%m/%Y
```

```
%H:%M:%S')
data['timestamp']
```

FEATURE EXTRACTION

```
column_1 = data.iloc[:, 0]
db = pd.DataFrame({
    "year": column_1.dt.year,
    "month": column_1.dt.month,
    "day": column_1.dt.day,
    "hour": column_1.dt.hour,
    "dayofyear": column_1.dt.dayofyear,
    "isoweek": column_1.dt.isocalendar().week,
    "weekofyear": column_1.dt.isocalendar().week,
    "dayofweek": column_1.dt.dayofweek,
    "weekday": column_1.dt.weekday,
    "quarter": column_1.dt.quarter,
})
dataset1=data.drop('timestamp',axis=1)
data1=pd.concat([db,dataset1],axis=1)
data1
data1.shape
```

DATA ANALYSIS

```
data1.info()
data1.isnull().sum()
data1.dropna(inplace=True)
data1.isnull().sum()
data1.head()
```

data1.columns[1]

'month'

DATA VISUALIZATION

sns.pairplot(data1,hue='act363')

<seaborn.axisgrid.PairGrid at 0x1bae014e7c0>

df1 = pd.DataFrame(data=data1, columns=['act13', 'act323', 'act379'])

df1.plot.kde()

plt.show()

df1 = pd.DataFrame(data=data1, columns=['act13', 'act363', 'act279'])

df1.plot.kde()

plt.show()

df1 = pd.DataFrame(data=data1, columns=['act13', 'act323', 'act302'])

df1.plot.kde()

plt.show()

TRAINING AND TESTING SPLIT

data1.head()

data1.shape

X=data1.iloc[:,[1,2,3,4,6,16,17]].values

#month,day,hour,dayofyear,weekofyear,latitude,longitude

X[4]

y=data1.iloc[:,[10,11,12,13,14,15]].values

#act379 act13 act279 act323 act363 act302

y[4]

array([0, 0, 1, 0, 0, 0], dtype=int64)

Splitting the data

from sklearn.model_selection import train_test_split

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=50)
X_train.shape
X_test.shape
```

CREATING AND TRAINING KNN MODEL

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, classification_report
knn = KNeighborsClassifier(n_neighbors=10)
knn.fit(X_train,y_train)
KNeighborsClassifier(n_neighbors=10)
acc1=knn.score(X_test,y_test)
acc1
knn.score(X_train,y_train)
y_pred=knn.predict(X_test)
# Display classification report
print(acc1)
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

CREATING AND TRAINING DECISION TREE MODEL

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report
dtree = DecisionTreeClassifier(max_depth=500, random_state=300)
dtree.fit(X_train,y_train)
DecisionTreeClassifier(max_depth=500, random_state=300)
y_pred=dtree.predict(X_test)
acc2=dtree.score(X_test,y_test)
```

acc2

0.9806763285024155

dtree.score(X_train,y_train)

y_pred

 # Display classification report

print(acc2)

print('Classification Report:')

print(classification_report(y_test, y_pred))

CREATING AND TRAINING RANDOM FOREST MODEL

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, classification_report

rfc = RandomForestClassifier(n_estimators=100)

rfc.fit(X_train, y_train)

RandomForestClassifier()

y_pred=rfc.predict(X_test)

X_test[0]

array([2.0, 28.0, 15.0, 59.0, 9, 22.723873, 75.828416], dtype=object)

X_test[10]

array([6.0, 3.0, 10.0, 154.0, 22, 22.714156, 75.856339], dtype=object)

y_pred[10]

array([0, 0, 0, 1, 0, 0], dtype=int64)

acc3=rfc.score(X_test,y_test)

acc3

rfc.score(X_train,y_train)

 # Display classification report

print(acc3)

print('Classification Report:')

print(classification_report(y_test, y_pred))

Creating and Training AdaBoost Model

In [55]:

```python
# Reshape y_train to a 1-dimensional array
y_train_reshaped = np.argmax(y_train, axis=1)
# Reshape y_test to a 1-dimensional array
y_test_reshaped = np.argmax(y_test, axis=1)
# Creating & Training AdaBoost Model
from sklearn.ensemble import AdaBoostClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report
# Define base estimator
base_estimator = DecisionTreeClassifier(max_depth=1)
# Initialize AdaBoost classifier
adaboost = AdaBoostClassifier(base_estimator=base_estimator,
n_estimators=100, random_state=2)
# Train AdaBoost model
adaboost.fit(X_train, y_train_reshaped)
# Evaluate AdaBoost model
acc_adaboost = adaboost.score(X_test, y_test_reshaped)
acc_adaboost_train = adaboost.score(X_train, y_train_reshaped)
print("Accuracy of AdaBoost on test set:", acc_adaboost)
print("Accuracy of AdaBoost on training set:", acc_adaboost_train)
# Prediction with AdaBoost
y_pred_adaboost = adaboost.predict(X_test)
# Display classification report
print(acc_adaboost)
print('Classification Report:')
```

```
print(classification_report(y_test, y_pred))
```

GETTING CRIME MAP WITH LOCATION DETAILS

```
import folium
import pandas as pd

# Load your dataset (replace 'your_dataset.csv' with the actual file path)
#data = pd.read_csv('your_dataset.csv')

# Create a base map
crime_map = folium.Map(location=[data['latitude'].mean(),
data['longitude'].mean()], zoom_start=12)

# Add markers for each crime incident
for index, row in data.iterrows():
    folium.Marker([row['latitude'], row['longitude']],
                  popup=f"Latitude: {row['latitude']}<br>\
                  Longitude: {row['longitude']}",
                  icon=folium.Icon(color='red')).add_to(crime_map)

# Save the map as an HTML file
crime_map.save('crime_map.html')
```

PREDICTION

```
test_vector = np.reshape(np.asarray([2., 28., 15., 59. ,  9. ,
22.723873,75.828416]),(1,7))
p =np.array(rfc.predict(test_vector)[0])

print(p)
```

```
label = ['Robbery','Gambling','Accident','Violence','Kidnapping','Murder']
print (label[3])
[0 0 0 1 0 0]
Violence
import matplotlib.pyplot as plt; plt.rcdefaults()
import numpy as np
import matplotlib.pyplot as plt
objects = ('KNN','DecisionTree','RF','Adaboost')
y_pos = np.arange(len(objects))
performance = [acc1,acc2,acc3,acc_adaboost]
plt.bar(y_pos, performance, align='center', alpha=0.5)
plt.xticks(y_pos, objects)
plt.ylabel('Accuracy Level')
plt.title('Accuracy of Algorithms')
plt.show()
test_vector = np.reshape(np.asarray([  2.      , 28.      , 21.      , 59.      ,  9.      ,
22.73726 , 75.875987]),(1,7))
p =np.array(rfc.predict(test_vector)[0])
print(p)
label = ['Robbery','Gambling','Accident','Violence','Kidnapping','Murder']
if p[0] == 1:
   print("Robbery")
if p[0] ==2:
   print("Gambling")
if p[0] ==3:
   print("Accident")
if p[0] ==4:
   print("Violence")
```

