Crime Analysis using Data Mining and Data Analytics

S. Bodare[1], S. Kurkute[2], M. Akash[3], R. Pawar[4]

[1,2,3,4]*Student at Svpm College of Engineerin, Malegaon(BK)*

---***---

Abstract - *Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our 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 the 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 is one of the most important social problems in the country, affecting public safety, children development, and adult socioeconomic status. Understanding what factors cause higher crime is critical for policy makers in their efforts to reduce crime and increase citizens life quality. We tackle a fundamental problem in our paper: crime rate inference at the neighborhood level. Traditional approaches have used demographics and geographical influences to estimate crime rates in a region.*

Key Words: Data Mining, Linear regression, Crime rate analysis,Data Analytics

1. INTRODUCTION

The built environment based crime research until now has rarely focused on robbery. The reason for this could be the difficulty in knowing exactly where a robbery happened, or the precautions against robbery seemed less applicable than other crimes such as burglary and thus the research being perceived more worthwhile. However, street robbery is a daily fear in urban life and reducing the type of settings that is most likely to take place should be the aim for achieving urban environments that are used effectively, day and night, for a sustainable urban development[3].

Regression is one of a powerful tool for predicting the future values based on previous data sets. The use of prediction in crime analysis can make crime free region. The proposed technique can predict those regions which are sensitive to criminal activities like theft, murder, rape, Anti-social behavior, Domestic violence etc. The scheme uses the data sets taken from the Indian government website and can be used to predict the number of criminal activities in future. The data will then supplied with other factors like criminal type, age, month and year to the regression model which will predict the future value for the criminal activities. . This predicted number of activities will indicate that the region is high sensitive or low sensitive. If the predicted number will be high then the region is high sensitive and if low then the region is low sensitive. Understanding how to control crime is important because exposures to violence and crime have been unusually high in the India for several decades and, while declining, they remain high.

Sensing technologies and large-scale computing infrastructures have produced a variety of big data in urban spaces. These heterogeneous data convey rich knowledge about city dynamics and enable us to address many urban challenges. For example, human mobility data could help improve the efficiency of transportation systems such as estimating real-time traffic flow and forecasting travel time for road segments. With the similar motivation, we employ such modern urban data for crime rate inference. In the criminology literature, researchers have studied the relationship between crime and various features (social, demographics, and geographic factors). Examples are historical crime records, education, ethnicity, income level, unemployment, and spatial proximity.

In data mining, newer types of data are used. For example, studies use twitter to predict crime, and cellphone data to evaluate crime and social theories at scale. Overall, existing work on crime prediction can be categorized into three paradigms. Time-centric paradigm. This line of work focuses on the temporal dimension of crime incidents. For example, Mohler et. al. propose to useaself-exciting point process to model the crime and gain insights into the temporal trends in the rate of burglary. Existing works in the fields of criminology, sociology, psychology and economics tend to mainly explore relationships between criminal activity and socio-economic variables such as education, ethnicity, income level, and unemployment [5]. In another study, Ratcliffe investigates the temporal constraints on crime, and propose an offender travel and opportunity model. His findings suggest that a proportion of offending is driven by the availability of opportunities presented in the routine lives of offenders. Place-centric paradigm. Theories of crime can be divided into those that seek to explain the development of criminal offenders, and those that seek to explain the development of criminal events[4].

Crime rate is increasing now-a-days in many countries. In today's world with such higher crime rate and brutal crime happening, there must be some protection against this crime. Here we introduced a system by which crime rate can be predict. Crime data must feed into the system. We introduced data mining algorithm to predict crime.

Regression Analysis, plays an important role in analyzing and predicting crimes. Regression Analysis will cluster co-offenders, collaboration and dissolution of organized crime groups, identifying various relevant crime patterns, hidden links, link prediction and statistical analysis of crime data. This system will prevent crime occurring in society.

Crime data is analyzed which is stored in the database. Data mining algorithm will extract information and patterns from database. System will group crime. Regression Analysis will be done based on places where crime occurred, type of crime and the timing crime took place. This wil help to predict crime. Admin will enter crime details into the system which is required for prediction. Admin can view criminal historical data. Crime incident prediction depends mainly on the historical crime record and various geospatial and demographic information.

1.1 System Architecture

A description of the program architecture is presented. Subsystem design or Block diagram, Package Diagram, Deployment diagram with description is to be presented.
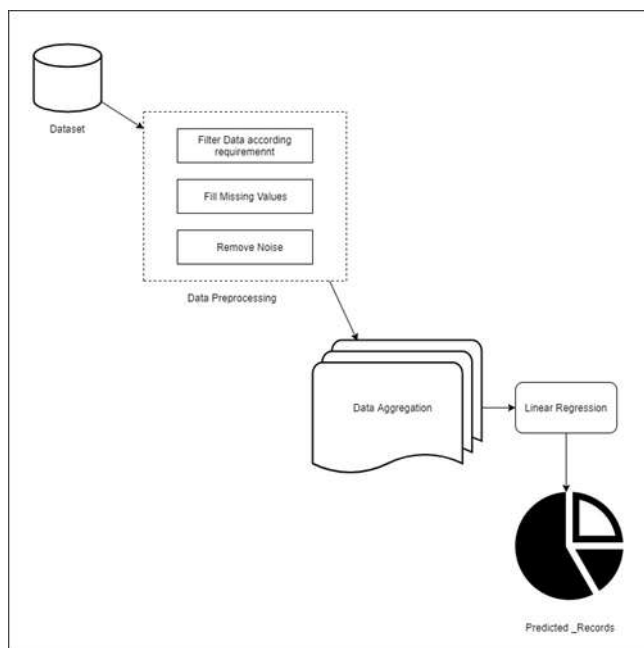

Fig.System Architecture

1.2 Dataset

A data set is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. Several characteristics define a data set's structure and properties. These include the number and types of the attributes or variables, and various statistical measures.

