

# Review of Predictive Modeling on Crime Pattern against Women

Priya Gandhi<sup>1</sup>, Shayog Sharma<sup>2</sup>

<sup>1</sup>M.tech Scholar, CSE Deptt, Geeta Engineering College, Naultha, Panipat.

<sup>2</sup>Astt. Prof, CSE Deptt, Geeta Engineering College, Naultha, Panipat.

**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. Predictive analytics is applied to many research areas, including meteorology, security, genetics, economics, and marketing. Predictive Modeling is use for betterment of society. This research uses predictive modelling for predicting crimes against women in India in various states and districts. These predictions e used by Administration to take decisive action based on predictions and analysis to make the situation better. In this paper, we used data from “Open Government Data Platform” which has crime records of year 2001-2013 for districts and states in India. This research identifies the various categories against crime trend and predicting future until 2019 .Moreover it focuses on the percentage of crimes rate and analysis of upcoming crimes - state wise and city wise in India.

**Keywords**—Modeling Technologies, Crime Crime Against Women.

## I. INTRODUCTION

Predictive modeling is a process used in predictive analytics to create a statistical model of future behavior. Data mining help to finding predictive information from huge database and companies use predictive modeling tools for strategic decision making.[2]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.

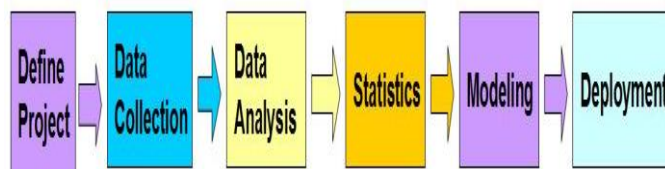


Figure 1. Predictive Analytics Process

Predictive analytics are applied in many research areas, including crime, business growth, security, genetics, economics and marketing.

## II. ADVENTAGES 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.

1. **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, in branch or online.
2. **A second major benefit** of using predictive models is that they generally make better forecasts than their human counterparts do. How much better depends on the problem at hand and can be difficult to quantify.
3. **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. [4]

## III. 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.

#### IV. PREDICTIVE MODELLING ISSUES:

Following are a few here so you can learn from our mistakes and avoid them in your own analyses:

##### 1. Failing to consider enough variables

When deciding which variables to audition for a model, you want to include anything you have on-hand that you think could possibly be predictive. Weeding out the extra variables is something that your modeling program will do, so don't be afraid to throw the kitchen sink at it for your first pass.

##### 2. Not hand-crafting some additional variables

Any guide-list of variables should be used as just that – a guide – enriched by other variables that may be unique to your institution. If there are few unique variables to be had, consider creating some to augment your dataset. Try adding new fields like “distance from institution” or creating riffs and derivations of variables you already have.

##### 3. Selecting the wrong Y-variable

When building your dataset for a logistic regression model, you'll want to select the response with the smaller number of data points as your y-variable. A great example of this from the highered world would come from building a retention model. In most cases, you'll actually want to model attrition, identifying those students who are likely to leave (hopefully the smaller group!) rather than those who are likely to stay.

##### 4. Not enough Y-variable responses

Along with making sure that your model population is large enough (1,000 records minimum) and spans enough time (3 years is good), you'll want to make sure that there are enough Y-variable responses to model. Generally, you'll want to shoot for at least 100 instances of the response you'd like to model.

##### 5. Building a model on the wrong population

To borrow an example from the world of fundraising, a model built to predict future giving will look a lot different for someone with a giving history than someone who has never given before. Consider which population you'd eventually like to use the model to score and build the model tailored to that population, or consider building two models, one for each sub-group.

##### 6. Judging the quality of a model using one measure

It's difficult to capture the quality of a model in a single number, which is why modeling outputs provide so many model fit measures. Beyond the numbers, graphic outputs like decile analysis and lift analysis can provide visual insight into how well the model is fitting your data and what the gains from using a model are likely to be.

If you're not sure which model measures to focus on, ask around. If you know someone building models similar to yours, see which ones they rely on and what ranges they shoot for. The take-home point is that with all of the information available on a model output, you'll want to consider multiple gauges before deciding whether your model is worth moving forward with.

#### V. 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. 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.

Here we identify the hypothesis questionnaires against women crime.

**Kinds of analysis done is as follows:**

1. Which states have high number of crimes committed against women?
2. Which Crimes against women are committed most?
3. Which state has how much contribution to different crimes?
4. Which district has how much contribution to different crimes?
5. Which state has how much contribution to arrests?

6. Which crime against women has how much contribution to arrests?
7. Which state has how much arrests under different crimes against women?
8. Which type of crimes are more committed by different genders?
9. Which states has more number of male/female arrests under crime against women?
10. Distribution of age and gender of persons arrested under different categories of crimes against women.

In this paper, crime detect based on linear model, the projected incidence of crimes likes rapes, dowry deaths, abduction & kidnapping is performed for each of the states and cities using 'Linear Regression Algorithm' and error function. This is used to build a table of different crime heads for all the states predicting the number of crimes till the year- 2019 prediction. In which python's scikit-learn library's Ordinary least squares Linear Regression used to fit this linear model.

