Priya Gandhi et al. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 5, Issue 1, March 2018, pp. 284-288

Approach of Predictive Modeling on Crime Against Women Problem

Priya Gandhi, Shayog Sharma

Department of Computer Science & Engineering, Geeta Engineering College, Naultha, Panipat, India.

Abstract: **Predictive analytics is the branch of data mining concerned with the prediction of future probabilities and trends. The central element of predictive analytics is the predictor, a variable that can be measured for an individual or other entity to predict future behavior. For example, an insurance company is likely to take into account potential driving safety predictors such as age, gender, and driving record when issuing car insurance policies. Predictive Modelling can be used for betterment of society. This research uses predictive modelling for predicting crimes against women in India in various states and districts. These predictions can be used by Administration to take informed action based on predictions and analysis to make the situation better. We used data from Open Government Data Platform which has crime records of year 2001-2013 for districts and states. We applied Linear Regression to predict the future trends for different cities and stated. We also analyzed the person arrested under these crimes, gender wise arrests, age wise arrests, contribution of different states and cities in different crimes, arrests made under different crimes in different cities, etc**

Keywords: **Predictive Modelling, Crime, Women, Analytics.**

I. INTRODUCTION

Predictive modeling is a process used in predictive analytics to create a statistical model of future behavior. Predictive analytics is the area of data mining concerned with forecasting probabilities and trends.

Enterprises can gain significant long-term benefits by applying predictive analytics to their operational and historical data.

Unlike traditional business intelligence practices, which are more backward-looking in nature, Predictive analytics approaches are focused on helping companies glean actionable intelligence based on historical data, if applied correctly, predictive analytics can enable companies to identify and respond to new opportunities more quickly, predictive analytics is especially useful in situations where companies need to make quick decisions with large volume of data.

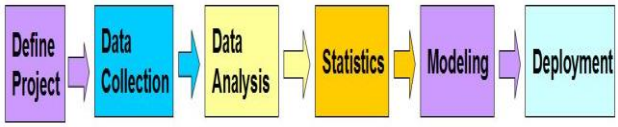


Figure 1: Predictive Analytics Process

The benefits and challenges of Predictive Analytics

In recent years, Predictive Analysis has emerged as an important solution offering enterprises a potentially cost effective model to ease their future business needs and accomplish business prediction based on final variables. Some key Functional benefits of Predictive Analysis benefits below worth considering.

One benefit is speed. When predictive models are used as part of an automated decision-making system, millions of customers can be evaluated and dealt with in just a few seconds. If a bank wants to produce a list of credit card customers who might also be good for a car loan, a predictive model allows this to be undertaken quickly and at almost zero cost. Trawling through all the bank's credit card customers manually to find the good prospects would be completely impractical. Similarly, such systems allow decisions to be made in real time while the customer is on the phone, online.

A second major benefit of using predictive models is that they generally make better forecasts than their human counterparts. How much better depends on the problem at hand and can be difficult to quantify. However, in my experience, I would expect a well-implemented decision-making system, based on predictive analytics, to make decisions that are about 20–30% more accurate than their human counterparts. In our credit scoring example this translates into granting 20–30% fewer loans to customers who would have defaulted or 20–30% more loans to good customers who will repay, depending upon how one decides to use the model. To put this in terms of raw bottom line benefit, if a bank writes off $500m in bad loans every year, then a reasonable expectation is that this could be reduced by at least $100m, if not more, by using predictive analytics. If we are talking about a marketing department spending $20m on direct marketing to recruit 300,000 new customers each year, then by adopting predictive analytics one would expect to spend about $5m less to recruit the same number of customers. Alternatively, they could expect to recruit about 75,000 more customers for the same $20m spend.

A third benefit is consistency. A given predictive model will always generate the same prediction when presented with the same data. This isn't the case with human decision makers. There is lots of evidence that even the most competent expert will come to very different conclusions and make different decision about something depending on their mood, the time of day, whether they are hungry or not and a host of other factors.

Challenges In Predictive Analytics

1. **Variable Cleaning** - Variable cleaning refers to fixing problems with values of variables themselves, including incorrect or miscoded values, outliers, and missing values. Fixing these problems can be critical to building good predictive models, and mistakes in variable cleaning can destroy predictive power in the variables that were modified. The variables requiring cleaning should have been identified already during Data Understanding. This section describes the most common solutions to those data problems.
2. **Incorrect Values** - Incorrect values are problematic because predictive modeling algorithms assume that every value in each column is completely correct. If any values are coded incorrectly, the only mechanism the algorithms have to overcome these errors is to overwhelm the errors with correctly coded values, thus making the incorrect values insignificant.
   How do you find incorrect values and fix them? For categorical variables, the first step is to examine the frequency counts.

