Crime Rate Analysis and Hotspot Prediction

A Credit Project Report

Bachelor of Technology
in
Computer Science Engineering

by

Diya Soni
(19EUCCS018)

Jiya Verma
(19EUCCS029)

Under the Supervision of

Mr. Dhirendra Singh

Department of Computer Science Engineering

RAJASTHAN TECHNICAL UNIVERSITY, KOTA

Kota - 324010, India

Declaration

I hereby declare that the project work, which is being presented to the Department of Computer Science and Engineering, Rajasthan Technical University Kota, entitled as **"Crime Rate Analysis and Hotspot Prediction"** was carried out and written by us with our correct and complete knowledge carried under the guidance of **Dr. C.P. Gupta**, Professor, Department of Computer Science and Engineering and supervised by **Mr. Dhirendra Singh**, Department of Computer Science and Engineering, Rajasthan Technical University Kota.

The results contained in this report have not been submitted in part or in full to any other University or Institute for the award of any degree or diploma to the best of our knowledge.

Mr. Dhirendra Singh
Department of Computer Science and Engineering
Rajasthan Technical University, Kota

Diya Soni(**19EUCCS018**)
Jiya Verma(**19EUCCS029**)
Department of Computer Science and Engineering,
Rajasthan Technical University, Kota

Date:

Certificate

This is to certify that the final semester project, entitled as **"Crime Rate Analysis and Hotspot Prediction"** has been successfully carried out by **Diya Soni (Enrollment No. 19EUCCS018)** and **Jiya Verma (Enrollment No. 19EUCCS029)**, under my guidance partially fulfilling the criteria of **"Bachelors of Technology in Computer Science and Engineering"** from the Department of Computer Science and Engineering, Rajasthan Technical University Kota, for the academic year 2019-23.

Mr. Dhirendra Singh
Department of Computer Science and Engineering
Rajasthan Technical University, Kota

Diya Soni(**19EUCCS018**)
Jiya Verma(**19EUCCS029**)
Department of Computer Science and Engineering,
Rajasthan Technical University, Kota

Date:

Acknowledgements

I am thankful to my project's supervisor **Mr. Dhirendra Singh** for his continuous support, conviction, encouragement, and invaluable advice in credit project work. I also like to thank **Dr. C.P. Gupta** for helping me throughout the literature review and presentation preparation process.

Diya Soni (19EUCCS018)
Jiya Verma (19EUCCS029)

Abstract

Crime always had an adverse effect on our society and daily lives but due to lack of technology and without the availability of proper structured data it was impossible to predict and analyze Crime pattern. Now with the advancement in machine learning technology and with the help of analyzing algorithms it has now become feasible to analyze the crime pattern and predict the crime.

In this article, we present a comprehensive study with an experimental analysis of machine learning approaches for crime rate analysis, using this our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. The results of the analysis and hotspot prediction will be presented in a visually intuitive manner, such as mapping crime incidents and highlighting hotspot area on a map of the city.

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of machine learning information can be extracted with the help of previously unknown, useful information from an unstructured data.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Crime is increasing considerably day by day. Increase in Crime rate is one of the major issues our society is facing in daily life. With advancement in Technology and unpredictable behavior of criminals, it is becoming difficult to predict crime patterns.

Major crimes occurring on everyday basis are like kidnapping, theft, murder, rape, gambling, etc. Law enforcement agencies collect crime data information with the help of local police authorities and organize and store the information as record. Crime is highly unpredictable due to its randomness and non-uniformity.

But the occurrence of any crime is often correlated to factors like poverty and employment. Due to this rapid increase in crime rates the analysis of crime has become a necessity. Crime analysis basically consists of study of procedures and methods that intend to reduce the risks associated with crime. Its practical approach aimed at identifying and analyzing crime patterns. But due to huge volumes of crime data available, it becomes difficult to analyze the crime data efficiently.

Here our Analysis comes into picture and with computation support to make the analysis process more robust and effective. Advanced systems need to be incorporated in place of traditional and old systems for predicting crimes because traditional methods fail when crime data has higher dimensions and is rather very complex in nature. Therefore a better crime prediction and analysis tool was needed for identifying modern crime patterns effectively. This paper introduces some methodologies with the help of which it can be predicted at what place and time which type of crime has a higher probability of occurrence. Classification helps in extracting features and predicting future trends in crime data based on similarities.

Methodologies used in this study are, K-Neighbors Classifier, Extra Class classifier, Decision Tree Classifier, Support Vector Machines (SVM) and Artificial Neural Network (ANN).

1.2 Problem Statement

As we are aware of the increasing crime rate in today's scenario. Government is facing many challenges to solve the crimes. With the increasing population and crime rate also increasing, it's a serious issue for the government to make strategic decisions ,make laws for controlling it and maintaining law and order in a country . It's the government's duty to ensure the safety of citizens of the country. The use of Machine learning and deep learning algorithms in modern crime analysis would make the task easier and problems would be solved in more efficient and in less time.

1.3 Goal

Much of the current work is focused in two major directions:

- Predicting surges and hotspots of crime
- Understanding patterns of criminal behavior that could help in solving criminal investigations

1.4 Objective

The objective of our work is to:

- Predicting the most probable type of crime which tends to takes place at particular location and time.
- Predicting hotspots of crime.
- Analyzing crime pattern.
- Classify crime based on location.

1.5 Methodology

In the past couple of decades it has been used as a most common tool in nearly any task that requires information extraction from large datasets. Machine learning is also extensively used in scientific operations similar to bioinformatics, drug, and astronomy. One common point of all of these operations is that, in variation to more traditional uses of computers, in these cases, due to the complexity of the patterns

that need to be detected, an individual programmer cannot give a crystal clear, fine detailed specification of how similar tasks should be executed. Taking illustration of previous records as reference , numerous expertise decisions are acquired and are learned during training and processing from our experience (rather than following crystal clear instructions given to us). Machine learning tools are concerned with endowing programs with the capability to learn and acclimatise.

