An Empirical Investigation of Socio-Economic Parameters Associated with Rape Crime in Karnataka using Regression Analysis

Ravva Venkata Subba Aravind[1] Rashad Mohamed Khan[2] Shashank Prakash Patil[3]

[1,2,3]Department of Computer Science and Engineering

[1,2,3]Ramaiah Institute of Technology, Bengaluru, India,

Abstract— Crime is a socially harmful act that brings disgrace to society. The crime rate in India has always been on the rise. Due to the increase in population in the state of Karnataka, the crime rate has escalated. Rape is a heinous crime that has plagued society. It is a dastardly crime and highly deplorable and unacceptable. It poses a challenge and is a worrying factor for the Government. The purpose of the paper is to conduct an empirical investigation of the relationship between various economic and demographic factors associated with rape. Regression analysis is carried out to understand the co-relation of various social factors and identify significant variables to curb crime effectively.

Keywords: crime, rape, regression, analysis, social, Karnataka

I. INTRODUCTION

Crime is an evil that degrades the quality of life and affects everybody in a civilized society. Crime sabotages the potential of the country to promote peace and development from a wider perspective [1]. High crime rates can significantly impact and negatively influence foreign and domestic investments [2], including high productive and skilled labor. Therefore, it is imperative to study what incites people to commit a crime. Historically, crimes have been solved by legal enforcement specialists. However, lately, computer analysts have aided legal enforcement in solving crimes by discovering crime patterns. The crime rate is affected by several Economic factors such as income level, unemployment rate and Gross Domestic Product (GDP), Consumer Price Index (CPI), etc. in India [3].

It is critical to understand the geographical distribution of rape crime as it varies with factors such as urbanization, socio-economic parameters, education, and spatial-temporal factors. Therefore, to understand the pattern of these crimes, clustering plays an important role [4]. Clustering of crime helps in the identification of regions with a high concentration of crime. Therefore, the spatial and temporal distribution of crime like rape is vital in identifying effective control measures to prevent it.

Comprehensive details about the data collection sources and their interpretation have been mentioned in the paper. The following figure showcases rape crime hotspot districts from the data available for the year 2019. Bangalore, Mysore, Bellary and Belgaum belong to the hotspot cluster reporting the highest number of cases in Karnataka in the previous year.

Fig. 1: Hotspot districts in 2019

II. LITERATURE SURVEY

Early theories of crime highlight the effect of poverty and social deprivation on crime rates [5]. Among economists, Fleisher thoroughly studied criminal behavior. He argued that crime rates are associated with unemployment and low-income levels [6]. Dubey opines that that economic, socio-cultural and political factors play an important role in crime and crime control practices in India. They are of the opinion that the increasing crime rate has been influenced by the constant political tussle in India as well as due to the financial crisis [7].

Gary Becker presented a model based on the cost of crime. He explained the economics of crime in terms of the cost and benefits of a particular crime. He stated that the cost of different punishments to an offender could be made comparable by converting them into their monetary equivalent or worth [8]. Dutta and Husain investigated the impact of variables like police force and arrest rate. Socio-economic variables like poverty and urbanization on crime in India were also studied. They concluded that these parameters are likely to have a significant impact on crime rates [9].

Gumus studied the effect of per capita income, income inequality, population, and presence of black population on the crime rate in the US and stated that these all are important determinants of the crime rate. Police expenditures and unemployment rate also have an impact on crime but not as much as other factors. Shingleton created a multi-variable regression model using predetermined

environmental factors affecting crimes. This model is used to predict future violence levels using statistical analysis [10].