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.

Moreover, prediction of this report for all over india identified the very highest percentage of reported crime under the category of "Cruelty by Husband on his relatives".

Basically it depend upon the two hypotheses - crime actually risen in numberes and women feel more empowered and raise their voice againt crime i.e about domestic abuse etc.

On the others hand 'Assault on women', 'Kidnapping' are also increased.

## VI. RELATED WORK

There has been countless of work done related to crimes. Large datasets have been reviewed, and information such as location and the type of crimes have been extracted to help police department to prevent crime and against follow up law enforcements. Existing methods have used these databases to identify crime hotspots based on locations. There are several maps applications that show the exact crime location along with the crime type for any given city .Even though crime locations have been identified, there is no information available that includes the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future. In this paper [10] cluster wise crime raking are identified, accordingly crime is divided into three types heinous crime, non-heinous crime and special & local laws violation. Second estimation is to find which area categories are more sensitive towards, areas categories which are considered are slums, residential, commercial, VIP zones, travel points and markets. On the other hand, In this paper [9] here use semi-supervised learning technique for knowledge discovery from the crime records and to help increase the predictive accuracy Crime Type and suspect sex, age, victim age group and Weapon find after analyse the Crime Pattern.

Another research titled "The utility of hotspot mapping for predicting spatial patterns of crime" looks at the different crime types to see if they differ in their prediction abilities [7]. Moreover, another title "A Survey on Crime Data Analysis of Data Mining Using Clustering Techniques". [4] In which Classification of Crime, traffic Violations, Fraud, Drug Offences, Violent Crime, Cyber Crime Clustering can also be considered the most important unsupervised learning technique; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. There are so many techniques used in clustering, in this paper discuss about K-means algorithm, Ak-Mode Algorithm, Expectation-Maximization Algorithm is used ISSN:2327782. Other existing works explore relationships between the criminal activity and the socio-economics variables such as education, ethnicity, income level, and unemployment [1]. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015 3 despite all of the existing work, none of them consider the three elements (location, time, crime type) together. In addition, It is very little research that can accurately predict where crimes will happen in the future [7]. In our study, we provide a data-mining model for crime prediction based on crime types and using spatial and temporal criminal hotspots.

D.E. Brown constructed a software framework called ReCAP (Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques. In 2009, Li Ding et al.[11] propose an integrated system called PerSearch that takes a given description of a crime, including its location, type, and the physical description of suspects(personal characteristics or vehicles) as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling, social network analysis, crime profile, and physical matching. Essentially, geographic profiling determines where" the suspects are, while other components determine the suspects. Data mining as an analysis and knowledge discovery tool has immense potential for crime data analysis. Predictive Model is tool to predict the future trend and aware against the upcoming crime.

Moreover, 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. All these facts confirm that the field is not yet mature and needs further investigations.

## VII. CONCLUSION AND FUTURE SCOPE

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. In fact, perhaps the most important measure of a crime forecasting technique may be whether it aids in crime control and prevention.

#### VIII. REFERENCES

- [1] K.S.Arthisree, M.E, A.Jaganraj, ArulmiguMeenakshi, "Identify Crime Detection Using Data Mining Techniques" ,International Journal of Advanced Research in Computer Science and Software Engineering, Amman College Of Engineering, Thiruvannamalai district, Near Kanchipuram, India,volume 3,Issue 8,August 2013.
- [2] Applied Predictive Analytics "Principles and Techniques for the Professional Data Analyst" by Dean Abbott
- [3] B.Chandra and M. P.Gupta, "Crime Data Mining for Indian Police Information System Manish Gupta", Indian Institute of Technology Delhi, HauzKhas, New Delhi - 110 016,India.
- [4] D.E. Brown, "The regional crime analysis program (RECAP): A frame work for mining data to catch criminals" , In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 2848-2853, 1998.
- [5] Ram Singla, Deepak Dembla, YogeshChaba and KevikaSingla CITM, "An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis",A Data Mining Tool C,Faridabad.
- [6] Malathi. ,Dr. S. Santhosh Baboo Reader, "An Enhanced Algorithm to Predict a Future Crime using Data Mining" ,International Journal of Computer Applications (09758887),Volume 21 No.1, May 2011.
- [7] TongWang1, Cynthia Rudin1, Daniel Wagner, and Rich Sevieri, "Learning to detect patterns of crime", JECET, 1:124131, 2012.
- [8] Revatthy Krishnamurthy,J. Satheesh Kumar, "Survey Of Data Mining Techniques On Crime Data Analysis" , International Journal of Data Mining Techniques and Applications Vol 01, Issue 02, December 2012.
- [9] ShyamVaranNath Florida, "Crime Pattern Detection Using Data Mining", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops) (WI-IATW06).
- [10] Priyanka Gera, Rajan Vohra, "City Crime Profiling Using Cluster Analysis" International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014.
- [11] Malathi. A and Dr. S. Santhosh Baboo, "Enhanced Algorithms to Identify Change in Crime Patterns", International Journal of Combinatorial Optimization Problems and Informatics, Vol. 2, No.3, Sep-Dec 2011, pp. 32- 38, ISSN: 200