**CSE3505- Foundation of Data Analytics**

***J Component – Project Report***

***CRIME ANALYSIS AND PREDICTION***

*By*

20BCE1375 R. Shruthi

B.Tech CSE with Specializaton

*Submitted to*

**Dr.R.Priyadarshini,**

Assistant Professor Senior,

SCOPE, VIT, Chennai

**School of Computer Science and Engineering**

****

*November 2022*

****

*BONAFIDE CERTIFICATE*

Certified that this project report entitled “Crime Analysis and Prediction” is a bonafide work of R. Shruthi 20BCE1375 who carried out the J-component under my supervision and guidance. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for award of any degree or diploma and the same is certified.

**Dr.R.Priyadarshini,**

Assistant Professor Senior,

SCOPE, VIT, Chennai

**INDEX**

|  |  |
| --- | --- |
| 1 | Title |
| 2 | Bonafide |
| 3 | Contents |
| 4 | Abstract |
| 5 | Introduction |
| 6 | Literature Review |
| 7 | Architecture |
| 8 | Design |
| 9 | Model Implementation |
| 10 | Performance Analysis |
| 11 | Conclusion |
| 12 | Future Scope |

**ABSTRACT**

Crime Analysis is very important for a constitution. With growing technologies and population, crime has almost become a part of our living. Sometimes people make this crime as their job to earn a living. Pit-pocket, thievery, car theft, drug peddling has become a part of today’s times. Crime analysis is a methodical approach for identifying and analysing 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 analytics, we can analyse 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 analysed 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 project, we have the efficiency to display the trend of crime in every specific district in our country, we can also predict the future crime trend in our country. Previous year crimes in every district is analysed and trends are predicted from it. Some crimes can be systematic while some may be very rare. The trend differs from state to state and analysing them can help the government in increasing security and awareness in maximum crime districts.

There are a lot of department these days like the police department, the CBI, the RAW , intelligence and many other cybercrime departments which use this data analysis to detect where maximum crime is happening and how to reduce it or catch criminals in these prominent areas. In today’s world security is an aspect which is given higher priority by all political and government worldwide and aiming to reduce crime incidence. As data mining is the appropriate field to apply on high volume crime dataset and knowledge gained from data mining approaches will be useful and support police force. Crime analysis is a function of identifying the patterns and analysing the rapidly increasing crime trends. It helps in highlighting the events and incidents that may need further investigation. This information helps police departments and other crime departments to solve the crimes faster and effectively. The challenge faced by police departments is difficulty in analysing the large volume of crime data and criminal behaviour. In order to identify the patterns in crime Data analytics plays an important role.

**INTRODUCTION**

In present scenario criminals are becoming technologically sophisticated in committing crime and one challenge faced by intelligence and law enforcement agencies is difficulty in analysing large volume of data involved in crime and terrorist activities therefore agencies need to know technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement. So appropriate field needs to be chosen to perform crime analysis and as data analysis refers to extracting knowledge from large amounts of data, data mining is used here on high volume crime dataset and knowledge gained from data mining approaches is useful and support police forces. To perform crime analysis appropriate data mining approach, need to be chosen and for that the trend of the crime in each district needs to be known. So we take linear regression to predict the future crime in a district by knowing the previous trend of those crimes. In this project linear regression technique of data mining used to extract useful information from the high volume crime dataset and to interpret the data which assist police in identify and analyse crime patterns to reduce further occurrences of similar incidence and provide information to reduce the crime. In this project linear regression is implemented using open source data mining tool which are analytical tools used for analysing data. Among the available open source data mining suite such R, PowerBi and Python, we have used Python and its libraries here to perform the analysis. Also for crime analysis dataset used is Crime in India from 2001-2014. In this project all the crimes committed by human crimes are analysed and future trend is predicted.

The main objectives of crime analysis include:

1. Extraction of crime patterns by analysis of available crime and criminal data

2. Prediction of crime based on distribution of existing data and anticipation of crime rate using linear regression

3. Detection of crime

It is important to analyse crime due to following reasons:

1. Analyse crime to inform law enforcers about general and specific crime trends in timely manner.

2. Analyse crime to take advantage of the plenty of information existing in justice system and public domain.

3. Provide security and increase the awareness among people regarding the crimes in their districts and how to defend themselves.

