Analysis Of Crimes Against Women In India Using Regression

R. Devakunchari, Bhowmick S, Bhutada S P, Shishodia Y

Abstract: Women security is one of the biggest concerns of the country right now. Crimes against women in the form of murders, rapes, dowry threats, have been on the up in the past decade. It has existed in the past but only in the last decade, has it become a source of discussion and concern. Several data analysis projects have been sanctioned and used for crime detection, prediction and prevention; few have been done about crime against women in India and almost none with regards to different crimes against women. This paper describes an in-detail analysis of crimes against^women from 2002 to 2011. This study uses regression as well as visualization for analyzing the crime patterns over the previously mentioned years and helps predict the possible age group to target with the awareness drives, the frequency of different crimes in different states and thus, evaluate the effectiveness of the current security measures in all the states of the country. The knowledge gained from these then can be given to the police and the various crime agencies to help them take better decisions regarding prevention of crimes against women in India. This approach can be quite effective and can be used globally.

I. INTRODUCTION

Sexual and Gender-Based Violence (SGBV), collectively, also known as Violence Against Women (VAW), are violent acts that are predominantly committed against women and girls. This type of violence is gender-based, meaning that these crimes are committed against women and girls just because of the sole reason that they are female. Violence against women can be classified into several categories and these include violence carried out by "individuals" as well as "states". Some of the forms of violence carried out by individuals are: rape, domestic violence, sexual harassment, female infanticide, etc; as well as harmful customary or traditional practices such as dowry violence, honor killings, female genital mutilation, and forced marriage. Some forms of violence are inflicted by the state such as war rape, sexual violence and slavery during conflict, forced abortion, violence by the police and authoritative personnel, stoning and flogging. Often crimes like forced prostitution and trafficking in women are perpetrated by organized crime.

This paper analyses these crimes based upon the age groups of the offenders and states they belong to by analyzing the data available. This analysis involves several steps like data processing, data cleaning, data modeling and analysis and finally, communication.

The rest of the paper is organized as follows. Section II is about the related work in this field and Section III proposes the model we have built for the analysis. Section IV talks about the results of this analysis. The paper ends with conclusion in Section V.

II. RELATED WORK

Many research papers have already been published with regards to crime data mining. Huge datasets have been assembled and then analyzed which has helped police and law enforcement agencies solve such crimes faster. Hackett M. [1], explains how the dowry deaths and cruelty depend on the level of a state's development. It also shows how greater the levels of gender equality, health and education in a state, the lower the rates of Dowry Deaths. This leads to the theory that social development change can lead to reduction in crime rates. Thaikkat R. [2], demonstrates how often crimes are committed by people known to the victim and how the crime rate has multiplied compared to the beginning of the century. Sureka R. et. al. [3] provides an insight on how the perspective of dalit women on spousal violence is a contributor to crimes against women in rural India. Das P. et. al. [4] identify states which are densely affected with crimes. Mukherjee C. [5] et.al. reveal the clear-cut patterns regional patterns. Ropmay D. [6] analyses the crimes against women in Meghalaya, a state which follows the matrilineal system.

III. PROPOSED SYSTEM

This section talks about the datasets used, tools used and the model built. The objective of this paper is to identify certain age groups or certain states of India which tend to commit each crime and also to build a model which will help predict the number of arrests that can be expected to be made for each crime, state wise and age - group wise.

Data Collection and Preprocessing

Input data plays a very important role in the field of any data pattern analysis, especially crime. In our proposed model, the datasets are used for deducing the hidden knowledge and patterns from it and then used for ratifying the predictive model built.

