

PAPER • OPEN ACCESS

Analysis of crimes committed against scheduled tribes

To cite this article: Vivek P Khadse *et al* 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **263** 042089

View the [article online](#) for updates and enhancements.

Related content

- [Ionization-recombination instability in a photo-ionized nebula](#)
Manasi Buzar Baruah, Samiran Chatterjee and Madhurjya P Bora
- [A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based brain-computer interface](#)
Joan Fruitet, Dennis J McFarland and Jonathan R Wolpaw
- [The Grayson micro-ruling machine](#)
W Stone

Analysis of crimes committed against scheduled tribes

Vivek P Khadse, P Akhil, Christopher Anto and Lydia J Gnanasigamani

School of Computer Science and Engineering, VIT University, Vellore-632014, India

E-mail:lydiajane.g@vit.ac.in

Abstract. One of the curses to the society is a crime which has a deep impact on the society. Victims of crimes are the one who is impacted the most. All communities in the world are affected by crime and the criminal justice system, but largely impacted communities are the backward classes. There are many cases reported of crime committed against scheduled tribes from the year 2005 till date. This paper states the analysis of Crimes Committed against Scheduled Tribes in the year 2015 in various states and union territories in India. In this study, Multiple Linear regression techniques have been used to analyze the crimes committed against scheduled tribes' community in India. This study compares the number of cases reported to the police station and rate of crime committed in different states in India. It also states the future prediction of the crime that would happen. It will also predict the number of cases of crime committed against the scheduled tribe that can be reported in future. The dataset which has been used in this study is taken from official Indian government repository for crimes which include different information of crimes committed against scheduled tribes in different states and union territories measured under the population census of the year 2011. This study will help different Indian states and union territory government to analyze and predict the future crimes that may occur and take appropriate measures against it before the actual crime would occur.

1. Introduction

There is a challenge in recent years faced by law enforcement analyst to bridge the gap between intelligence analysis and crime committed. Crime data is the primary data type used for analyzing crime. There are many issues about this data which can lead to misuse and misinterpretation of statistics. For example, crime represented by police crime data may not represent all the crimes occurred in the society. This is because some people don't like to report the crime happened to them because they are afraid of the future consequences and how society will see them [11, 12]. Those crimes which are not registered under police stations are known as unreported crimes while the crimes that are reported are known as registered/reported crimes. For example, rape cases are registered less than they occur. An areas police station performance can be measured only with the help of registered case and solved cases. It is always important to register crimes happened then only government officials can take necessary steps to decrease the rate [10].

In our study we are taking crimes happened against scheduled tribes during the year 2012-2015. Scheduled tribe people living in Indian state and union territories are considered. We have fetched the details of cases registered by scheduled tribe people and how many cases were solved for each year. After that, we will try to predict the future crime rates that are going to happen for them. So that the government officials can take necessary plans to decrease this rate.

We have used multiple linear regression models to find the relationship between different variables and thereby train the data for predicting the future rates. We will compare the number of ST



population that has been affected by crimes and number of solved cases against each state and union territories in India.

Sathyadevan et al. [11] have focused on analyzing the trends and occurrence of crime patterns. They use previous Crime Analysis crime data and analyzed it to find the crime prone areas and also predicted the probability of occurrence of crime in those areas using data mining concepts. They have mainly focused on each day crime factors. The author has tried to detect the pattern and he uses the pattern to predict, anticipate and prevent crime. They collected the unstructured data from various online sources such as news, RSS feeds, and blogs and used Mongo DB to store data. The collected data is classified using Naïve Bayes classifier which helps in finding which news article belongs to which crime type, then the pattern is identified and Apriori algorithm is used to find the frequent crime pattern. Decision tree concept is used for prediction and heat maps are used for visualization. The regions and places with highest crime records are predicted in this paper based on the historical records. The output may vary every time with the change in historical records.

