CRIME DATA ANALYSIS

Mr. B.S.Murthy [1], G.Rajesh,

[1]**Assistant professor(HOD) , MCA DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- suryanarayanamurthy.b@gmail,com
[2]**PG Student of MCA, Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- geddamrajesh100@gmail.com

ABSTRACT

Recognizing the patterns of a criminal activity of a place is paramount in order to prevent it . Law enforcement agencies can work effectively and respond faster if they have better knowledge about crime patterns in different geological points of a city. The aim is to use machine learning with the help of python to classify a criminal incident by type depending on its occurrence at a given time and location. The survey is done using crime records of Sanfrancisco.

Crime is one of the biggest and dominating problem in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. This requires keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future.

1 INTRODUCTION

Criminal activities are present in every region of the world affecting quality of life and socio-economical development. As such, it is a major concern of many governments who are using different advanced technology to tackle such issues. Crime Analysis, a sub branch of criminology, studies the behavioral pattern of criminal activities and tries to identify the indicators of such events. Machine learning agents work with data and employ different techniques to find patterns in data making it very useful for predictive analysis. Law enforcement agencies use different patrolling strategies based on the information they get to keep an area secure. A machine learning agent can learn and analyze the pattern of occurrence of a crime based on the reports of previous criminal activities and can find hotspots based on time, type or any other factor. This technique is known as classification and it allows to predict nominal class labels. Classification has been used on many different domains such as financial market, business intelligence, healthcare, weather forecasting etc.

Literature Survey

1.Kamoun et. reviews the defensive usage of AI/MLS in cybersecurity and then presents a survey of its offensive use. Inspired by the System-Fault-Risk (SFR) framework, we categorize AI/MLSpowered cyberattacks by their actions into seven categories. We cover a wide spectrum of attack vectors, discuss their practical implications and provide some recommendations for future research.

2.Fatima Dakalbab et. al investigates AI strategies in crime prediction. They conduct a systematic literature review (SLR). This review evaluates the models from numerous points of view, including the crime analysis type, crimes studied, prediction technique, performance metrics and evaluations, strengths and weaknesses of the proposed method, and limitations and future directions. They review 120 research papers published between 2008 and 2021 that cover AI approaches for crime prediction.

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

The dataset used for this is real and authentic. The dataset is acquired from UCI machine learning repository website. The title of the dataset is 'Crime and Communities'. It is prepared using real data from socio-economic data from 1990 US Census, law enforcement data from the 1990 US LEMAS survey and crime data from the 1995 FBI UCR. This dataset contains a total number of 147 attributes and 2216 instances.

Proposed System & alogirtham

From a large list of attributes, only eighteen attributes are chosen for Exploratory Data Analysis. The chosen attributes are namely state, HousVacant, PctHouseOccup, PctHouseOwnCC, PctVacantBoarded, PctVacMore6Mos, PctUnemployed, PctEmploy, murdperPop, rapesperPop, robbperPop, assaultperPop, burglperPop, larcperPop, autoTheftperPop, arsonsperPop, nonviolperpop and violentcrimesperpop.