Retrieval Number B10270182S219/19©BEIESP
Journal Website: www.ijeat.org

124

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

The data used for this paper has been collected from the National Crime Records Bureau (NCRB) [7] crime data, which provides data and documents for public use. The collected data contains crime information, in the form of number of arrests, of all the 29 states and 7 union territories. Initially, the dataset classified crimes under several different categories like 'kidnapping and abduction', 'rape', 'insult to the modesty of women', 'dowry death', 'cruelty by husband or his relatives', 'assault on women with intent to outrage her modesty', 'importation of girls from foreign country', 'immoral traffic', 'dowry prohibition act', 'commission of sati' and 'indecent representation of women'. But for every year in the dataset, the data for the crime types 'dowry prohibition act', 'indecent representation of women', 'immoral traffic' do not hold any consequential value. So they have been removed and the remaining 8 categories have been considered for analyzing the dataset. Thus the dataset has been readied for the proposed crime data analysis.

Tools Used

For this analysis, R programming language has been used with RStudio, a free and open-source integrated development environment (IDE). Also, Tableau, a data visualization software, was used to create interactive visualizations.

System Design

This could predict what future trends may look like in different states which will allow for a better and more concentrated approach on these states system. Each module attempted to answer a different question or attempted to build a predictive model. Module 1 attempted to answer the question "Which Age Group has a tendency to commit each Crime?" through interactive visualization. Here, the data for different states was summarized so as to give only 2 dependent variables to work with. It gave several different answers for different types of crimes against women.

Module 2 attempted to answer the question "People from which State have a tendency to commit each Crime?" through interactive visualization. Here, the data for different age groups was summarized so as to give only 2 dependent variables to work with. It gave several different answers for different types of crimes against women. Module 3 attempted to build a model which would help predict the crime rate across the country for various different crimes against women, depending upon the data up to the previous year i.e. 2001 to 2012. The model would give an estimate in terms of number of arrests that are expected to be made in the upcoming year (2012). This predicted value would then be compared with the actual data for the year 2012 to see the accuracy. Figure 1 shows the structure of the entire system consisting of all the modules and their steps to final result.
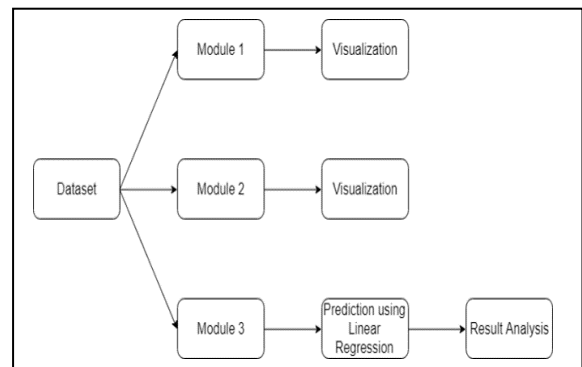

Fig. 1: System Modules

Implementation

The visualization of the acquired data was done using Tableau software for table and graph modulations. The csv files were input into the software and then parameters required were input into the software. The software then converted the said csv dataset into a graph of aforementioned parameters.
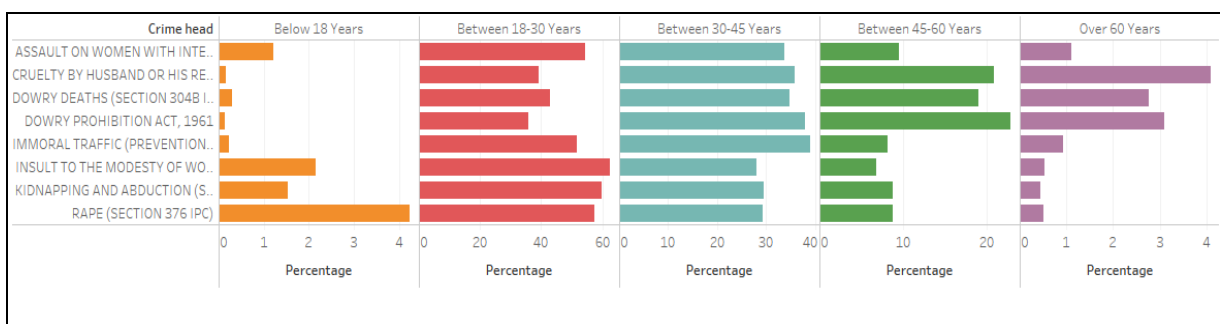

Fig. 2: Trend among age group for different crimes

Dataset for Crimes against Women in India by Age for the year 2012 was fed into Tableau and parameters were set. The first question answered was "Which particular age group has a tendency to commit a certain crime?". This question was answered with the help of Figure 2, which shows the percentage-wise trend among each age group for different crimes. The second question answered was "Which particular state has a tendency to commit a certain crime?". This question was answered with the help of Figure 3, which shows the percentage-wise trend among each state for different crimes.

