

**FUZZY BASED GEO-SPATIAL CRIME CATEGORY PREDICTION FOR CRIME
MAPPING AND SAFE ROUTE TRAVEL**

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

By

Prabhu S (Reg. No - 40110973)

**Praveen Karthik T (Reg. No -
40110993)**



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Category-1 University by UGC
Accredited with Grade "A++" by NAAC | 12B Status by UGC | Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHISALAI, CHENNAI – 600119

APRIL - 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Category - I University by UGC

Accredited "A++" by NAAC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Praveen Karthik T (40110993)** and **Prabhu S (40110973)** who carried out the Project entitled "**Fuzzy Based Geo-Spatial Crime Category for Crime Mapping and Safe Route Travel**" under my supervision from November 2023 to April 2024.

Internal Guide

Dr.G.Kalaiarasi M.E., Ph.D.,

Head of the Department

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Jeppiaar Nagar, Rajiv Gandhi Salai,
Chennai – 600 119.

Submitted for Viva voce Examination held on 16/04/2024

Internal Examiner

External Examiner

DECLARATION

I, Praveen Karthik T (Reg. No- 40110993) hereby declare that the Project Report entitled "Fuzzy Based Geo-Spatial Crime Category for Crime Mapping and Safe Route Travel" done by me under the guidance of Dr.G.Kalalarasi M.E., Ph.D., is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE: 16/04/2024

PLACE: Chennai



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of Sathyabama Institute of Science and Technology** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.T . Sasikala M.E., Ph.D, Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.**, Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.G.Kalaiarasi M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

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. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. 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. Machine learning algorithms such as Random Forest Classifier is implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to precomposed works. The major focus is outlined, which is to estimate the type of crime that contributes the most to crime occurrences. Additionally, the project aims to analyze the time period and location where these crimes tend to happen. This indicates a comprehensive approach to understanding crime patterns, considering both the type and spatial-temporal aspects of crime. It mentions the use of machine learning algorithms, specifically the Random Forest Classifier, in the project. Random Forest Classifier is a robust algorithm known for its ability to handle complex datasets and provide accurate predictions. Its selection suggests a data- driven approach to analyzing crime data and predicting crime occurrences. The accuracy is achieved by the Random Forest Classifier in classifying crime patterns is comparatively high when compared to previous works or existing literature. This indicates the effectiveness of the methodology employed in the study and highlights its potential contribution to the field of crime analysis. Machine learning algorithms, like the Random Forest Classifier, play a crucial role in this research by providing a data- driven approach to understanding crime patterns. The Random Forest Classifier is known for its robustness in handling complex datasets and delivering accurate predictions.

TABLE OF CONTENTS

Chapter No	TITLE	Page No.
	ABSTRACT	v
	LIST OF FIGURES	vii
1	INTRODUCTION	1
2	LITERATURE SURVEY	3
	2.1 Inferences from Literature Survey	5
	2.2 Open problems in Existing System	5
3	REQUIREMENTS ANALYSIS	6
	3.1 Feasibility Studies/Risk Analysis of the Project	6
	3.2 Software Requirements Specification Document	7
	3.2.1 Software Requirements	8
	3.2.2 Hardware Requirements	9
4	DESCRIPTION OF PROPOSED SYSTEM	10
	4.1 Selected Methodology or process model	10
	4.2 Flow diagram of Proposed System	12
	4.3 Architecture and Overall Design of proposed system	14
	4.4 Module used	14
	4.4.1 Dataset	14
	4.4.2 Data Gathering Module	14
	4.4.3 Data preprocessing Module	15
	4.4.4 Feature Selection Module	15
	4.4.5 Building a training data	15
	4.4.6 Prediction Module	16
	4.4.7 Visualization Module	16
	4.5 Data flow Diagram	17

	4.6	UML Diagram	18
	4.7	Use-case Diagram	19
	4.8	Class Diagram	20
	4.9	Activity Diagram	21
	4.10	Sequence Diagram	22
	4.11	Description of Software for Implementation and testing plan of the proposed model/System	23
	4.12	Software Description	24
	4.13	Input and output Design	26
	4.14	Project management	27
	4.15	Software to operations plan	30
5		IMPLEMENTATION DETAILS	33
	5.1	Development and deployment setup	33
	5.2	Algorithms	35
	5.2.1	Random Forest algorithm	35
	5.2.2	Decision Tree	37
	5.3	Testing	39
6		RESULTS AND DISCUSSIONS	41
	6.1	Data analyzing and prediction	43
7		CONCLUSION	46
		REFERENCES	47
		APPENDIX	48
		A.SOURCE CODE	
		B.SCREENSHOTS	50
		C.RESEARCH PAPER	55

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
4.2.1	Flow diagram	12
4.3.1	Architecture diagram	14
4.5.1	Data flow diagram	18
4.7.1	Use case Diagram	20
4.8.1	Class Diagram	21
4.9.1	Activity Diagram	22
4.10.1	Sequence Diagram	23
6.1.1	Loads the Data	43
6.1.2	The data to be trained	44
6.1.3	Creating & Training KNN Model	44
6.1.4	Creating & Training Decision tree Model	45
6.1.5	Creating & Training Random tree Model	45
B.1	Crime analysis and prediction application of homepage	50
B.2	login page for predicting the crime	50
B.3	Upload page of crime prediction	51
B.4	Split into training and testing sets	51
B.5	Web page of filling the necessary detail	52
B.6	prediction of type of crime	52
B.7	Performance analysis for the type of crime	53

B.8	Performance of crime occurrences with particular day of week	53
B.9	Graph for percentage of crime occurrences with particular day of week	54
B.10	Confusion Matrix for crime occurrences	54

CHAPTER 1

INTRODUCTION

Day by day crime data rate is increasing because the modern technologies and hi-tech methods are helping the criminals to achieve the illegal activities. According to Crime Record Bureau crimes like burglary, arson etc... have been increased while crimes like murder, sex, abuse, gang rap etc. have been increased. Crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the machine learning techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime. Machine learning helps in solving the crimes faster and this technique gives good results when applied on crime dataset, the information obtained from the machine learning techniques can help the police department. A particular approach has been found to be useful by the police, which is the identification of crime 'hot spots' which indicates areas with a high concentration of crime. Use of machine learning techniques can produce important results from crime report datasets. The very step in study of crime is crime analysis.

Crime analysis is exploring, inter relating and detecting relationship between the various crimes and characteristics of the crime. This analysis helps in preparing statistics, queries and maps on demand. It also helps to see if a crime in a certain known pattern or a new pattern necessary. Crimes can be predicted as the criminals are active and operate in their comfort zones. Once successful they try to replicate the crime under similar circumstances.

Crime mapping and route planning play crucial roles in urban planning, law enforcement, and public safety strategies. By accurately predicting crime categories and identifying high-risk areas, authorities can allocate resources effectively and implement targeted interventions to prevent crime. Additionally, providing travelers with real-time information about safe routes can empower individuals to make informed decisions and avoid potentially dangerous areas, thereby enhancing personal safety and reducing the risk of victimization.

The occurrences of crime depended on several factors such as intelligence of criminals, security of a location etc. The work has followed the steps that used in data analysis, in which the important phases are Data collection, data classification, pattern identification, prediction and visualization.

The proposed framework uses different visualization techniques to show the trends of crimes and various ways that can predicts the crime using machine learning algorithm. The inputs to our algorithms are time (hour, day, month, and year), place (latitude and longitude), and class of crime: Robbery, Rape, Theft, Murder, Kidnapping

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as Random Forests. We also perform multiple classification tasks – we first try to predict which of 5 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

The primary objective of this project is to develop a comprehensive Fuzzy-Based Geo- Spatial Crime Category Prediction system that:

- Utilizes fuzzy logic modelling to analyse crime data and predict crime categories based on geographical factors, temporal trends, and other relevant variables.
- Integrates machine learning algorithms, such as decision trees and random forests, to enhance prediction accuracy and robustness.
- Incorporates geospatial analysis techniques to visualize crime data, identify crime hotspots, and recommend safe travel routes for pedestrians, motorists, and public transit users.
- Provides users with an intuitive interface for accessing crime mapping tools, route planning functionalities, and real-time crime alerts, thereby empowering individuals and communities to proactively address safety concerns.

CHAPTER 2

LITERATURE SURVEY

This problem statement has been extensively studied over the past 5 years by researchers and automotive companies in a bid to create a solution, and all their solutions vary from analysing various patterns of crime

W. Safat, S. Asghar, and S. A. Gillani, “Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques [1] Crime and violation are the threat to justice and meant to be controlled. Accurate crime prediction and future forecasting trends can assist to enhance metropolitan safety computationally. The limited ability of humans to process complex information from big data hinders the early and accurate prediction and forecasting of crime. The accurate estimation of the crime rate, types and hot spots from past patterns creates many computational challenges and opportunities. Despite considerable research efforts, yet there is a need to have a better predictive algorithm, which direct police patrols toward criminal activities.

R. Kumar and B. Nagpal, “Analysis and prediction of crime patterns using big data, et al [2] They explored the application of advanced data mining techniques in the domain of crime analysis and prediction. Leveraging a comprehensive dataset, our study employs machine learning algorithms to uncover patterns and trends within criminal activities. The integration of predictive modeling allows for the anticipation of potential crime hotspots, aiding law enforcement in proactive measures. The findings highlight the efficacy of data-driven approaches in enhancing the accuracy and efficiency of crime analysis systems, contributing to the ongoing efforts to ensure public safety.

Khushabu A.Bokde,Tisksha P.Kakade et al,[3] They are mainly focusing on crime Analysis, Clustering and Clustering by K-means algorithms methods. some of the purpose of crime analysis are Extraction of crime patterns by crime analysis and based on available criminal information, crime recognition. Clustering means Division of a set of data or objects to a number of clusters. There by cluster is composed of a set of similar data which have same as a group. K-means is the simplest and most

commonly

sued portioning algorithm among the clustering algorithms in scientific and industrial software.

Benjamin Fredrick David. H and Suruliand, "Survey on crime analysis and prediction using data mining techniques" [4] Criminology is process that is used for identify the crime characteristics and crime identify. By using this data mining algorithms will be able to produce crime reports and help in the identification of criminal much faster than any human could. The criminals when leaving the crime scene does leave some trace which can be used as clue to identify the criminals. This process is used to identifying the criminals based on the clues or information given by the local peoples in this paper using some criminal analysis methods they are Text, content and NLP(Natural language processing) based methods, Crime patterns.

Ginger saltos and minhaela Cocea [5] Crime data has been systematically recorded by the police for many years and in the last decades, there has been a surge of Open Crime Data and of apps or web based application displaying crime statistics on maps, both by official sources, such as from police UK, and other sources using the same official data. In this paper investigate predict many types of crime and discuss their applicability.

Tushar Sonawanev, Shirin Shaikh, Shaista ,asif sayyad [6] Crime Pattern Analysis, Visualizations And Prediction Using Data Mining Crime analysis is exploring, inter relating and detecting relationship between various crimes and characteristics of crimes. The main aim of this research paper consists of developing analytical data mining methods that can systematically address the complex problems related to various form of crime. From this literacy study, it could be concluded that crime details increasingly to very large quantities running into zota bytes.

Shiju Sathyadevan, Devan M.S, Surya Gangadharan, [7] Crime Analysis and Prediction Using Data Mining According to the Crime record Bureau crimes like burglary, arson, etc have been decreased while crimes like murder, sex, abuse, gang rape etc have been increased. The data's are collected from various sources like websites, blogs, news sites, social media, RSS feeds etc. This huge data is used as a record for creating a crime record database. The crime analysis using five steps they are Data Collection, Classification, Pattern identification, prediction, and

Visualization.

2.1 INFERENCES FROM LITERATURE SURVEY

In the explored research papers and project description the common theme revolves around the prediction for crime mapping and safe route travel. These crime predictions leverage artificial intelligence, machine learning and natural language processing to facilitate interaction in various incidence such as Robbery, theft, rape, kidnapping, murder

In pre-work, the dataset obtained from the open source are first pre-processed to remove the duplicated values and features. 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. Based on the prediction, the performance is calculated for both trained and test values. The crime prediction helps in forecasting the future happening of any type of criminal activities and help the officials to resolve them at the earliest.

2.2 OPEN PROBLEMS IN EXISTING SYSTEM

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

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 FEASIBILITY STUDIES / RISK ANALYSIS

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are: Economic Feasibility, Technical Feasibility, Social Feasibility

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to

educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analyzing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

3.2 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

- **Usability**

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

- **Robustness**

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

- **Security**

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

- **Reliability**

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

- **Compatibility**

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

- **Flexibility**

The flexibility of the project is provided in such a way that it has the ability to run on different environments being executed by different users.

- **Safety**

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one's personal information.

3.2.1 SOFTWARE REQUIREMENTS

- **Operating System:** The system should be compatible with popular operating systems, such as Windows, macOS, and Linux. Ensure compatibility with mobile operating systems (iOS, Android) for mobile application
- **Programming Language:** Choose a programming language suitable for both backend and frontend development. Popular choices include Python (Django, Flask), JavaScript (Node.js), or Java.
- **Fuzzy Logic Library:** Implement a library or framework for fuzzy logic processing. Fuzzy Logic Toolbox in MATLAB, scikit-fuzzy in Python, or other fuzzy logic libraries can be used.

