CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING

Roopa[1], Prof. Thouseef Ulla Khan[2]

[1]PG Scholar (MCA), Dept of MCA, Vidya Vikas Institute of Engineering And Technology, Mysore ,Karnataka, India
[2]Assistant Professor, Dept of MCA, Vidya Vikas Institute of Engineering And Technology, Mysore ,Karnataka, India

---***---

Abstract – *Crime is one of our society's most serious and pervasive problems, and preventing it is a critical duty. This necessitates keeping note of all offences and creating a database for future reference. The present issue is keeping a reliable crime record and analysing this data to aid in the prediction and resolution of future crimes. The objective of this paper is to analyze dataset, which consist of numerous crimes and predicting the type of crime, which may happen in future depending upon various conditions. In this project, we will be using the technique of machine learning and data science for crime prediction of Indian crime data set. Crime analysis and prediction is a methodical way to spotting crime. This algorithm can anticipate and depict crime-prone areas. Using the notion of machine learning, we may extract previously unknown, meaningful information from unstructured data. The extraction of new information is anticipated using current datasets. Crime is a perilous and widespread societal issue that affects people all around the world. Crime has an impact on people's quality of life, economic prosperity, and the nation's reputation. To safeguard their communities from crime, modern technology and novel techniques to enhancing crime analytics are required. We present a system that can analyse, identify, and forecast various crime probabilities in a given location. This study describes many sorts of criminal analysis and crime prediction using machine learning approaches.*

Key Words: Decision trees, linear regression, and k-means clustering

1. INTRODUCTION

The crime data rate is growing on a daily basis because current technology and high-tech ways assist criminals in carrying out their illicit actions. Burglary, arson, and other crimes, according to the Crime Record Bureau have escalated, as have crimes such as murder, rape, abuse, gang rap, and so on. Data on crime will be gathered from numerous blogs, news sites, and websites. The massive amount of data is utilized to create a record. A database of crime reports. The knowledge gained via data mining techniques will be useful in lowering crime by making it easier to discover the perpetrators and the regions most affected by crime.

When applied to a crime dataset, data mining techniques produce good results. The information generated from data mining techniques can assist the police department. The discovery of criminal "hot spots," which show regions with a high concentration of crime, has been proven valuable by the police. Data mining approaches can yield significant findings from crime report databases. Crime analysis is the first phase in the study of crime. Criminal analysis is the exploration, interrelationship, and detection of relationships between numerous crimes and crime variables. This analysis aids in the creation of statistics, queries, and maps on demand. It also aids in determining whether a crime has occurred in a certain recognized location.

1.1 Objectives

The prediction using data mining techniques that is prediction rules. Frequent patterns are extracted based on the criteria's like crime type. Prediction is done based on the previous year datasets. The prediction report consists of all the datasets from the year 2012-2020.the year wise comparison is shown based on the state wise datasets. The clustering algorithm can be perform based on every datasets based on each year wise comparison is made.

1.2 Scope

The primary goals of crime evaluations are as follows: 1. Identifying crime tendencies by study of existing crimes and criminal information 2. Using geographic distribution to forecast crime of available information as well as prediction ofcrime total utilizing various datamining techniques 3. Criminal detection

2. Existing System:

This algorithm can forecast high-risk areas for crime and show crime-prone areas. Using the concept of data mining, we may draw previously undiscovered, pertinent information from unstructured data.

Disadvantages:

- They exhibited lower prediction accuracy using this technique.

- The outcomes of this approach are not ideal.

3. Planned system:

The system under consideration is a web-based application. An advanced criminal mechanism of detection whose main goal is to forecast crimes and their tendencies. The proposed system employs a data mining approach known as "Prediction Rules" for crime pattern detection, as well as automation for early crime pattern prediction, which helps to avert crimes. Predicts crime trends based on past crime information, date, and location.

Advantages:

- Conducting criminal analysis and identifying trends in crime.

- Disseminate knowledge to help with the creation of crime reduction and preventive measures.