Module 1

The Figure 2 show the trend among each of the age groups for different crimes, with the value depicted by x-axis being percentage.

It was seen that the age group of below 18 years tended to commit rape while criminals with their ages between 18 and 30 years tended to insult the modesty of women and kidnap and rape them. Most of the criminals of the age group of 30 to 60 years tended to be arrested for the crimes of "Cruelty by Husband and/or his Relatives" and "Dowry Prohibition Act". Most criminals over the age of 60 were arrested for the crime of "Cruelty by Husband and/or his Relatives".

Module 2

The Figure 3 showed the trend among each of the states for different crimes, with the value depicted by x-axis being percentage. Criminals belonging to the states of J&K and Arunachal Pradesh had the tendency to Assault women with the intent to outrage her modesty (54% and 38.1% respectively). The states of Andhra Pradesh, Assam, Bihar, Chandigarh, Chhattisgarh, Gujarat, Haryana, Maharashtra, Madhya Pradesh, Punjab, Uttar Pradesh, Uttarakhand and West Bengal, all had the tendency to commit the crime "Cruelty by Husband and/or his Relatives".
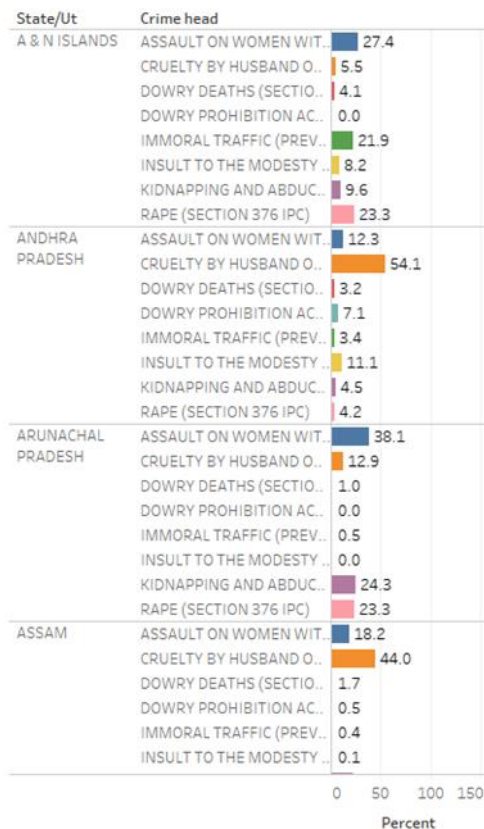

Fig.3: Trend among each state for different crimes (percentage)

For this particular crime, Gujarat and Maharashtra had the highest percentage of offenders with 83.4 % and 66.7% respectively. Goa had the highest percentage of offenders commit the crime of "Immoral Traffic" while half of the offenders from Manipur were arrested for Kidnapping and Abduction. Over more than half to two-thirds of the offenders from Mizoram and Meghalaya were arrested for Rape. Bihar had 19.8 % of its offenders being caught for Dowry Deaths which translated to 3994 arrests. This gave a glimpse of how high the crime rate was in that state.

Module 3

Several models were built and this model was chosen after testing for all the different models and their efficiency. The different models were similar in the values of Multiple R-squared and Adjusted R-squared, both close to 1(0.999). The difference between the residual standard errors led to choosing Model 1 over the other. The lesser the value of the residual standard error, the more accurate is the predicted data.