II. PROBLEM IDENTIFICATION

Predictive Modeling on crime against women

This research is focused on using Predictive analytics for predicting crimes against women in India. Crime has always been a crucial problem in India and is affecting India's development. According to the National Crime Records Bureau of India, reported incidents of crime against women increased 6.4% during 2012, and a crime against a woman is committed every three minutes. 65% of Indian men believe women should tolerate violence in order to keep the family together, and women sometimes deserve to be beaten. There is also lack of awareness, inadequate infrastructure and insufficient programs in place for their safety and protection. This motivates us to analyze crime against women in the country. This research uses data from Open Government Data (OGD) Platform India. Although women may be victims of any of the general crimes such as 'murder','robbery', 'cheating', etc. only the crimes which are directed specifically against women are characterized as '**crimes against women**'. This research considers crime like rape, dowry deaths, cruelty by husband etc. under the category of crimes against women. Goal of this research is to predict the no of crimes against women that would happen in different states and districts of India and create visualization for better understanding of this situation. This will make administration enable to strategies better. We hope that this research will help in finding solutions for preventing these kinds of crimes in India.

Data obtained from Open Government Data (OGD) Platform India has data of years 2001-2013 for crimes against women under different categories for different states and districts. Also we obtained data about arrests made under these crimes during years 2001-2012 for different states. We trained a simple Linear Regression model on crime data to capture the trend and made predictions for future years. We made predictions for years 2014-2019. We also analyzed the crime data using different kinds of visualizations like heat maps, bar-charts, line graphs, pie charts etc.

Kinds of analysis done is as follows:

- Contribution of different states in different crimes against women.
- Contribution of different districts in different crimes against women.
- Persons arrested in different states under different crimes against women.
- Trend of different crimes in India and different states.
- Which crime happens most in crimes against women category.
- Which state has most crimes in this category.
- Gender-wise analysis of Persons arrested under crimes against women.
- Age-wise analysis of Persons arrested under crimes against women.

III. LITERATURE REVIEW

Datasets

For this research we have used six datasets. All are from Open Government Data (OGD) Platform India (data.gov.in).

1. Crime against Women during 2001-2012
2. Crime against Women during 2013
3. District-wise crimes committed against Women during 2001-2012
4. District-wise crimes committed against Women during 2013
5. Age and sex wise persons arrested under crime against women during 2012
6. Persons arrested under crime against Women during 2001-2012.

Datasets 1 and 2 are contributed by Ministry of Home Affairs, Department of States, National Crime Records Bureau (NCRB).

Datasets 3 to 6 are contributed by Ministry of Home Affairs, Department of States, National Crime Records Bureau (NCRB).

Variable Selection:

In general, for applying Linear Regression model we assume that variables are independent. But if variables are dependent then we choose only those variables which are independent for applying this model. Here we have only one variable for predicting the crime rate which is the crime rate of previous years. So we don't need variable selection in this case.

Here we projected different values of different crimes committed against women using this model i.e "**Cruelty by Husband or**

Relatives",“assault on women with intent to outrage her modesty", “Dowry Deaths", "Immoral Traffic", “Indecent Representation of Women","Insult to the Modesty of Women",“Kidnapping & Abduction",“Rape",“Total Crimes Against Women".

IV. PROPOSED METHODOLOGY

we predict scores on one variable from the scores on a second variable. The variable we are predicting is called the criterion variable and is referred to as Y. The variable we are basing our predictions on is called the predictor variable and is referred to as X. When there is only one predictor variable, the prediction method is called simple regression.

Prediction Model for predicting crimes committed against women under different categories in different States.