- Recognize and examine recurring criminal trends to prevent similar incidents from happening again.

- To create a data-cleaning algorithm that purges the crime dataset of unnecessary information so that it may be explored.

4. System Design

Designing a machine learning system's software architecture, infrastructure, algorithms, and data to meet specific needs is known as machine learning systems architecture.By outlining the intricacies of how the programme should be created, the software design will be used to assist in the development of software for web apps.
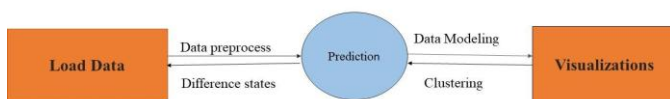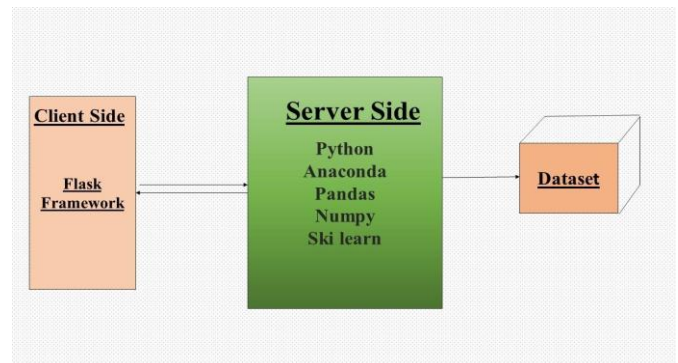
Fig: 4.2.1 Context diagram

Fig: 4.2.1Architecture Design

5. Detailed design

By outlining the specifics of how the application should be constructed, the software design will be utilised to assist in the software development of an android application. Use case models, sequence diagrams, and other supplementary requirement data are included in the software design specifications, which are narrative and graphical documentation of the software design

5.1 Diagram of Use Cases:

An example of a behavioural diagram in the Unified Modelling Language (UML) is a use case diagram, which is based on and defined by use case studies. Its purpose is to present a graphical depiction of a system's functioning in terms of actors, their objectives (represented as use cases), and any connections between those use cases. The main objective of a use case diagram is to show which system actions are taken for a particular actor. You may display the parts that each player in the system plays
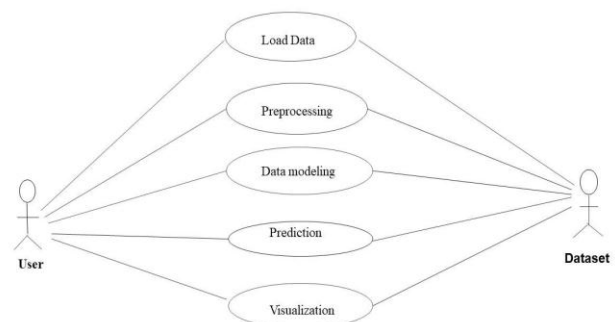
Fig 5.2.1 Use Case Diagram

6. Implementation

To forecast the future values of the dependent variable, a regression algorithm is created to identify the past link between an independent and a dependent variable. Regression uses the historical relationship between variables to forecast how they will behave going forward.

Investigative work on the sites of high and low frequency crimes was done using the K-means clustering technique.

6.1 Algorithm Implementation

6.1.1Regression Algorithm

A regression algorithm is designed to find the historical relationship between an independent and a dependent variable to predict the future values of the dependent variable. In order to forecast future behaviour, a regression models the historical connection between variables. The Algorithm uses the linear regression techniques based on the data set collected for the project. The linear regression technique helps in predicting the future behavior of road CRIME with help of the statistical methods. The algorithm find the mean and variance value of the dependent variables, and apply the formula $Y=b0+b1*x$ to predict the future behavior.

6.1.2 K-means Clustering

K-means clustering algorithm was used to investigate the high and low-frequency CRIME locations. The algorithm follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. M These centroids should be placed in a cunning way because of different location causes different result.