```
if p[0] ==5:
    print("Kidnapping")
if p[0] ==6:
    print("Murder")
[1 0 0 0 0 0]
Robbery
In [60]:
test_vector = np.reshape(np.asarray([2.       , 28.       , 10.       , 59.       ,  9.       ,
22.769531,75.888772]),(1,7))
p =np.array(rfc.predict(test_vector)[0])
print(p)
label = ['Robbery','Gambling','Accident','Violence','Kidnapping','Murder']
if p[0] == 1:
    print("Robbery")
if p[1] ==1:
    print("Gambling")
if p[2] ==1:
    print("Accident")
if p[3] ==1:
    print("Violence")
if p[4] ==1:
    print("Kidnapping")
if p[5] ==1:
    print("Murder")
[0 0 1 0 0 0]
Accident
In [61]:
```

FOR IDENTIFYING THE RISK PROBABILITY OF A PARTICULAR

AREA

CLUSTERING OF DATASET USING KMEANS ALGORITHM

```
import pandas as pd
from sklearn.cluster import KMeans
# Group by latitude and longitude, count the number of occurrences (crime count)
crime_count = data1.groupby(['latitude', 'longitude']).size().reset_index(name='crime_count')
# Extract latitude and longitude columns
coordinates = crime_count[['latitude', 'longitude']]
# Use KMeans clustering on crime counts
num_clusters = 3
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
crime_count['cluster'] = kmeans.fit_predict(coordinates)
# Calculate mean and standard deviation of crime counts
mean_count = crime_count['crime_count'].mean()
std_dev = crime_count['crime_count'].std()
# Define thresholds based on standard deviations from the mean
low_risk_threshold = mean_count - 0.5 * std_dev
medium_risk_threshold = mean_count + 0.5 * std_dev
# Categorize clusters based on thresholds
bins = [-1, low_risk_threshold, medium_risk_threshold, float('inf')]
labels = ['Low Risk', 'Medium Risk', 'High Risk']
crime_count['risk_category'] = pd.cut(crime_count['crime_count'], bins=bins, labels=labels, duplicates='drop')
# Combine clusters for each risk level
risk_mapping = {'Low Risk': 'Low Risk', 'Medium Risk': 'Medium Risk', 'High Risk': 'High Risk'}
```

crime_count['risk_category'] = crime_count['risk_category'].map(risk_mapping)

 # Map risk categories to a single cluster for simplicity

combined_mapping = {'Low Risk': 0, 'Medium Risk': 1, 'High Risk': 2}

crime_count['cluster'] = crime_count['risk_category'].map(combined_mapping)

 # Check unique values in 'risk_category'

unique_risk_categories = crime_count['risk_category'].unique()

print("Unique Risk Categories:", unique_risk_categories)

 # Print information about each cluster and its risk category

for cluster_id, group in crime_count.groupby('cluster'):

 total_crime_count = group['crime_count'].sum()

    # Check if there are any rows in the group

    if not group.empty:

        risk_category = group['risk_category'].iloc[0] # Assuming risk category is the same for all points in a cluster

        print(f"Cluster {cluster_id} - Total Crime Count: {total_crime_count} - Risk Category: {risk_category}")

        print(group[['latitude', 'longitude', 'crim e_count', 'risk_category']])

        print("\n")

    else:

        print(f"Cluster {cluster_id} - Empty Cluster)

Map for risk identification

import pandas as pd
import folium
from sklearn.cluster import KMeans
 # Create a folium map centered around the mean latitude and longitude
map_center = [crime_count['latitude'].mean(), crime_count['longitude'].mean()]
crime_map = folium.Map(location=map_center, zoom_start=12)

 # Add markers for each data point with popup text
for index, row in crime_count.iterrows():

```
    folium.Marker(
        location=[row['latitude'], row['longitude']],
        popup=f"Latitude: {row['latitude']} <br>\ Longitude: {row['longitude']}<br>\ Crime Count: {row['crime_count']}"
    ).add_to(crime_map)

# Save the map as an HTML file
crime_map.save('map1.html')
```

PREDICTION

```
def predict_risk_category(latitude, longitude, crime_count):
    # Predict the cluster for the given latitude and longitude
    prediction = kmeans.predict([[latitude, longitude]])[0]

    # Map the predicted cluster to the risk category
    risk_category = {0: 'Low Risk', 1: 'Medium Risk', 2: 'High Risk'}.get(prediction, 'Unknown')

    # Use the provided crime count to further refine the risk category
    if crime_count < low_risk_threshold:
        risk_category = 'Low Risk'
    elif low_risk_threshold <= crime_count <= medium_risk_threshold:
        risk_category = 'Medium Risk'
    else:
        risk_category = 'High Risk'

    return risk_category

# Get latitude, longitude, and crime count input from the user
latitude_input = float(input("Enter latitude: "))
longitude_input = float(input("Enter longitude: "))
crime_count_input = int(input("Enter crime count: "))

# Call the function to predict risk category
predicted_risk_category = predict_risk_category(latitude_input, longitude_input, crime_count_input)

# Print the predicted risk category
print(f"Predicted Risk Category: {predicted_risk_category}")
```

CONVERTING INTO PICKLE FILE

```
import pickle
pickle.dump(rfc, open('model.pkl','wb'))
model = pickle.load(open('model.pkl','rb'))
```

FRONTEND IMPLEMENTATION

INDEX.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Crime Prediction System</title>
  <link rel="stylesheet" type="text/css" href="{{ url_for('static',
filename='css/styles.css') }}">
</head>
<body>
  <div class="banner">
    <div class="navbar">
      <img src="https://shirin-2486.github.io/Crime-prediction-system/logo.png"
class="logo">
      <ul>
        <li><a href="">Home</a></li>
        <li><a href="/risk">Predict</a></li>
        <li><a href="/crime">Crime</a></li>
      </ul>
    </div>
    <div class="content">
      <div class="centered">
```

```
    <h1>CRIME PREDICTION</h1>
    <p>Check your safety with us</p>
    </div>
    <div class="instruction">
      <h2>User Instructions</h2>
      <p>1.To Identify the probability of risk level in your area navigate to the
Predict page.</p>
      <p>  If you are unaware of latitude and longitude details,please check out
the specified map.</p>
      <p>2.To know about the criminal activity in your area navigate to Crime
page.</p>
      <p>  If you are unaware of the details , please check out the specified
map.</p>
    </div>
   <div>
  </div>
  <style>
   body {
   background-image: url("https://wallpapercave.com/wp/wp3007988.jpg");
   background-repeat: no-repeat;
   background-attachment: fixed;
   background-size: 100% 100%;


   }
  </style>
</body>
</html>
```