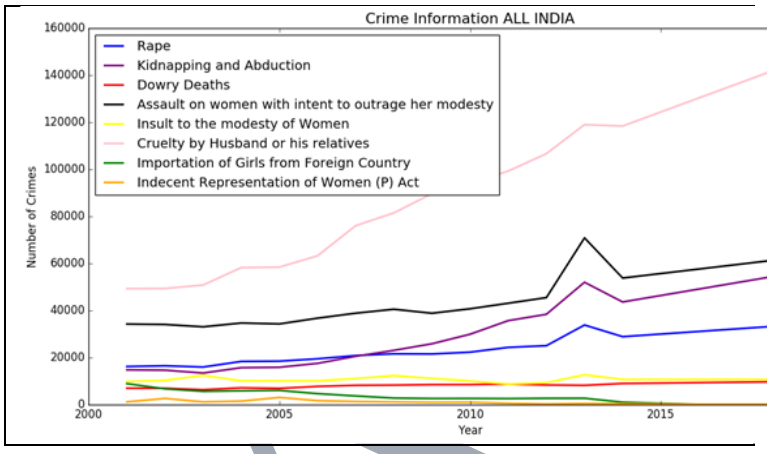
We have the data for years 2001 to 2013 for crimes committed against women under different categories. This data is plotted as a scatter plot and a linear regression line is then fit on the available data. Based on this linear model, the projected incidence of crimes likes rapes, dowry deaths, abduction & kidnapping is performed for each of the states. This is then used to build a table of different crime heads for all the states predicting the number of crimes till the year 2019. We used python's scikit-learn library's Ordinary least squares Linear Regression to fit this linear model

. This optimizes R-squared error. This model very effectively captures the trend in different categories of crimes happening in different states.

V. SIMULATION AND ANALYSIS

Data visualizations is an important tool to get an idea of data and analysis for getting inferences from it. But creating visualizations that are really that are really informative is a sophisticated task. Here we create visualizations of Crime Against women data to answer some of the questions. We used different kinds of graphs, plots like scatterplot, heat map, bar chart, piecharts etc to visualize this datasets.

Trend of crimes against women happening in different states and all over india.



For this we created a line graph of different kinds of crimes from year 2001 to 2019 i.e including predicted data for Indian and its states.

- Most reported crimes are under the category of "Cruelty by Husband or his relatives". Also increase in this category is very high in comparison to other categories. Here we can make two hypotheses that
  (1) These crimes has actually risen in numbers or
  (2) Now women feel more empowered and raise their voice against these kinds of crimes i.e about domestic abuse etc.
- Other crimes like Assault on women, Kidnapping are also increasing but at a slower rate than "Cruelty by husband or his relatives" whereas crimes like "Insult to modest of women", "Dowry Deaths", "Importation of girls", "Indecent representation of girls" are either flat or are on decline which is good sign of improvement.
- Also we can see that Rape is almost a flat line but due to media hype it seems that rape are increasing fast.

Share of percentage of crimes committed for future years.

Following table shows the normalized percentages of crimes committed against women across years 2001-2013 for actual data and for 2014-2019 for predicted data.

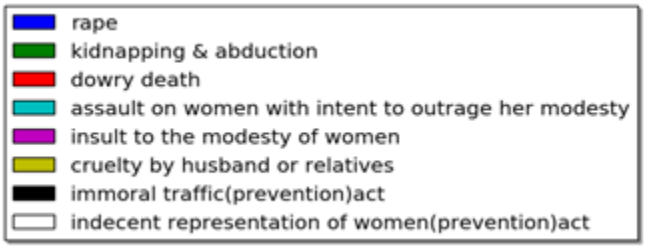
- Crimes whose percentage is predicted to be decreased are "Rape", "Dowry Deaths", "assault on women", "insult to modesty of women", "immoral traffic", and "indecent representation of women".

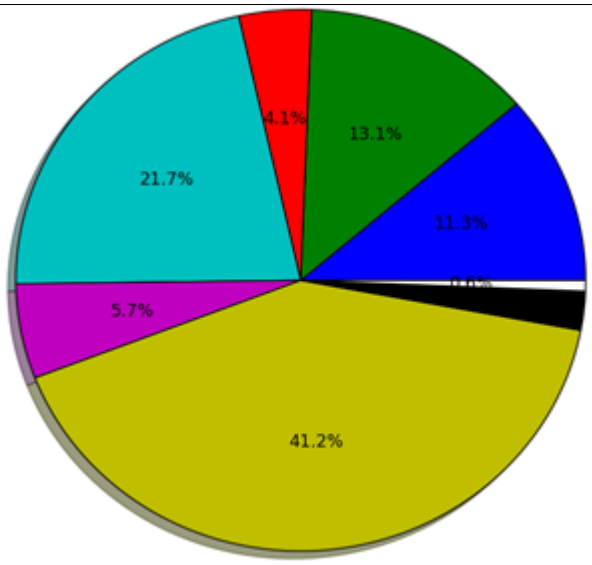
- Whereas "Kidnapping and abduction" , "Cruelty by husband or his relatives" is going to increase.
- Highest increase is going to be in "Cruelty by husband or his relatives" and will remain highest committed crime.