6.1.3 Decision Tree:

For both prediction and classification, a decision tree is employed. A function, which is intervals formed by splits on the individuals' attributes value, may be trained for classification purposes.

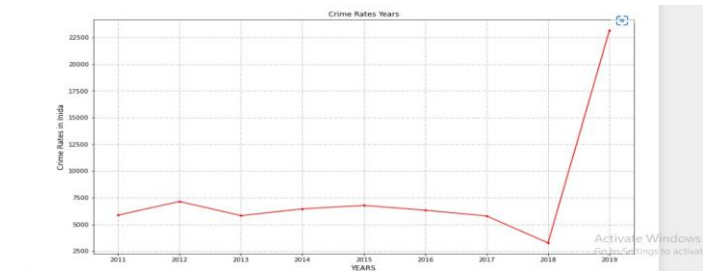| X (year)	Y(value)	A1=(x-mean of x)	B1=(y—mean of y)	A1 *B1	(A1)2	(b1)2
2008	3496	-4	-329	1316	16	108241
2010	3500	-3	-325	975	9	105625
2010	3987	-2	162	-324	4	26244
2011	2987	-1	-838	838	1	702244
2012	3019	0	-806	0	0	649636
2013	3999	1	174	174	1	30276
2014	4015	2	190	380	4	36100
2015	4786	3	961	2883	9	923521
2016	4018	4	193	772	16	37249
2021	4445	5	620	3100	25	384400

Fig 6.4.10 To display the crime rates year in India

7. Testing

This chapter gives the various test cases performed to check for the effective execution of the venture. Testing is a procedure of cross verification of the designed system model under active state and various inputs. There are several ways to carry out this approach. The main objective of software development life cycle is to produce a product with no errors or very few errors. In the processes of achieving hassle free software we plan testing and test cases. Software testing is done for the success of the application. The testing is done mainly to check whether the product meet the requirement of the user properly. It is used to check the bugs and errors in the system or to find out the defects of the system.

7.1 Test causes with positive scenarios:

TC No	Positive scenario	Required Input	Expected output	Actual output	Test Result
1	Enter Prediction values	Enter a valid values	Should predicted successfully	predicted successfully	Pass
2	Enter clustering values	State,year, type	Should cluster successfully	cluster successfully	Pass
3	Enter Prediction values	Enter a valid values	Should predicted successfully	Database error	fail
4	Enter clustering values	State,year, type	Should cluster successfully	Database error	Fail

Conclusion

Due to a variety of variables, including a growth in poverty, unemployment, corruption, etc., crime rates in India are rising daily. The suggested paradigm is extremely beneficial to both the investigating authorities and the taking the required actions as a police officer to lower crime. The initiative aids in the examination of these crimes by using

various interactive visualisation techniques, crime networks Future development of this project will focus on teaching bots to identify crime hotspots using machine learning acquiring skills. Given that machine learning and data mining are comparable, an advanced machine learning concept can be applied to improve prediction. It is possible to increase the data's dependability, correctness, and privacy prediction.

REFERENCES:

[1] Ginger Saltos and Mihaela Coacea, 2017 International Journal of Information Technology and Decision Making, An Exploration of Crime Prediction Using Data Mining on Open Data.

[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using 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 and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018

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

[5] Raj Kumar and Sakkarai Pandi, "Crime Analysis and Prediction Using Data Mining Techniques," International Journal of Recent Trends in Engineering & Research, 2019.

[6] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of machine learning using python, International Journal of Advanced Research in computer science, 2015.

[7] Kalyani Kadam and Ayisheshim Almaw, "Survey Paper on Crime Prediction Using Ensemble Approach," International Journal of Pure and Applied Mathematics, 2018.

[8] Review on Crime Analysis and Prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and Technology, 2018, by Dr. M. Sreedevi, A. Hardhat Vardhan Reddy, and Ch. Venkata Sai Krishna Reddy

[9] International journal of engineering, Science and mathematics, 2017. K.S.N. Murthy, A.V.S. Pavan Kumar, and Gangu Dharmaraju