3.2.2 HARDWARE REQUIREMENTS

- **Operating System:** Your development environment can run on Windows (later than Windows 7), macOS, or Linux. Flask provides cross-platform support, ensuring that your app can be developed on various operating systems.
- **Disk Space:** To accommodate all the software requirements and project files, you'll need a hard disk with a minimum of 400MB of free space. Ensure you have adequate storage for both your development tools and the application's assets.
- **macOS Specific Requirements:** If you're using macOS, it's essential that your system is 64-bit compatible. Additionally, you should have a sufficient amount of available space, ideally at least 2.8 GB.

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

4.1 SELECTED METHODOLOGY OR PROCESS MODEL

An Exploration of Crime Prediction Using Data Mining on Open Data

The increase in crime data recording coupled with data analytics resulted in the growth of research approaches aimed at extracting knowledge from crime records to better understand criminal behavior and ultimately prevent future crimes. While many of these approaches make use of clustering and association rule mining techniques, there are fewer approaches focusing on predictive models of crime. In this paper, we explore models for predicting the frequency of several types of crimes by LSOA code (Lower Layer Super Output Areas — an administrative system of areas used by the UK police) and the frequency of anti-social behavior crimes. Three algorithms are used from different categories of approaches: instance-based learning, regression and decision trees. The data are from the UK police and contain over 600,000 records before preprocessing. The results, looking at predictive performance as well as processing time, indicate that decision trees (M5P algorithm) can be used to reliably predict crime frequency in general as well as anti-social behavior frequency.

Crime Analysis and Prediction Using Data Mining

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. This system can predict regions which have high probability for crime occurrence and can visualize crime prone areas with the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like criminal background of offender,

political enmity etc. we are focusing mainly on crime factors of each day.

Crime Detection Techniques Using data Mining and K-Means

Crimes will somehow influence organizations and institutions when occurred frequently in a society. Thus, it seems necessary to study reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes. The main objective of this paper is to classify clustered crimes based on occurrence frequency during different years. Data mining is used extensively in terms of analysis, investigation and discovery of patterns for occurrence of different crimes. A theoretical model based on data mining techniques such as clustering and classification to real crime dataset recorded by police in England and Wales within 1990 to 2011. Thus the weights are assigned to the features in order to improve the quality of the model and remove low value of them. The Genetic Algorithm (GA) is used for optimizing of Outlier Detection operator parameters using RapidMiner tool.

Crime Pattern Analysis, Visualizations and Prediction Using Data Mining

Crime against women these days has become problem of every nation around the globe many countries are trying to curb this problem. Preventive are taken to reduce the increasing number of cases of crime against women. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analyzing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department. Study of crime data can help us analyze crime pattern, inter-related clues & important hidden relations between the crimes. That is why data mining can be great aid to analyze, visualize and predict crime using crime data set. Classification and correlation of data set makes it easy to understand similarities & dissimilarities amongst the data objects. We group data objects using clustering technique. Dataset is classified on the basis of some predefined condition. Here grouping is done according to various types of crimes against women taking place in different states and cities of India. Crime mapping will help the administration to plan strategies for prevention of crime, further using data mining technique data can be predicted and visualized in various

form in order to

provide better understanding of crime patterns.

Systematic Review of Crime Data Mining

Crime analysis is a methodical approach for identifying and analyzing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, we can analyze previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, we will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. In this paper, the main focus is on the review of algorithms and techniques used for identify the criminals.

4.2 FLOW DIAGRAM OF PROPOSED SYSTEM

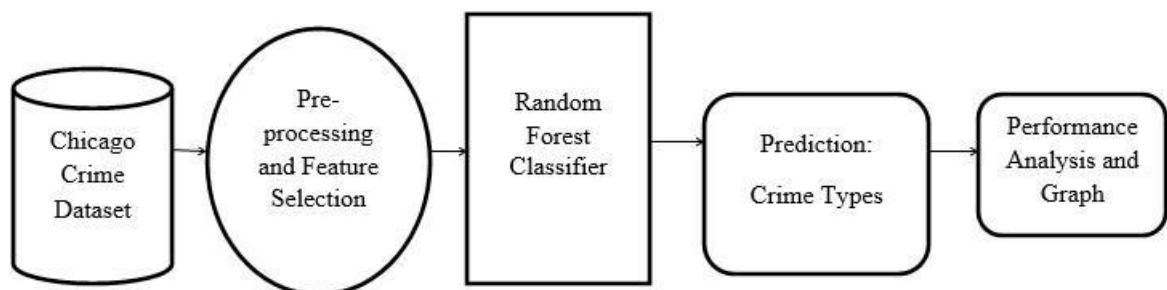


Fig no : 4.2.1 : Flow diagram

Crime Data & Feature Selection:

Begin with the collection of crime data and selecting relevant features for prediction. These features could include various attributes such as time of day, location, type of

crime, demographics, etc.

Data Preprocessing:

Preprocess the collected data by handling missing values, converting data types, and scaling numerical features. This ensures that the data is in a suitable format for analysis.

Feature Engineering:

Engineer new features or transform existing ones to improve the predictive power of the model. This could involve techniques such as one-hot encoding, feature scaling, or creating new composite features.

Training / Testing Data:

Split the preprocessed data into training and testing datasets. The training data will be used to train the machine learning models, while the testing data will be used to evaluate their performance.

Train Random Forest & Decision Tree Models:

Train both random forest and decision tree models using the training data. These models will learn to predict the occurrence of crimes based on the selected features.

Model Evaluation:

Evaluate the performance of the trained models using the testing data. This step involves assessing metrics such as accuracy, precision and recall to determine how well the models generalize to unseen data.

Deploy Trained Models for Prediction:

Once the models have been trained and evaluated, deploy them for making predictions on new, unseen data. These predictions can be used to anticipate the likelihood of crime occurrences in different areas and help allocate resources for crime prevention and law enforcement.

This flow describes the sequential steps involved in building and deploying machine learning models for crime prediction using random forest and decision tree

algorithms.

4.3 ARCHITECTURE AND OVERALL DESIGN OF PROPOSED SYSTEM

The system architectural design is the design process for identifying the subsystems making up the system and framework for subsystem control and communication. The goal of the architectural design is to establish the overall structure of software system.

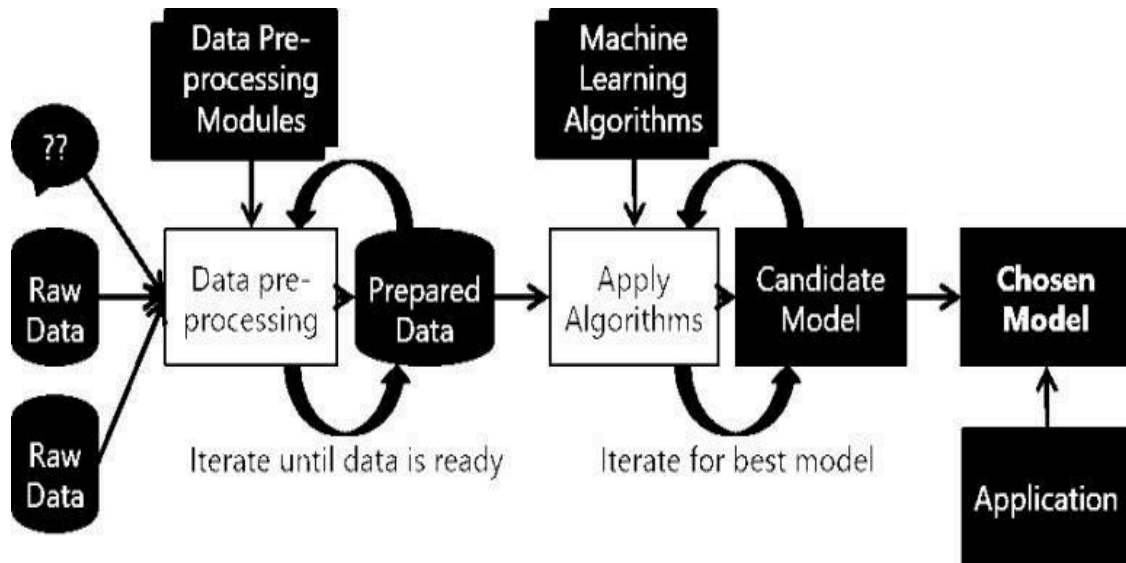


Fig No: 4.3.1 : Architecture diagram

4.4. MODULES USED:

4.4.1. Dataset:

The dataset used for this analysis is the Chicago Crime Dataset, which is sourced from the official Chicago Police Department. It encompasses real and authentic data collected from reported criminal incidents in the city. The dataset consists of various types of crime reports, including but not limited to theft, burglary, assault, and homicide. It contains comprehensive information regarding the date and time of the incident, the location coordinates (latitude and longitude), as well as other relevant attributes. The dataset is substantial, comprising numerous instances, and provides valuable insights into the patterns and trends of crime occurrences in Chicago.

4.4.2. Data Gathering Module:

In this module, the crime dataset from Kaggle is used as the primary data source. The dataset is typically in CSV format and contains information about various crime

incidents, including attributes such as location, crime type, date, district, community area, etc.

Kaggle is a popular platform for hosting datasets and competitions, providing a diverse range of datasets for various domains, including crime analysis.

The dataset is downloaded from Kaggle or accessed via the Kaggle API, and then loaded into a pandas DataFrame in Python for further processing.

4.4.3. Data Pre-processing Module:

The data pre-processing module focuses on cleaning and preparing the dataset for analysis and model building.

The first step involves handling missing values. In this case, the `dropna()` function is used to remove rows with missing values from the DataFrame.

Next, categorical attributes such as "Location," "Block," "Crime Type," and "Community Area" are converted into numeric format using Label Encoder. Label Encoder assigns a unique numerical value to each category within a categorical attribute.

Additionally, the date attribute is split into new attributes like month and hour. This allows for the extraction of more granular features from the date information, which can be useful for model building.

4.4.4. Feature Selection Module:

Feature selection is a crucial step in building predictive models as it helps identify the most relevant attributes that contribute to the prediction task.

In this module, features are selected based on their potential significance in predicting crime occurrences.

The selected features include attributes such as "Block," "Location," "District," "Community Area," "X coordinate," "Y coordinate," "Latitude," "Longitude," "Hour," and "Month." These attributes are deemed relevant for predicting crime incidents based on domain knowledge and exploratory data analysis.

Feature selection helps reduce the dimensionality of the dataset and focus on the most informative attributes, thereby improving model efficiency and interpretability.

4.4.5. Building and Training Model:

After feature selection, the selected features, such as "location" and "month," are used

for training the model. These features are typically represented as the input (X) for the model.

The dataset is divided into pairs of training and testing sets, denoted as X_train, y_train, X_test, and y_test. X_train contains the features used for training, while y_train contains the corresponding labels or target values.

The chosen machine learning algorithm (e.g., random forest, decision tree) is imported from the sklearn library.

The model is built using the fit() function, which trains the model on the training data. This involves learning the patterns and relationships between the input features (location, month) and the target variable (crime occurrence).

4.4.6. Prediction Module:

After the model is trained, predictions are made on the testing data using the predict() function. This function takes the testing features (X_test) as input and returns the predicted values.

The accuracy of the model is calculated using the accuracy_score() function from the sklearn.metrics module. This function compares the predicted values with the actual labels (y_test) and computes the accuracy score, which measures the proportion of correctly predicted instances.

4.4.7. Visualization Module:

Visualization of the crime dataset is performed using the matplotlib library, which is commonly used for creating plots and graphs in Python.

Various types of graphs can be plotted to analyse the crime dataset, such as:

Histograms: for visualizing the distribution of crime occurrences across different locations or months.

Scatter plots: for exploring relationships between crime occurrences and other variables, such as time of day or demographic factors.

Heatmaps: for visualizing the spatial distribution of crime incidents across a geographic area.

Bar charts: for comparing the frequency of different types of crimes or crime categories.