APP.PY

```
import numpy as np
import pandas as pd
from flask import Flask, request, jsonify, render_template
from sklearn.cluster import KMeans
import pickle

app = Flask(__name__, template_folder='templates', static_url_path='/static',
static_folder='static')

# Load the KMeans model and other necessary data
kmeans_model = pickle.load(open('kmeans_model.pkl', 'rb'))
model = pickle.load(open('model.pkl', 'rb'))

low_risk_threshold = 10.812758300436213
medium_risk_threshold = 21.499741699563785

@app.route('/')
def home():
    return render_template('index.html')

@app.route('/crime')
def crime():
    return render_template('crime.html')
@app.route('/predict_crime', methods=['POST'])
def predict_crime():
    '''
    For rendering results on HTML GUI
    '''
```

```
    float_features = [float(x) for x in request.form.values()]
    test_vector = np.reshape(np.asarray(float_features), (1, 7))
    p = np.array(model.predict(test_vector)[0])
    print(p)
    label = ['Robbery', 'Gambling', 'Accident', 'Violence', 'Kidnapping', 'Murder']
    if p[0] == 1:
        print("Robbery")
        pred = "Robbery"
    if p[1] == 1:
        print("Gambling")
        pred = "Gambling"
    if p[2] == 1:
        print("Accident")
        pred = "Accident"
    if p[3] == 1:
        print("Violence")
        pred = "Violence"
    if p[4] == 1:
        print("Kidnapping")
        pred = "Kidnapping"
    if p[5] == 1:
        pred = "Murder"
    output = pred

    return render_template('crime.html', prediction_text='Crime:
{}'.format(output))
```

```
@app.route('/risk')
def risk():
    return render_template('risk.html')

@app.route('/predict_risk', methods=['POST'])
def predict_risk():
    '''
    For rendering results on HTML GUI
    '''
    float_features = [float(x) for x in request.form.values()]

    # Assuming the order is latitude, longitude, crime count
    latitude, longitude, crime_count = float_features[:3]

    # Use the predict_risk_category function
    risk_category = predict_risk_category(latitude, longitude, crime_count)

    return render_template('risk.html', prediction_text='Risk Category:
{}'.format(risk_category))
def predict_risk_category(latitude, longitude, crime_count):
    # Predict the cluster for the given latitude and longitude
    prediction = kmeans_model.predict([[latitude, longitude]])[0]

    # Map the predicted cluster to the risk category
    risk_category_mapping = {0: 'Low Risk', 1: 'Medium Risk', 2: 'High Risk'}
    risk_category = risk_category_mapping.get(prediction, 'Unknown')

    # Refine the risk category based on crime count
```

```
    if crime_count < low_risk_threshold:
        risk_category = 'Low Risk'
    elif low_risk_threshold <= crime_count <= medium_risk_threshold:
        risk_category = 'The area you are looking forward is safe '
    else:
        risk_category = 'The area you are looking forward is quite risky,So You
Must Take The Required Safety Measures'

    return risk_category


if __name__ == "__main__":
    app.debug = True
    app.run(host='0.0.0.0', port=8000)
```

RISK.HTML

```
<!DOCTYPE html>
<html >
<head>
 <meta charset="UTF-8">
 <title>Risk Identification</title>
 <link rel="stylesheet" type="text/css" href="{{ url_for('static',
filename='css/styles.css') }}">
<link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet'
type='text/css'>
<link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet'
type='text/css'>
```

```
<link
href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300'
rel='stylesheet' type='text/css'>
</head>
<body>
   <div class="navbar">
    <img src="https://shirin-2486.github.io/Crime-prediction-system/logo.png"
class="logo">
    <ul>
      <li><a href="/">Home</a></li>
      <li><a href="/risk">Predict</a></li>
      <li><a href="/crime">Crime</a></li>
    </ul>
   </div>
   <div class="content">
    <h1>Risk Identification</h1>
    <p>Check your safety with us</p>
    <div>
     <center>
      <div class="login">

        <form action="{{ url_for('predict_risk')}}" method="post">
         <br>
         <input id="fuel" type="text" name="latitude" placeholder="latitude"
required="required" />
         <br>
         <br>
         <input id="fuel" type="text" name="longitude"
```

```
placeholder="longitude" required="required" />
                <br>
                <br>
                <input id="fuel" type="text" name="count" placeholder="count"
required="required" />
                <br>
                <br>
               <button id = 'sub' type="submit" class="btn btn-primary btn-block btn-
large" >Predict </button>
               </form>
               <br>
               <br>
               <br><br><h3 style="color:#040324;">{{ prediction_text }}<h3>
               <p>      </p>
               <h4>Crime map</h4>
               <iframe src="https://shirin-2486.github.io/Crime-prediction-
system/map1.html" width="75%" height="600" frameborder="0"
style="border:0"></iframe>
          </center>

    <style>
    body {
    background-image:linear-gradient(rgba(19, 197, 206, 0.5), rgba(19, 209, 180,
0.5)),
                      url("https://blog.ccasociety.com/wpcontent/uploads/2020/04/cyber
-crime-image.jpg");
    background-repeat: no-repeat;
    background-attachment: fixed;
```

```
background-size: 100% 100%;
}
#research {
    font-size: 18px;
    width: 210px;
    height: 23px;
    top: 23px;
  }
  #box {
    border-radius: 60px;
    border-color: 45px;
    border-style: solid;
    font-family: cursive;
    text-align: center;
    background-color: rgb(168, 131, 61);
    font-size: medium;
    position: absolute;
    width: 700px;
    bottom: 9%;
    height: 850px;
    right: 30%;
    padding: 0px;
    margin: 0px;
    font-size: 14px;
  }
  #fuel {
    width: 210px;
    height: 43px;
```

```
    text-align: center;
    border-radius: 14px;
    font-size: 20px;
    padding-left: 10px;
    padding-right: 10px;
}
#fuel:hover {
    background-color: coral;
}
#research {
    width: 210px;
    height: 43px;
    text-align: center;
    border-radius: 14px;
    font-size: 18px;
}
#research:hover {
    background-color: coral;
}
#resea {
    width: 210px;
    height: 43px;
    text-align: center;
    border-radius: 14px;
    font-size: 18px;
}
#resea:hover {
    background-color: coral;
```

```
        }
        #sub {
            width: 120px;
            height: 43px;
            text-align: center;
            border-radius: 14px;
            font-size: 18px;
            background-color: brown;
        }
        #sub:hover {
            background-color: darkcyan;
        }
        h3,h4,h5,h6,h7,h8, h9 {
            font-size: 3em;
            text-align: center;
            color: #050433;
            font-weight: 100;
            text-transform: capitalize;
            letter-spacing: 4px;
            font-family: 'Roboto', sans-serif;
        }
    </style>
</body>
</html>
```