**LITERATURE REVIEWS**

Paper-1:

Data mining and decision support systems have an important role to play in assisting human inference in this forensic domain that creates one of the most challenging decision-making environments. Technologies range widely and include social network analysis, geographical information systems, and data mining technologies for clustering crimes, finding links between crime and profiling offenders, identifying criminal networks, matching crimes, generating suspects, and predicting criminal activity.

Paper-2:

Crime continues to remain a severe threat to all communities and nations across the globe alongside the sophistication in technology and processes that are being exploited to enable highly complex criminal activities. Data mining, the process of uncovering hidden information from Big Data, is now an important tool for investigating, curbing and preventing crime and is exploited by both private and government institutions around the world.

Paper-3:

Most prediction techniques are used for retrospective forecasting, i.e., predicting the future through historical data. Historical crime data are used alone or together with crime attractors and generators (which can be demographic, environmental, etc.) in diverse types of prediction models. Apart from static data, such as demographics or socio-economic variables, as predictors, researchers have recently included dynamic space and time features, thus giving a boost to predicting crime occurrences. Although current crime prediction models show increasing accuracy, little emphasis has been placed on drawing the empirical and technical landscape to outline strengths and opportunities for future research, but also to identify weaknesses and threats.

Paper-4:

Crimes are a social irritation and cost our society deeply in several ways. Any research that can help in solving crimes quickly will pay for itself. About 10% of the criminals commit about 50% of the crimes. The system is trained by feeding previous years record of crimes taken from legitimate online portal of India listing various crimes such as murder, kidnapping and abduction, dacoits, robbery, burglary, rape and other such crimes. As per data of Indian statistics, which gives data of various crime of past 14 years (2001-2014) a regression model is created and the crime rate for the following years in various states can be predicted. Data used is supervised, semi-supervised and unsupervised learning technique on the crime records for knowledge discovery and to help in increasing the predictive accuracy of the crime. This work will be helpful to the local police stations in crime suppression.

Paper-5:

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 behaviour 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. Here we use 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 behaviour crimes. Three algorithms are used from different categories of approaches: instance-based learning, regression and decision trees.

Paper-6:

In data mining we often have to learn from biased data, because, for instance, data comes from different batches or there was a gender or racial bias in the collection of social data. In some applications it may be necessary to explicitly control this bias in the models we learn from the data. This paper is the first to study learning linear regression models under constraints that control the biasing effect of a given attribute such as gender or batch number. Then we analytically derive linear models that minimize squared error while controlling the bias by imposing constraints on the mean outcome or residuals of the models. Experiments with discrimination-aware crime prediction and batch effect normalization tasks show that the proposed techniques are successful in controlling attribute effects in linear regression models.

**ARCHITECTURE**

DISPLAY THE TREND IN GRAPHS

PROCESS THE DATASET FOR FILTERING VALUES

GET DISTRICT NAME, STATE NAME, YEAR OR CRIME NAME ACCORDING TO MODULE

REMOVING REDUNTANT AND NULL VALUES

DATA CLEANING

CRIME DATASET

Here, the data analysis is done using linear regression.

Choosing appropriate dataset: The dataset containing the number of crimes in all the districts of our country is chosen. The dataset is open source and taken from kaggle.

Link for dataset: <https://www.kaggle.com/datasets/rajanand/crime-in-india>

Data cleaning: The data is cleaned of null values and columns which are essentials are taken out and a new dataset with desired data is made. The new dataset will have all values of number of crimes in districts without any null or not applicable values. Some null values are removed or replaced with a minimum value of 1.