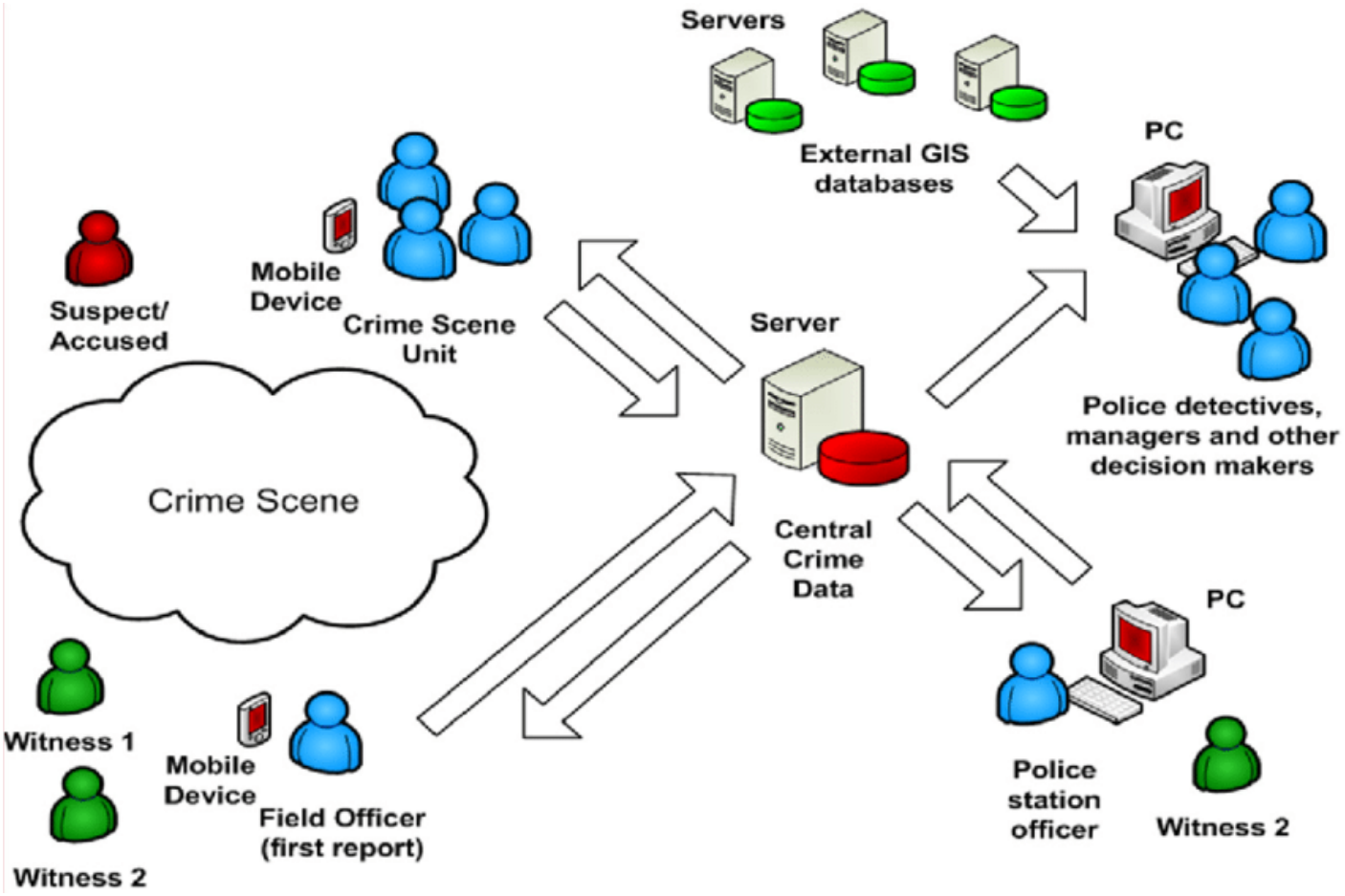


Fig:3.1 System Architecture

IMPLEMENTATION

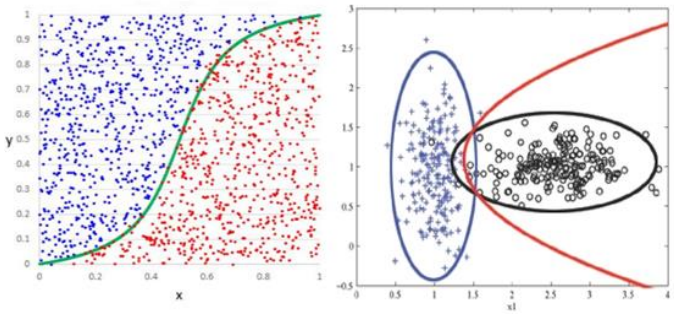
MODULES

1. **Decision Tree Classifier:** Decision tree classification model forms a tree structure from dataset. Decision tree is built by dividing a dataset into smaller pieces. At each step in the algorithm, a decision tree node is splitted into two or more branches until it reaches leaf nodes. Leaf nodes indicates the class labels or result. At each step, decision tree chooses a feature that best splits the data with the help of two functions: Gini Impurity and Information Gain. Gini Impurity measures the probability of classifying a random sample incorrectly if the label is picked randomly according to the distribution in a branch

.

Fig 4.4 Decision Tree

2. **Gaussian Naive Bayes:** Gaussian Naive Bayes is a supervised classifier that uses naive assumption that there is no dependency between two features. This classifier is implemented by applying Bayesian Theorem.