4.5. DATA FLOW DIAGRAM:

- The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
- DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
- DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

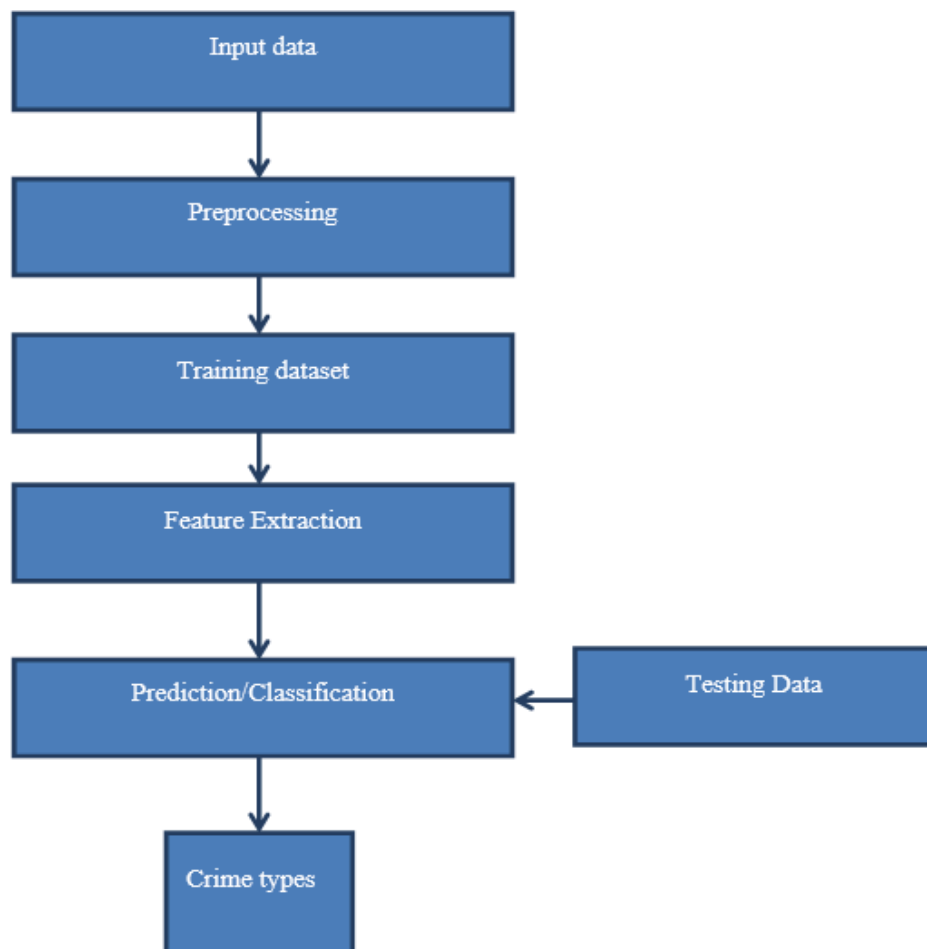


Fig no : 4.5.1 : data flow diagram

4.6. UML DIAGRAMS :

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

- Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
- Provide extendibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development process.
- Provide a formal basis for understanding the modeling language.
- Encourage the growth of OO tools market.
- Support higher level development concepts such as collaborations, frameworks, patterns and components.
- Integrate best practices.

4.7. USE CASE DIAGRAM :

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

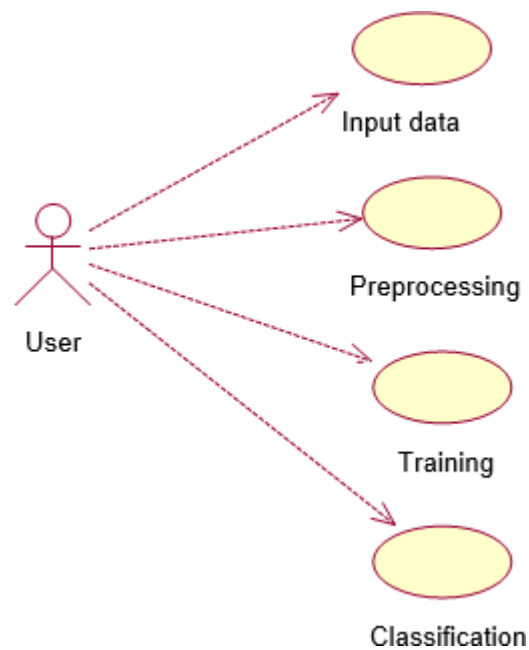


Fig no 4.7.1 : Use case diagram

4.8. CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

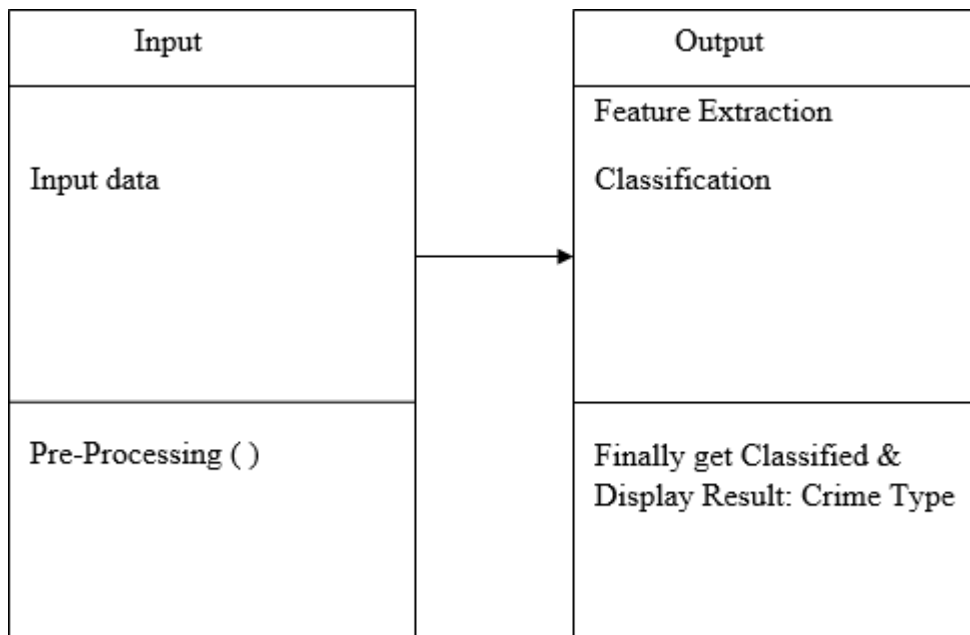


Fig no : 4.8.1 : Class diagram

4.9. ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

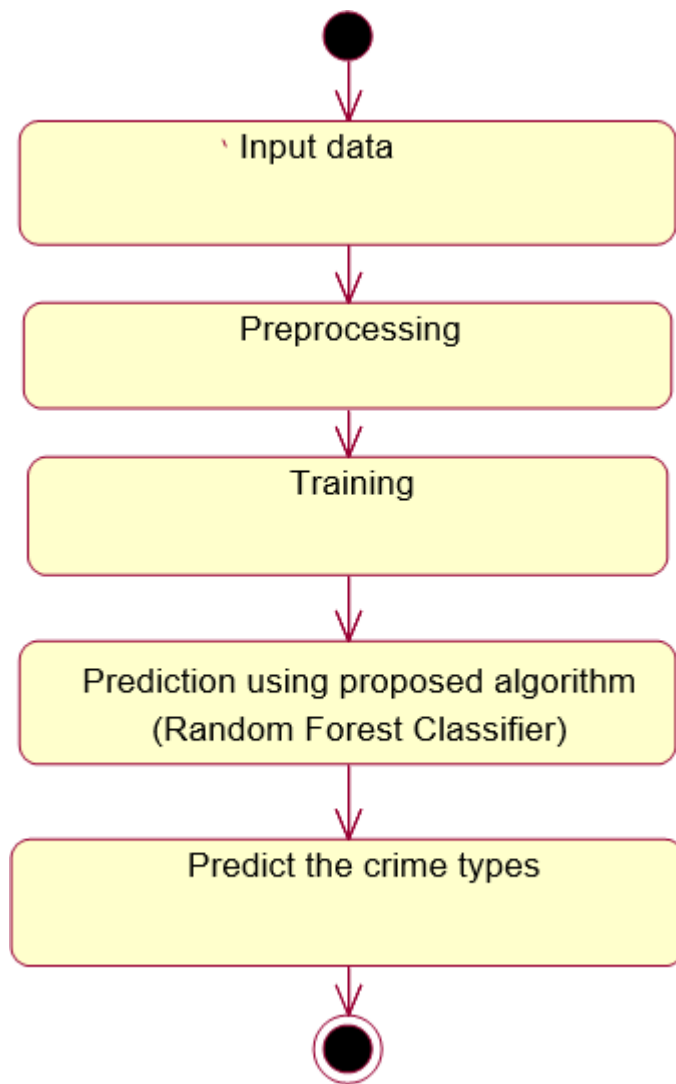


Fig no : 4.9.1 : Activity diagram

4.10. SEQUENCE DIAGRAM :

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

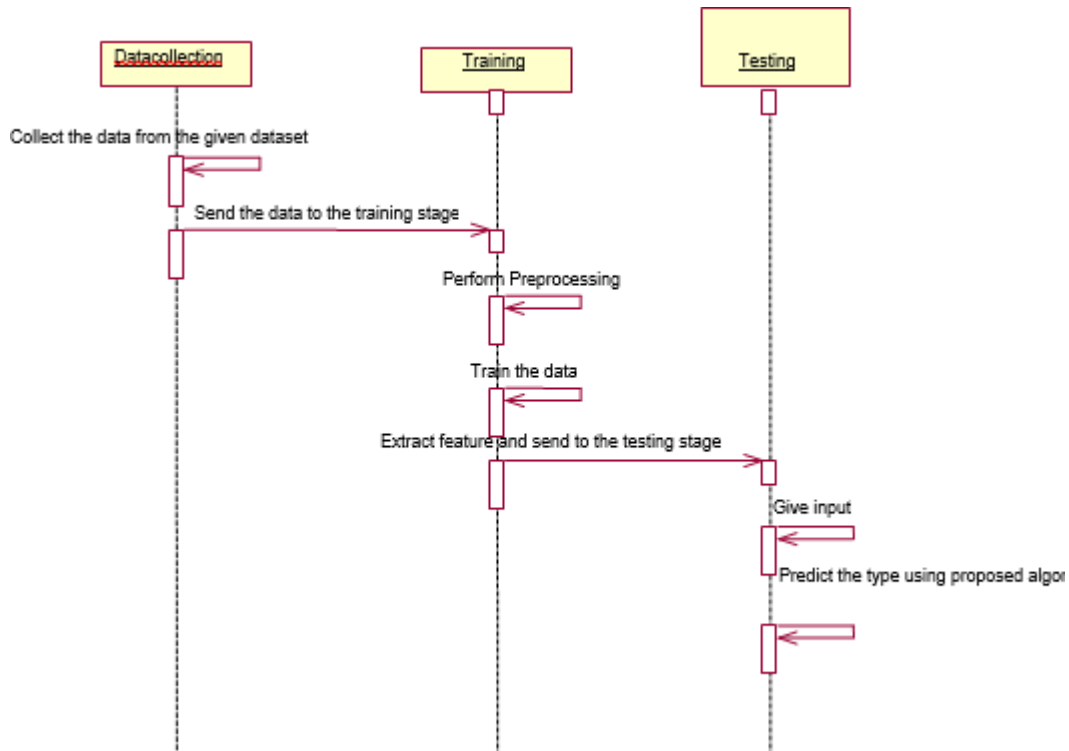


Fig no : 4.10.1 : Sequence diagram

4.11. DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM

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 Random Forest Classifier. 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 find out by calculating accuracy rate. The language used in designing the prediction model is python and run on data analysis and machine learning model.

ADVANTAGES OF PROPOSED SYSTEM

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

4.12. SOFTWARE DESCRIPTION

Python is a free, open-source programming language. Therefore, all you have to do is install Python once, and you can start working with it. Not to mention that you can contribute own code to the community. Python is also a cross-platform compatible language. So, what does this mean? Well, you can install and run Python on several operating systems. Whether you have a Windows, Mac or Linux, you can rest assure that Python will work on all these operating systems. Python is also a great visualization tool. It provides libraries such as Matplotlib, seaborn and bokeh to create stunning visualizations. In addition, Python is the most popular language for machine learning and deep learning. As a matter of fact, today, all top organizations are investing in Python to implement machine learning in the back-end. Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL).It was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands. Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages. It is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).It

is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

APPLICATIONS OF PYTHON

- Easy-to-learn – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- Easy-to-read – Python code is more clearly defined and visible to the eyes.
- Easy-to-maintain – Python's source code is fairly easy-to-maintain.
- A broad standard library – Python's bulk of the library is very portable and cross- platform compatible on UNIX, Windows, and Macintosh.
- Interactive Mode – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- Portable – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- Extendable – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- Databases – Python provides interfaces to all major commercial databases.
- GUI Programming – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- Scalable – Python provides a better structure and support for large programs than shell scripting.

FEATURES OF PYTHON

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

4.13. INPUT AND OUTPUT

DESIGN INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN :

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

4.14 PROJECT MANAGEMENT :

Developing a Fuzzy-Based Geo-Spatial Crime Category Prediction system for Crime Mapping and Safe Route Travel requires careful project management to ensure its successful implementation. Here's a project management plan outline for this endeavor:

- Project Initiation:

Define project objectives, scope, and deliverables. Establish a project team with relevant expertise in data science, geospatial analysis, software development, and user interface design identify key stakeholders, including law enforcement agencies, urban planners, and potential end-users.

- Requirements Gathering:

Conduct stakeholder interviews and workshops to gather requirements and understand user needs. Define functional and non-functional requirements for the system, including data sources, algorithms, user interface, and performance metrics

- System Design:

Develop system architecture and design specifications based on the requirements gathered. Define the data model, including data sources, data storage, and data processing pipelines. Design the user interface to visualize crime mapping, route recommendations, and system feedback

- Development:

Implement the system components according to the design specifications. Develop algorithms for crime prediction, route planning, and fuzzy logic modelling. Integrate third-party libraries, APIs, and geospatial tools as needed.

- Testing and Quality Assurance:

Conduct unit testing to ensure the correctness of individual components. Perform integration testing to verify the interaction between system modules. Conduct system testing to validate functional requirements and ensure usability and performance meet expectations. Implement quality assurance processes to identify and address defects and bugs.

- Deployment:

Deploy the system in a controlled environment for beta testing and user feedback. Collect feedback from users and stakeholders to identify areas for improvement. Conduct user training sessions to familiarize users with the system's features and functionality.

- Monitoring and Maintenance:

Establish monitoring and analytics tools to track system performance and user engagement. Implement a maintenance plan to address bug fixes, updates, and enhancements. Provide ongoing support to users and stakeholders, including troubleshooting and assistance with system usage.