CRIME.HTML

```
<!DOCTYPE html>
<html>
<head>
```

```
  <title>Crime Prediction System</title>
  <link rel="stylesheet" type="text/css" href="{{ url_for('static',
filename='css/styles.css') }}">
</head>
<body>
    <div class="navbar">
      <img src="https://shirin-2486.github.io/Crime-prediction-system/logo.png"
class="logo">
      <ul>
        <li><a href="/">Home</a></li>
        <li><a href="/risk">Predict</a></li>
        <li><a href="/crime">Crime</a></li>
      </ul>
    </div>
    <div class="content">
      <h1>Crime Identification</h1>
      <p >Check your safety with us</p>
      <div>
    <center>
      <div class="login">
        <form action="{{ url_for('predict_crime')}}"method="post">
          <br>
          <input id="first" type="text" name="month" placeholder="month"
required="required" />
          <br>
          <br>
          <input id="second" type="text" name="day" placeholder="day"
required="required" />
```

```
        <br>
        <br>
        <input id="third" type="text" name="hour" placeholder="hour"
required="required" />
        <br>
        <br>
        <input id="fourth" type="text" name="dayofyear"
placeholder="dayofyear" required="required" />
        <br>
        <br>
        <input id="fuel" type="text" name="weekofyear"
placeholder="weekofyear" required="required" />
        <br>
        <br>
        <input id="resea" type="text" name="latitude" placeholder="latitude"
required="required" />
        <br>
        <br>
        <input id="research" type="text" name="longitude"
placeholder="longitude" required="required" />
        <br>
        <br>
        <button id = 'sub' type="submit" class="btn btn-primary btn-block btn-
large" >Predict </button>
      </form>
      <h3 style="color:rgb(248, 247, 254);">{{ prediction_text }}<h3>
            <h4>Crime map</h4>
            <iframe src="https://shirin-2486.github.io/Crime-prediction-
```

```
system/map2.html" width="75%" height="600" frameborder="0"
style="border:0"></iframe>
      </div>
    </center>
  <style>
     body {
      background-image:linear-gradient(rgba(13, 13, 13, 0.5), rgba(9, 9, 9, 0.5)),
                url("https://cyfor.co.uk/wp-content/uploads/2014/11/cyfor-
123.jpg");
    background-repeat: no-repeat;
    background-attachment: fixed;
    background-size: 100% 100%;
    }
     #research {
              font-size: 18px;
              width: 210px;
              height: 23px;
              top: 23px;
            }
            #box {
              border-radius: 60px;
              border-color: 45px;
              border-style: solid;
              font-family: cursive;
              text-align: center;
              background-color: rgb(168, 131, 61);
              font-size: medium;
              position: absolute;
```

```
    width: 700px;
    bottom: 9%;
    height: 850px;
    right: 30%;
    padding: 0px;
    margin: 0px;
    font-size: 14px;
}
#fuel {
    width: 210px;
    height: 43px;
    text-align: center;
    border-radius: 14px;
    font-size: 20px;
    padding-left: 10px;
    padding-right: 10px;
}
#fuel:hover {
    background-color: coral;
}
#research {
    width: 210px;
    height: 43px;
    text-align: center;
    border-radius: 14px;
    font-size: 18px;
}
```

```
#research:hover {
   background-color: coral;
}
#resea {
   width: 210px;
   height: 43px;
   text-align: center;
   border-radius: 14px;
   font-size: 18px;
}
#resea:hover {
   background-color: coral;
}
#sub {
   width: 120px;
   height: 43px;
   text-align: center;
   border-radius: 14px;
   font-size: 18px;
   background-color: brown;
}
#sub:hover {
   background-color: darkcyan;
}
#first {
   border-radius: 14px;
   height: 25px;
   font-size: 20px;
```

```
            text-align: center;
        }
        #second {
            border-radius: 14px;
            height: 25px;
            font-size: 20px;
            text-align: center;
        }
        #third {
            border-radius: 14px;
            height: 25px;
            font-size: 20px;
            text-align: center;
        }
        #fourth {
            border-radius: 14px;
            height: 25px;
            font-size: 20px;
            text-align: center;
        }
  </style>
</body>
</html>
```

ANALYSE.HTML

```
<!DOCTYPE html>
<html>
<head>
```

```html
  <title>Crime Prediction System</title>
  <link rel="stylesheet" type="text/css" href="{{ url_for('static',
filename='css/styles.css') }}">
</head>
<body>
    <div class="navbar">
      <img src="https://shirin-2486.github.io/Crime-prediction-system/logo.png"
class="logo">
      <ul>
        <li><a href="/">Home</a></li>
        <li><a href="/risk">Predict</a></li>
        <li><a href="/crime">Crime</a></li>
        <li><a href="/anlyse">Analysis</a></li>
      </ul>
    </div>
    <div class="content">
        <h1>Crime Analysis</h1>
        <p>  </p>
        <p>  </p>
        <center>
        <p class="text">Crime occurrence by type</p>
        <img
          src="https://shirin-2486.github.io/Crime-prediction-system/crime 1.png"
class="image">
        <p>   </p>
        <div class="overlay">
          </div>
          <p class="text">Month based Total no of crimes</p>
```

```
<img
  src="https://shirin-2486.github.io/Crime-prediction-system/crime 5.png"
class="image">
<p>   </p>
<div class="overlay">
  </div>
  <p class="text">Crime type occurrences over month</p>
<img
  src="https://shirin-2486.github.io/Crime-prediction-system/crime 3.png"
class="image">
  <div class="overlay">
  </div>
</center>
<style>
    body {
    background-image:linear-gradient(rgba(22, 21, 21, 0.5), rgba(9, 9, 9,
0.5)), url("https://wallpapercave.com/wp/wp2581376.jpg" );
    background-repeat: no-repeat;
    background-attachment: fixed;
    background-size: 100% 100%;

    }
    .content p{
        text-align: center;
        font-size: 18px;
        font-family: sans-serif;
        text-transform: none;
    }
```

```
.content img {
  border-radius: 8px;
  display: block;
  margin-left: auto;
  margin-right: auto;
  width: 50%;
}

        </style>
</body>
</html>
```