Firstly, We collect data set from Indian government website as data.gov.in.[1] The given dataset is in well trained manner. the data file is in CSV format. The dataset contains different types of fields such as state, crime head, and crime taken by gender in different age as Femaleagebelow18, Maleagebelow18, female age between 20-25 and so on.

1.3 Data Analysis

In today's business, data analysis is playing a role in making decisions more scientific and helping the business achieve effective operation. Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing mainly on business information. Here, We analysis the given dataset. Different Operations are performed in data analysis such as Filter data according requirement, Fill missing values, Remove noise. So well Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision making.[3]

1.3 Linear Regression

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regressions. Linear regression is mostly used as, If the goal is prediction, or forecasting, or error reduction,[clarification needed] linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.

1.4 Prediction Record

Here, final output is predicted which is in the form of pie chart.

2. Inference Model

Linear regression is a basic and commonly used type of predictive analysis.

2.1 Math and Equations

In the regression there is simple Mathematics equations

which is useful to solve our algorithm:

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a_0 = \overline{y} - a_1\overline{x}$$

$$y = a_0 + a_1 x$$

2.2 Algorithm

The algorithm as below Steps:-

1. SUB Regress(x,y,n,al,a0,syx,r2)
2. sumx=0:sumxy=0:st=0
3. sumy=0:sumx2=0:sr=0
4. DOFOR i=1,n
5. sumx=sumx+xi
6. sumy=sumy+yi
7. sumxy=sumxy+xi*yi
8. sumx2=sumx2+xi*xi
9. END DO
10. xm=sumx/n
11. ym=sumy/n
12. a1=(n*sumxy-sumx*sumy)/(n*sumx2-sumx*sumx)
13. a0=ym-a1*xm
14. DOFOR i=1,n
15. st=st+(yi−ym)2
16. sr=sr+(yi−al∗xi−a0)2
17. END DO 18. syx (sr/(n−2))1/2
19. r2=(st-sr)/st
20. END Regress

3. Result

In this project we predict different states in India. We know that India is one of biggest country in world which have 7th largest country. In that India has 1.35 billion population. It take second largest population in world. Very big population make many problems in that Crime is one of biggest problem for India. We contribute little bit help for Indian peoples. From our project people can understand which state have highest crime and take awareness campaign in particular area.

3.1 Different State fine from Database

In this topic we going to collect dataset from www.data.gov.in[1]. In that dataset we got all states in India and number of male, females which have their age ex. Male below 18 year, Male between 18_30 year etc. Separate the state with different crime like Rape, Assault, Dowry etc.
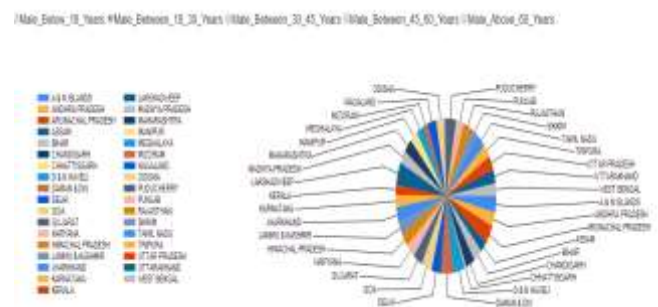
Crime Head Wise Male Details

3.2 Graphs For Indication

Peoples can not understand prediction easily for that we include graphs for every attributes. Following graph indicate that Rape cases in India which conducted by different age group and different colours shows different states.

3.3 Final Table For Prediction

CONCLUSION

We handle different attribute relates to crime dataset. With the help of that attribute, we predict which attribute conducted crime with respect to states. So with the help of predicted output, we take different precautions where the maximum amount of crime committed.

REFERENCES

[1] Indian Government Portal –https://data.gov.in

[2] Hongjian Wang, Huaxiu Yao, Daniel Kifer, Corina Graif, Zhenhui Li,Non-Stationary Model for crime rate inference using modern urban data, 2332-7790 (c) 2017 IEEE.1-13

[3] Wikipedia-http://Wikipedia.com

[4] S. D. Johnson,"A brief history of the analysis of crime concentration", European Journal of Applied Mathematics, 21 (4-5) (2010) 349–370, http://journals.cambridge.org/article_S0956792510000082

[5] Chhaya Chauhan, Smriti Sehgal, "A REVIEW: CRIME ANALYSIS USING DATA MINING TECHNIQUES AND ALGORITHMS" 2017 IEEE