Awal et al. [3] use a linear regression model to predict future crime rates in Bangladesh. The data is collected from Bangladesh State Police Department and the model is used in this dataset. Crime forecasting is done after training the dataset to forecast the occurrence of various crimes. Machine learning and data mining techniques are used to forecast future crime trends. The dataset contained 840 instances of crime where each instance belongs to different regions of the state. The author has used a linear regression model for predicting each of the crime categories such as theft, burglary, robbery, dacoity, women and child repression. After training the linear regression the author forecasts various crimes for months in the year 2016. The paper states that the forecast results obtained for various crimes would aid police in reducing the future crimes of Bangladesh state.

Crimecast [4, 7] is used to predict crimes going to happen. And based on the prediction it will also predict the probable future crimes that are going to happen. This is done with the help of simulating probabilistic model implementation and artificial neural networks. Here they prepared dataset which contains the name of the crime, time period and location. Then calculate the probability of crime occurrence is done. After that hotspot detection is done which is the main part of this paper which has the ANN. It considers that criminals won't move from their localized area. They also considered other factors such as weather and season which plays a vital role in happening crimes.

2. Experimental Methodology

Regression analysis processes for estimating the relationship between variables. Regression analysis contains many different techniques for modeling and analyzing different data, Regression is a relationship between one dependent and one or more independent variables. It helps to understand how a value of dependent variable changes when any one of the independent variables is changed [2,3] while remaining dependent variables remains unchanged. Regression analysis is mostly used for forecasting and for prediction [1, 5]. It can also be used to understand which independent variables are related to dependent variable and thereby explore their relationships [8, 9]. Some of the methods for regression analysis are linear regression and ordinary least squares regression.

Regression models involve the following:

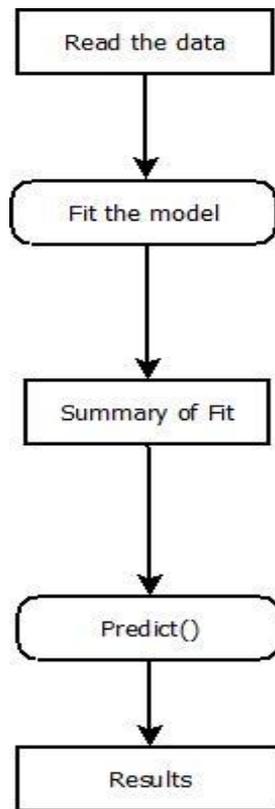


Figure 1 Architecture Structure of Application

Assumptions in multiple linear regressions are:

- Regression residuals should be distributed normally.
- A relationship assumed between the independent variable and the dependent variables which are considered as a linear.
- The residuals values approximately are rectangular in shape and homoscedastic.
- The absence of multicollinearity is assumed i.e. the independent variables are not too highly correlated.
- The task of fitting a single line through a scatter plot is the center of multilinear regressions result.
- The multiple linear regression results fit a line through a multi-dimensional space of data points.
- It identifies the effect of independent values on dependent values.
- Also, gives the understanding of the deviation of an independent variable from dependent variable.
- It can be used to determine different trends and future values.

For our study, we used r programming for predicting and analyzing the crime rates. In r there is a package util which contains different statistical functions. Lm() is the function which is used for linear modeling. It can be used to create multiple linear regression models. Its syntax is `lm(formula, data)`.

The first step is to read the entire data into R's workspace. After that load, the utils package in R. Then apply linear modeling to the different variables present in the data. After applying linear modeling take some data for training. After creating the training data its time to predict the crime rates

that may be going to happen so that we can take necessary steps to decrease the crime rates. Predict function is used to predict future crime rates. Its parameters are linear modeled data and the trained data. After the training, the prediction function will work to predict the future crime rates.