5 RESULTS AND DISCUSSION

SCREEN SHORTS

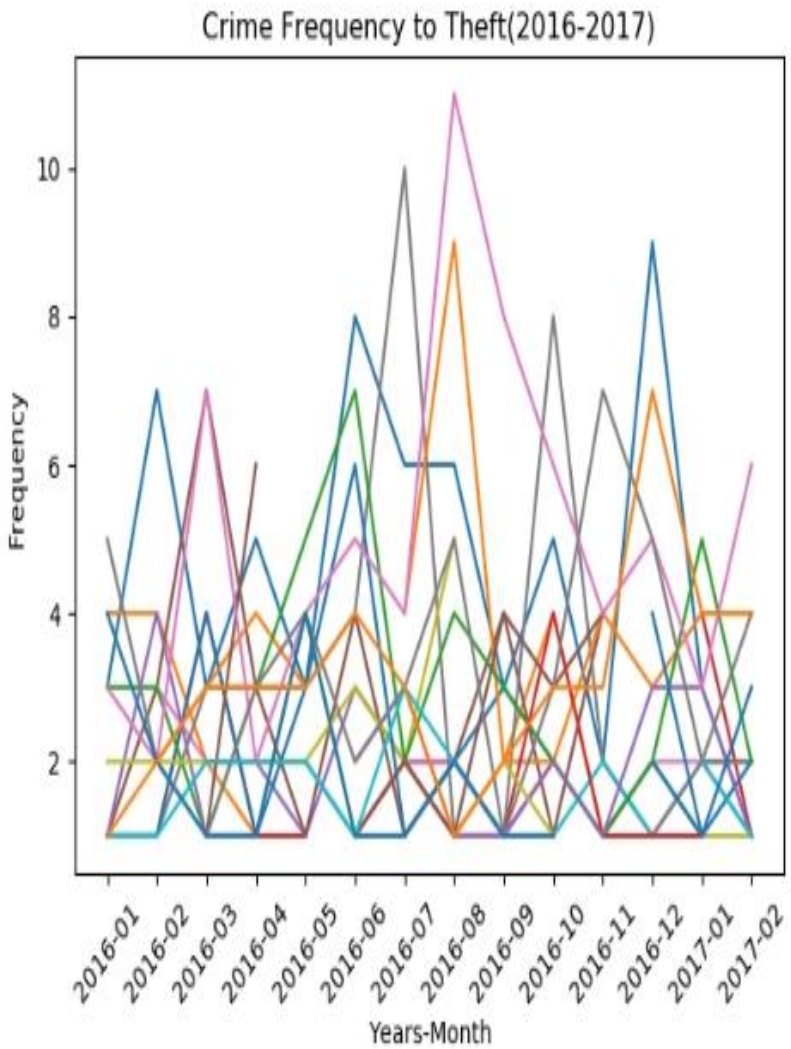


FIG 5.1

CRIME FREQUENCY TO