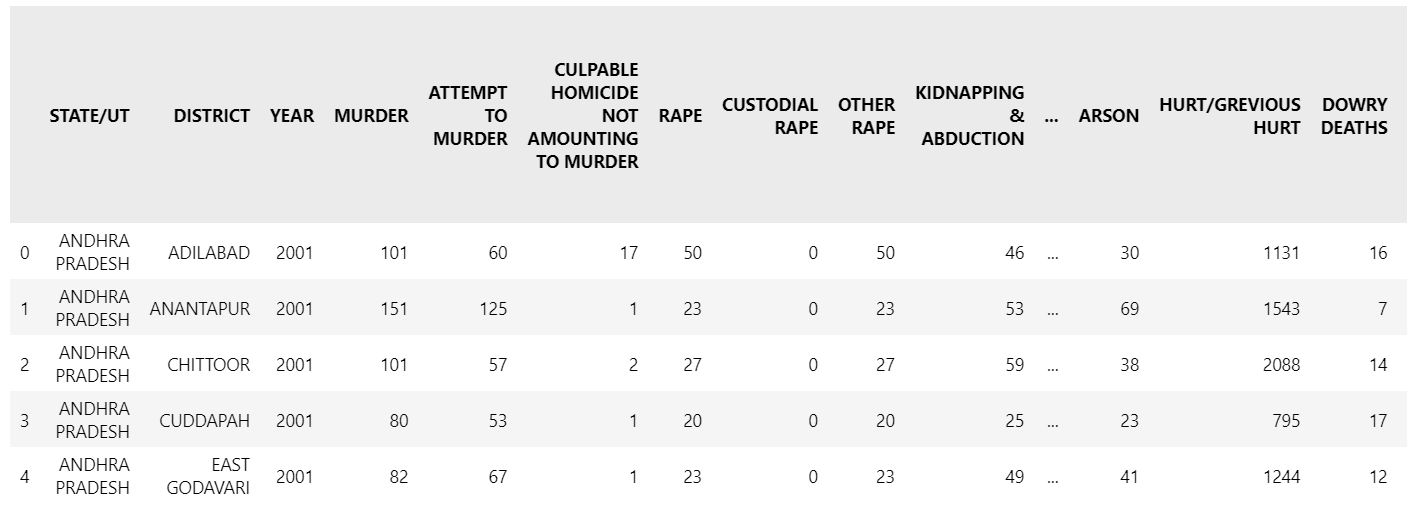
Process: The entire project is done using Python and its libraries. Data cleaning is also done using Python. Filtering of output is done on the basis of input crime, state and district values. 3 kinds of plot of trend in crime is plotted for each district. Line plot, bar plot and pie chart are plotted.

Regression: Perform linear regression to predict the future number of crimes in each district. User can choose the crime and the future year and get to know the number the future number of crimes that will happen. Analysis can be done on the basis of the trend obtained.

**DESIGN**

**Dataset used**

Crime dataset used for crime analysis is an offences recorded by the police in India and police force area from 2001 to 2014.



**COLUMNS IN DATASET**

Index(['STATE/UT', 'DISTRICT', 'YEAR', 'MURDER', 'ATTEMPT TO MURDER', 'CULPABLE HOMICIDE NOT AMOUNTING TO MURDER', 'RAPE', 'CUSTODIAL RAPE', 'OTHER RAPE', 'KIDNAPPING & ABDUCTION', 'KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS', 'KIDNAPPING AND ABDUCTION OF OTHERS', 'DACOITY', 'PREPARATION AND ASSEMBLY FOR DACOITY', 'ROBBERY', 'BURGLARY', 'THEFT', 'AUTO THEFT', 'OTHER THEFT', 'RIOTS', 'CRIMINAL BREACH OF TRUST', 'CHEATING', 'COUNTERFIETING', 'ARSON', 'HURT/GREVIOUS HURT', 'DOWRY DEATHS', 'ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY', 'INSULT TO MODESTY OF WOMEN', 'CRUELTY BY HUSBAND OR HIS RELATIVES', 'IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES', 'CAUSING DEATH BY NEGLIGENCE', 'OTHER IPC CRIMES', 'TOTAL IPC CRIMES'], dtype='object')