STYLE.CSS

```
@import
url("https://fonts.googleapis.com/css2?family=Open+Sans:wght@200;300;400;
500;600;700&display=swap");

* {
    margin: 0;
    padding: 0;
    box-sizing: border-box;
    font-family: sans-serif;
}
.instruction{
    width:700px;
    height:320px;
    background:linear-gradient(rgba(0, 0, 0, 0.75), rgba(0, 0, 0, 0.75)50%);
    position:absolute;
    left: 400px;
```

```
    bottom: -350px;
    text-align: center;
    padding: 10px;
    border: radius 10px;
}
.instruction p{
    text-align: center;
    font-size: 18px;
    font-family: sans-serif;
    text-transform: none;
}
.login h3{
    font-color:white;
    font-size: 50px;
    font-weight:500px ;
    margin: 0 20px;
}
html {
    scroll-behavior: smooth;
}
.banner {
    height: 100vh;
    background-image: linear-gradient(rgba(0, 0, 0, 0.75), rgba(0, 0, 0, 0.75)),
url(pic.jpg);
    background-size: cover;
    background-position: center;
}
.navbar {
```

```
    width: 85%;
    margin: auto;
    padding: 35px 0;
    display: flex;
    align-items: center;
    justify-content: space-between;
}
.logo {
    width: 120px;
    cursor: pointer;
}
.navbar ul li {
    list-style: none;
    display: inline-block;
    margin: 0 20px;
    position: relative;
}
.navbar ul li a {
    text-decoration: none;
    color: #fff;
    text-transform: uppercase;
}
.navbar ul li::after {
    content: '';
    height: 3px;
    width: 100%;
    background: #009688;
    position: absolute;
```

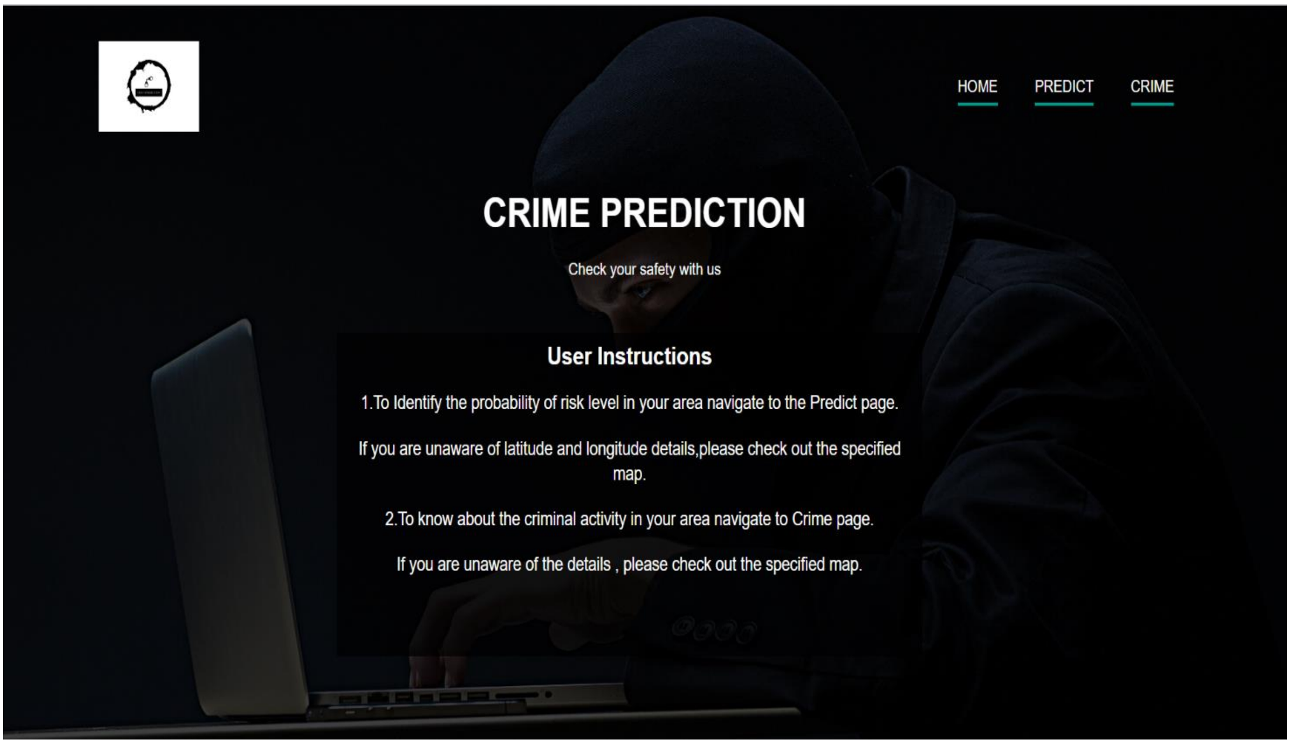
```
    left: 0;
    bottom: -10px;
}
.content {
    width: 100%;
    position: absolute;
    top: 25%;
    transform: translateY(-50);
    text-align: center;
    color: #fff;
}
.content h1 {
    font-size: 40px;
    margin: 0 20px;
}

.content p {
    margin: 20px auto;
    font-weight: 100;
    line-height: 25px;
}
button {
    width: 200px;
    padding: 15px 0;
    text-align: center;
    margin: 20px 10px;
    font-weight: bold;
    text-transform: uppercase;
```

```
    border-radius: 25px;
    border: 2px solid #009688;
    color: white;
    background-color: #009688;
    cursor: pointer;
    position: relative;
    overflow: hidden;
}
span {
    background: #009688;
    height: 100%;
    width: 100%;
    border-radius: 25px;
    position: absolute;
    left: 0;
    bottom: 0;
    z-index: -1;
    transition: 0.5s;
}
button:hover span {
    width: 100%;
}
button:hover {
    border: none;
}
```