The inputs to our algorithms are time(hour, day, month, time), place(latitude and longitude)

Classes of crime are:

- Act 379- Theft
- Act 13- Gambling
- Act 279 Accident
- Act 323- Violence
- Act 302- Murder
- Act 363- Kidnapping

The output would be the most probable crime to occur on the basis of previous record. We try out multiple bracket algorithms, such as KNN (K- Nearest Neighbors), Decision Trees, Logistic regression, and Random Forest for verifying their results.

1.6 Thesis Organization

The thesis work has been organized in the following chapters:

- Chapter 1: This chapter presents an introduction, on "Crime", "Machine Learning Algorithms" and "Methodology", for crime rate analysis and hotspot prediction. Subsequently, problem formulations, objectives and goals for the research are mentioned. ˆ
- Chapter 2: This chapter presents a literature review that details a comprehensive overview of the existing body of knowledge and research related to crime rate analysis, hotspot detection, and predictive modeling. It serves to establish the theoretical foundation of our study and demonstrates our familiarity with the relevant literature.
- Chapter 3: This chapter on 'data collection and preprocessing' provides a detailed account of the data acquisition process, including the sources of data and the steps taken to prepare the data for analysis.
- Chapter 4: The chapter includes focusing on examining and understanding the characteristics and patterns present in the collected data. It involves applying statistical techniques and visualizations to gain insights into the dataset.
- Chapter 5 : The chapter on predictive modeling focuses on the development and evaluation of models for crime rate prediction. It involves applying machine learning algorithms and statistical techniques to build models that can forecast future crime rates based on the available data.
- Chapter 6: This chapter presents the outcomes of the crime rate analysis and hotspot prediction project. It focuses on summarizing and interpreting the results obtained from the data analysis, modeling, and evaluation stages.

- Chapter 7: This chapter presents the summary of the research and the future research directions, for this research work.

Chapter 2

Literature Review

Researchers working on the project "Crime Rate Analysis and Hotspot Prediction" typically follow a systematic and iterative approach to ensure the effectiveness and accuracy of their analysis and predictions. They provide valuable insights into the application of various techniques, algorithms, and data sources for understanding crime patterns, identifying hotspots, and predicting future crime rates. The literature review synthesizes and discusses these works to establish a foundation for the project and highlight the gaps and opportunities for further research.

- In 2014, Bogomolov, Andrey, et al. implemented "Once upon a crime: Towards crime prediction from demographics and mobile data." which focuses on the prediction of crime using demographics and mobile data. It explores the potential of utilizing these data sources to forecast crime patterns. The study presents insights into the relationship between demographic factors and crime rates, as well as the role of mobile data in crime prediction.[1]

- In 2001 , Breiman, Leo. Implemented "Random Forests." This influential paper introduces the Random Forest algorithm, a popular ensemble learning

technique used for classification and regression tasks. It discusses the principles and advantages of Random Forests in handling high-dimensional data and dealing with complex relationships. The study provides insights into the application of Random Forests in crime rate analysis and prediction.[2]

- In 2002, Friedman, Jerome H. implemented "Stochastic gradient boosting." which focuses on the stochastic gradient boosting algorithm, a powerful machine learning technique used for predictive modeling. It provides an in-depth explanation of the algorithm and its application in various domains, including crime prediction. The study highlights the advantages and effectiveness of stochastic gradient boosting in predictive modeling tasks.[3]

- In 2008 , Kianmehr, Keivan, and Alhajj, Reda. Implemented "Effectiveness of support vector machine for crime hot-spots prediction." which examines the effectiveness of support vector machine (SVM) in predicting crime hotspots. It discusses the application of SVM as a machine learning algorithm for identifying locations with high crime rates. The study highlights the performance and capabilities of SVM in crime hotspot prediction.[4]

- In 2011 , Toole, Jameson L., et al. implemented "ACM Transactions on Intelligent Systems and Technology." that focuses on the analysis of urban crime patterns using mobile phone data. It explores the potential of utilizing mobile phone data to understand crime dynamics, spatial patterns, and hotspot identification. The study provides insights into the integration of mobile phone data with crime analysis techniques.[5]

- In 2013, Wang, Tong, et al. implemented "Machine Learning and Knowledge Discovery in Databases." that discusses the application of machine learning

techniques in knowledge discovery and crime analysis. It explores how machine learning algorithms can be applied to crime data to identify patterns, correlations, and predict future crime rates. The study provides insights into the use of machine learning in crime rate analysis.[6]

- In 2011, Yu, Chung-Hsien, et al. implemented "Crime forecasting using data mining techniques." which discusses the application of data mining techniques in crime forecasting. It explores how various data mining algorithms can be utilized to analyze crime data and predict future crime rates. The research provides an understanding of the effectiveness of different data mining techniques in crime prediction.[7]

Chapter 3

Data Collection and Pre-processing

3.1 Data Source

The dataset we have taken from the Kaggle website. The various crimes have been uploaded with the particular data, the crimes are being taken only in the India. (https://www.kaggle.com crime -analysis-in-India/data) The dataset is now supplied to machine learning model on the basis of this dataset the model is trained. In the first step of accumulating information, data from previously, current datasets from online sources are gathered together. These datasets are merged to form a common dataset, on which analysis will be done.

3.2 Dataset description

Below image shows the description of dataset. We have taken location(latitude and longitudes), timestamps (hour, weekdays, weekyear, dayofyear, month, year).

timestamp	act379	act13	act279	act323	act363	act302	latitude	longitude
28-02-2018 21:00	1	0	0	0	0	0	22.73726	75.87599
28-02-2018 21:15	1	0	0	0	0	0	22.72099	75.87608
28-02-2018 10:15	0	0	1	0	0	0	22.73668	75.88317
28-02-2018 10:15	0	0	1	0	0	0	22.74653	75.88714

FIGURE 3.1: Dataset After Preprocessing

3.3 Training Dataset

Training data is raw data used to train the machine learning model. Training dataset are fed to machine learning algorithms to teach them how to make predictions or perform a desired task. There are three steps of train the dataset-

1. Feed – feed the data to machine learning model.
2. Tag – tag the data with desired output. Model changes it to text vectors – it's a number that shows data features.