THEFT 2016 - 2017

http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#ix-indexer-is-deprec

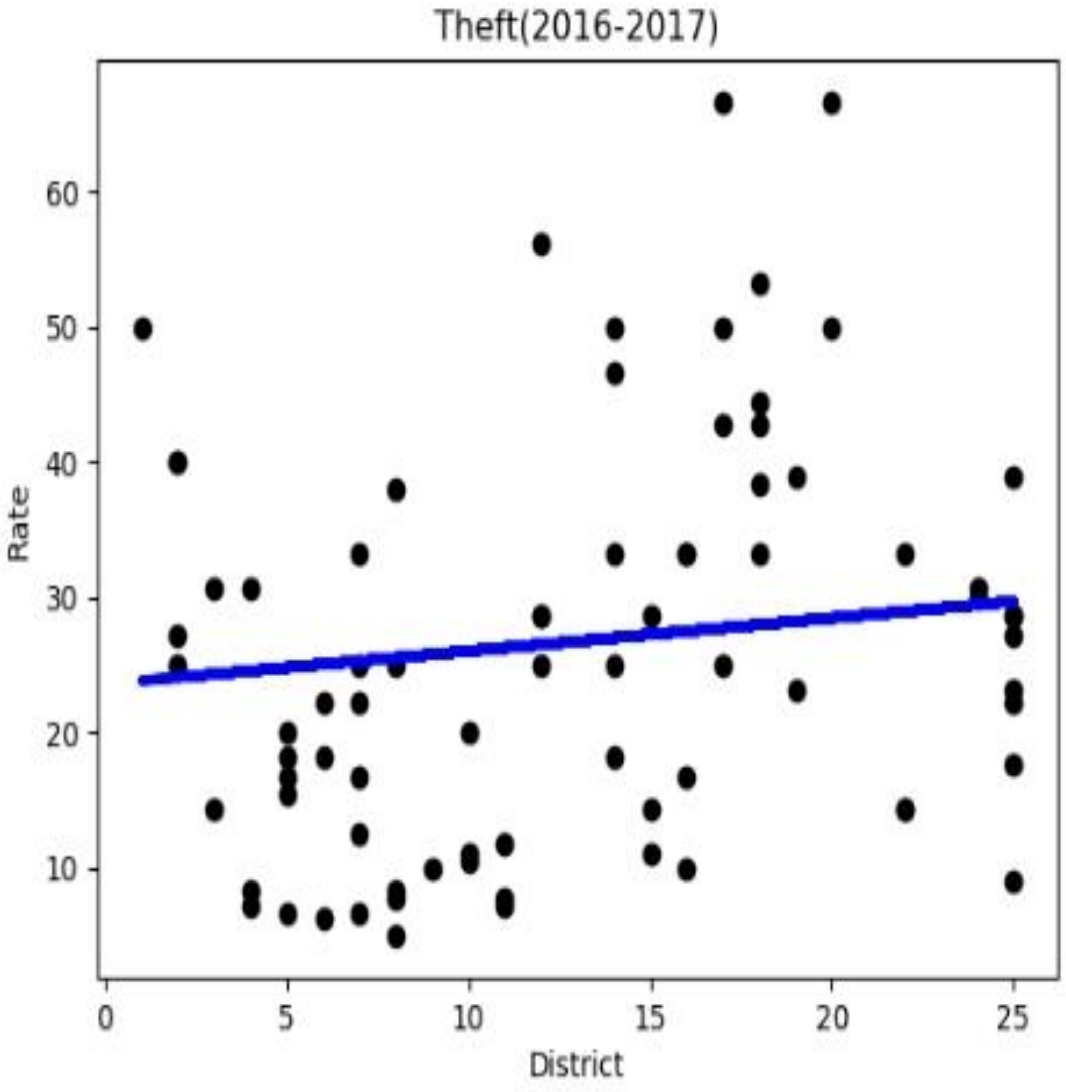


FIG 5.2 THEFT 2016-2017

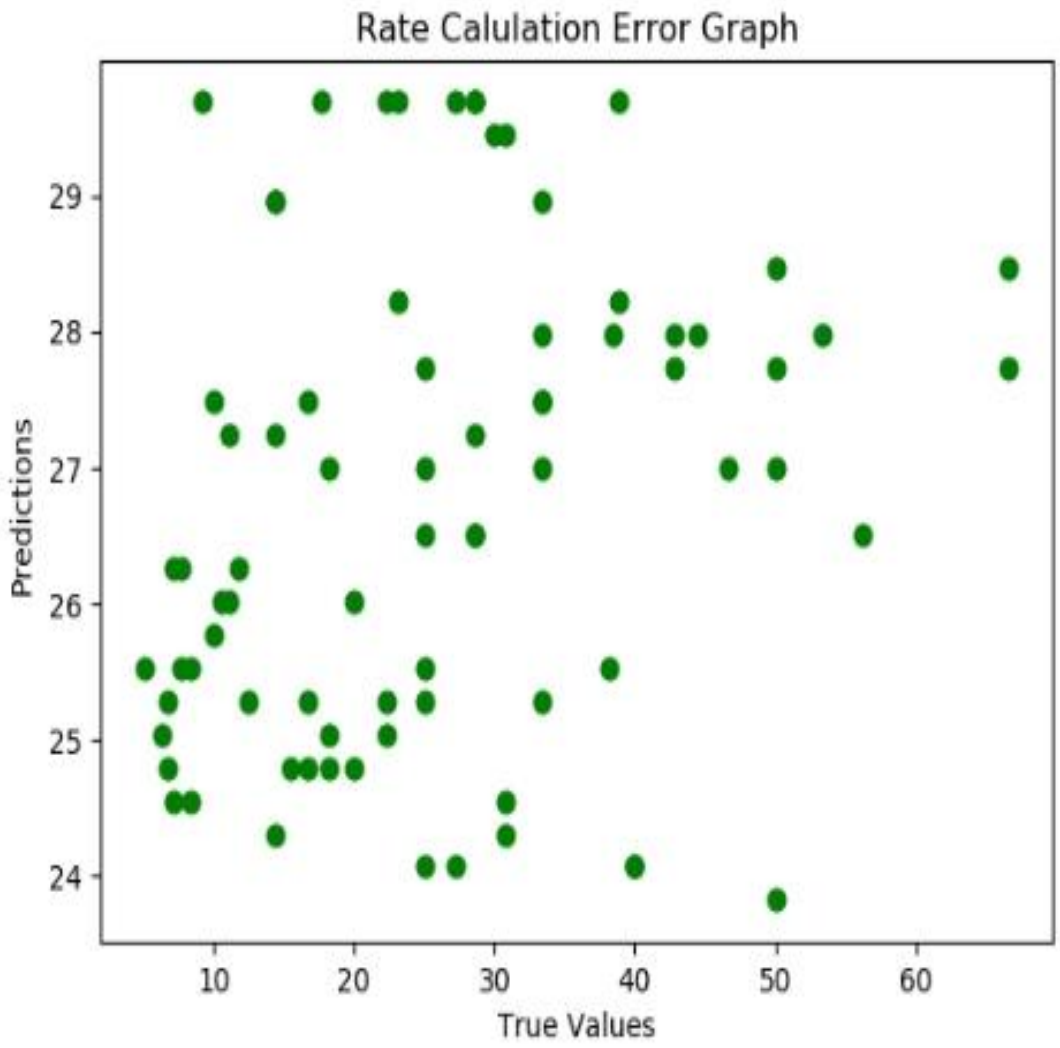