III. MATERIALS AND METHODS

A. Study Area

India is divided into 28 states and 9 union territories with Karnataka being the 6th largest state area wise. The latest data available on National Crime Records Bureau show that out of the total rape cases in India which have been reported, four out of five rape victims belong to these 10 states - Haryana, Jharkhand, Odisha, Rajasthan, Uttar Pradesh, Madhya Pradesh, Maharashtra, Kerala, Assam and Delhi. The number of total reported rape cases in these 10 states has almost doubled in the last 10 years - from 12,772 in 2009 to 23,173 in 2019. The rest 26 states have reported almost the same numbers as they did in 2009 [11]. Karnataka, however, recorded one of the lowest crime rates in rape at 1.6 per lakh of population. Only Tamil Nadu and Bihar have a lower rate [12]. According to the Census 2011, Karnataka has 30 districts, 176 sub-districts (Taluks) and 29,340 villages and its total population is 6,10,95,297. An average number of rape cases from 2015-2019 for each district was tabulated and the same is mapped in the following figure.
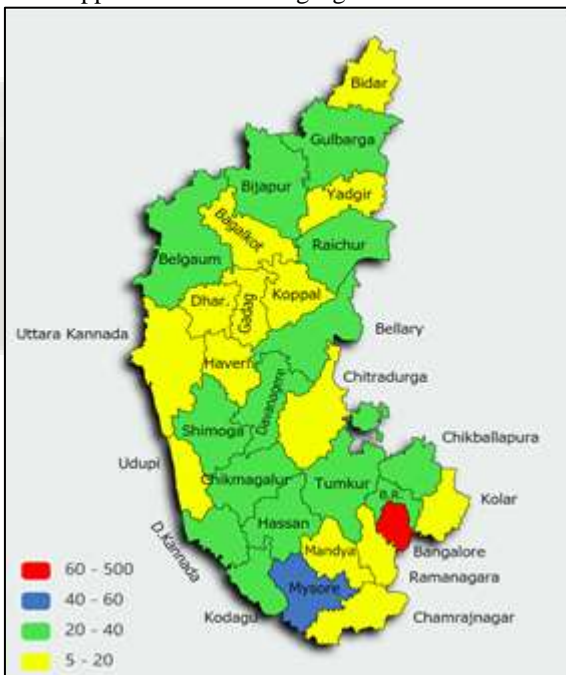

Fig. 2: Average cases (2015-19) in each district

Bangalore city has the highest average of 162.4 cases followed by Mysuru at 43.2 cases per year. Yadgir district has the lowest average of 8.8 cases per year.

IV. DATA COLLECTION

The statistics of rape cases in Karnataka are obtained from the official website of Karnataka State Police [13]. Annual crime reports from 2015 to 2019 are studied and taken for analysis. The crime report of 2019 mentions 37 distinct divisions comprising districts and cities jurisdiction whereas crime reports of 2015-18 mentions 36 distinct divisions. In contrast, there are 30 districts mentioned in 2011 census, hence the districts and cities mentioned in the crime report numbering

36 and 37 have been distributed accordingly. Belgavi city and Belgavi district are grouped under singular division Belgaum, Hubli-Dharwad city and Dharwad are grouped under Dharwad, D.K and Mangaluru city are grouped under Dakshina Kannada, Mysuru district and Mysuru city are grouped under Mysore, Kolar and KGF are grouped under a singular division Kolar.

V. MODEL USED

Data Mining provides for various algorithms which can be employed to predict and forecast crime [14]. In this paper, Regression Analysis is applied for forecasting. Regression Analysis, broadly, comprises of simple linear regression modelling and multiple linear regression modelling. A simple linear regression model works on one independent variable and only one dependent variable [15]. Likewise, a multiple linear regression model works on more than one independent variables [16]. As, in this specific study, a number of factors and parameters associated with rape are being analysed, multiple linear regression model is employed. The extent of dependency of a particular parameter and its negative or positive effect on crime is studied and analysed.

The Linear Regression Model can be expressed as

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \dots\dots + \theta_n X_n \quad \text{….eqn. 1}$$

where $\theta_0, \theta_1, \theta_2, \theta_3, \dots\dots, \theta_n$ are regression parameters.

The model is built on the following parameters: the number of rape cases (y) is the dependent variable, independent variables(x) are rural population percentage, urban population percentage, literacy rate, sex ratio, scheduled caste (SC) percentage, scheduled tribe (ST) percentage, marginal workers percentage, main workers percentage, population density of district, police station density area wise (per 100 sq km), police station density population wise (per 1lakh population).