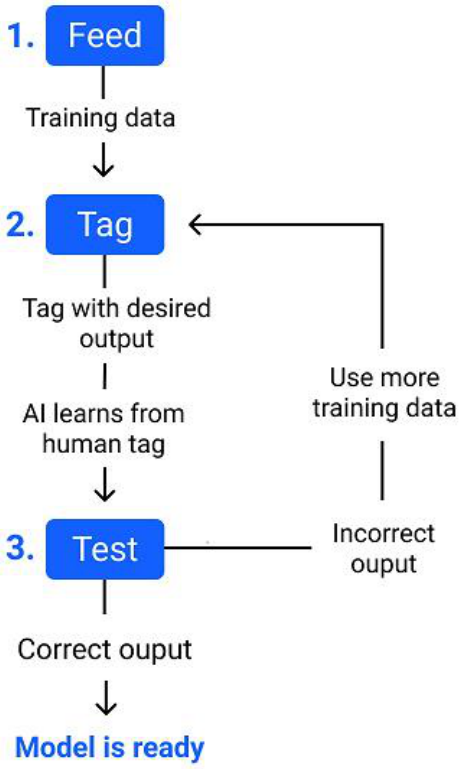

FIGURE 3.2: Training of Dataset

Train Test Split

Use train_test_split to split your data into a training set and a testing set.

```
In [82]: X=df1.iloc[:,[6,7,8,9,10,11,12,13,14,15]].values

In [83]: X

Out[83]: array([['22.73726', '75.875987', 2, ..., 59, 2, 1],
                ['22.720992', '75.876083', 2, ..., 59, 2, 1],
                ['22.736676', '75.883168', 2, ..., 59, 2, 1],
                ...,
                ['22.531931', '75.769126', 7, ..., 184, 1, 3],
                ['22.719569', '75.857726', 7, ..., 184, 1, 3],
                ['22.686437', '76.032055', 7, ..., 184, 1, 3]], dtype=object)
```

FIGURE 3.3: Input Features

3. Test- now test the model by feeding it testing data. Algorithms trained to associate feature vectors with tags based on manually tagged samples, then learn to make predictions when processing unseen data.

3.4 Testing the Data Set

In the Machine Learning models which are most popularly used nowadays one of the important and very difficult tasks is to decide the ratio of training and testing data. It highly depends on the dimension and nature of the dataset used. The presence of noisy values and missing values are cross checked before dividing a dataset. The improper ratio leads to challenges like over fitting and underfitting.If the training data contains all values which are significantly close to testing data values, then it will face the problem of over fitting and if in case the training dataset is very small and considers a lesser number of values for training, then it faces the problem of under fitting. We divided our dataset into the ratio 80:20 to be used for training and testing respectively and predicted results of classifier are compared with results of the testing data set values for estimating validation accuracy.

```
In [84]: y=df1.iloc[:,[0,1,2,3,4,5]].values

In [85]: y

Out[85]: array([['1', '0', '0', '0', '0', '0'],
                ['1', '0', '0', '0', '0', '0'],
                ['0', '0', '1', '0', '0', '0'],
                ...,
                ['0', '0', '1', '0', '0', '0'],
                ['0', '0', '1', '0', '0', '0'],
                ['0', '0', '1', '0', '0', '0']], dtype=object)
```

FIGURE 3.4: Output Results

3.5 Data preprocessing techniques:

3.5.1 Data Cleaning

1. Identification and Handling of Missing Values:

    - Missing value detection: Missing values were identified by examining each variable in the dataset. This was done using descriptive statistics and visualization techniques.
    - Removal: In cases where a significant portion of a variable's data was missing or the missingness was random, the corresponding records or variables were removed from the dataset to ensure data integrity.

2. Data Formatting and Standardization

    - Date and Time Formatting: The date and time variables were standardized to a consistent format to ensure accurate temporal analysis. This involved converting different date and time representations (e.g., MM/DD/YYYY, YYYY-MM-DD, HH:MM AM/PM) into a standardized format (e.g., YYYY-MM-DD HH:MM).
    - Geographical Standardization: Geographic variables, such as latitude and longitude, were checked for consistency and formatted uniformly to facilitate spatial analysis. Any inconsistencies or errors in the coordinates were addressed through geocoding or by cross-referencing with reliable sources.

3.5.2 Feature Engineering

1. Feature Selection:

- Statistical Significance: Statistical tests such as correlation analysis (e.g., Pearson correlation) or chi-square tests were used to identify features that exhibited a strong relationship with the target variable. Features with low statistical significance were excluded from further analysis.
- Domain Knowledge: Expert knowledge or domain-specific insights were leveraged to identify features that were deemed relevant in explaining crime rates and hotspot prediction. This involved collaborating with law enforcement professionals or crime analysts to understand the factors known to influence crime patterns.

Chapter 4

Exploratory Data Analysis

In the Exploratory Data Analysis (EDA) phase of the project, the following analyses were conducted to gain insights into the crime data:

4.1 Summary Statistics

- Class Distribution: The distribution of crime types or categories was examined to identify the prevalence of different types of crimes in the dataset.
- Frequency Analysis: The frequency of crime occurrences over time (e.g., hourly, daily, monthly) was analyzed to detect any patterns or anomalies.

4.2 Visualizations

- Histograms: Histograms were used to visualize the distribution of numerical variables, providing insights into the data's skewness, central tendency, and spread.

- Bar Charts: Bar charts were employed to display the frequency or proportion of different crime types, helping to identify the most common or rare occurrences.

- Box Plot: A box plot, also known as a box-and-whisker plot, is a graphical representation of the distribution of a variable or multiple variables in a dataset. It provides a summary of the data's central tendency, spread, and skewness.

- Pair Plot: A pair plot, also known as a scatterplot matrix, is a visualization technique that displays the pairwise relationships between multiple variables in a dataset. It is particularly useful when working with numerical variables.Pair plots provide insights into the correlations, patterns, and potential outliers in the data. They help identify any linear or nonlinear relationships between variables and can be used to guide further analysis or modeling decisions.

- Time Series Plots: Time series plots were utilized to visualize the trends and seasonality in crime rates over time, allowing for the identification of temporal patterns.

- Heatmaps: Heatmaps were used to illustrate the spatial distribution of crimes by plotting them on a map, providing a visual representation of crime hotspots.