FIG 5.3 RATE CALCULATION ERROR GRAPH

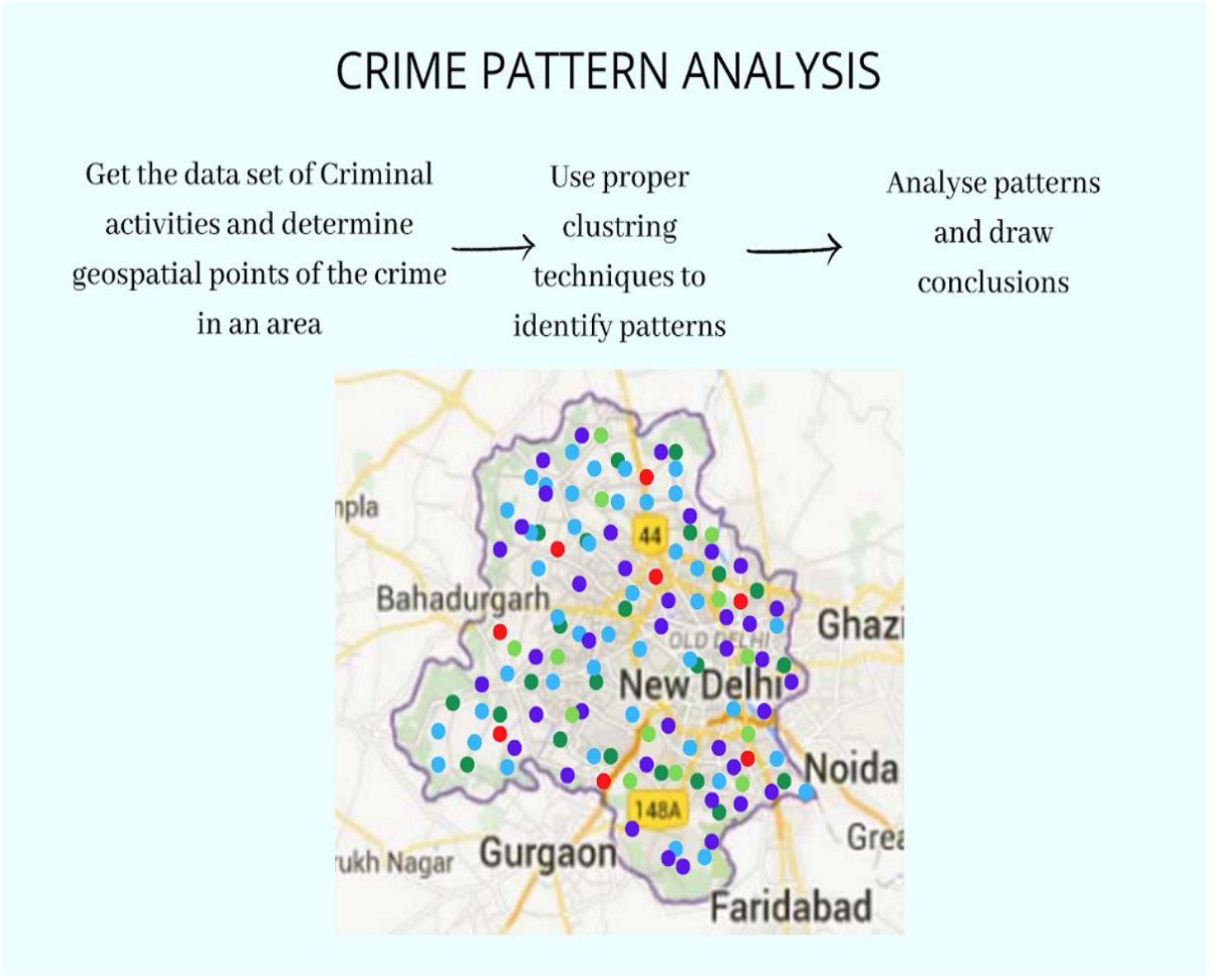
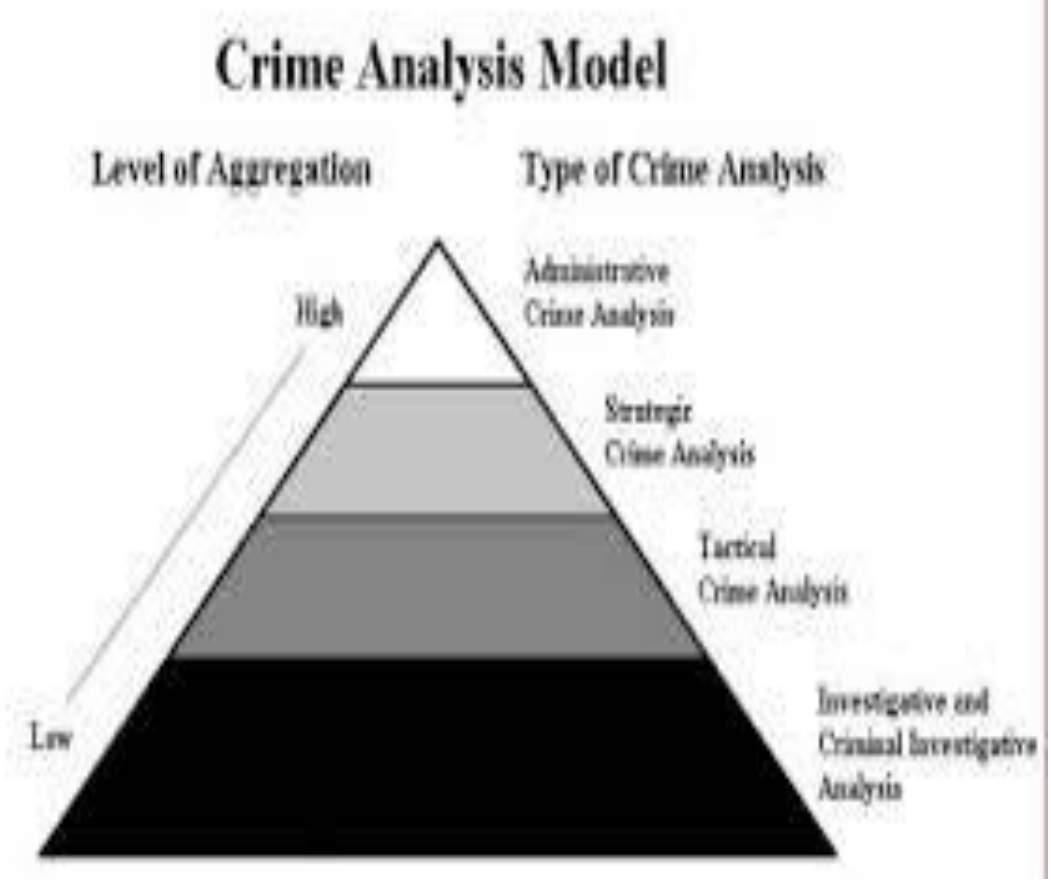