A.3 SAMPLE SCREENSHOT

HOME PAGE

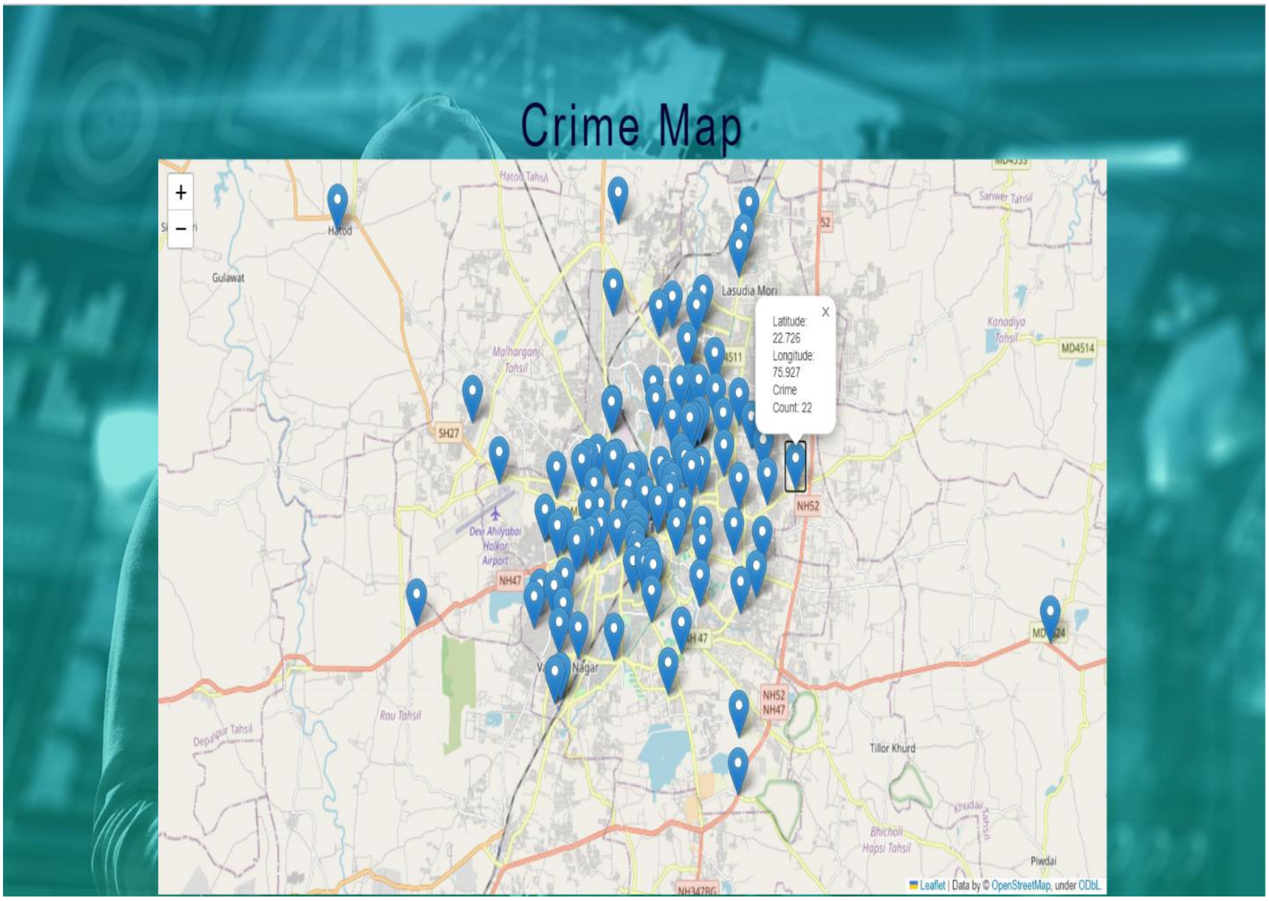


A.3.1 HOME PAGE

PREDICT PAGE

Before entering values

A.3.2 PREDICT PAGE



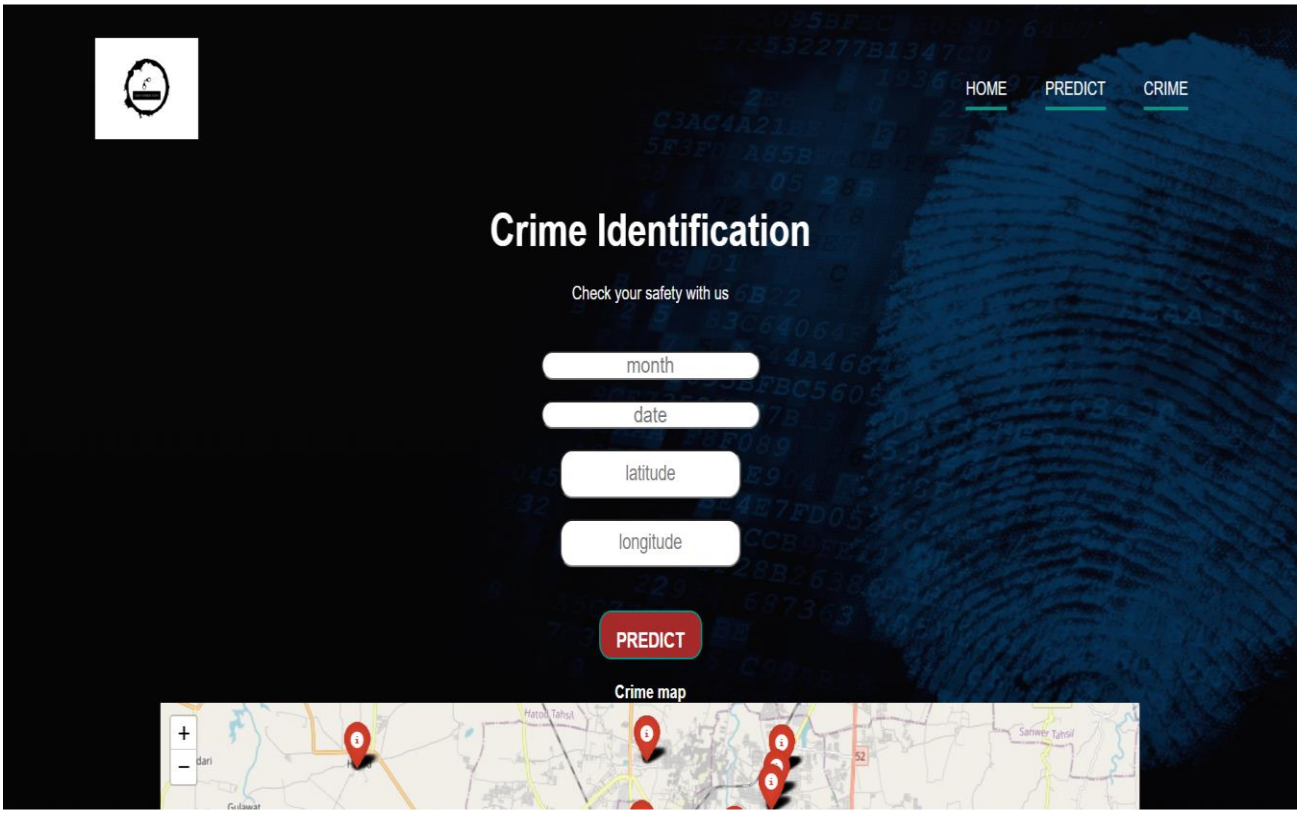
A.3.3 MAP WITH DETAILS

After entering values

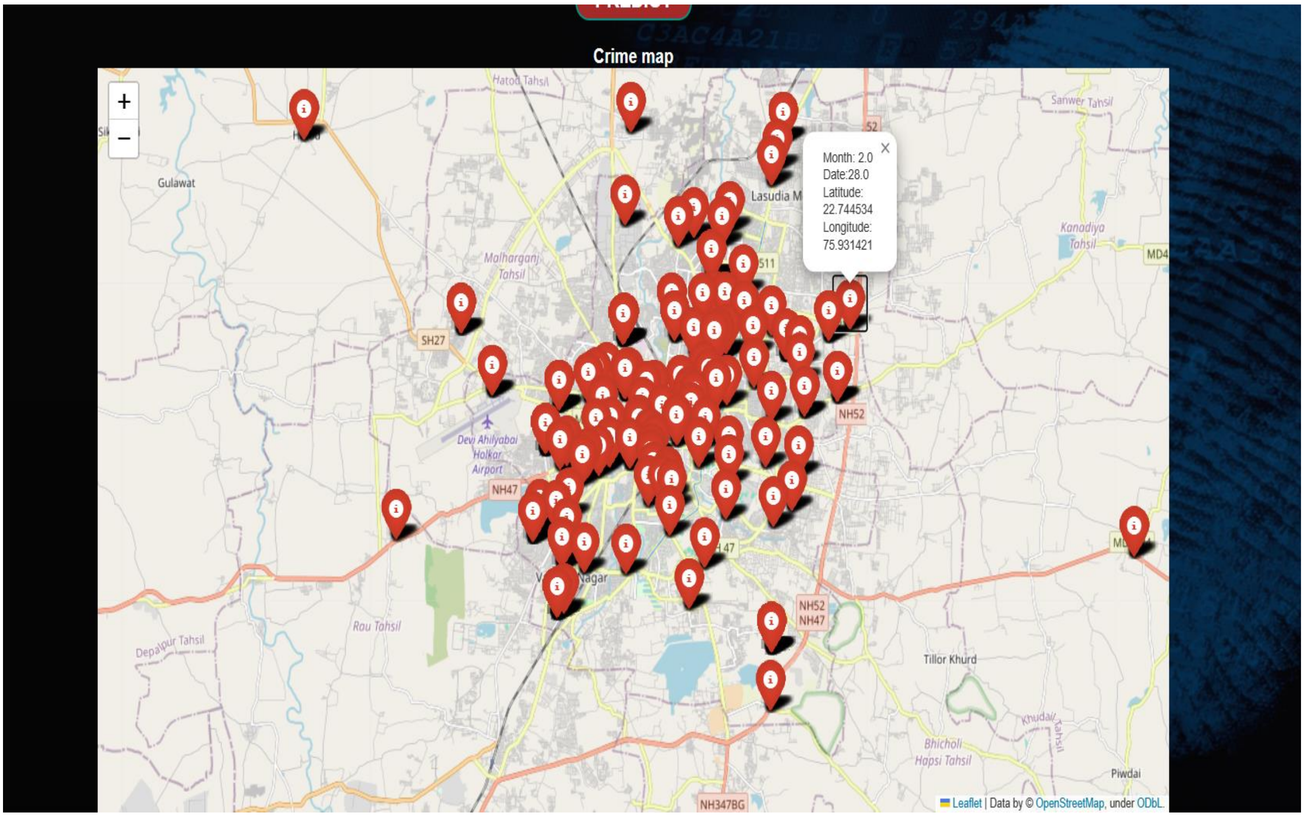
A.3.4 RESULT

CRIME PAGE

Before entering input values



A.3.5 CRIME PAGE



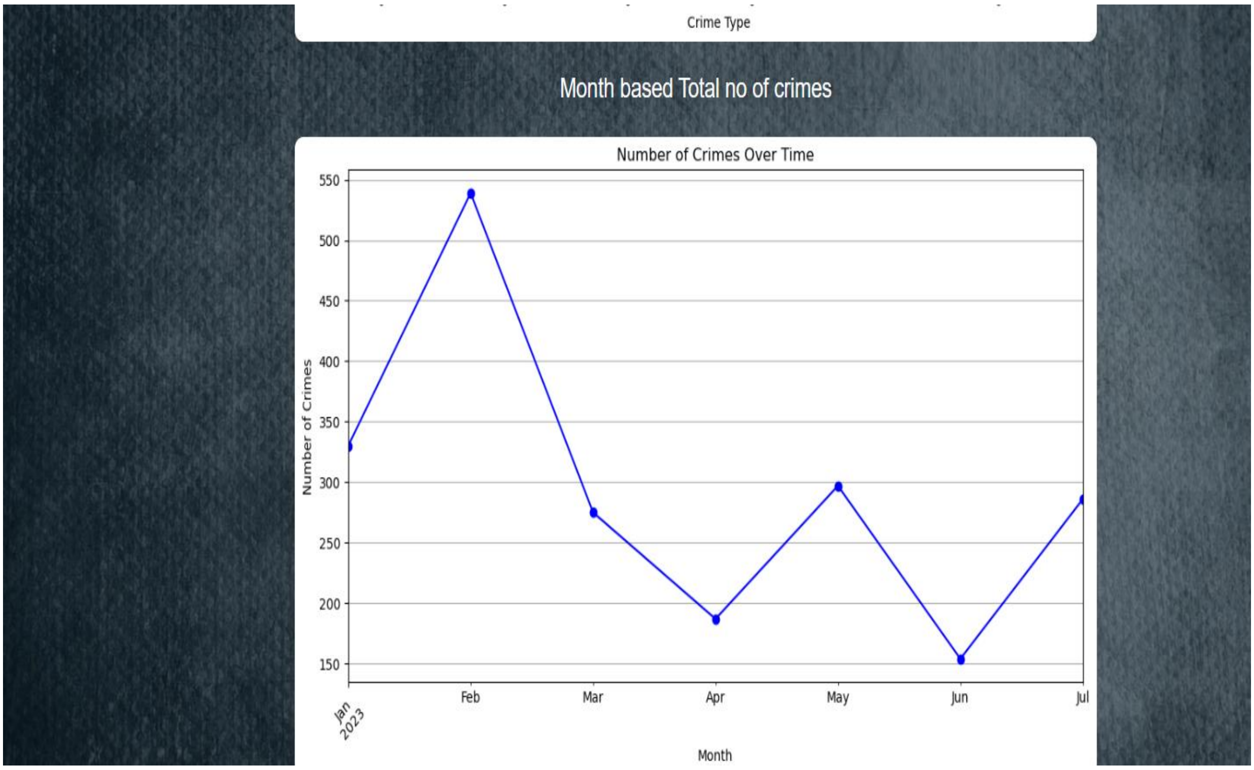
A.3.6 MAP WITH DETAILS

After entering values

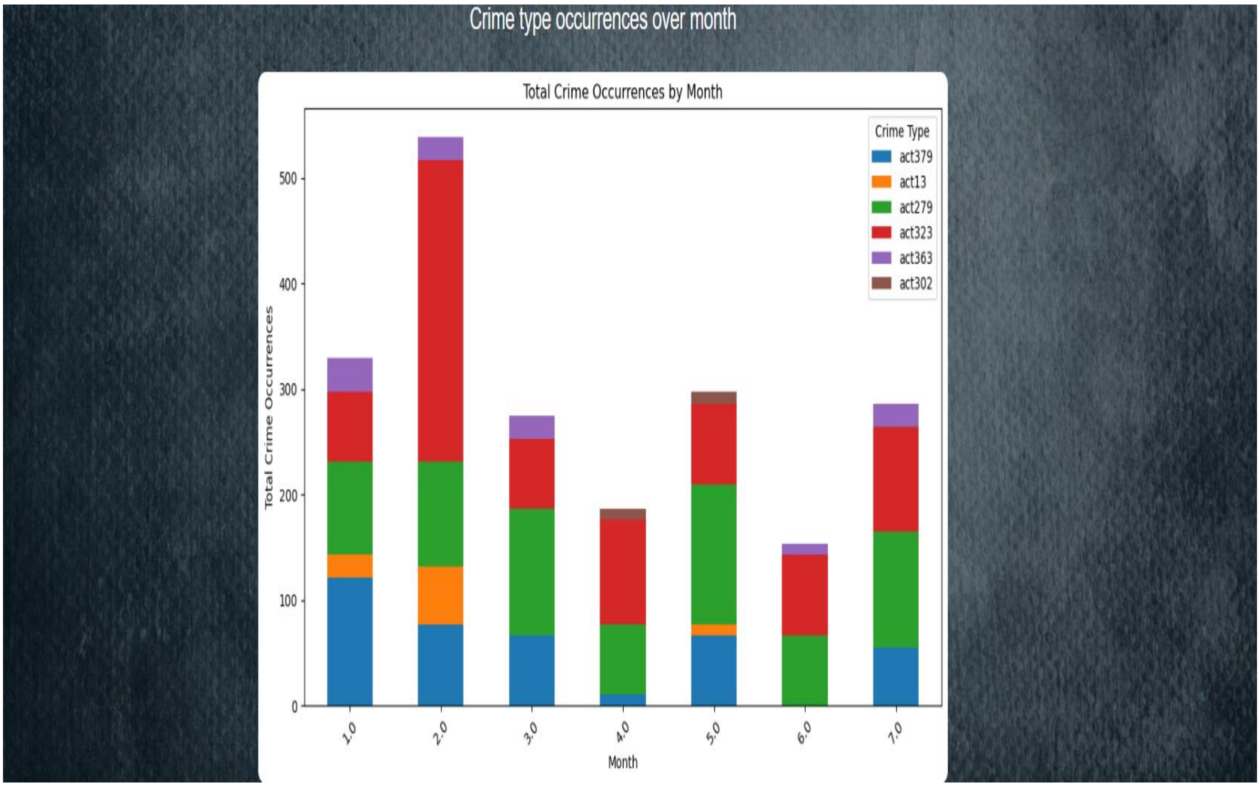
A.3.7 RESULT

ANALYSIS PAGE

A.3.8 Distribution of crime occurrences by type



A.3.9 Number of Crimes over time



A.3.10 Crime type occurrences over month

A.4 PLAIGARISM REPORT

REFERENCES

[1] Nasridinov A, Park YH. A Study on Performance Evaluation of Machine Learning Algorithms for 76 Crime Dataset. ACM (2014).

[2] N. Mahmud, K. I. Zinnah, Y. A. Rahman and N. Ahmed, "Crimecast: A crime prediction and strategy direction service," 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2016, pp. 414-418, doi: 10.1109/ICCITECHN.2016.7860234.

[3] Ahishakiye, A., Taremwa, and D., Omulo, E.O., 2017. A Performance Analysis of Business Intelligence Techniques on Crime Prediction. International Journal Computer and Informatio Technology, 6(03).