4.3 Spatial and Temporal Analysis:

- Spatial Clustering: Spatial clustering algorithms, were applied to group similar crime locations together and identify crime hotspots or high-density areas.

- Kernel Density Estimation (KDE): KDE was used to estimate the intensity of crime occurrences across the geographical area, visualizing areas with higher crime rates as density peaks.

- Temporal Patterns: Patterns of crime occurrence were analyzed based on various time dimensions, such as hourly, daily, weekly, or seasonal trends. This analysis helps identify peak crime hours, days, or periods.

4.4 Correlation Analysis

- Correlation Matrix: A correlation matrix was computed to measure the pairwise relationships between numerical variables. It helps identify potential correlations and dependencies that exist in the dataset.
- Heatmap: A heatmap of the correlation matrix was created to visually represent the strength and direction of the relationships between variables.

These EDA techniques allowed for a comprehensive understanding of the crime data, revealing important patterns, correlations, and factors influencing crime rates.

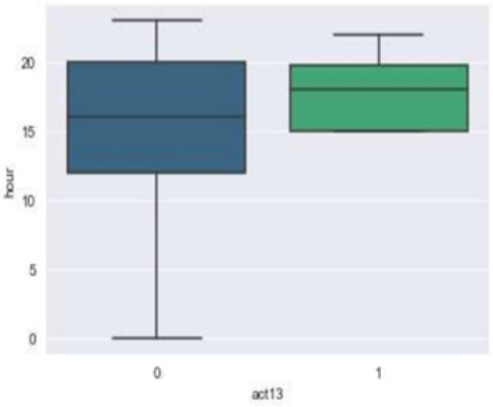

Fig 5-Crime analysis Plot of gambling

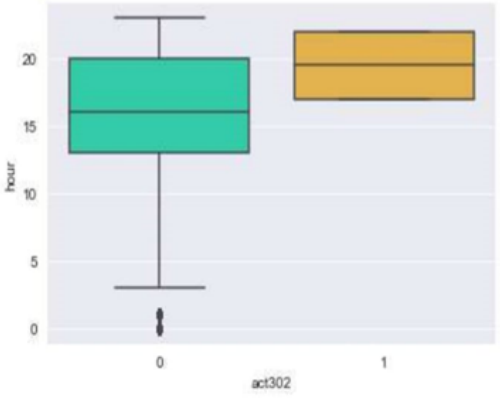

Fig 6-Crime analysis Plot of murder

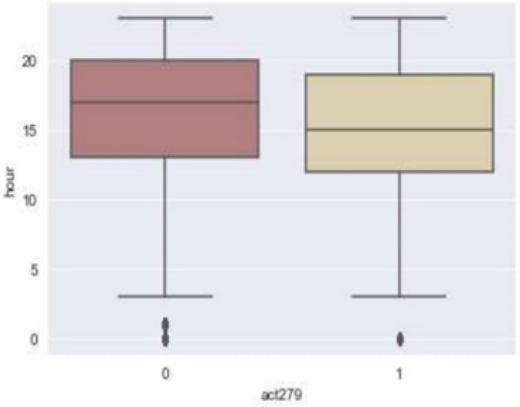

Fig 7-Crime analysis Plot of gambling

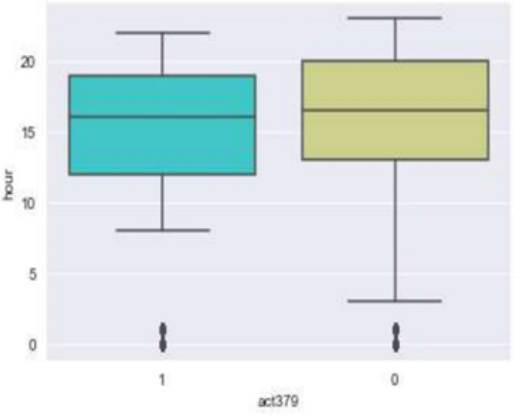

Fig 8-Crime analysis Plot of gambling

FIGURE 4.1: Results after Analysis

FIGURE 4.2: Pair Plot

Chapter 5

Predictive Modelling

Implementation

This project is implemented using machine learning, with the help of python language a machine is build which would use different parameter to generate the analysis of crime rate and predict the most probable crime which might occur in a particular location and timestamp. The dataset used here is publically available and extracted from kagglewebsite, which is first processed and converted into suitable format to generate the desired outcome having different features vector like timestamp, latitude, longitude which would acts as input for the machine. Various algorithms were applied in the machine and was trained against training dataset and then results of the accuracy was generated using testing data

For proper functioning and results various algorithms were used. Following are the algorithms used:

5.1 KNN (K-Nearest neighbors)

It is Classification algorithm based on supervised learning which uses k nearest neighbor used in pattern recognition which uses similarity measurement technique to classify the new sample pointbased on previously classified data points in its neighbor, k in knn refers to number of nearest neighbors in majority voting process.

Some frequently used distance functions.

Camberra :

$$d(x,y) = \sum_{i=1}^{m} \frac{|x_i - y_i|}{|x_i + y_i|} \quad (2)$$

Minkowsky :

$$d(x,y) = \left(\sum_{i=1}^{m} |x_i - y_i|^r \right)^{1/r} \quad (3)$$

Chebychev :

$$d(x,y) = \max_{i=1}^{m} |x_i - y_i| \quad (4)$$

Euclidean :

$$d(x,y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (5)$$

Manhattan / city - block :