**TOOL USED**

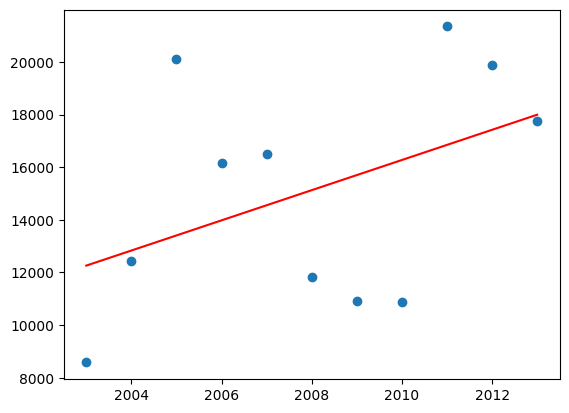
Python:

Python is a data analysis tool used widely for data analytics. There are many inbuilt functions which are helpful in cleaning data as well as processing it. Performing regression models is also possible using it. Making future predictions, processing and organizing data is easily possible using Python. Python libraries used are Pandas, Numpy, Matplotlib for plotting graphs and Linear regression.

**LINEAR REGRESSION**

Simple linear regression is an approach for predicting a response using a single feature. It is assumed that the two variables are linearly related. Hence, we try to find a linear function that predicts the response value(y) as accurately as possible as a function of the feature or independent variable(x). Here variable x is the year in which we want to predict the number of crimes and y is the number of crimes which occurs in that specific year.

We are predicting the total number of crimes that will occur in subsequent years by obtaining the equation from the graph.

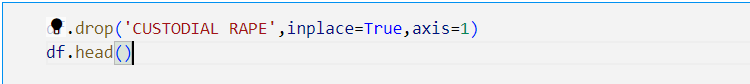


**MODEL IMPLEMENTATION**

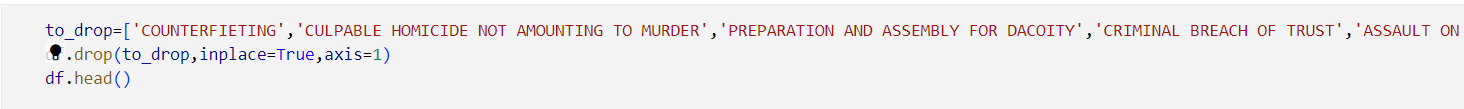
Loading dataset:

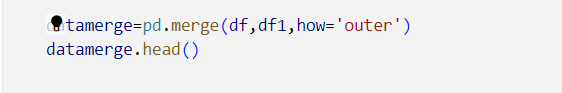


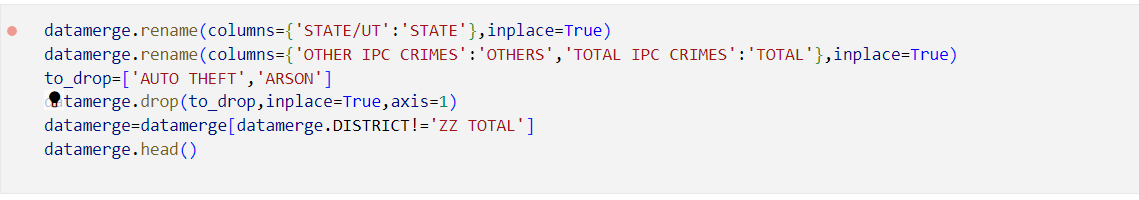
Data cleaning:





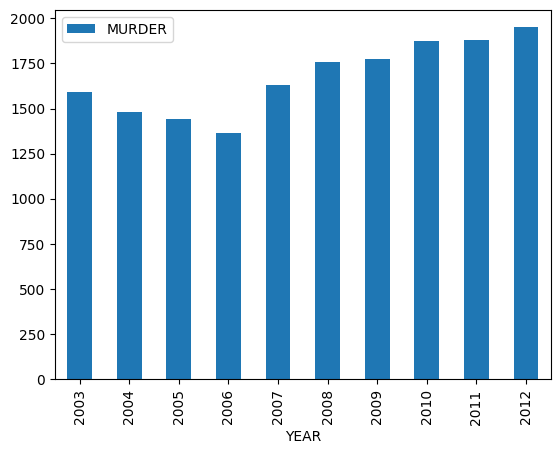


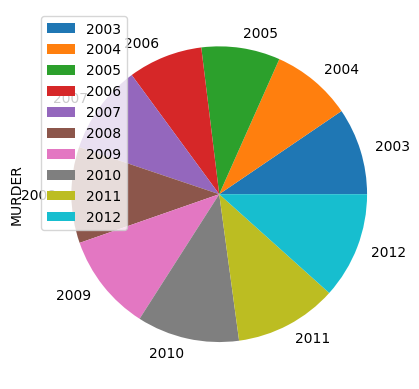


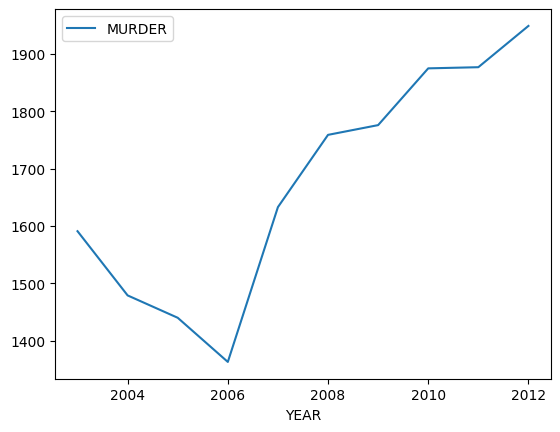


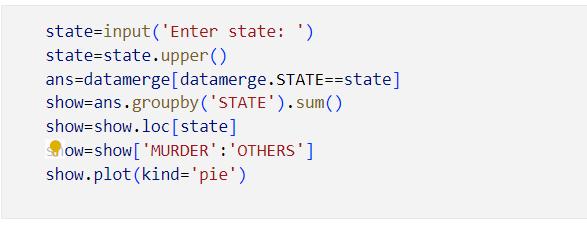


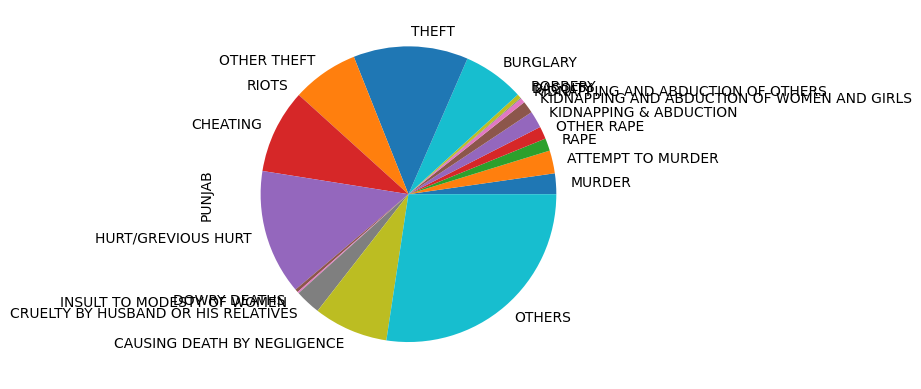






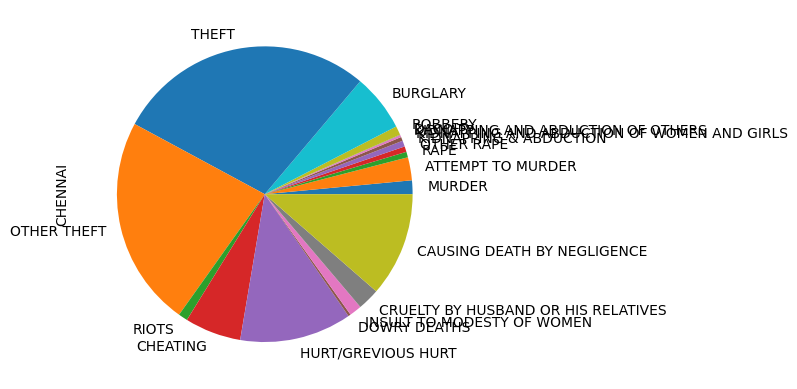


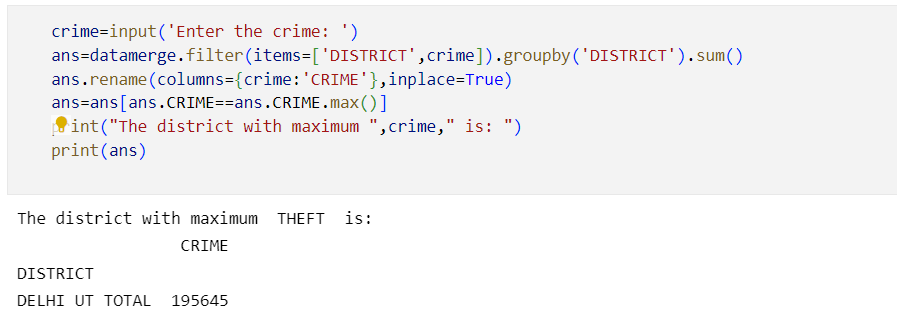




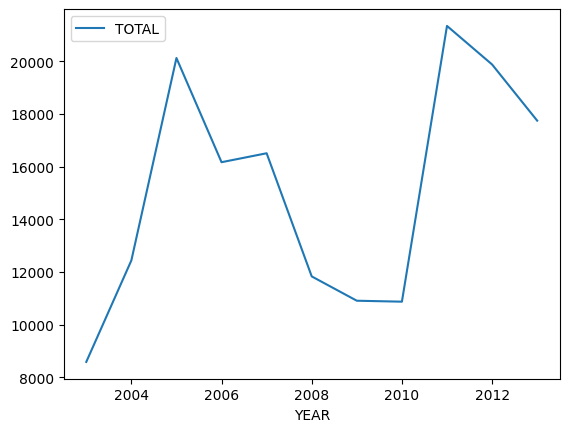


The most freuently occured crime is: THEFT

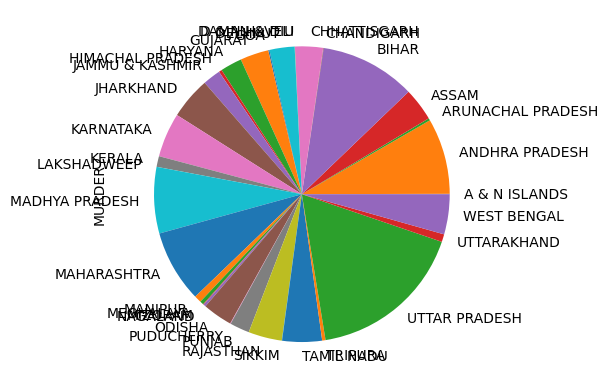


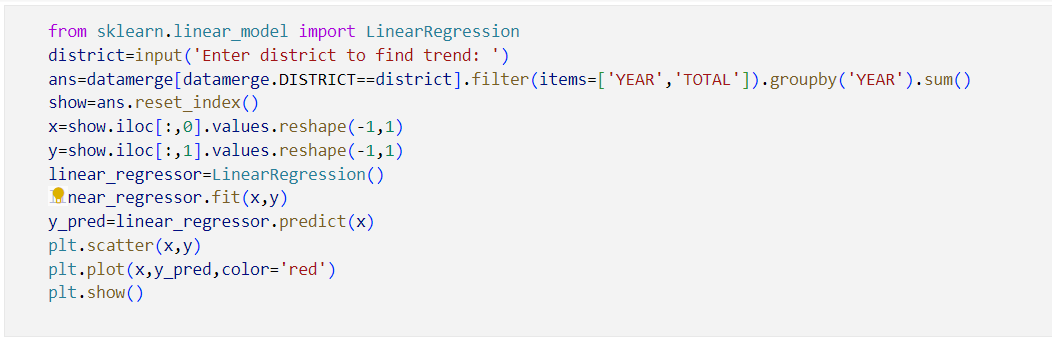


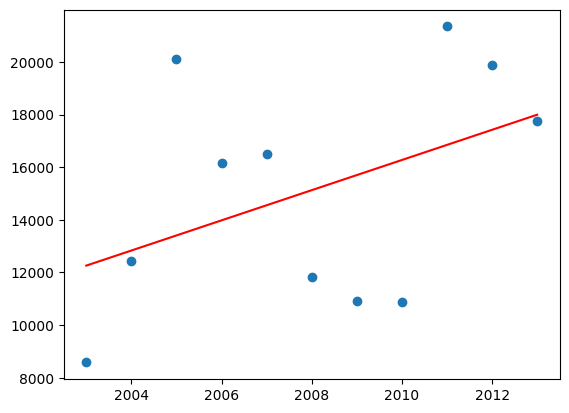












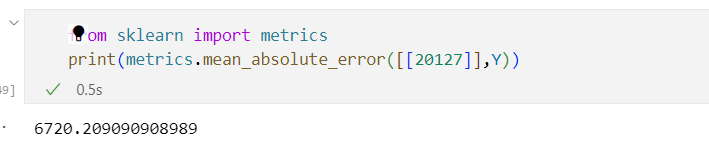
Prediction model:

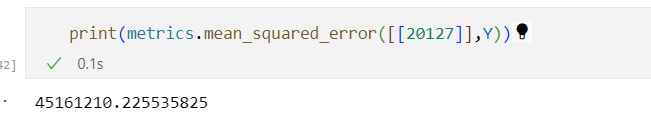


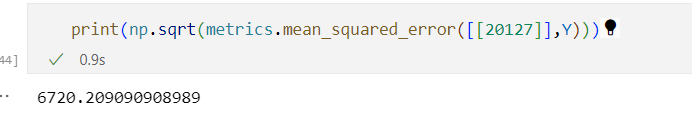
**PERFORMANCE ANALYSIS**

The performance metrics used for linear regression are:

1. Mean Squared error
2. Absolute Mean error
3. Root Mean Squared error







**CONCLUSION**

In this project, we have applied classification and regression algorithms to get the better results that can help police department, law enforcement officers to improve their work. It will increase the efficiency in solving the crimes faster. At present, data mining is playing an important role in crime control and criminal suppression in many countries. In this paper, data mining technique is used to forecast future crime trends of India. For this purpose, linear regression model is trained by crime data of previous