[4] A.Anisha., C.Dhanashree, and A.Arpita ,. Application for analysis and prediction of crime data using data mining. International journal of advanced computational engineering and networking, 2017.

[5] Varshitha D N, Vidyashree K P, Aishwarya P, Janya T S, K R Dhananjay Gupta, Sahana R, 2017, Paper on Different Approaches for Crime Prediction system, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCETEIT – 2017 (Volume 5 – Issue 20).

[6] A. Almehmadi, Z. Joudaki, and R. Jalali, ‘‘Language usage on Twitter predicts crime rates,’’ in Proc. 10th Int. Conf. Secur. Inf. Netw., Oct. 2017, pp. 307–310, doi: 10.1145/3136825.3136854.

[7] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, ‘‘Crime pattern detection, analysis & prediction,’’ in Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Apr. 2017, pp. 225–230, doi: 10.1109/ICECA.2017.8203676.

[8] Ayisheshim Almaw , Kalyani Kadam . Survey Paper on Crime Prediction using Ensemble Approach. International Journal of Pure and Applied Mathematics ,Volume 118 No. 8 2018, 133-139.

[9] Z. Li, T. Zhang, Z. Yuan, Z. Wu, and Z. Du, ‘‘Spatio-temporal pattern analysis and prediction for urban crime,’’ in Proc. 6th Int. Conf. Adv. Cloud

Big Data (CBD), Aug. 2018, pp. 177–182, doi: 10.1109/CBD.2018.00040.