FIG 5.4 CRIME PATTERN ANALYSIS

FIG 5.5



CRIME ANALYSIS MODEL

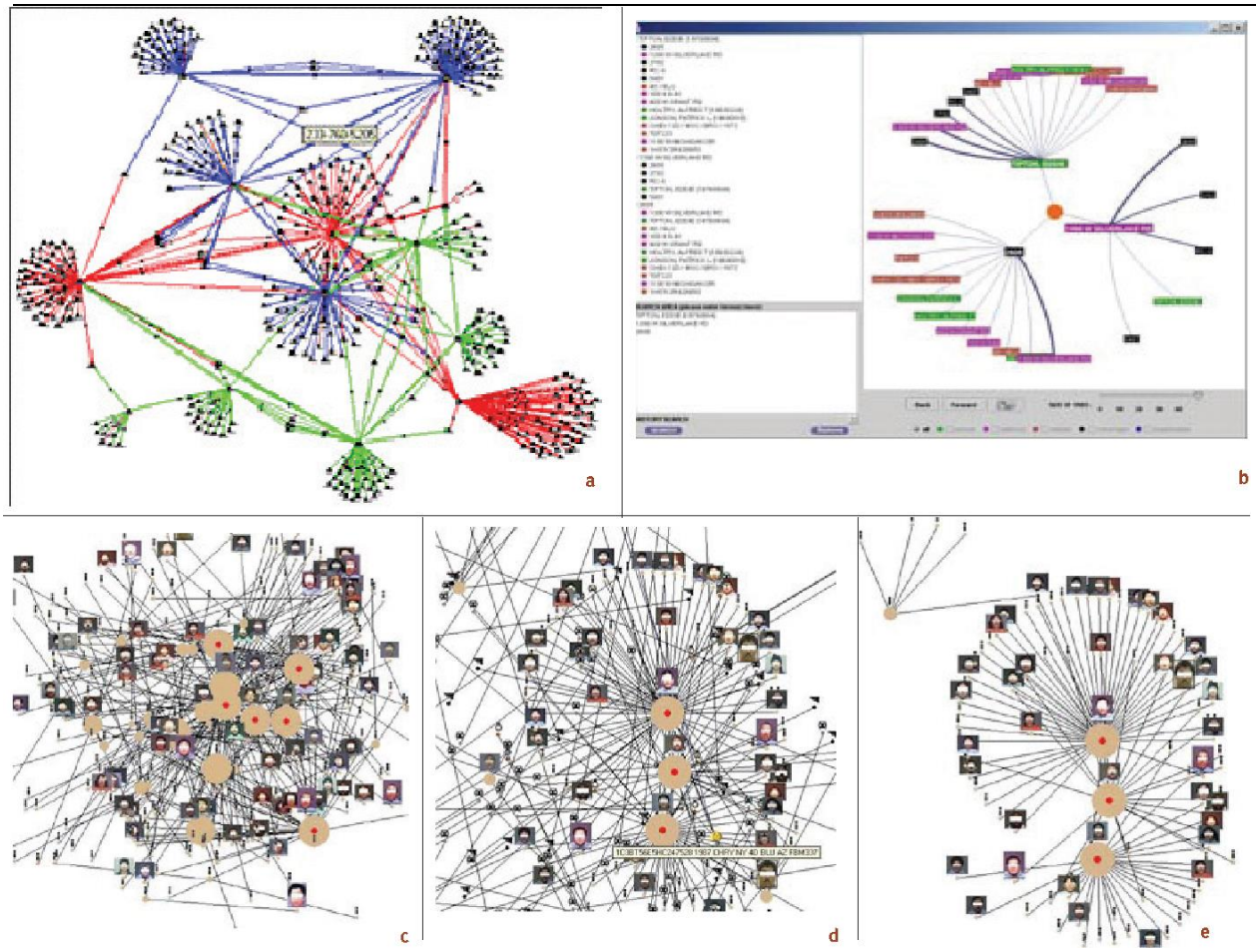


FIG 5.6 CRIMINAL NETWORK ANALYSIS

ıber of violent crimes has been in flux.