3. Results and Discussion

After running the Multiple Linear regression on the dataset containing different values we can compare the R-square value which states the success rate of the regression model.

```

Residuals:
    Min       1Q   Median       3Q      Max
-118.865  -2.842   -0.209   11.259   53.782

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.84173    5.81755    0.488  0.62844
Total_Crimes_I_2015  0.03617    0.01316    2.748  0.00964 **
Total_Crimes_I_2014  4.46373    1.77103    2.520  0.01674 *
Total_Crimes_I_2013 -1.69440    2.52309   -0.672  0.50654
Total_Crimes_I_2012  0.79743    1.68898    0.472  0.63994
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.85 on 33 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.9811,    Adjusted R-squared:  0.9788
F-statistic:  428 on 4 and 33 DF,  p-value: < 2.2e-16
    
```

Figure 2 Summary of the regression model for No. of cases registered for crime against ST

We can clearly see the R Square value to be 0.9788 i.e. 97% accuracy with low Standard Error.

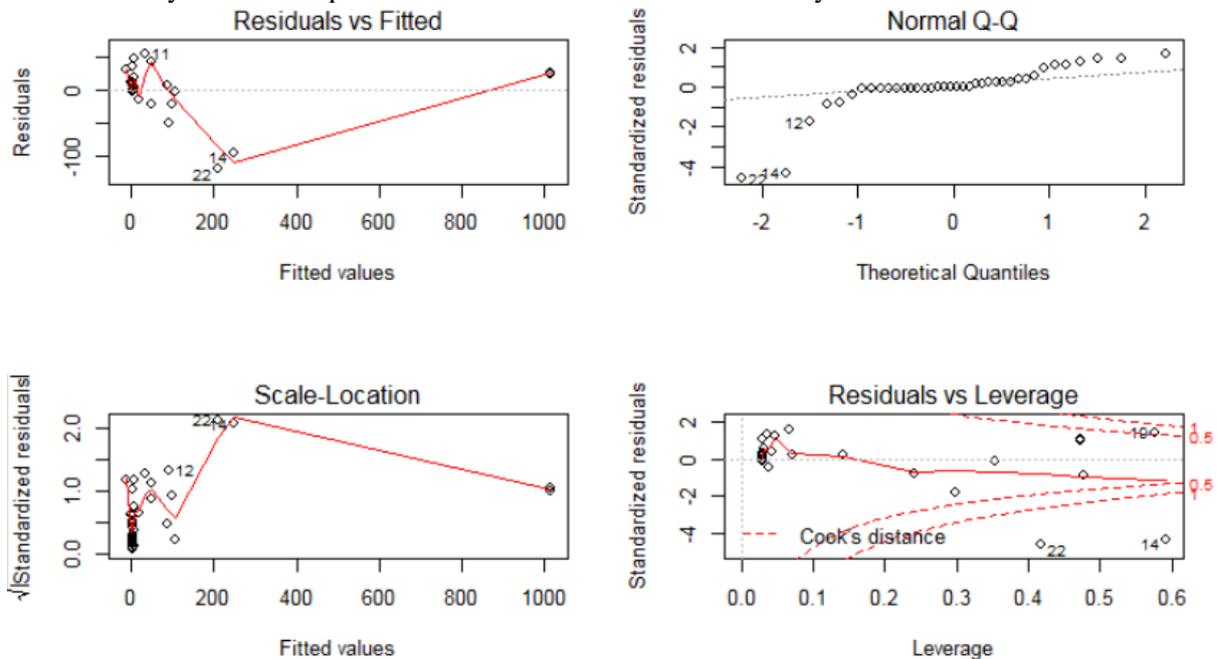


Figure 3 Residual and Normal Q Plot for No. of Cases registered for crime against ST