$$d(x,y) = \sum_{i=1}^{m} |x_i - y_i| \quad (6)$$

FIGURE 5.1: Distance Function

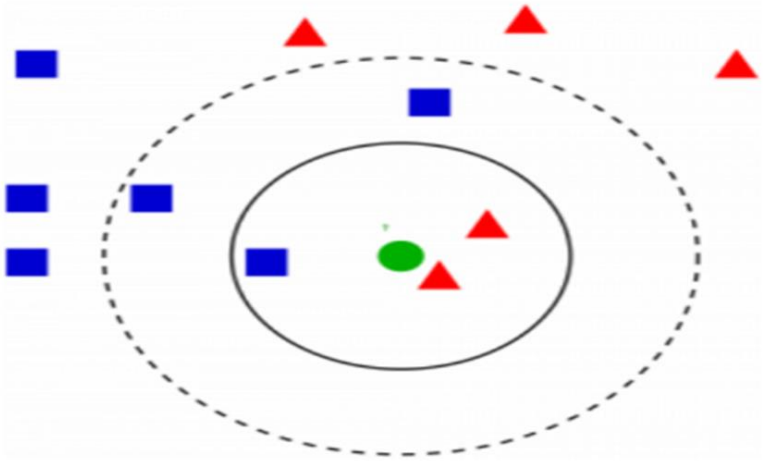

FIGURE 5.2: Example of KNN

5.2 Decision Tree

It is a type of predictive learning algorithm which is used in both classification and regression, this algorithm works intuitively,in which test on attribute is represented by a node and outcome is defined by branch and each leaf node represents a class label,decision on each node is taken where each node share parent child relation. Splitting is done till stage is reached with homogenous subset

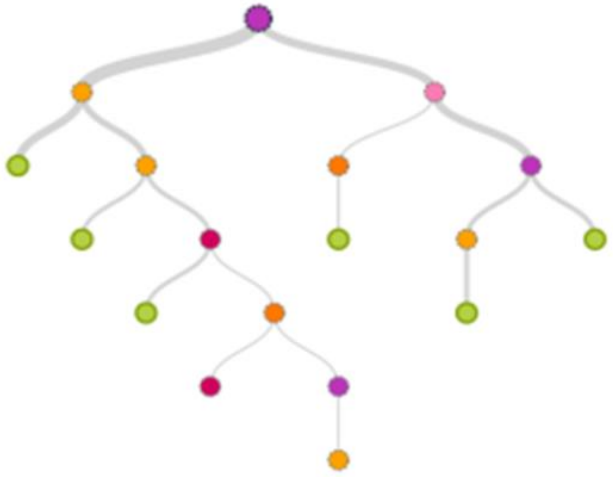

FIGURE 5.3: Decision Tree

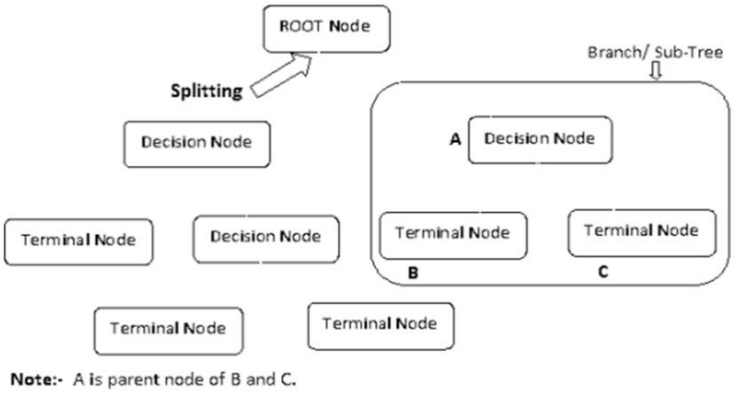

FIGURE 5.4: Example of Decision Tree

5.3 Random forest

Random forest algorithm uses randomness to generate minimum variance, it is based on learning method that combine various classifier to make best prediction on test data.

A random forests classifier is an ensemble classifier, which aggregates a family of classifiers h(x—1),h(x—2),..h(x—k). Each member of the family, h(x—), is a classification tree and k is the number of trees chosen from a model random vector.

$$y = \text{argmax}_{p \in \{h(x_1)..h(x_k)\}} \{ \sum_{j=1}^{\kappa} (I(h(x|\theta_j) = p)) \}$$

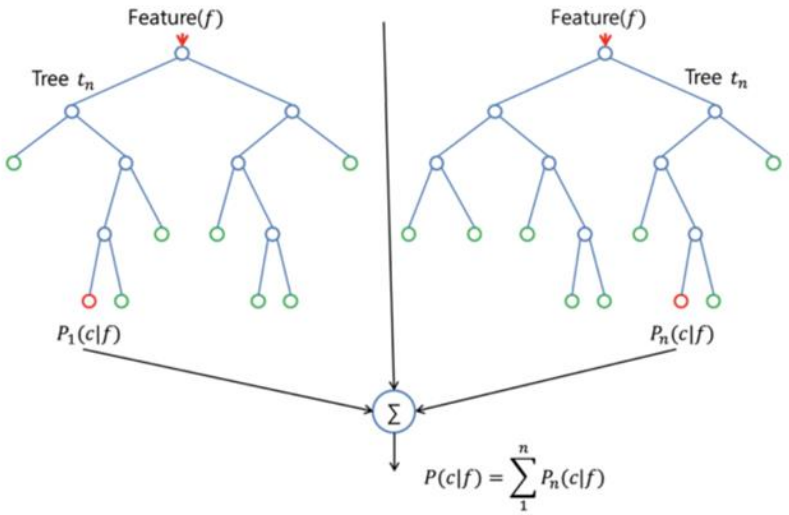

FIGURE 5.5: Example of Random Forest

5.4 SVM - Support Vector Machines

SVM algorithm is mainly used in classification problems ,it is an type of supervised learning which separates the data in two categories creating a N-dimensional hyper plane, svm uses extreme points to create this hyper plane, these points act as support vectors and thus this algorithm is called as support vector machine. hyper plane is the most accurate decision boundary which can be used to separate the data points. Below diagram shows the svm algorithm separating two different classes.