[10] Pratibha Kumari ,Akanksha Gahalot ,Uprant ,Suraina Dhiman ,Lokesh Chouhan. Crime Prediction and Analysis . Conference: 2020 2nd International Conference on Data, Engineering and Applications (IDEA) .February 2020 ,DOI:10.1109/IDEA49133.2020.9170731.
[11] Shah, N., Bhagat, N. & Shah, M. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Vis. Comput. Ind. Biomed. Art 4, 9 (2021). https://doi.org/10.1186/s42492-021-00075-z.
[12] Sapna Singh Kshatri, Deepak Singh, Bhavana Narain, Surbhi Bhatia, Mohammad Tabrez Quasim, and Ganesh Ram Sinha. An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach. IEEE Access, 9:67488–67500, 2021.
[13] S. Singh, A. Gole, P. Kanherkar, P. R. Abhishek and P. Wankhede, "Crime Prediction using Machine Learning," *2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, Nagpur, India, 2022, pp. 179-185, doi: 10.1109/ICETEMS56252.2022.10093303.
[14] V. Mandalapu, L. Elluri, P. Vyas and N. Roy, "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions," in IEEE Access, vol. 11, pp. 60153- 60170, 2023, doi: 10.1109/ACCESS.2023.3286344.
[15]S. Khatun, K. Banoth, A. Dilli, S. Kakarlapudi, S. V. Karrola and G. C. Babu, "Machine Learning based Advanced Crime Prediction and Analysis," *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Erode, India, 2023, pp. 90-96, doi: 10.1109/ICSCDS56580.2023.10104655.