- Crimes like "Immoral traffic", "indecent representation of women" have negligible share of 0.074% and 0.019% respectively.

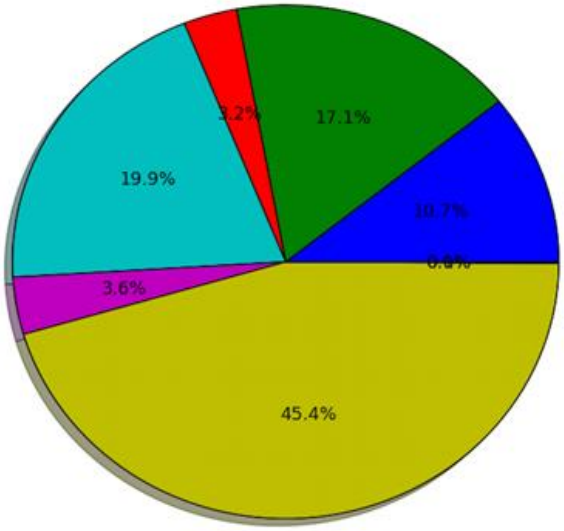
CRIME HEAD	actual	pred
rape	11.3035891394	10.7357238922
kidnapping & abduction	13.0827058947	17.1466510802
dowry death	4.11325463526	3.16767014147
assault on women with intent to outrage her modesty	21.7042798397	19.9032464916
insult to the modesty of women	5.66095032609	3.60184087027
cruelty by husband or relatives	41.1829651559	45.3501001583
immoral traffic(prevention)act	2.31503922061	0.0749626888963
indecent representation of women(prevention)act	0.637215788335	0.0198046770816

Following are piecharts showing the same percentages from above table.





First pie chart is for actual data.



This piechart is for predicted data percentages of different crimes over years 2014-2019.

VI. CONCLUSIONS

In this research, we modelled data of previous year from 2001 to 2013 to predict crime against women district and state wise for years 2014 to 2019. We also analysed the persons arrested under crime against women. There are lot inferences we've drawn from the analysis like in 2013 there are many rape cases reported in comparison to other years which could be due to protests of December 2012. We made lot of visualizations for getting these inference.

We believe these visualizations will help in better understanding the crimes scene in the country not only on crimes against women

but as a whole. Moreover, predictive modelling is a very powerful tool if used correctly for crime prediction.

Prediction provide possibilities for crime forecasting, both in terms of crime rates and pattern as well as the evaluation of crime policies. While their use has never been tested in this domain, evidence from others strongly suggests that they may yield better forecasts of future crime levels and certainly would permit more active experimentation on public safety strategies. These forecasts help in determining the amount of crime fighting resources needed and how they should be allocated across the jurisdiction.

VII. FUTURE WORK

Future work will be to apply predictive modelling not just considering the number of crimes in different districts/states but also considering population, unemployment ratio, also the various attributes regarding the crime occurred like date, time, location etc. can immensely improve the prediction model. For more data about crime is needed which is very sensitive information and not easily available, we need to convince the administration to open source the data in anonymized form. Using this level of information we can do predictive modelling to predict not just trend but also the date, time, exact location of crime which will have a large effect in the way policing in done. Also doing this level of analysis we can find correlation between different attributes like crime rate, unemployment ratio, population etc. This will help us in better understanding the reason behind the crime.

This study can be extended for other crimes like burglary, murder which are not specific to come against women.

VIII. REFERENCES

[1]. Predictive Analytics: Opportunities, Challenges and Use Cases by Lityx.

[2]. Applied Predictive Analytics "Principles and Techniques for the Professional Data Analyst" by Dean Abbott

[3]. Understanding Environmental Factors that Affect Violence in Salinas, California Clarke, Jason A.; Onufer, Tracy L. Dec,2009.

[4]. Malathi, Dr. S. Santhosh Baboo ,2011.

[5]. Poisson-Based Regression Analysis of Aggregate Crime Rates D. Wayne Osgood,2000.

[6]. Forecasting Crime: A City Level Analysis John V. Pepper Department of Economics University of Virginia.Ch-6.

[7]. Analysis of Criminal Behaviour Using a Logistic Regression Model Maria Jofre Department of Industrial Engineering, University of Chile.

[8]. Investigating Factors Associated with Burglary Crime Analysis using Logistic Regression Modeling Daniel Antolos Embry-Riddle Aeronautical University,2011