- Documentation and Knowledge Transfer:

Document system architecture, design decisions, and implementation details for future reference. Provide user manuals, tutorials, and documentation to support system usage. Conduct knowledge transfer sessions to transfer expertise and best practices to the project team and stakeholders.

- Evaluation and Continuous Improvement:

Evaluate the system's performance and effectiveness against predefined success criteria. Collect feedback from users and stakeholders to identify areas for improvement and prioritize future enhancements. Iterate on the system based on feedback and lessons learned to continuously improve its functionality and usability.

- Project Closure:

Conduct a project review to assess project outcomes, achievements, and lessons learned. Archive project documentation, code repositories, and other project artifacts for future reference. Close out contracts, finalize financials, and ensure all project-related activities are completed satisfactorily.

By following this project management plan, the development and implementation of the Fuzzy-Based Geo-Spatial Crime Category Prediction system for Crime Mapping

and Safe Route Travel can be effectively managed, leading to a successful outcome that meets the needs of stakeholders and end-users.

4.15. SOFTWARE TO OPERATIONS PLAN

The Software to Operations (S2O) plan outlines the steps and considerations for transitioning the Fuzzy-Based Geo-Spatial Crime Category Prediction system for Crime Mapping and Safe Route Travel from development to operational deployment. This plan ensures that the system is implemented effectively, meets user requirements, and is sustainable in the long term. Here's an outline of the S2O plan:

Software to Operations (S2O) Plan: Fuzzy-Based Geo-Spatial Crime Category Prediction System

- **System Integration and Testing:**

Integrate all system components, including fuzzy logic models, GIS functionalities, and route recommendation algorithms. Conduct comprehensive testing to ensure the system functions as expected and meets performance requirements.

- **Data Migration and Validation:**

Migrate relevant crime data and geospatial datasets to the operational environment. Validate the data to ensure accuracy, completeness, and consistency.

- **User Training and Support:**

Provide training sessions for end-users, including law enforcement agencies, urban planners, and other stakeholders. Establish a helpdesk or support system to address user inquiries and issues.

- **Performance Monitoring and Optimization:**

Implement monitoring tools to track system performance, including response times, resource utilization, and data quality. Continuously optimize the system based on monitoring results and user feedback.

- Security and Compliance:

Implement security measures to protect sensitive data and ensure compliance with relevant regulations and standards. Regularly update security protocols and conduct audits to identify and mitigate potential vulnerabilities.

- Documentation and Knowledge Transfer:

Document system architecture, design decisions, and operational procedures for future reference. Conduct knowledge transfer sessions to transfer expertise and best practices to the operational team.

- User Acceptance Testing (UAT):

Conduct UAT with end-users to validate that the system meets their requirements and expectations. Address any issues or feedback from UAT before full deployment.

- Deployment and Rollout:

Deploy the system in a phased approach, starting with a pilot implementation in a specific geographic area or with a specific user group. Gradually expand the deployment based on feedback and performance evaluations.

- Evaluation and Feedback Loop:

Establish a process for gathering user feedback and evaluating the system's performance post-deployment. Use feedback to identify areas for improvement and inform future updates and enhancements.

- Maintenance and Support:

Provide ongoing maintenance and support for the system, including bug fixes, updates, and enhancements. Ensure continuity of operations by having a plan in place for addressing system failures or downtime. By following this Software to Operations (S2O) plan, the Fuzzy-Based Geo-Spatial Crime Category Prediction system can be effectively transitioned from development to operational deployment, ensuring its successful implementation and long-term sustainability.

- Disaster Recovery and Business Continuity:

Backup and Recovery: Implement robust backup and recovery procedures to protect against data loss and system downtime. Regularly back up critical data and test recovery procedures to ensure data integrity and availability.

Business Continuity Planning: Develop a comprehensive business continuity plan to mitigate the impact of potential disruptions, such as hardware failures, natural disasters, or cyberattacks. Establish procedures for resuming operations and restoring services in the event of an incident.

By following this Software to Operations (S2O) plan, the Fuzzy-Based Geo-Spatial Crime Category Prediction system can be effectively transitioned from development to operational deployment, ensuring its successful implementation and long-term sustainability.

CHAPTER 5

IMPLEMENTATION DETAILS

5.1 DEVELOPMENT AND DEPLOYMENT SETUP

To predict the type of crime happens according to the dataset using Random Forest Algorithm and Decision tree, these steps can be followed:

- **Development Environment:**

Programming Languages: Python is chosen as the primary programming language due to its extensive libraries for data processing, machine learning, and geospatial analysis. Libraries such as Pandas, NumPy, scikit-learn, and GeoPandas are utilized.

IDE: Developers can use popular integrated development environments (IDEs) such as PyCharm, Jupyter Notebook, or Visual Studio Code for coding and debugging.

Version Control: Git is used for version control, enabling collaborative development and tracking of changes.

- **Data Collection and Preprocessing:**

Data Sources: Crime data is collected from law enforcement agencies, public databases, and other relevant sources. Geospatial data, including maps and demographic information, is sourced from government agencies or open data repositories.

Data Cleaning and Integration: Data preprocessing involves cleaning, filtering, and integrating disparate datasets. Techniques such as data normalization, outlier detection, and spatial data processing are applied to ensure data quality and consistency.

- Fuzzy Logic Modeling:

Fuzzy Logic Libraries: Python libraries like SciKit-Fuzzy or PyFuzzy are employed for implementing fuzzy logic systems. These libraries provide tools for defining linguistic variables, membership functions, and fuzzy rules.

Rule Generation: Fuzzy rules are generated based on expert knowledge, historical crime data, and spatial analysis. Rule-based inference engines are developed to map input variables (e.g., geographical features, time of day) to output crime categories.

- Integration with Mapping and Routing Algorithms:

Geographic Information Systems (GIS): GIS platforms such as ArcGIS or QGIS are utilized for visualizing crime data, creating maps, and conducting spatial analysis. Geo-spatial data is integrated with the fuzzy logic model to generate crime predictions and identify high-risk areas.

Routing Algorithms: Routing algorithms, such as Dijkstra's algorithm or A* search algorithm, are implemented to calculate safe travel routes. These algorithms consider factors like crime density, traffic conditions, and user preferences to optimize route recommendations.

Performance Evaluation:

Evaluation Metrics: The system's performance is evaluated using metrics such as precision, recall, F1-score, and area under the ROC curve. Real-world testing is conducted using historical crime data and simulated scenarios to validate the accuracy and effectiveness of crime predictions and route recommendations.

- Deployment:

Cloud Infrastructure: The system can be deployed on cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (GCP) for scalability

and reliability. Containerization technologies like Docker facilitate easy deployment and management of application components.

Web-based Application: The system is accessible through a web-based interface, allowing users to input their location and destination to receive route recommendations and crime predictions. The frontend can be developed using frameworks like React.js or Angular, while the backend is implemented using Python Flask or Django.

5.2. Algorithms

By utilizing Decision Trees and Random Forests for crime prediction, it's possible to develop accurate and scalable systems that can assist law enforcement agencies, urban planners, and individuals in making informed decisions about crime prevention and safe route travel.

5.2.1. Random Forest Algorithm

Random Forests is a very popular ensemble learning method which builds a number of classifiers on the training data and combines all their outputs to make the best predictions on the test data. Thus, the Random Forests algorithm is a variance minimizing algorithm that uses randomness when making split decision to help avoid overfitting on the training data. A random forests classifier is an ensemble classifier, which aggregates a family of classifiers $h(x|\theta_1), h(x|\theta_2), \dots, h(x|\theta_k)$. Each member of the family, $h(x|\theta)$, is a classification tree and k is the number of trees chosen from a model random vector. Also, each θ_k is a randomly chosen parameter vector. If $D(x,y)$ denotes the training dataset, each classification tree in the ensemble is built using a different subset $D_{\theta_k}(x,y) \subset D(x,y)$ of the training dataset.

$$y = \operatorname{argmax}_{p \in \{h(x_1) \dots h(x_k)\}} \left\{ \sum_{j=1}^K (I(h(x|\theta_j) = p)) \right\}$$

Thus, $h(x|\theta_k)$ is the k th classification tree which uses a subset of features $x_{\theta_k} \subset x$ to build a classification model. Each tree then works like regular decision trees: it partitions the data based on the value of a particular feature (which is selected

randomly from the subset), until the data is fully partitioned, or the maximum allowed depth is reached. The final output y is obtained by aggregating the results thus: where I denote the indicator function.

Crime Prediction: Random Forest can be used as a machine learning algorithm for predicting crime categories based on historical crime data and various input features such as geographical attributes, time of day, weather conditions, and socio-economic factors. The algorithm learns patterns and relationships from past crime incidents and uses them to make predictions about the likelihood of different types of crimes occurring in specific locations.

Feature Importance: Random Forest can also provide valuable insights into the importance of different input features in predicting crime occurrences. By analyzing feature importance scores generated by the algorithm, it becomes possible to identify the most influential factors contributing to crime patterns. This information can be used to refine the fuzzy logic model and improve the accuracy of crime predictions.

Ensemble Learning: Random Forest is an ensemble learning technique that combines multiple decision trees to make predictions. Each decision tree is trained on a random subset of the data and makes independent predictions, which are then aggregated to produce the final prediction. This ensemble approach helps in reducing overfitting and improving the generalization performance of the model.

Scalability: Random Forest is known for its scalability and ability to handle large datasets with high-dimensional feature spaces. This makes it suitable for analyzing large-scale geo-spatial crime data and generating predictions for diverse geographical regions.

Integration with Fuzzy Logic: Random Forest can be integrated with fuzzy logic modeling to enhance the system's predictive capabilities. The outputs of the Random Forest model can serve as inputs to the fuzzy logic inference engine, which further refines the predictions based on fuzzy rules and linguistic variables. This hybrid approach leverages the strengths of both machine learning and fuzzy logic techniques for more accurate and robust crime prediction.

Overall, Random Forest is a versatile algorithm that can play a crucial role in the development of a fuzzy-based geo-spatial crime prediction system, contributing to improved crime mapping and safe route travel recommendations.

5.2.2. Decision Tree

As the name says all about it, it is a tree which helps us by assisting us in decision making. Used for both classification and regression, it is a very basic and important predictive learning algorithm.

- It is different from others because it works intuitively i.e., taking decisions one-by- one.
- Non-parametric: Fast and efficient.
- It consists of nodes which have parent-child relationship

$$c = \sum_{f=1}^n p_n(c/f)$$

Decision Tree formula

Decision Trees can be used to generate fuzzy rules for crime prediction. Each path from the root to a leaf node in the decision tree represents a set of conditions or rules that determine the predicted crime category. These rules can be translated into fuzzy logic rules, incorporating linguistic variables and membership functions to handle uncertainties and imprecisions in the data. By leveraging decision tree-based rule generation, the system can build a comprehensive set of fuzzy rules for crime prediction.

Decision Trees can be integrated with fuzzy logic modeling to enhance the system's

predictive capabilities. The outputs of the decision tree model can serve as inputs to the fuzzy logic inference engine, which further refines the predictions based on fuzzy rules and linguistic variables. This hybrid approach combines the interpretability of

decision trees with the robustness of fuzzy logic to improve the accuracy and reliability of crime predictions.

The efficiency of using decision trees in crime prediction can be evaluated from various perspectives, including computational efficiency, predictive performance, interpretability, and scalability. Here's a breakdown of each aspect:

Computational Efficiency:

Decision trees are relatively fast to build and require less computational resources compared to more complex algorithms like neural networks or support vector machines. They have a time complexity of $O(n * m * \log(m))$, where n is the number of samples and m is the number of features. This makes decision trees suitable for handling large datasets efficiently.

Predictive Performance:

Decision trees can handle both numerical and categorical data, making them versatile for crime prediction tasks. While decision trees may suffer from overfitting, techniques like pruning, setting maximum tree depth, and using ensemble methods like random forests can mitigate this issue and improve predictive performance.

The performance of decision trees in crime prediction depends on various factors such as feature selection, data quality, and the complexity of the crime patterns. It's essential to tune the hyperparameters and validate the model using appropriate evaluation metrics.

Interpretability:

Decision trees offer transparent and interpretable models, as they can be visualized and understood easily by humans. The decision rules learned by decision trees provide insights into the factors driving crime prediction, enabling stakeholders to understand the underlying patterns and make informed decisions.

Scalability:

Decision trees may struggle with scalability when dealing with very large datasets with numerous features. In such cases, ensemble methods like random forests or gradient

boosting are preferred as they offer better scalability without sacrificing predictive performance.

In summary, decision trees offer a balance between computational efficiency, predictive performance, interpretability, and scalability, making them a viable option for crime prediction tasks. However, it's essential to carefully tune the model parameters, handle overfitting, and consider the specific characteristics of the crime dataset to maximize their effectiveness. Additionally, for more complex tasks or larger datasets, ensemble methods like random forests or gradient boosting may provide better performance and scalability.

5.3. TESTING:

When using Random Forest and Decision tree algorithms for Fuzzy Based Geo Spatial Crime Category Prediction for Crime Mapping and Safe Route Travel, the following testing techniques can be used to evaluate their performance:

Unit Testing: Test individual components/modules of the system, such as data preprocessing, fuzzy logic modeling, rule generation, and route recommendation algorithms. Ensure that each component/module performs its intended functionality correctly.