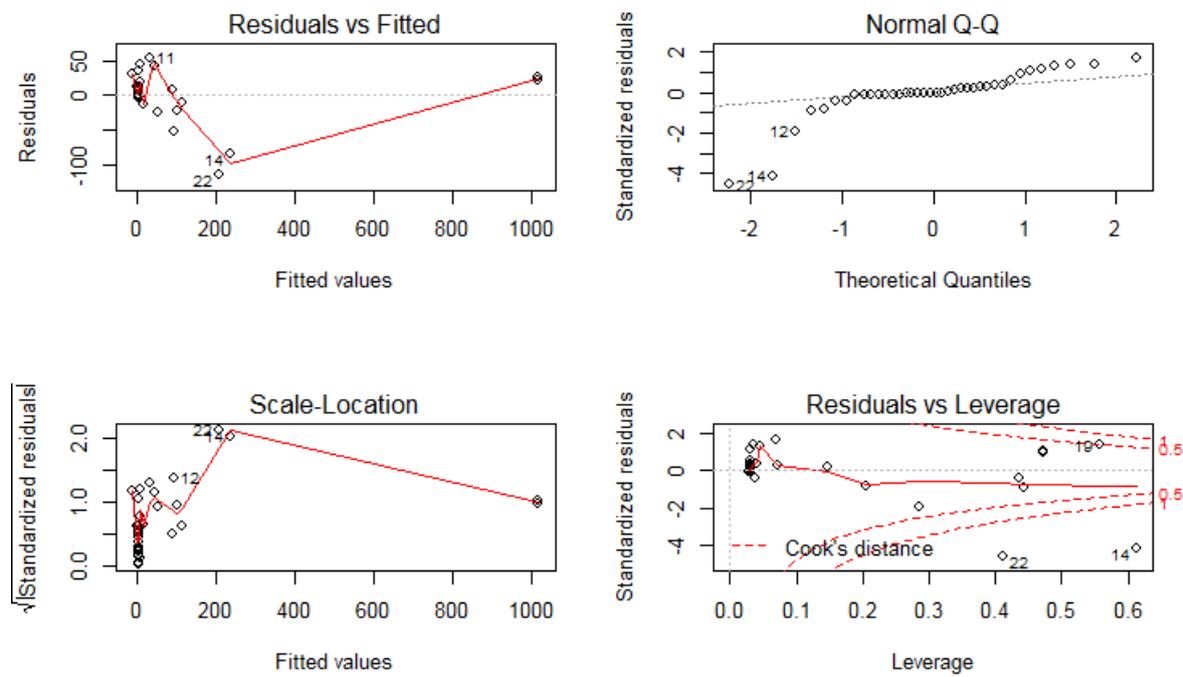


Figure 4 Residual and Normal Q Plot for No. of Victims (ST) suffered against crime

By seeing the above two figures, we can clearly see that the no. of victims and cases registered against the crime is same across 4 years of period across the country. There are some states and union territories where crime rate is negligible. But overall we can observe the crime is near about constant for 4 years.