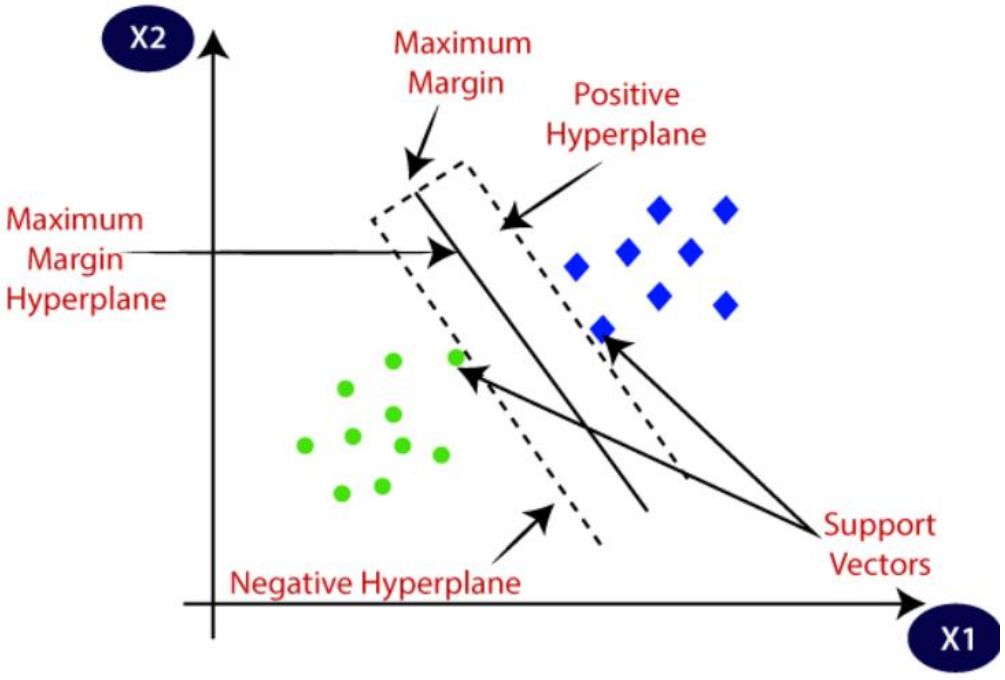

FIGURE 5.6: SVM Classifier

5.5 Artificial Neural Network

ANN is inspired from functioning of neurons in brains which process information in form of electrical signal. Ann uses artificial neurons which has inputs and produces

single output, where feature act as input and the desired result is obtained using output of neurons which is calculated using weighted sum of inputs and passing them to activation function. To adjust the weight assigned for each connection back propagation is used which uses error from the training data and obtained results, and assigning new weights on basis of cost function with given state of respected connection.

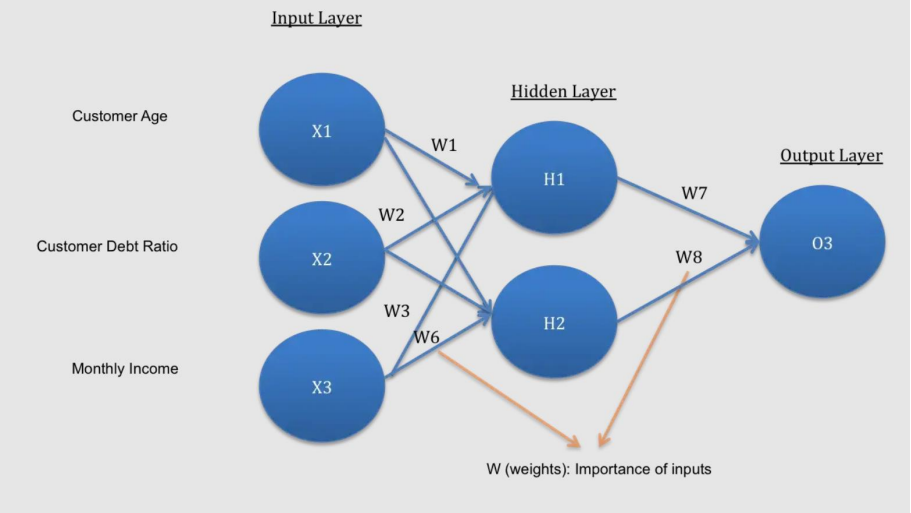

FIGURE 5.7: Architecture of ANN

Chapter 6

Results and Findings

- Evaluation of the developed predictive models for crime rate prediction.
- Presentation of model performance metrics
- Evaluation of the effectiveness and accuracy of the hotspot detection approach employed in the project.
- Comparison of different predictive models and determination of the most reliable model for crime rate prediction.
- Mapping and visualization of identified crime hotspots or high-density areas.

6.1 Result of Classifiers

Result of all classifier are shown below :

Result of all the classifier

```
In [109]: print('KNN Classifier -',classifier.score(X_test,y_test))

          print('Random Forest Classifier -',Rclassifier.score(X_test,y_test))

          print('SVM Classifier -',svc_model.score(X1_test,y1_test))

          print('Decision Tree Classifier -',Dclassifier.score(X_test,y_test))
```

```
KNN Classifier - 0.9710144927536232
Random Forest Classifier - 0.9806763285024155
SVM Classifier - 0.7922705314009661
Decision Tree Classifier - 0.9710144927536232
```