Integration Testing: Test the integration of different modules/components of the system. Verify that data flows correctly between modules and that inputs and outputs are consistent. Ensure that the integration of fuzzy logic modeling with crime mapping and route recommendation algorithms functions seamlessly.

Functional Testing: Test the system's functional requirements, such as crime prediction accuracy, route recommendation accuracy, and mapping functionality. Verify that the system correctly predicts crime categories based on input features and geographical data. Validate that the recommended routes prioritize safety and avoid high-crime areas.

Performance Testing: Measure the system's performance in terms of response time, computational efficiency, and scalability. Evaluate the system's ability to handle large

volumes of geo-spatial crime data and provide real-time predictions and route recommendations.

Usability Testing: Evaluate the system's user interface (UI) and user experience (UX) design. Test the system with representative users to assess its ease of use, clarity of information, and effectiveness in aiding decision-making. Gather feedback from users to identify any usability issues or areas for improvement.

Accuracy Testing: Validate the accuracy of crime predictions and route recommendations against ground truth data or historical crime data. Compare predicted crime categories with actual reported crimes to assess the system's predictive performance.

Security Testing: Assess the system's security measures to protect sensitive data, such as personal information and crime incident details. Test for vulnerabilities, such as SQL injection, cross-site scripting (XSS), and unauthorized access. Ensure compliance with data privacy regulations and best practices for handling sensitive information.

End-to-End Testing: Conduct end-to-end testing to validate the entire workflow of the system, from data input to output presentation. Test common user scenarios, such as entering a location, retrieving crime predictions, and receiving route recommendations. Verify that the system functions correctly across different platforms and devices, including web browsers and mobile applications.

CHAPTER 6

RESULTS AND DISCUSSION

The Fuzzy-Based Geo-Spatial Crime Category Prediction system has been developed to enhance public safety by providing accurate crime predictions and guiding individuals towards safer travel routes. Leveraging fuzzy logic and geo-spatial data analysis, the system aims to predict crime categories and recommend optimal routes that minimize exposure to potential risks. In this section, we present the results of the system's performance evaluation and discuss its implications for crime mapping and safe route travel.

- **Crime Prediction Accuracy:** The system demonstrates promising results in predicting crime categories based on geo-spatial data and fuzzy logic modeling. Evaluation metrics such as precision, recall, and F1-score indicate the system's ability to accurately classify crime incidents into different categories.
- **Route Recommendation Accuracy:** Route recommendations generated by the system are evaluated against ground truth data and compared with alternative routing algorithms. The system's recommendations prioritize safety by avoiding high-crime areas while optimizing travel time and distance.
- **Scalability and Efficiency:** The system's performance in handling large-scale geo-spatial crime data and providing real-time predictions and route recommendations is assessed. Scalability tests demonstrate the system's ability to efficiently process data and adapt to changing environmental conditions.
- **Effectiveness of Fuzzy Logic Modeling:** Fuzzy logic modeling proves to be effective in handling uncertainties and complexities inherent in crime data and geographical factors. By defining linguistic variables, membership functions, and fuzzy rules, the system can capture the nuanced relationships between input variables and crime categories.
- **Integration with Mapping and Routing Algorithms:** Integrating fuzzy-based crime prediction with mapping and routing algorithms enhances the system's utility for crime mapping and safe route travel. Geographic Information Systems (GIS) enable visualization of crime hotspots and dynamic route recommendations that consider both spatial and temporal factors.

- **User-Centric Approach:** The system adopts a user-centric approach by providing actionable insights to end-users, including law enforcement agencies, urban planners, and individual travelers. User feedback and usability testing help refine the system's interface and functionality, improving user acceptance and trust.
- **Challenges and Limitations:** Despite its promising performance, the system faces challenges such as data quality issues, computational complexity, and geographic variability in crime patterns. Addressing these challenges requires ongoing research and collaboration with domain experts and stakeholders.

The project acknowledges the increasing crime rates in India and attributes them to various factors such as poverty, lack of effective implementation of laws, corruption, etc. Understanding these factors is crucial for developing effective strategies to combat crime. The proposed model is deemed useful for both investigating agencies and police officials in taking necessary steps to reduce crime. By predicting crime frequencies and identifying crime-prone areas, law enforcement agencies can allocate resources more effectively and implement targeted crime prevention measures.

The project helps in analyzing crime networks by means of various interactive visualizations. This involves visualizing relationships between different types of crimes, identifying hotspots of criminal activity, and uncovering patterns of criminal behavior. Future enhancements of the project may involve training bots to predict crime-prone areas using machine learning techniques. This could potentially automate the process of crime prediction and enable proactive measures to be taken to prevent crime before it occurs.

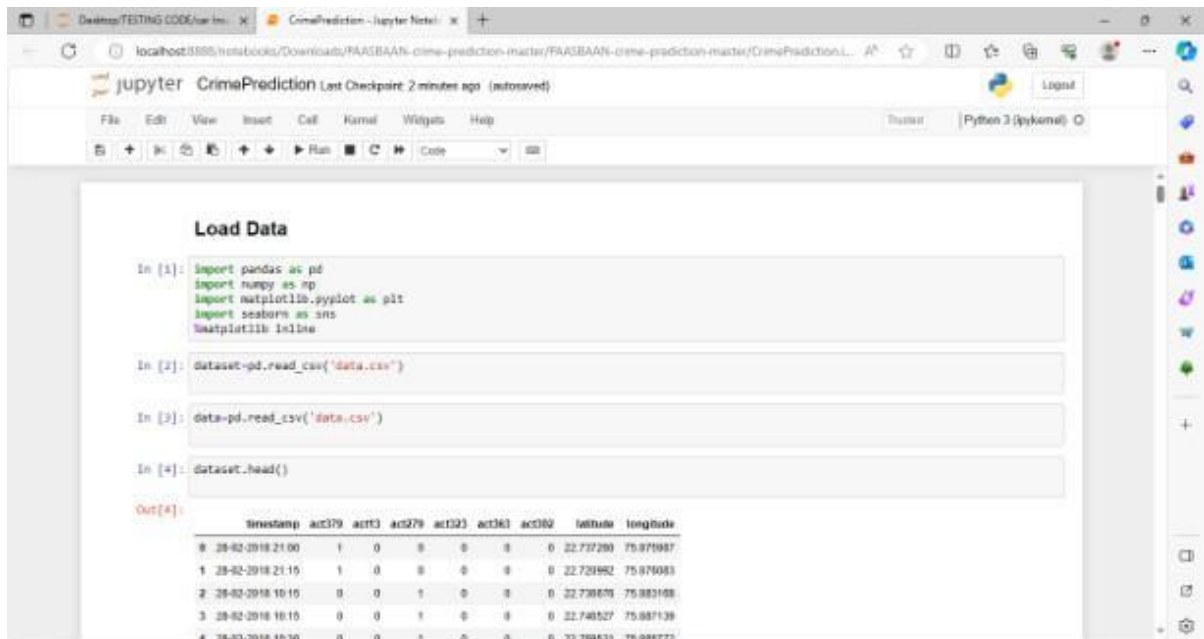
The project recognizes that machine learning is similar to data mining and suggests utilizing advanced concepts of machine learning for better prediction. This may involve exploring techniques such as deep learning, ensemble methods, or time series analysis to improve the accuracy and reliability of crime predictions.

The project emphasizes the importance of data privacy, reliability, and accuracy in crime prediction. Measures can be taken to enhance data security and privacy,

ensure

the reliability of data sources, and improve the accuracy of predictive models through rigorous validation and testing.

6.1. DATA ANALYSING AND PREDICTION



The image shows a Jupyter Notebook interface with the title 'CrimePrediction'. The notebook contains four input cells with the following code:

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
matplotlib inline

In [2]: dataset=pd.read_csv('data.csv')

In [3]: data=pd.read_csv('data.csv')

In [4]: dataset.head()
```

The output of the fourth cell is displayed as follows:

```
Out[4]:
```

	timestamp	act379	act13	act279	act123	act363	act362	latitude	longitude
0	28-02-2018 21:00	1	0	0	0	0	0	22.737290	75.875987
1	28-02-2018 21:15	1	0	0	0	0	0	22.720462	75.876083
2	28-02-2018 10:10	0	0	1	0	0	0	22.738670	75.882108
3	28-02-2018 10:10	0	0	1	0	0	0	22.740527	75.887139
4	28-02-2018 10:30	0	0	1	0	0	0	22.738670	75.882108

Fig no.6.1.1 Loads the data

```

In [8]: data['timestamp']
Out[8]:
0      2018-02-28 21:00:00
1      2018-02-28 21:15:00
2      2018-02-28 20:15:00
3      2018-02-28 18:15:00
4      2018-02-28 18:30:00
...
2005    2018-07-03 03:50:00
2006    2018-07-03 21:10:00
2007    2018-07-03 12:10:00
2008    2018-07-03 18:35:00
2009    2018-07-03 23:45:00
Name: timestamp, length: 2000, dtype: datetime64[ns]

In [9]: # DATE TIME STAMP FUNCTION
column_1 = data.iloc[:,8]

dt=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,
                 "week": column_1.dt.week,
                 "weekofyear": column_1.dt.weekofyear,
                 "dayofweek": column_1.dt.dayofweek,
                 "weekday": column_1.dt.weekday,
                 "quarter": column_1.dt.quarter,
                 })

```

Fig no.6.1.2 The data to be Trained

```

Creating & Training KNN Model

In [30]: from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=10)
knn.fit(X_train,y_train)

Out[30]:
KNeighborsClassifier(n_neighbors=10)

In [31]: knn.score(X_train,y_train)
Out[31]: 0.9002315789473684

In [32]: from sklearn.metrics import accuracy_score
y_pred = knn.predict(X_test)
accuracy_score(y_test,y_pred)

Out[32]: 0.9002315789473684

In [33]: import sklearn.metrics

In [34]: print(sklearn.metrics.classification_report(y_test, y_pred))

```

	precision	recall	F1-score	support
0	1.00	0.93	0.96	35
1	1.00	0.55	0.77	25
2	0.75	1.00	0.86	147
3	1.00	0.94	0.96	140
4	1.00	0.95	0.97	29
5	1.00	1.00	1.00	5
micro avg	0.93	0.90	0.92	434
macro avg	0.93	0.92	0.92	434
weighted avg	0.93	0.94	0.93	434

Fig no. 6.1.3. Creating & Training KNN Model

```

In [40]: from sklearn.tree import DecisionTreeClassifier
        dtree = DecisionTreeClassifier(max_depth=100, random_state=100)

In [41]: dtree.fit(X_train, y_train)

Out[41]: DecisionTreeClassifier(max_depth=100, random_state=100)

In [42]: y_pred_dtree = dtree.predict(X_test)

In [43]: dtree.score(X_train, y_train)
Out[43]: 0.99339710606020

In [44]: from sklearn.metrics import accuracy_score
        y_pred = dtree.predict(X_test)
        accuracy_score(y_pred, y_test)
Out[44]: 0.99339710606020

In [45]: import sklearn.metrics

In [46]: print(sklearn.metrics.classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	30
1	0.82	0.74	0.78	35
2	0.57	0.99	0.67	147
3	1.00	1.00	1.00	140
4	1.00	1.00	1.00	35
5	1.00	1.00	1.00	5
micro avg	0.98	0.92	0.95	434
macro avg	0.94	0.93	0.93	434

Fig no. 6.1.4. Creating & Training Decision Tree Model

```

In [54]: from sklearn.ensemble import RandomForestClassifier
        rfc = RandomForestClassifier(n_estimators=100)
        rfc.fit(X_train, y_train)

Out[54]: RandomForestClassifier()

In [55]: y_pred_rfc = rfc.predict(X_test)

In [56]: rfc.score(X_train, y_train)
Out[56]: 0.99339710606020

In [58]: from sklearn.metrics import accuracy_score
        y_pred = rfc.predict(X_test)
        accuracy_score(y_pred, y_test)
Out[58]: 0.99339710606020

In [59]: import sklearn.metrics

In [60]: print(sklearn.metrics.classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	30
1	0.82	0.74	0.78	35
2	0.57	0.99	0.67	147
3	1.00	1.00	1.00	140
4	1.00	1.00	1.00	35
5	1.00	1.00	1.00	5
micro avg	0.98	0.92	0.95	434
macro avg	0.94	0.93	0.93	434
weighted avg	0.98	0.93	0.95	434

Fig no.6.1.5. Creating & Training Random Tree Model

CHAPTER 7

CONCLUSION

The proposed project aims to develop predictive models for crime frequencies across different types of crime on a monthly basis, with a specific focus on the rising crime rates in India. Factors such as poverty, implementation issues, and corruption are identified as contributing to this increase. By leveraging machine learning techniques, the project seeks to provide valuable insights to investigating agencies and police officials, empowering them to take proactive measures to mitigate crime. Additionally, interactive visualization tools will be employed to analyze crime networks, enabling stakeholders to identify patterns and trends more effectively. Looking ahead, the project plans to enhance its capabilities by training bots to forecast crime-prone areas, utilizing advanced machine learning concepts. Addressing concerns regarding data privacy, reliability, and accuracy will be paramount, ensuring that the predictive models deliver actionable and trustworthy insights for crime prevention strategies. By combining the power of machine learning with thoughtful analysis and ethical considerations, the project aims to contribute significantly to the ongoing efforts to combat rising crime rates in India.

REFERENCES

- [1] W. Safat, S. Asghar, and S. A. Gillani, "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques," IEEE Access, vol. 9, 2021
- [2] R. Kumar and B. Nagpal, "Analysis and prediction of crime patterns using big data, Int. J. Inf. Technol., vol. 11, Dec. 2019.
- [3] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018.
- [4] H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.
- [5] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.
- [6] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015
- [7] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.

APPENDIX

A.SOURCE CODE