RURPO	Rural Population Percentage
URBPOP	Urban Population Percentage
LITRATE	Literacy Rate
SEXRATIO	Sex Ratio
SCPOP	Scheduled Caste (SC) Population Percentage
STPOP	Scheduled Tribe (ST) Population Percentage
MARWORK	Marginal Workers Percentage
MAINWORK	Main Workers Percentage
POPDEN	Population Density
POLDENAREA	Police Station Density Area Wise
POLDENPOP	Police Station Density Population Wise

Table 1: Determinants of Crime

Parameter	Data Type/ Data Scale	Number of Distinct Values	Values
Districts	Nominal	30	Belgaum, Bagalkot, Bijapur, Bidar, Raichur, Koppal, Gadag, Dharwad, Uttara Kannada, Haveri, Bellary, Chitradurga,

			Davengere, Shimoga, Udupi, Chikmagalur, Tumkur, Bangalore, Mandya, Hassan, Dakshina Kannada, Kodagu, Mysore, Chamrajnagara, Gulbaraga, Yadgir, Kolar, Chikkaballapur, Bangalore Rural, Ramnagara
Year	Date		2015, 2016, 2017, 2018, 2019
Number of Rape Cases	Interval		
Rural Population	Interval		
Urban Population	Interval		
Literacy Rate	Interval		
Sex Ratio	Interval		
SC Percentage	Interval		
ST Percentage	Interval		
Marginal Workers	Interval		
Main Workers	Interval		
Population Density	Interval		
Number of Police Stations per 100 sq. km	Interval		
Number of Police Stations per 1 lakh population	Interval		

Table 2: District-wise attribute

VI. RESULTS AND DISCUSSIONS

A. Regression Analysis

According to Field, "Regression Analysis enables us to predict future outcomes based on the predictor variables" [17].

A multi-linear regression model is developed to determine the relationship between dependent variable (y) and independent variables (x). The model develops regression coefficients associated with each independent parameter. The following table contains coefficients and p-values associated with all the parameters.

Covariates	Coefficient	p-value
RURPO	1.3094	0.040
URBPOP	1.1367	0.107
LITRATE	0.4712	0.106
SEXRATIO	-0.1149	0.710
SCPOP	0.1988	0.440
STPOP	0.3655	0.048
MARWORK	-1.5661	0.040
MAINWORK	-1.3585	0.075
POPDEN	0.3189	0.190
POLDENAREA	-0.0926	0.775
POLDENPOP	-0.0785	0.754

Table 3: Coefficient estimates

Based on the p-value, only five covariates were found to be significantly associated and hence taken into consideration for formulating the regression equation. Consequently, the regression equation will have six regression parameters and can be expressed as:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \theta_5 X_5 \ldots \text{eqn. 2}$$

Substituting the values for parameters, the new equation evolves into:

$$Y = \theta_0 + \theta_1 (RURPO) + \theta_2 (STPOP) + \theta_3 (MARWORK) + \theta_4 (MAINWORK) + \theta_5 (URBPOP) \quad \ldots \text{eqn. 3}$$

From Table III, the coefficient values for different covariates can be obtained. Hence, substituting the values of coefficients into eqn. 3, a new equation is evolved termed as eqn. 4.

$$Y \text{ (Number of Rape Cases)} = \theta_0 + (1.3094)*(RURPO) + (0.3655)*(STPOP) + (-1.5661)*(MARWORK) + (-1.3585)*(MAINWORK) + (1.1367)*(URBPOP) \ldots \text{eqn. 4}$$

Evident from the above equations and p-values of different parameters as seen in Table III, only rural population percentage, ST population, marginal workers percentage, main workers percentage and literacy rate contribute significantly in determination of the number of rape cases in Karnataka. Among the significant covariates, some covariates exhibit positive coefficients, while some covariates exhibit positive coefficients. Positive co-efficient indicate higher the value of covariate, higher will be the number of rape cases. Negative coefficient indicates higher the value of covariate, lower will be the number of rape cases. According to the eqn. 4 and values in Table III, among five significant variates, three variates namely rural population percentage, urban population percentage and ST population have positive coefficients and two variates namely marginal workers percentage and main workers percentage have negative coefficients.

B. Interpretation of Results

The regression equation formed as equation 4 has explicitly defined factors which hold significant impact on the response variable. Clearly, with increase in rural and urban population percentage, ST population percentage, the number of rape cases increase, while with increase in marginal and main workers percentage, number of rape cases decrease. According, the coefficients associated with the significant covariates can be interpreted in the following manner:

- Increase in 1% of rural population in a district will result in increase 1.3094 units in rape cases.
- Increase in 1% of ST population in a district will result in increase 0.3655 units in rape cases.
- Increase in 1% of urban population in a district will result in increase 1.1367 units in rape cases.
- Increase in 1% of marginal workers in a district will result in decrease in 1.5661 units in rape cases.
- Increase in 1% of main workers in a district will result in decrease in 1.3585 units in rape cases.

