

Crime Rate Prediction using Data Mining Algorithms

Bodare. S, Kurkute. S, Misal.A, Pawar. R, Prof.Khalate.Y.R

S.V.P.M COE, Malegaon (bk)

Corresponding author's email id: misalakash55@gmail.com

DOI: <http://doi.org/10.5281/zenodo.2629982>

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.

Keywords: Data Mining, Linear regression, Crime rate analysis

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 cell phone 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. proposes to use a self-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 will 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.

ARCHITECTUR DESIGN

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

diagram with description is to be presented.

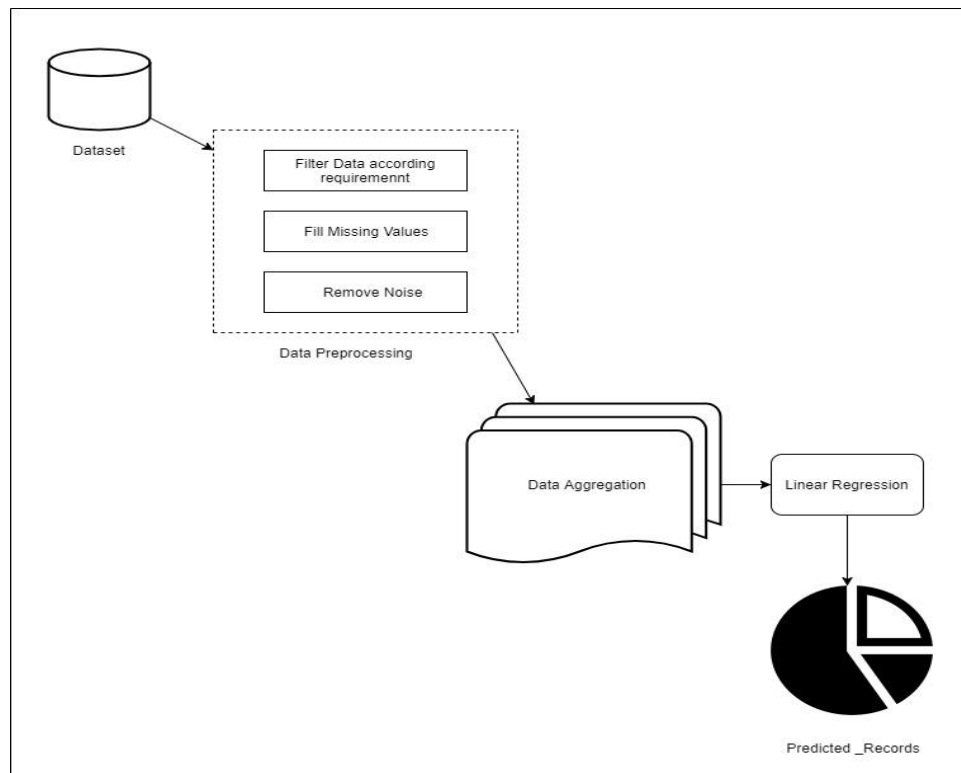


Figure: - 1 Architecture Diagram

A. 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 dataset from Indian government website 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 Female age below 18, Male age below 18, female age between 20-25 and so on.

B. 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]

C. Linear Regression:

In statistics, line is regression is a linear approach to modeling the relationship between as calar 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

D. Prediction Records:

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

INFERENCE MODEL

A. Linear Regression:

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

1. Does a set of predictor variables do a good job in predicting an outcome (dependent) variable?
2. Which variables in particular are significant predictors of the outcome variable, and in what way do they– indicated by the magnitude and sign of the beta estimates–impact the outcome variable?

These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is

defined by the formula $y = c + b \cdot x$, where
 y = estimated dependent variable score, c
 = constant, b = regression coefficient, and
 x = score on the independent variable.[2]

A. Math and Equations

In the regression there is a simple
 Mathematics equation which is useful to
 solve our algorithm:

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

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

$$y = a_0 + a_1 x$$

B. Algorithm

Steps:-

1. SUB Regress($x, y, n, a_1, a_0, syx, r^2$)
2. $sumx=0; sumxy=0; st=0$
3. $sumy=0; sumx^2=0; sr=0$
4. DOFOR $i=1, n$
5. $sumx=sumx+x_i$
6. $sumy=sumy+y_i$
7. $sumxy=sumxy+x_i \cdot y_i$
8. $sumx^2=sumx^2+x_i \cdot x_i$
9. END DO
10. $xm=sumx/n$

$$11. ym=sumy/n$$

$$12. a_1=(n \cdot sumxy - sumx \cdot sumy) / (n \cdot sumx^2 - sumx \cdot sumx)$$

$$13. a_0=ym-a_1 \cdot xm$$

$$14. DOFOR i=1, n$$

$$15. st=st+(y_i-ym)^2$$

$$16. sr=sr+(y_i-a_1 \cdot x_i-a_0)^2$$

$$17. END DO \quad 18. syx (sr/(n-2))^{1/2}$$

$$18. r^2=(st-sr)/st$$

$$19. END Regress$$

C. Example belongs to algorithm

X_i -column shows person arrested under
 crime against Robbery.

Y_i -column shows suspects towards
 Robbery. (Y_i)

Last two columns-deviations scores

Last two rows-sums and mean for
 regression analysis.

Table 1.

Crime	X_i	y_i	(x_i-x)	(y_i-y)
1	95	85	17	8
2	85	95	7	18
3	80	70	2	-7

4	70	65	-8	-12
5	60	70	-18	-7
Sum	390	385		
Mean	78	77		

Also, calculate Square of deviation:-

And, finally compute Product of the Deviation

Table: 2

Student	X_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$
1	95	85	136
2	85	95	126
3	80	70	-14
4	70	65	96
5	60	70	126
Sum	390	385	470
Mean	78	77	

Table: 3

Crime	X_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	95	85	289	64
2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630
Mean	78	77		

$$B1 = 0.644$$

The regression equation: $Y = B0 + B1X$

Solve for the regression slope ($B0$):

$$B0 = \bar{y} - B1 * \bar{x}$$

$$B0 = 77 - (0.644)(78)$$

$$B0 = 26.768$$

First, we solve for the regression coefficient ($B1$):

$$B1 = \frac{\sum [(x_i - \bar{x})(y_i - \bar{y})]}{\sum [(x_i - \bar{x})^2]}$$

$$B1 = 470/730$$

Therefore, the regression equation is:

$$Y = 26.768 + 0.644X.$$

$$Y = B_0 + B_1X$$

$$Y = 26.768 + 0.644X = 26.768 + 0.644 * 80$$

$$Y = 26.768 + 51.52 = 78.288$$

Coefficient of Determination:

Whenever you use a regression equation, you should ask how well the equation fits the data,

$$R^2 = \{ (1/N) * \sum [(x_i - \bar{x}) * (y_i - \bar{y})] / (\sigma_x * \sigma_y) \}^2$$

$$R^2 = [(1/5) * 470 / (12.083 * 11.225)]^2$$

$$R^2 = (94 / 135.632)^2 = (0.693)^2 = 0.48$$

A coefficient of determination equal to 0.48 indicates that about 48% crime rates for particular crime For Certain Region.

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. Andrey Bogomolov†, Bruno Lepri§, Jacopo Staiano†, Nuria Oliver‡, Fabio Pianesi§, Alex Pentland], "Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data"
6. Chhaya Chauhan, Smriti Sehgal, "A REVIEW: CRIME ANALYSIS USING DATA MINING TECHNIQUES AND ALGORITHMS" 2017 IEEE

AUTHORS' PROFILE

[1] Bodare. S

Department: Computer Science Engineering

College: S.V.P.M COE, Malegaon (bk)

[2] Kurkute. S

Department: Computer Science Engineering

College: S.V.P.M COE, Malegaon (bk)

[3] Misal. A

Department: Computer Science Engineering

College: S.V.P.M COE, Malegaon (bk)

[4] Pawar. R

Department: Computer Science Engineering

College: S.V.P.M COE, Malegaon (bk)

[5] Prof. Khalate. Y.R

Department: Computer Science Engineering

College: S.V.P.M COE, Malegaon (bk)

Cite this Article

Bodare. S, Kurkute. S, Misal. A, Pawar. R **Crime Rate Prediction using Data Mining Algorithms** Journal of Data Mining and Knowledge Engineering, 4(1) 21- 29
<http://doi.org/10.5281/zenodo.2629982>