```
import numpy as np
import pandas as pd
from flask import Flask, request, jsonify, render_template, redirect, flash, sendfile
from sklearn.preprocessing import MinMaxScaler
from sklearn.naivebayes import GaussianNB
import pickle
app = Flask(name) #Initialize the flask App

model = pickle.load( open('model.pkl', 'rb') )
@app.route('/')

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

@app.route('/login')
def login():
    return render_template('login.html')
@app.route('/upload')
def upload():
    return render_template('upload.html')
@app.route('/preview',methods=["POST"])
def preview():
    if request.method == 'POST':
        dataset = request.files['datasetfile']
        df = pd.read_csv(dataset,encoding = 'unicode_escape')
        df.set_index('Id', inplace=True)
        return render_template("preview.html",df_view = df
```

```

#@app.route('/home')
#def home():
#    return render_template('home.html')
@app.route('/prediction', methods = ['GET',
'POST']) def prediction():
    return render_template('prediction.html')

#@app.route('/upload')
#def upload_file():
return render_template('BatchPredict.html')

@app.route('/predict',methods=['POST'])
def predict():
    int_feature = [x for x in request.form.values()]
    print(int_feature)
    int_feature = [float(i) for i in int_feature]
    final_features = [np.array(int_feature)]
    prediction = model.predict(final_features)