VII. DISCUSSION

The above interpretation of results indicate that few covariates have a significant and real impact on the number of rape cases. It is not unknown that population growth contributes to increase in crime, however, it is noticed that in specific both rural population and urban population percentage increase contribute in increase of cases. Similarly, increase in workers percentage contributes in decreasing crime. Though few covariates are not significant, it is inferred that their impact can be studied by the positivity or negativity of their coefficients. Accordingly, increase in police density per areas as well as per population helps in reducing crime.

VIII. CONCLUSION

The interpretation of results from the regression models stipulate important covariates which influence the rape crime rate. These factors help in reducing the crime in future across Karnataka. Stringent policing and proactive citizenry will help prevention of crime.

- Further, to narrow down and identify important covariates and factors associated with increasing rape crime, the following can be undertaken in future:
- Disintegration of research: Contract the research area into smaller sub-divisions and regions and identify local factors associated in that region. This will prove to be an effective strategy as analysing a smaller area will also introduce new local factors which have influence on the increasing crime rate. Analysing the demography of an area will also help in understanding neighbouring regions' properties
- Analysis of low crime incidence regions: Study the data and factors associated with crime rate in areas and regions which have a considerably low crime rate. Analyse these regions and incorporate results in other districts with high crime rate.

REFERENCES

[1] Arthur, J.A. (1991) 'Socio-economic predictors of crime in rural Georgia'. Criminal Justice Review, 16: 29-41.

[2] Becker, G.S., (1968) 'Crime and Punishment: An Economic Approach', Journal of Political Economy, 76 (2): 169-217.

[3] Nath, S.V. (2006). Crime Pattern Detection Using Data Mining. *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops*, 41-44.

[4] Antolos, Daniel, "Investigating Factors Associated with Burglary Crime Analysis using Logistic Regression Modeling" (2011). Dissertations and Theses. 15.

[5] Shaw, Clifford R. and McKay, Henry D. (1942) Juvenile Delinquency and Urban Areas. The University of Chicago Press,

[6] Fleisher, B., (1963) 'The Effect of Unemployment on Juvenile Delinquency'. Journal of Political Economy, 71(6): 543-555

[7] DUBEY AND AGGARWAL (2015) Crime, Crime Rates and Control Techniques: A Statistical Analysis, lawoctopus.com

[8] BECKER, G.S., (1968) 'Crime and Punishment: An Economic Approach', Journal of Political Economy BOURGUIGNON, EHESS AND DELTA (1999), Crime, Violence and Inequitable Development ,Annual World Bank Conference on Development Economics, Washington, D.C., April 28-30, 1999.

[9] DUTTA, M. AND HUSAIN, Z. (2009). Determinants of crime rates: Crime Deterrence and Growth in post liberalized India.

[10] Shingleton et al., "Crime Trend Prediction Using Regression Models for Salinas, California", Dissertations and Theses, Naval Postgraduate School, Monterey, California, 2012.

[11] Rai, Dipu. "India's 10 most dangerous states for women". India Today, 10th October, 2020

[12] Bharadwaj, K.V Aditya. "Karnataka registers the highest number of sedition cases in the country". The Hindu. October 1, 2020.

[13] Karnataka State Police www.ksp.gov.in.

[14] Bharati, M. & Ramageri, Bharati. (2010). Data mining techniques and applications. Indian Journal of Computer Science and Engineering. 1.

[15] Kumari, Khushbu & Yadav, Suniti. (2018). Linear regression analysis study. Journal of the Practice of Cardiovascular Sciences. 4. 33. 10.4103/jpcs.jpcs_8_18.

[16] Kaya Uyanık, Gülden & Güler, Neşe. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences. 106. 234–240. 10.1016/j.sbspro.2013.12.027.

[17] Field, A. (2009). Discovering statistic using SPSS (3rd ed.). London: SAGE Publications Ltd.