years. After training linear regression, total amount of crime is predicted for the year 2005. From the experimental result it is also observed that, most of the crimes are increasing with the growth of population. Thus the knowledge discovered from crime data analysis may assist police department and various law enforcement agencies to forecast, prevent or solve the future crime trends of India. A future plan is to forecast the location of crime occurrence, so that prior actions can be taken to prevent crime.

**FUTURE SCOPE**

Along with the present scope of our project, which is prediction of the number of crimes that is likely to be committed in the upcoming years, we can also predict the estimated time for the crime to take place as a future scope. Along with this, one can try to predict the location of the crime. We will test the accuracy of frequent-crimes and prediction based on different test sets. So the system will automatically learn the changing patterns in crime by examining the crime patterns. Also the crime factors change over time. By shifting through the crime data we have to identify new factors that lead to crime. Since we are considering only some limited factors full accuracy cannot be achieved. For getting better results in prediction we have to find more crime attributes. Our software predicts the crimes likely to be performed using the attribute year. We will use Linear Regression Algorithm with association rule mining for this purpose. This will determine the next crime a criminal is about to commit. Also the k-clustering algorithm can be used to get more specific accurate values of prediction. We can also increase privacy and some other security measures to protect data set that we are using. Along with this, this work can be further extended to predict who will commit a crime and this can be done using Face recognition. The system will detect if there is any suspicious change in the behaviour or usual movements. For example, if a person is moving back and forth in same region over and over might indicate that he is a pickpocket and it will also track person over time.

**REFERENCES**

1. Data mining and crime analysis by Giles Oatley and Brian Ewart.
2. Data mining applications in crime by Hossein Hassani, Xu Huang, Emmanuel S. Silva, Mansi Ghodsi.
3. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention, by Neil Shah, Nandish Bhagat, and Manan Shah.
4. Crime pattern detection, analysis & prediction by Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma.
5. An Exploration of Crime Prediction Using Data Mining on Open Data by Ginger Saltos and Mihaela Cocea.
6. Controlling Attribute Effect in Linear Regression by Toon Calders, Asim Karim, Faisal Kamiran, Wasif Ali, Xiangliang Zhang.