IV. EXPERIMENTAL RESULTS

Model 1 was built using the data from the years 2003-2011 (inclusive of both) while Model 2 was built using the data from the years 2005, 2006, 2007, 2009, 2010 and 2011. Model 2 was built using only these years as these were found to be most significant years out of all. Significance was found out on the basis of the values of Estimate, Standard Error, t Value and Pr value.

Both the models still had a significant value of residual because of the fact that the data for crimes is highly unpredictable but a basic idea could still be generated. Using the selected model, values for the year 2012 were predicted. The predicted values were then compared with the actual data to find out its accuracy.

Table 1 and Table 2 describes the various statistics related to the models. It gives the values such as Residual Std. Error, Multiple R squared, Adjusted R squared, F-statistic and p-value for each of the models.

Table 1: Model 1 Statistics

Residual Std. Error	**352.4 on 309 degrees of freedom**
Multiple R-squared	0.9998
Adjusted R-squared	0.9998
F-statistic	1.897e+05 on 9 and 309 degrees of freedom
p-Value	<2.2e-16

Table 2: Model 2 Statistics

Residual Std. Error	363.7 on 312 degrees of freedom
Multiple R-squared	0.9998
Adjusted R-squared	0.9998
F-statistic	2.671e+05 on 6 and 312 degrees of freedom
p-Value	<2.2e-16

Table 3 describes the actual and predicted value given by the model selected. These values help to determine the accuracy of the selected model.

Table 3: Predicted vs Actual values for the year 2012

Predicted	Actual
1571.9004	1626
1032.1031	780

829.4085	842
1107.9373	1259
4862.6766	4822
102.8516	182

After this, the error percentage was calculated which came out to be 0.004415% of the actual data. This shows that the model built had very high accuracy.

V. CONCLUSION

This analysis was done by using datasets from 2002-2012. Several modules were kept in mind while performing the data analysis. The data which was been received was not efficient, and thus data cleaning, preprocessing was done extensively to make the data efficient for use. The prediction of the data is significantly higher than it would have been had the data been used as such. The objective of the analysis to provide efficient and clear solution was achieved, thus proving that the data is reliable enough to be used for framing new laws, preventing new crimes, bringing into place new strategies to curb these activities. As future work, newer datasets could be analyzed, so that new policies can be framed. Detection technologies can improve incident detection and engage safety resources for the public sooner. This will lead to improved response times and will help the general public. Technology will help increase accuracy and efficiency of incident response and reporting and, thus resources, like police force, can be allocated better. This in turn, will enhance incident investigation efficiency and help increase clearance rates. Analytics can also help discover and identify trends to improve operational effectiveness. Proactive policing can help stop crime before it actually happens.

REFERENCES

1 Hackett, M., Domestic Violence against Women: Statistical Analysis of Crimes across India. Journal of comparative family studies 42(2), pp. 267-288 (Report 10.2307/41604436). Retrieved from Research Gate, 2018.
2 Thaikkat, R., Analysis of Crimes Against Women in India between Years 2001 – 2015. International Journal of New Technology and Research (IJNTR) ISSN:2454-4116, Vol.3, Issue 4, pp. 63-64, 2018.
3 Sureka R., Mathur A., Anshu, Spousal Violence and Dalit Women: Theoretical Underpinnings and Retrospection. International Journal of Recent Scientific Research Vol. 7, Issue, 7, pp. 12342-12348, 2018.
4 Das, Priyanka & Das, Asit., Behavioral analysis of crime against women using a graph based clustering approach, pp. 1-6, 2017.
5 Mukherjee, C & Rustagi, P & Krishnaji, N., Crimes against Women in India: Analysis of Official Statistics. Economic and Political Weekly, vol. 36, Pp. 4070-4080, 2001.
6 Ropmay, D., A., Crimes against Women in Matrilineal Meghalaya, A Forensic Medical Perspective. J Indian Acad Forensic Med., Vol. 36, No. 4, 2018.
7 National Crime Records Bureau, Ministry of Home Affairs <http://ncrb.gov.in/> Accessed on December 6, 2018.