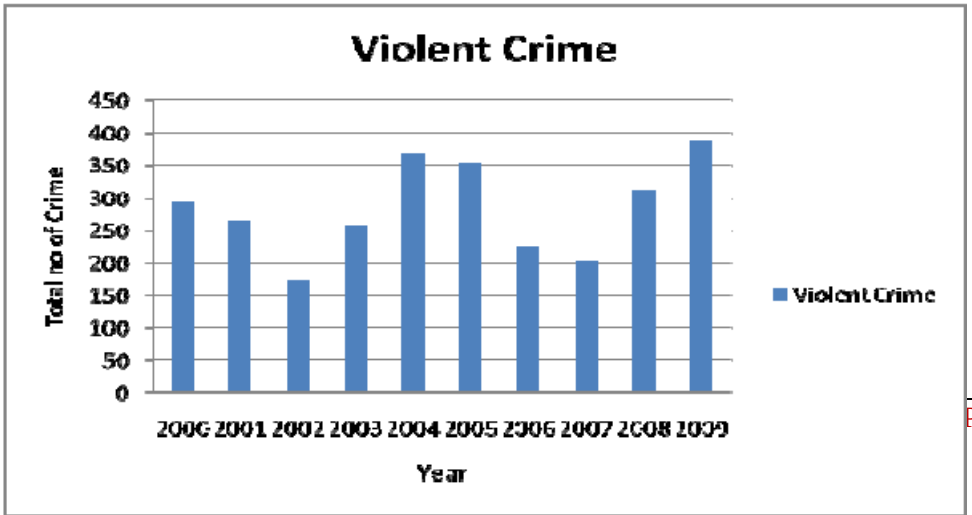


FIG 5.7 VIOLENT CRIME IN YEARS

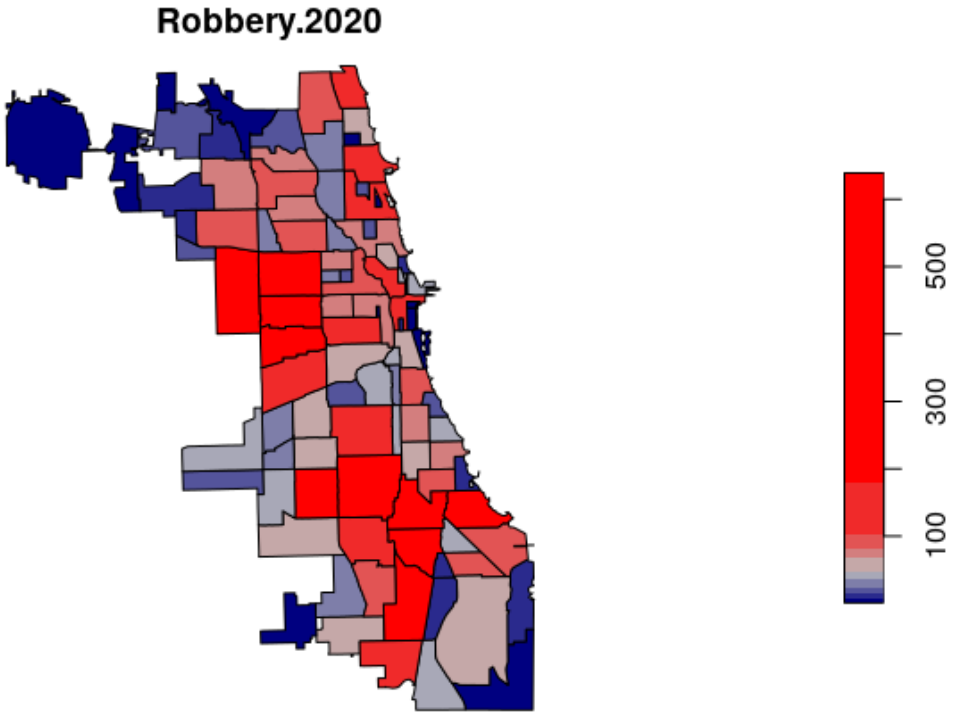


FIG.5.8 ROBBERIES 2020

6. CONCLUSION AND FUTURE WORK

CONCLUSION

Throughout the research it has been evident that basic details of a criminal activities in an area contains indicators that can be used by machine learning agents to classify a criminal activity given a location and date. Even though the learning agent suffers from imbalanced categories of the dataset, it was able to overcome the difficulty by oversampling and undersampling the dataset. Through the experiments, it can be seen the imbalanced dataset was benefitted by using ENN undersampling. Using the undersampled data, Adaboost decision tree successfully classified criminal activities based on the time and location. With a accuracy of 81.93%, it was able to outperform other machine

learning algorithms. Imbalanced classes are one of the main hurdles to achieve a better result. Though the machine learning agent was able to predictive model out of simply crime data, a demographic dataset would probably help to further improve the result and solidify it.

7. REFRENCES

1. Ginger Saltos and Mihaela Coacea, An Exploration of Crime Prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.

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

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. 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.

6. RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.

7. Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

8. Dr .M.Sreedevi, A.Harsha Vardhan Reddy, ch.Venkata Sai Krishna Reddy, Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology ,2018.

9. K.S.N .Murthy, A.V.S.Pavan kumar, Gangu Dharmaraju, international journal of engineering, Science and mathematics, 2017.

10. Deepika k.K, Smitha Vinod, Crime analysis in India using data mining techniques , International journal of Engineering and technology, 2018.