FIGURE 6.1: Result of Classifiers

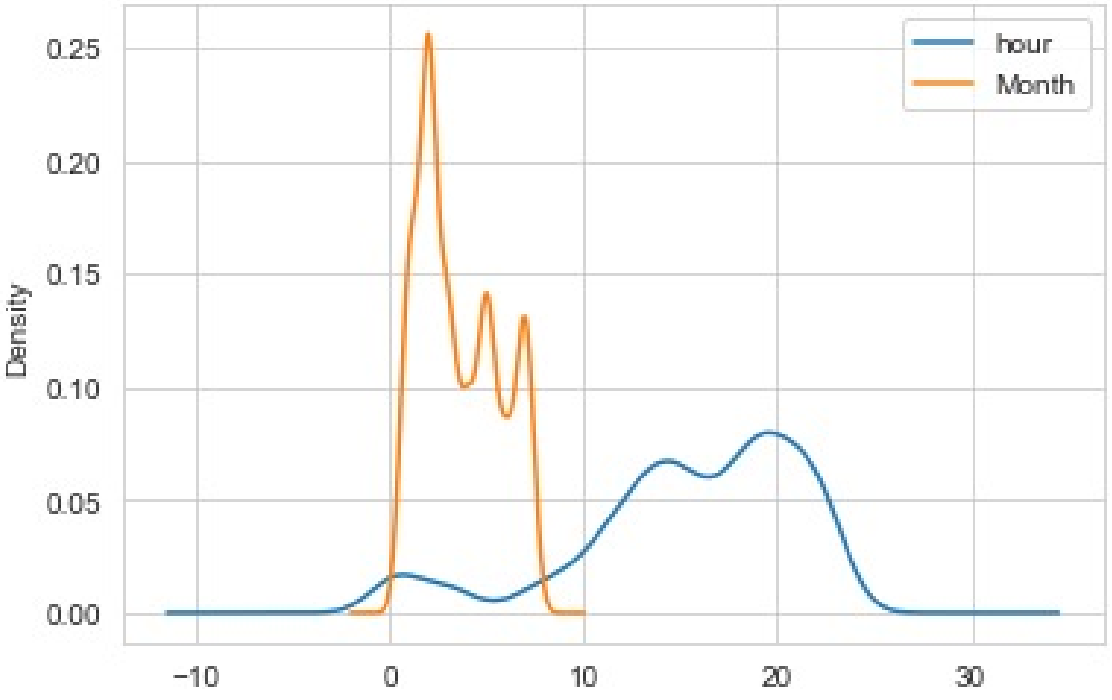

FIGURE 6.2: KDE Plot

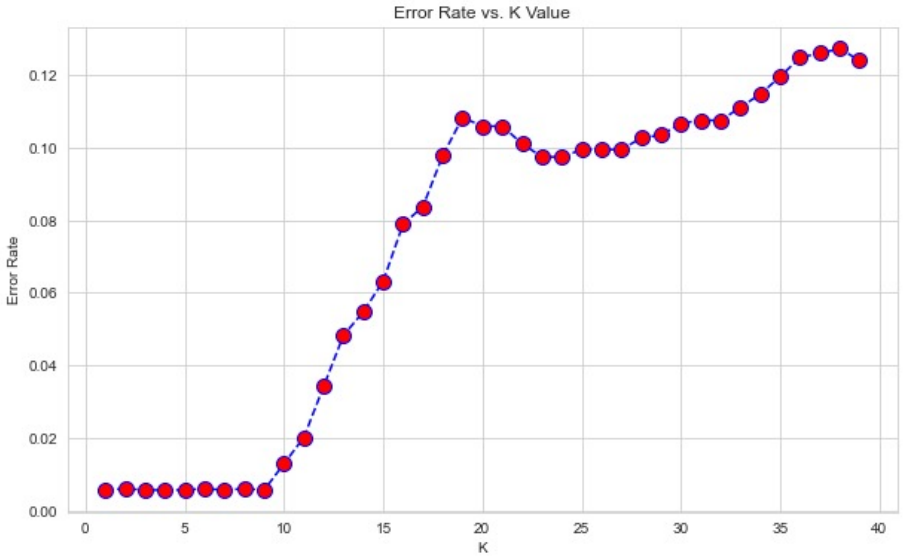

FIGURE 6.3: Estimating K values

	timestamp	act379	act13	act279	act323	act363	act302	latitude	longitude
1	28-02-2018 21:00	1	0	0	0	0	0	22.73726	75.875987
2	28-02-2018 21:15	1	0	0	0	0	0	22.720992	75.876083
3	28-02-2018 10:15	0	0	1	0	0	0	22.736676	75.883168
4	28-02-2018 10:15	0	0	1	0	0	0	22.746527	75.887139
5	28-02-2018 10:30	0	0	1	0	0	0	22.769531	75.888772

Fig 2-Original dataset

	act379	act13	act279	act323	act363	act302	latitude	longitude	Month	weekdays	hour	year	weekofyear	Dayofyear	weekday	quarter
1	1	0	0	0	0	0	22.73726	75.875987	2	28	21	2018	9	59	2	1
2	1	0	0	0	0	0	22.720992	75.876083	2	28	21	2018	9	59	2	1
3	0	0	1	0	0	0	22.736676	75.883168	2	28	10	2018	9	59	2	1
4	0	0	1	0	0	0	22.746527	75.887139	2	28	10	2018	9	59	2	1
5	0	0	1	0	0	0	22.769531	75.888772	2	28	10	2018	9	59	2	1

Fig 3-Data after processing

FIGURE 6.4: Data Processing

6.2 Hotspot Identification

Visualize the data on map of city using Folium. Here we plots points on Open Street Map with indicators for the type of event that occurred.