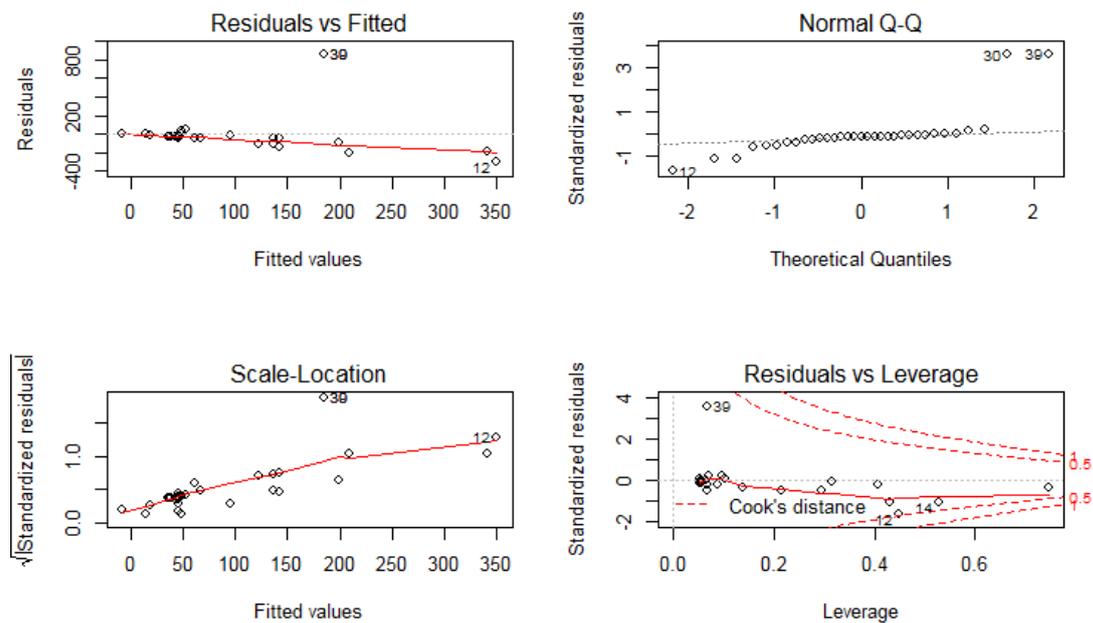


Figure 5 Residual and Normal Q Plot for the Rate of Crime against ST

From Fig. 5, we can see the rate of crime is decreasing in the Residual vs Fitted graph. But if we have a look at Normal Q-Q graph we can see the rate is constant overall and not actually decreasing.

4. Conclusions

After further evaluating the regression model and generating the regression model of different states data we can see that these are the states which mainly maximizes the crime against Scheduled Tribe (ST) community. Due to this overall rate of crime against ST in the country remains more over constant.

We can also see that for last 4 years the rate of crime is constant and not much appropriate and efficient measures have been taken against it. It's not only the duty of Central Government but also a duty of the local police to look into this issue seriously as for the last 4 years this backward tribe community is a victim of different types of crimes. By using this regression model we can predict the future crimes that too of specific type by analyzing appropriate data. Also, this regression model predicted results will help the Government and Local police to take appropriate action against it before an actual crime took place.

References

- [1] Malathi A and Santhosh Baboo S 2011 An enhanced algorithm to predict a future crime using data mining *International Journal of Computer Applications* **21**1-6
- [2] Frank E and Bouckaert R 2006 Naive Bayes for text classification with unbalanced classes *Knowledge Discovery in Databases: PKDD 2006* 503-510
- [3] Awal M A, Rabbi J, Hossain S I and Hashem M M A 2016 Using linear regression to forecast future trends in crime of Bangladesh *5th International Conference on Informatics, Electronics and Vision (ICIEV)* 333-338
- [4] Levine N 2004 CrimeStat III: a spatial statistics program for the analysis of crime incident locations *Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC*, 177-191
- [5] Yamuna S and Bhuvaneshwari N S 2012 Data mining techniques to analyze and predict crimes *Int. J. Engg. Science* **12**43-247
- [6] Buczak A L and Gifford C M 2010 Fuzzy association rule mining for community crime pattern discovery *In ACM SIGKDD Workshop on Intelligence and Security Informatics* **2**
- [7] Mahmud N, Zinnah K I, Rahman Y A and Ahmed N 2016 CrimeCast: A crime prediction and strategy direction service *19th International Conference on Computer and Information Technology (ICCIT)* 414-418
- [8] Brantingham P L and Brantingham P J 1993 Nodes, paths and edges: Considerations on the complexity of crime and the physical environment *Journal of Environmental Psychology* **13** 3-28
- [9] Chandra B, Gupta M and Gupta M P 2008 A multivariate time series clustering approach for crime trends prediction *IEEE International Conference on Systems, Man and Cybernetics* 892-896
- [10] Sherman L W, Gartin P R and Buerger M E 1989 Hot spots of predatory crime: Routine activities and the criminology of place *Criminology* **27**27-56
- [11] Software - The Stanford Natural Language Processing Group. 2014. Software. [ONLINE] Available at: <https://nlp.stanford.edu/software/>. [Accessed 16 March 2017]
- [12] Wikipedia. 2017. Crime Analysis. [ONLINE] Available at: https://en.wikipedia.org/wiki/Crime_analysis [Accessed 15 March 2017]
- [13] Sathyadevan S and Gangadharan S 2014 Crime analysis and prediction using data mining *First International Conference on Networks & Soft Computing (ICNSC)* 406-412