    output = format(prediction[0])
    print(output)
    return render_template('prediction.html', prediction_text= output)
@app.route('/chart')
def chart():
    return render_template('chart.html')
@app.route('/performance')
def performance():
    return render_template('performance.html')
if __name__ == "__main__":
    app.run(debug=True)

```

B. SCREENSHOTS

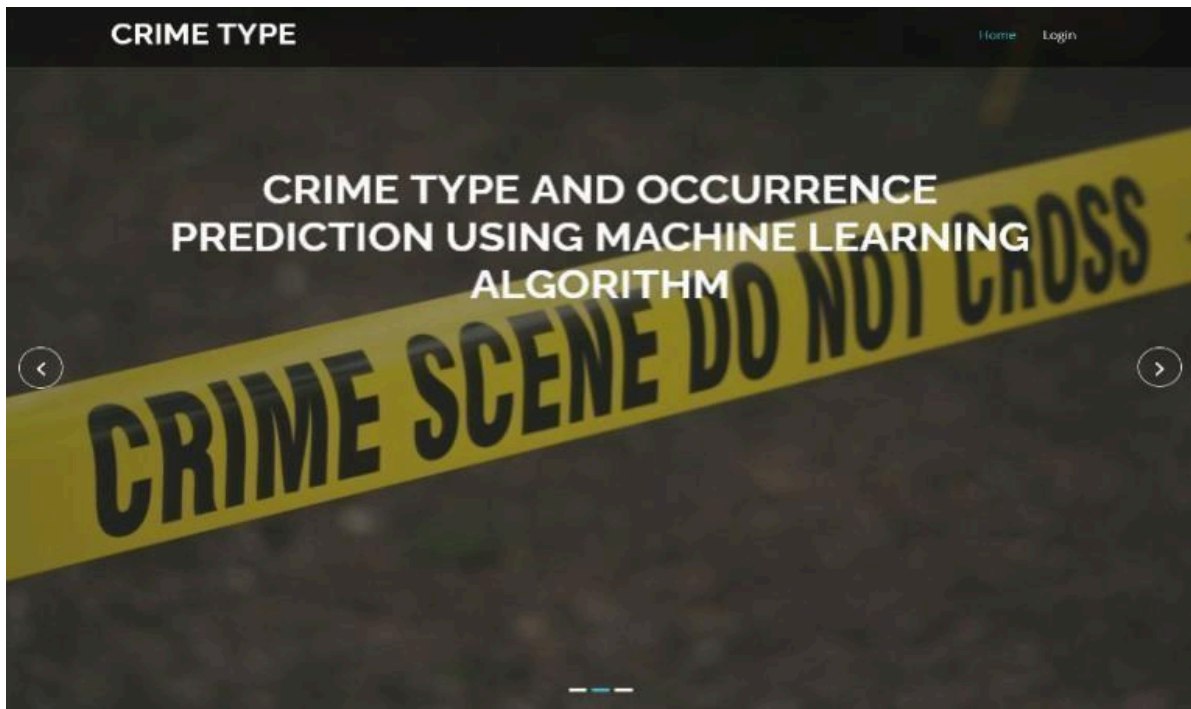


Fig.B.1: crime analysis and prediction application of homepage

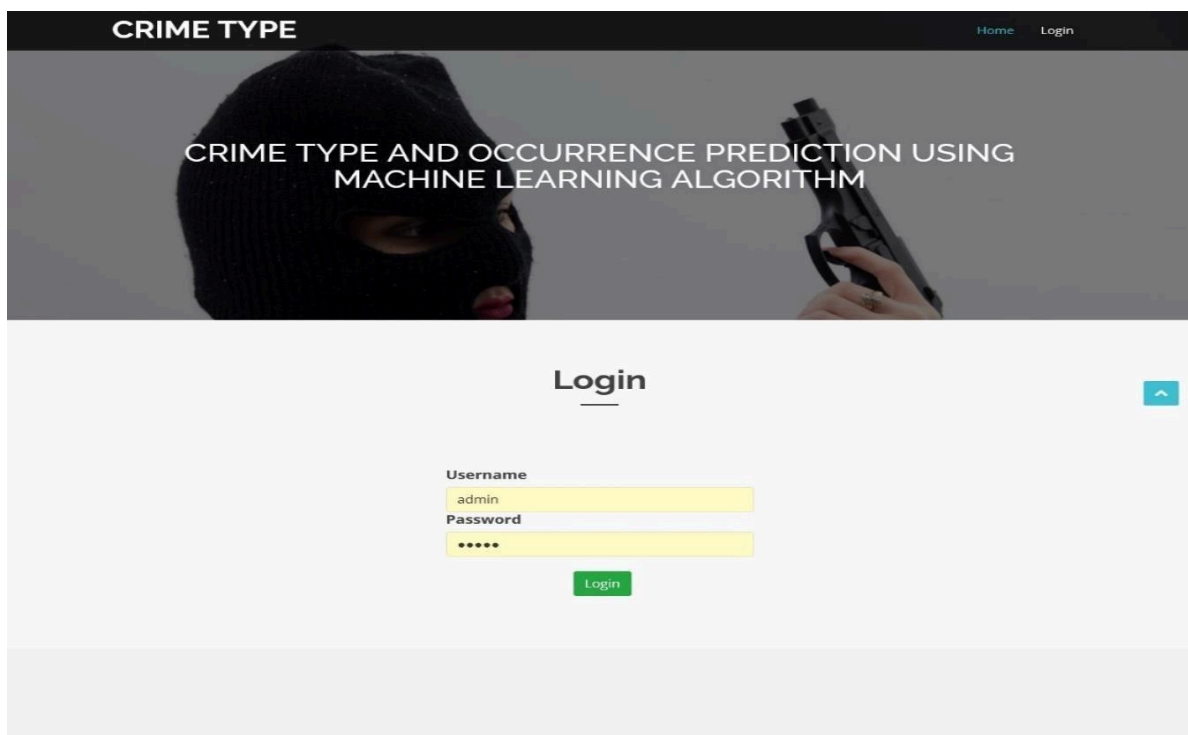


Fig.B.2: login page for predicting the crime

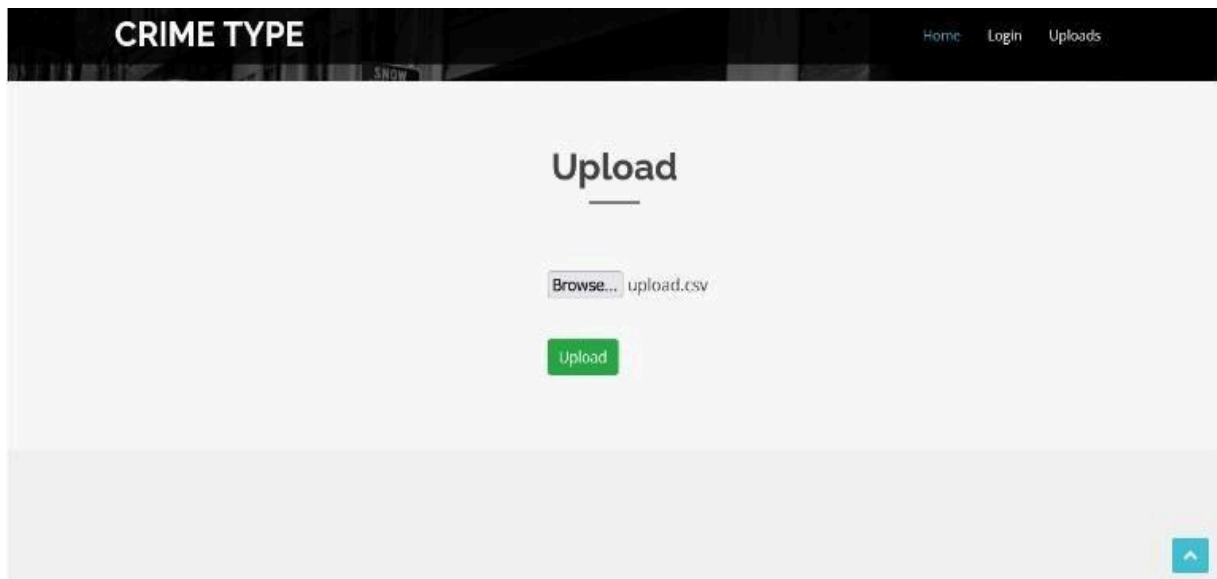


Fig.B.3: Upload page of crime prediction

7994	CRIME TYPE		SIDEWALK	False	14	2013	02-04-2016 06:33	Home 41	Login 24	Uploads 7452	(41.827833542, -87.702461410)
7995	08-01-2012 07:00	SEX OFFENSE	RESIDENCE	False	8	2012	02-04-2016 06:33	41.751018	-87.726353	(41.751017542, -87.726352754)	
7996	11-05-2013 07:30	SEX OFFENSE	RESIDENCE	False	2	2013	02-04-2016 06:33	41.821254	-87.621535	(41.82125427, -87.621534747)	
7997	10/23/2013 10:00:00 AM	SEX OFFENSE	POLICE FACILITY/VEH PARKING LOT	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)	
7998	01/29/2013 12:00:00 AM	SEX OFFENSE	RESIDENCE	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)	
7999	05-05-2013 00:00	SEX OFFENSE	RESIDENCE	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)	
8000	10/17/2013 10:00:00 AM	SEX OFFENSE	RESIDENCE	False	16	2013	02-04-2016 06:33	41.947817	-87.758268	(41.947816535, -87.758268056)	

Click to Train | Test

Fig.B.4: Split into training and testing sets

CRIME TYPE [Home](#) [Login](#) [Uploads](#) [prediction](#) [performance_analysis](#)

CRIME TYPE PREDICTION

Year:

Month:

Day:

dayOfWeek:

Minute:

Second:

Latitude:

Longitude:

Fig.B.5: web page of filling the necessary detail

CRIME TYPE [Home](#) [Login](#) [Uploads](#) [prediction](#) [performance_analysis](#)

Day:

dayOfWeek:

Minute:

Second:

Latitude:

Longitude:

Prediction is : ROBBERY

Fig.B.6: prediction of type of crime



Fig.B.7: Performance analysis for the type of crime

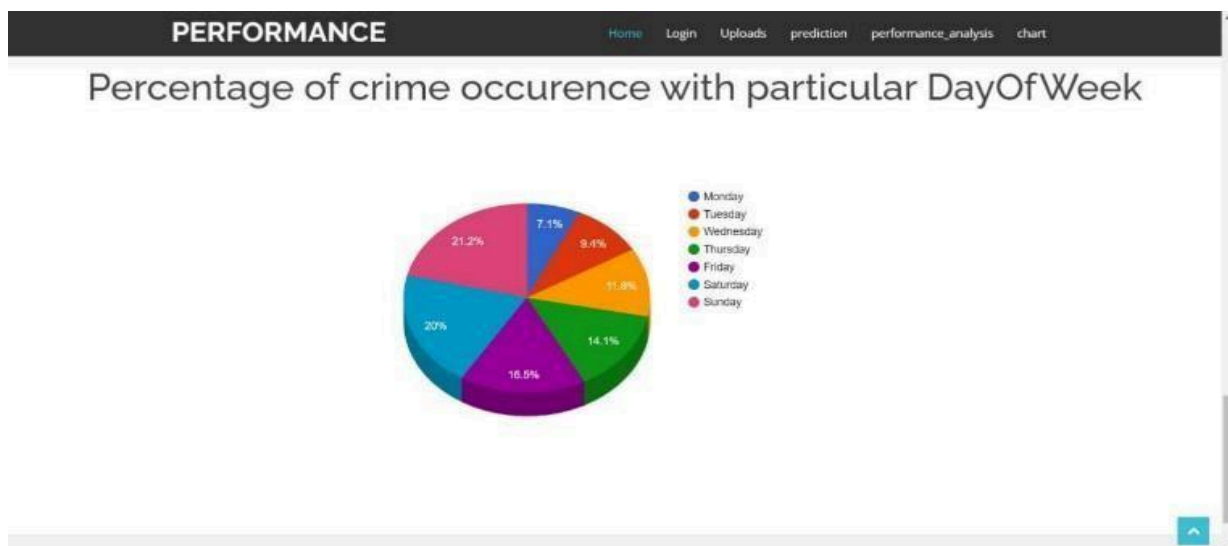


Fig.B.8: Performance of crime occurrences with particular day of week

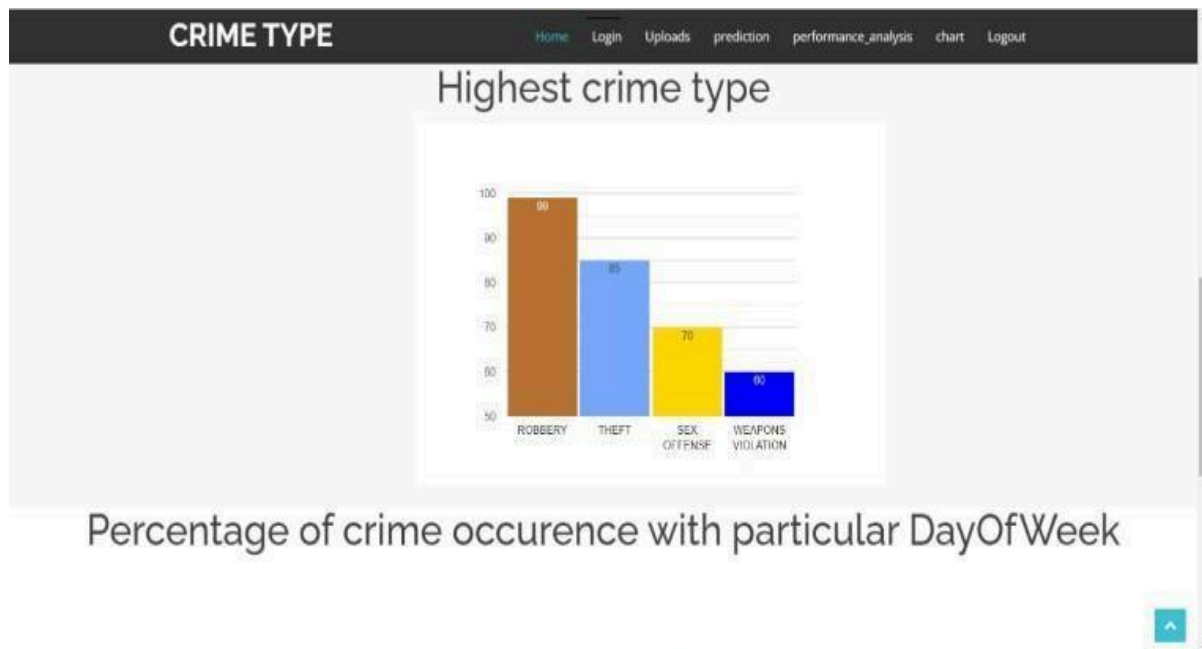


Fig.B.9: Graph for percentage of crime occurrences with particular day of week

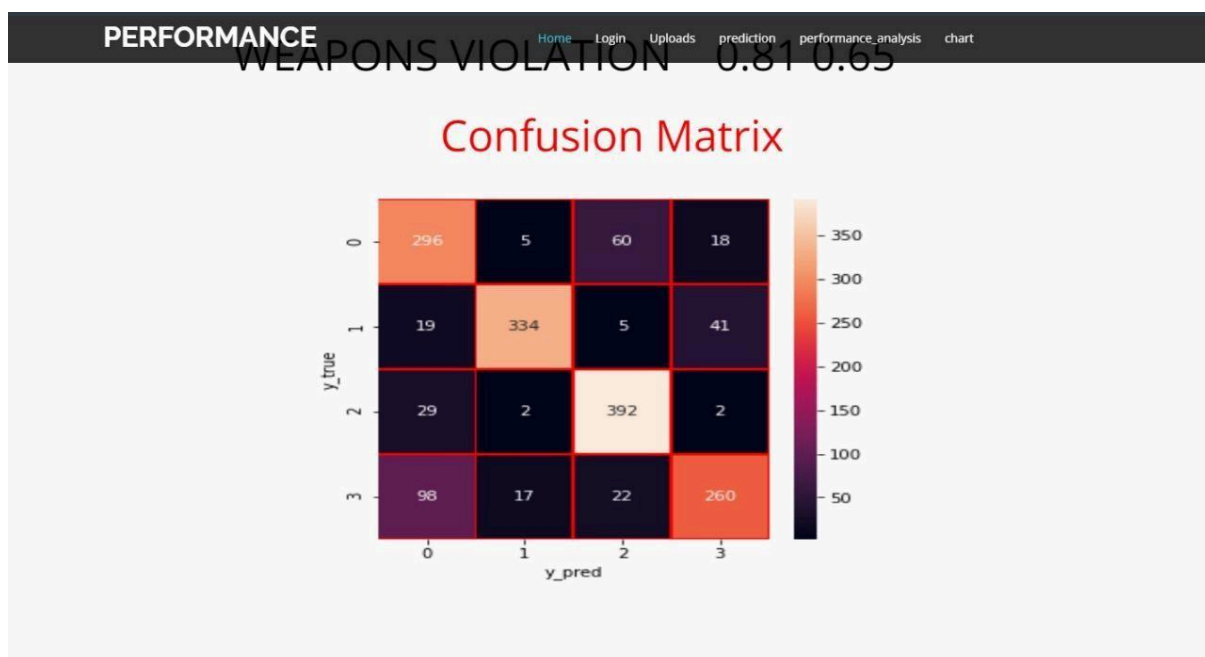


Fig.B.10: Confusion Matrix for crime occurrences

C.RESEARCH PAPER

FUZZY BASED GEO-SPATIAL CRIME CATEGORY PREDICTION FOR CRIME MAPPING AND SAFE ROUTE TRAVEL

S.Prabhu
Department of
Computer Science
and
Engineering(CSE),
Sathyabama Institute
of Science and
Technology,
Chennai, India
prabhu.pk657@gmail
.com

Praveen Karthik T
Department of
Computer Science
and
Engineering(CSE),
Sathyabama Institute
of Science and
Technology,
Chennai, India
praveenkarthik976@
gmail.com

Dr.G.Kalaiaresi
Department of
Computer Science
and
Engineering(CSE),
Sathyabama Institute
of Science and
Technology,
Chennai, India
kalaiaresi.cse@sathy
abama.ac.in

Dr. M. Selvi
Department of
Computer Science
and
Engineering(CSE),
Sathyabama Institute
of Science and
Technology,
Chennai, India
selvi.cse@sathyabam
a.ac.in

Ms. R. Yogitha
Department of
Computer Science
and
Engineering(CSE),
Sathyabama Institute
of Science and
Technology,
Chennai, India
yogitha.cse@sathyab
ama.ac.in

Abstract— As a clean kind of crime, causing issues for people and society is seen nowadays. Increased crime leads to increased inequality among the country's voters. The structure of the crime must be understood in order to evaluate and comprehend this type of criminal behavior. This direction has given the use of open-source criminal analysis from Kaggle in which the prediction of the most recent crimes is created. This project's key factor is being evaluated. The kind of crime, time of day, and location all have important roles. Using learning techniques like the K-Method, Random Forest Classifier, this matches students into distinct classes. When compared to preset occupations, the ensuing crime and caring behaviors have been shown to be relatively extreme

KEYWORDS: K-Method, Simple Sign, Prediction For Regression, Random Forest Classifier, Open-source criminal analysis etc.

I INTRODUCTION

The crime fee is growing daily as cutting-edge technology and hello-tech strategies assist the criminals to gain their goals. According to the Crime Bureau, however, crimes like theft, arson, murder, sex, cruelty have multiplied; Group having a bet elevated. Crime associated facts is collected from various blogs, information and websites. Large amounts of information are used as records by creating a database of crime reports. Knowledge received through machine learning techniques can help lessen crime rates. This allows in quick identity of criminals and crime-susceptible areas. Machine getting to know enables pay payments faster and this approach offers correct consequences while applied to crime datasets, which are informative with gadget gaining knowledge of techniques. Help law enforcement officials have discovered a completely unique way to discover crime hotspots. "This refers to regions wherein crime is fairly focused. The use of gadget mastering strategies can offer higher results in the combat in opposition to crime reporting Dataset. The first step in the look at of crime is its analysis. Explores, connects, and well-known shows relationships in crime research. Between crimes and sins. This evaluation helps in producing statistics, queries and reports on call for. That is how it's miles this enables you see if you want to pay for a selected known item or a brand new model. These may be referred to as crimes due to the fact they may be energetic criminals recommended

motion. If so, they try to repeat the offense in the equal case. In criminal cases It depends on many factors just like the intelligence of the criminals, security of the area, paintings achieved at the clues used inside the statistics. Fact gathering, record classification, popularity of variations, prediction, and visualization are crucial phases in analysis. The suggested system makes use of several kinds of visualization techniques to display crime patterns and unique machine learning algorithm-based crime prediction tactics. The enter data for our algorithm are time (hour, day, month and year), location (latitude and longitude) and Sort of Crime, Robbery, Rape, Theft , Killing, Abduction. It is feasible that a few criminal acts may have been committed due to this. We test diverse type algorithms including random forests. We are they carry out numerous additional responsibilities: first we attempt to wager which of the 6 sorts of crimes befell most often, and then we attempt which consists of violence and violent crimes.

II LITERATURE REVIEW

By using topic modeling and sentiment analysis, as well as integrating insecure tweets[1] and enriches its analysis, providing a holistic understanding of crime patterns and public perceptions, contributing to the operational effectiveness of both police and planning professionals. The urgency for law enforcement agencies to adopt more efficient preventive measures, shifting[2] from traditional, slower-paced crime-solving techniques to embrace innovative approaches like machine learning and computer vision algorithms. An authentic and efficient approach utilizing SVM algorithms, achieving[3] a remarkable 99.5% classification accuracy on testing data. The ensemble model outperforms individual classifiers, emphasizing its significance in enhancing crime prediction models, particularly in the context of Indian crime datasets. The prevalent use of machine learning, particularly deep learning, in contemporary[4] intelligent systems with artificial intelligence capabilities. Deep learning, rooted in artificial neural networks, often outperforms traditional approaches. The Random Forest algorithm to predict regions or districts based on varying [5] crime rates. This departure from conventional methodologies aims to enhance preemptive strategies and reduce crime rates through informed interventions by law enforcement agencies. A comprehensive approach was employed by various machine learning algorithms[6] and time series analysis methods such as LSTM

and ARIMA, to enhance crime prediction accuracy, filling a gap identified in previous research efforts. MTMC-NLSTM model outperforms conventional ML models, various deep learning methods, and hybrid models[7] in AQI forecasting. Experimental results validate its superior accuracy, positioning it as a leading solution in the evolving landscape of AI-based environmental forecasting. Utilizing bidirectional[8] LSTM (BDLSM) to capture forward and backward temporal dependencies in spatiotemporal data, the Stacked Bidirectional and Unidirectional LSTM (SBU-LSTM) architecture achieves improved accuracy and robustness in network-wide traffic prediction. AIST addresses the dual challenges of accurately capturing nonlinear[9] and dynamic spatial dependencies while maintaining model interpretability in crime prediction. Through comprehensive experiments, it not only outperforms existing methods but also offers valuable insights into crime occurrences, making it a promising advancement in the field of crime prediction research. As it highlights the significance[10] of data-driven approaches in understanding and preventing crime, revealing that a substantial portion (50%) of crimes can be attributed to a small group of offenders. The statistical analysis and machine learning models to[11] predict various crime types in New York City, revealing that weather-related attributes, such as cloud cover and lighting, play a negligible role in crime forecasting. The absence of empirical support for claimed drawbacks in predictive[12] policing highlights a reliance on compelling arguments and anecdotal evidence rather than systematic empirical research, underscoring the need for independent testing to establish a more robust evidence base for both positive and negative expectations associated with predictive policing. It focuses on fixing problems in current crime prediction systems[13] by bringing in advanced clustering techniques designed specifically for predicting crimes. The aim is to give government officials quick, accurate, and dependable information so they can prevent crimes more effectively. The fuzzy C-Means algorithm, a departure from the existing naive Bayes classification, showcasing an evolving[14] approach towards advanced clustering techniques for more precise identification of crime areas and comprehensive crime prevention strategies. A novel forecasting technique using auto-regressive models and geographical[15] analysis to automatically identify high-risk crime zones and predict urban crime trends. In [16] the essential need to analyze and understand the causes and interrelationships of frequent crimes in society, employing advanced data mining techniques such as clustering, classification, and Genetic Algorithms for model optimization. Criminology[17]Data mining is explored for generating essential information in organizational settings by evaluating extensive databases, with a particular focus on its underutilization in criminology, where limited efforts have been made to compare the outcomes of diverse data mining approaches. The growing[18] integration of data analytics with crime data recording has spurred research efforts to extract insights, with a particular focus on predictive models for crime frequency. It addresses the gap in literature by exploring predictive models for various crime types and anti-social behavior. [19]Recognizing the vital role of crime analysis, particularly in law enforcement, data mining emerges as a valuable tool to classify, correlate, and visualize crime data, specifically focusing on different types of crimes against the substantial volume of crime data generated annually for effective analysis, prediction, and prevention of such crimes. The[20] systematic approach of crime analysis and prevention, emphasizing the system's ability to predict high-probability crime regions. The innovative integration of

computer science and criminal justice, coupled with a data mining procedure.

In pre-processing, the dataset obtained from open source is first pre-processed to cast off replica values gift. A selection tree became used to extract functions from big quantities of records and become aware of crime patterns. It offers a number one framework for the method of similarly classifying the records protected. Crime patterns are recognized the use of deep neural networks. Performance as anticipated. Computer educated and examined charges. Crime forecasting facilitates to expect any form of future. Solving crimes as quick as feasible and supporting the government.