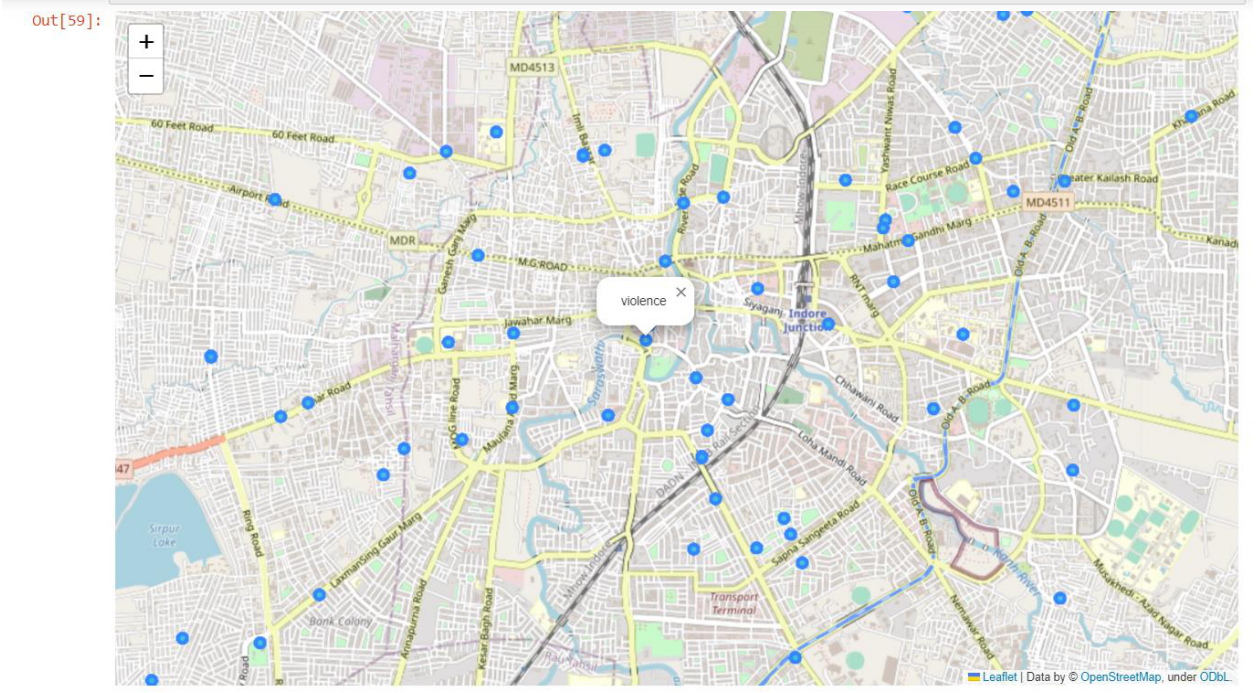

FIGURE 6.5: Map of Indore

While these indicators are useful and we can zoom in on the map to see them in detail, an additional heat map would improve the visibility of high-frequency areas where multiple indicators may overlap.

Out[60]:

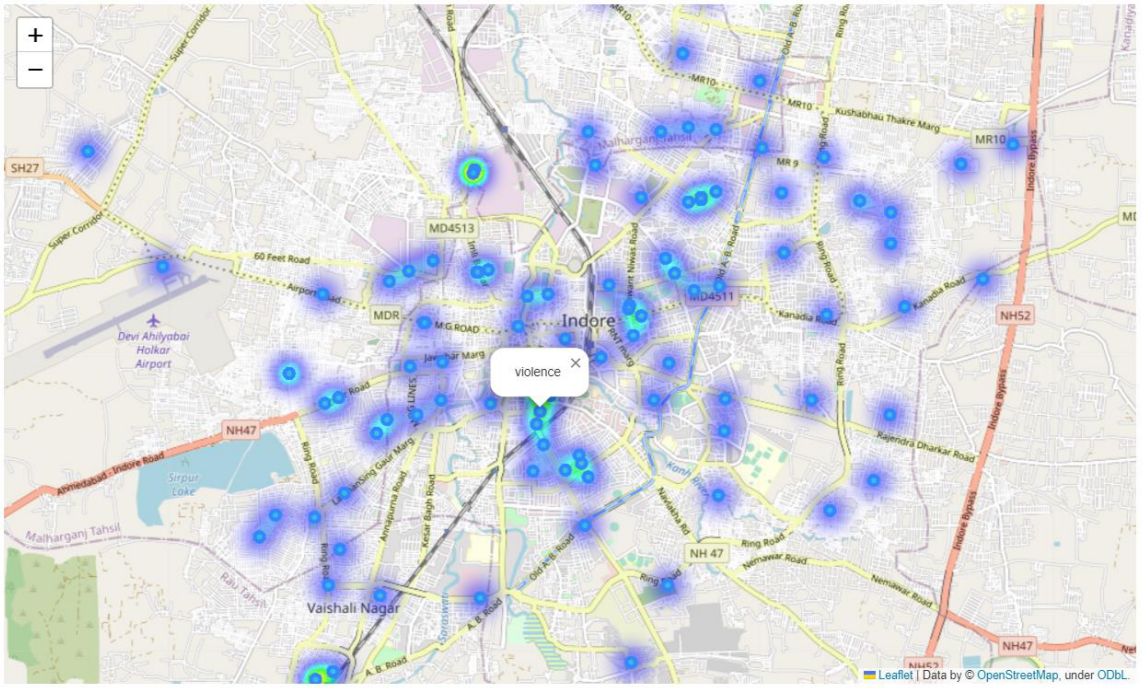

FIGURE 6.6: Heat Map

6.3 Applications

- Public safety and protection related to crime and better understanding of crime is beneficial in multiple ways.
- It can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime and more concerted efforts by citizens and authorities to create healthy neighborhood Environment
- With the advent of the big data era and the availability of fast, efficient algorithms for data analysis.
- Understanding patterns in crime from data is an active and growing field of research.

Chapter 7

Conclusions

7.1 Study Summary

The purpose of this study is to carry out crime analysis and its prevention through the application of data mining methods. The results obtained as a part of the implementation of the algorithm have revealed that these methods have a good application in crime prediction. Classification carried out by using a Random forest classifier as a data mining classification method has classified crime data at an accuracy rate of 98.06 . Furthermore, the viability of decision trees exceeds any other classifier because it specifies the results explicitly. These rules are easily understandable in human terms.

Machine learning and data mining can play an important role in crime analysis due to their decision-making power and increase the computational strength of the process. It is very important to make accurate decisions in order to achieve crime

prevention and law enforcement strategies. This can open doorways for law enforcement agencies to ensure better control over crimes and predict them accurately and efficiently.

7.2 Scope for Future Work

As a part of the future extension of this study, more models will be created for predicting the hot spots of crime so that police can be deployed to these locations at the right time.

The changes in the behaviour of the algorithm will also be recorded as more data gets added. We also plan into developing an algorithm that can help predict the social link networks of criminals, gangs, and suspects.

References

[1] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Once upon a crime: Towards crime prediction from demographics and mobile data. 09 2014.

[2] Leo Breiman. Machine learning, volume 45, number 1 - springerlink. *Machine Learning*, 45:5–32, 10 2001.

[3] Jerome Friedman. Stochastic gradient boosting. *Computational Statistics Data Analysis*, 38:367–378, 02 2002.

[4] Keivan Kianmehr and Reda Alhajj. Effectiveness of support vector machine for crime hot-spots prediction. *Applied Artificial Intelligence*, 22:433–458, 05 2008.

[5] Jameson L. Toole, Nathan Eagle, and Joshua B. Plotkin. Spatiotemporal correlations in criminal offense records. *ACM Trans. Intell. Syst. Technol.*, 2(4), jul 2011.

[6] Vislavath Srinath T.Prathima, Madan Vijay Karnati. Crime prediction using machine learning. *International Journal of Engineering and Techniques (IJET)*, 7, 2021.

[7] Jacky Yu, Max Ward, Melissa Morabito, and Wei Ding. Crime forecasting using data mining techniques. pages 779–786, 12 2011.