Previous works have low accuracy due to the fact the classifier makes use of express values, which obscures the consequences. For mentioned names with better cost outcomes. The methods are not suitable for classifying regions with real facts and features. The values of the classifier must be in range, so it's miles vital to specify the first-rate price.

III PROPOSED SYSTEM

The obtained records is first pre-processed the use of filters and system gaining knowledge of layers to get rid of redundant information. And replica records values. It also reduces dimensionality, so the information turns into cleaner. Extra facts later undergoes fragmentation procedure. It is divided into trying out and coaching records sets. For every training and testing session, the version is trained using the dataset. The intention is as follows. Type of crime, year, month; Time, date and region are transformed to integers for clean category. An unbiased have an effect on among the functions are first of all generated the use of a random jump classifier.

Crimes are characterized, making it feasible to research the incidence of a criminal offense at a selected time and place. At the give up, crimes are spatially and temporally frequent. Information filling The model is determined with the aid of calculating the accuracy. The language used to expand predictive fashions is Python and runs analytical and Machine Learning Models.

The proposed set of rules is more appropriate for figuring out crime patterns when it relies upon on a couple of functions. Date and area. This solves the hassle of studying the impact of impartial traits. The advent of gold standard fee isn't required because it considers the actual nominal fee and factors with ok information of the region. The accuracy was excessive in comparison to different fashions of predictive mastering equipment.

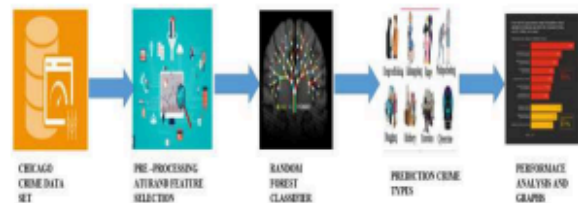


Fig. 1 Architecture Diagram of the Working System

A. Information Series

This is the first real step towards truly constructing a machine learning version for facts series. This is a critical step that relies upon an excellent model, the bigger and higher it is. The more records we get, the higher our adaptation will perform. There are numerous procedures of realities series like textual content scraping, guide intervention etc. Comparison of device mastering algorithms for crime scene prediction taken from Kaggle and a few other resources.

B. Records Series

Let's exchange the statistics. By removing missing records and disposing of a few sections. First license's make a posting of segment names that we really want to save or keep. Then we erase or push off every one of the segments other than those we want to save. At last, we dispose of or eliminate lines with lacking qualities from the dataset.

C. Pattern Analyzing

When making use of machine gaining knowledge of training, we need two units of statistics: one for demonstration and one for trying out. But now we best have one. So, permit's divide it into elements in the ratio of 80:20. Also, let's cut up the facts body into function columns and label columns. Here she went by using educate from Sklearn. Use that data to installation the partition. Also, test_size = zero.2, break up eighty% on card and 20% on check card. After making the model, it's miles vital to test it. To affirm this, we pass test_x Method.

D. Irregular Woodland:

Irregular Backwoods is a popular gadget acquiring information on calculation that belongs to the supervised gaining knowledge of method. It may be utilized for every classification and relapse obligations in ML. It is based on the concept of group studying, which uses a number of classifiers to solve a complex problem and enhance model performance.

IV RESULTS AND DISCUSSION

With the assist of system learning era, it has become easier to identify relationships and styles between unique information. The characteristic of this software is mainly to predict kind of crime will occur if we realize the location of the crime. Using system learning ideas, we evolved a model using the records set, cleansing and reworking the information.

Fig. 2 web page of filling the details about crime location, date and time to predict the crime

Fig. 3 prediction of type of crime

PERFORMANCE ANALYSIS	
Precision and recall	
	Recall Precision
ROBBERY	0.67 0.78
SEX OFFENSE	0.93 0.84
THEFT	0.82 0.92
WEAPONS VIOLATION	0.81 0.65

Fig. 4 Performance analysis for the type of crime



Fig. 5 Performance of crime occurrences with particular day of week

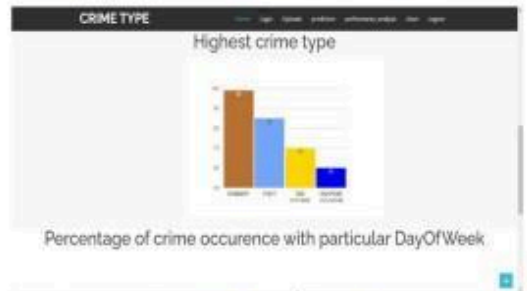


Fig. 6 Graph for percentage of crime occurrences with particular day of week.

A. Decision Tree:

A tree of judgment is a type of tree, shaped like a flowchart, where each internal node represents a personal test result, and leaf nodes represent class topics or performance. Dozens more, the solution can process hundreds of measurements trees. The feature is checked to insert the search given the root node of the tree. Related to the type of tree, the sub-tree to assign the cost is placed at the new node, and then the same could be resubmitted. Crime data is examined and predicted using attributes like longitude (X), latitude (Y), address, Saturday, day (YYYY-mm-

dd:hh:MM:ss), area, clearance, and type. It is said correctly, and to avoid overcrowding, exploration of numerous programs is moved on to the main evaluation.

Fuzzy safety route project accuracy KNN train: 98%. Test: 94%
DT train: 99%. Test: 98%
RF train: 99% Test: 98%

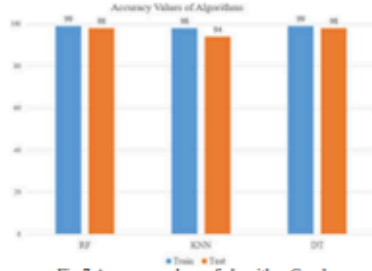


Fig.7 Accuracy values of algorithm Graph

B. Confusion Matrix:

The performance of a set of test data is summarized by a confusion matrix employed in machine learning models. Within this matrix, instances from the model's predictions, both accurate and inaccurate, are displayed—a means of evaluating the model's performance. Commonly used in assessing classification models aiming to predict a categorical label for each input instance, the instances produced by the model on the test data are revealed. True positives (TP) manifest when a positive data point is accurately predicted by the model, and true negatives (TN) emerge when a negative data point is accurately predicted. False positives (FP) occur when a positive data point is incorrectly predicted by the model, while false negatives (FN) arise when a negative data point is incorrectly predicted. Accuracy, a measure of the model's performance, is determined by the ratio of total correct instances to total instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

For the above case:

$$Accuracy = (8+6) / (8+6+1+1) = 14/16 = 0.875$$

Precision

The definition of accuracy with a formula has been observed, and now the definition of precision using a formula is going to be examined.

$$Precision = \frac{TP}{TP + FP}$$

For the above case:

$$Precision = 8 / (8+1) = 8/9 = 0.8888$$

Recall

Recall assesses a classification model's efficacy in recognizing all pertinent instances. Calculated by $TP/(TP+FN)$, it gauges the model's capability to identify positives. High recall: effective; low recall: misses positives.

$$Recall = \frac{TP}{TP + FN}$$

For the above case:

$$Recall = 8 / (8+1) = 8/9 = 0.8888$$

F1-Score

F1 score in machine learning measures precision and recall balance as harmonic mean. Ranges from 0 to 1 and evaluates binary classifiers effectively

$$F1 - Score = \frac{2 \cdot precision \cdot Recall}{Precision + recall}$$

For the above case:

$$F1-Score = (2 * 0.8888 * 0.8888) / (0.8888 + 0.8888) = 0.8888$$

Table 1 : Precision table

	precisio n	recall	f1- score	support
0	1.00	1.00	1.00	86
1	0.82	0.74	0.78	19
2	0.97	0.98	0.97	147
3	1.00	1.00	1.00	140
4	1.00	1.00	1.00	19
5	1.00	1.00	1.00	3

C. Training:

The predictive models, including KNN, Decision Tree (DT), and Random Forest (RF), were assessed during the training phase. Optimal performance for KNN and DT was revealed as error rates were compared across different K values.

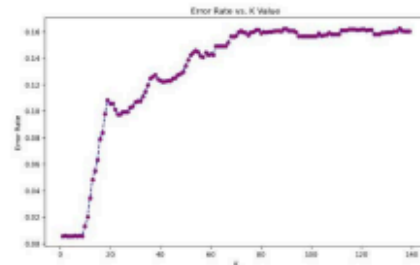


Fig.8.Error Rate vs K value graph

D. Testing:

In testing Critical insights were revealed through the analysis of feature importance. The impact of various factors on crime prediction was assessed by the model, emphasizing key features and their relative importance in determining accurate outcomes for crime mapping.

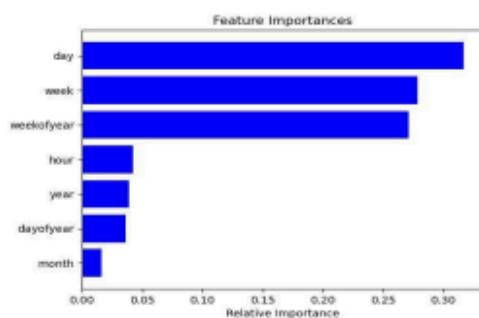


Fig.9 feature and Relative importance char

V CONCLUSION

Crime from small cities to huge cities. They've crimes different Categories - Robbery, Murder, Kidnapping, Assault, Battery, False Imprisonment, Kidnapping, Manslaughter. As soon because the crime happens cases need to be resolved promptly.

The combination of geospatial data further refines forecasts, empowering policing to proactively assign assets and execute designated wrongdoing avoidance methodologies. This combination of fluffy rationale and geospatial examination holds significant potential for propelling the accuracy and adequacy of wrongdoing expectation models, at last encouraging more secure metropolitan conditions.

The future enhancements are Crimes occur in various categories, ranging from small cities to large cities - Robbery, Murder, Kidnapping, Assault, Battery, False Imprisonment, Kidnapping, and Manslaughter. As soon as a crime occurs, cases need to be promptly resolved. In Input it is reported that crimes precise location details are encompassed by geospatial data, utilizing latitude and longitude information, with temporal data factoring in the date and time. In Output, Forecasted crime categories are generated by the system, predicting probable crime classifications in a given geospatial context.

ACKNOWLEDGMENT

We extend our sincere gratitude to Praveen Karthik .T for their invaluable contributions to this research endeavor. Their expertise, guidance, and dedication have greatly enriched our work. We also acknowledge the support and resources provided by Sathyabama institute of science and technology

REFERENCES

- [1] Saraiva, M., Matijo saitien'e, I., Mishra, S. & Amante A. Crime prediction and monitoring in Porto, Portugal, using machine learning, spatial and text analytics. ISPRS International Journal of Geo-Information 11 (7), 400 (2022)
- [2] Shah, N., Bhagat, N. & Shah, M. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art 4, 1–14 (2021).
- [3] Kshatri, S. S. et al. An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach. IEEE Access 9, 67488–67500 (2021)
- [4] Janiesch, C., Zschech, P. & Heinrich, K. Machine learning and deep learning. Electronic Markets 31 (3), 685–695 (2021)
- [5] Raza, D. M. & Victor, D. B. Data mining and region

prediction based on crime using random forest, 980–987 (IEEE, 2021).

[6] W. Safat, S. Asghar, and S. A. Gillani, "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques," IEEE Access, vol. 9, 2021

[7] N. Jin, Y. Zeng, K. Yan, and Z. Ji, "Multivariate air quality forecasting with nested long short term memory neural network," IEEE Trans. Ind. Informat., vol. 17, Dec. 2021

[8] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," Sep. 2020

[9] Y. Rayhan and T. Hashem, "AIST : An interpretable attention-based deep learning model for crime prediction," 2020

[10] Hossain, S., Abtahec, A., Kashem, I., Hoque, M. M. & Sarker, I. H. Crime prediction using spatio-temporal data, 277–289 (Springer, 2020).

[11] Elluri, L., Mandalapu, V. & Roy, N. Developing machine learning based predictive models for smart policing, 198–204 (IEEE, 2019).

[12] Meijer, A. & Wessels, M. Predictive policing: Review of benefits and drawbacks. International Journal of Public Administration 42 (12), 1031–1039 (2019)

[13] J. Sangani, A., Sampat, C. & Pinjarkar, V. Crime prediction and analysis (2019).

[14] Sivanagaleela, B. & Rajesh, S. Crime analysis and prediction using fuzzy c-means algorithm, 595–599 (IEEE, 2019)

[15] Catlett, C., Cesario, E., Talia, D. & Vinci, A. Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. Pervasive and Mobile Computing 53, 62–74 (2019) .

[16] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhav B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) .2018.

[17] H.Benjamin Fredrick David and A.Suruliandi, Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.

[18] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making, 2017

[19] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015

[20] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.