

A Review on Forecasting Crime against women in India using Machine Learning Approaches

Shivani Surve¹, Dr. Rocky Upadhyay², Dr. Sheshang Degadwala³, Dhairya Vyas⁴

¹Computer Engineering Department, Sigma Institute of Engineering, Vadodara, Gujarat, India

²Computer Engineering Department, Sigma Institute of Engineering, Vadodara, Gujarat, India

³Computer Engineering Department, Sigma Institute of Engineering, Vadodara, Gujarat, India

⁴Managing Director, Shree Drashti Infotech LLP, Vadodara, Gujarat, India

ABSTRACT

Crimes against women have become a global problem, and many governments are striving to curb them. The National Crime Records Bureau indicates that crimes against women have risen substantially. In June, NCW received the most crime complaints against women in eight months. The Indian government is interested in finding a solution to this problem and promoting social progress. Each year, crime reports generate a vast amount of data, which is collated. This information may help us evaluate and anticipate criminal behavior and reduce criminal activity. Data analysis involves assessing, cleansing, manipulating, and modelling data to draw conclusions and enhance decision-making. This research uses supervision learning to analyze the Indian women's criminal examination. The police department received crime reports. Anomalies, invalid locations, longitudes, and scopes were created in advance. The study was meant to breakdown women's crimes by kind and district and produce crime heat maps. The results help decision makers predict and prevent crimes against women. Applying Find the geographical criminal hotspot and the kind of crime, such as murder, rape, sexual assault, beating, dowry threats by the husband or his family, immoral trafficking, stalking, etc.

Keywords: Crime against women, Data Mining, Region, Machine Learning, Forecasting, Support Vector Machine, K-nearest Neighbor, Random Forest, Decision Tree, Navier Bayes.

Article Info

Publication Issue :

Volume 8, Issue 6

November-December-2022

Page Number : 436-442

Article History

Accepted: 20 Nov 2022

Published: 05 Dec 2022

I. INTRODUCTION

The overall crime rate has been steadily climbing for quite some time now. Because criminal activity is either methodical or fortuitous, it is impossible to predict it. Crimes committed against women are

sometimes referred to as gender-based forms of violence. Acts of criminality that are done against women and girls, either mostly or only. It is a crime that is fast growing at an alarming rate in many regions of the nation in the present day. The purpose of this study is to conduct a survey of the criminal

activity that is causing crime against women and to come up with effective preventative methods. As a result, it is imperative to perform an analysis of the various types of data pertaining to crimes committed against women to make projections regarding patterns and trends. This will enable law enforcement officials to take effective measures to reduce the number of crimes committed against women. Here are some examples of many sorts of crimes that may be committed in India, including homicide, rape, sexual assault, battering, dowry threats, cruelty committed by the husband or his family, bringing females into the nation illegally, immoral trafficking, etc. The application of data mining methods may improve the accuracy, performance, and speed with which crime predictions are made. Mining methods are used to assess criminal trends based on both recently collected data and historical information. Therefore, the primary obstacle that stands in our way is the

creation of a better and more effective instrument for the identification of crime patterns, which will allow us to properly identify crime patterns.

Problems Associated with Crime Prediction:

- Increase the amount of information in the criminal record that should be positioned to the left and studied.
- The analysis of in sequence raises concerns given that, in turn, incompatibility and deficiency exist.
- Restrictions on how much may be made by selling criminal records at the beginning of the rule The office of enforcement
- The specificity of the research location must be determined in advance for the agenda to be accurate.

II. LITERATURE STUDY

Sr. No	Title	Author	Publication	Year	Methods
1	Forecasting the Trends and Patterns of Crime in Bangladesh using Machine Learning Model	Al Amin Biswas, Sarnali Basak	2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) by IEEE Explore	2019	Linear Regression, Polynomial Regression, Random Forest.
2	Crimes Against Women in India using Regression.	R. Devakunchari, Bhowmick S, Bhutada S P, Shishodia z.	International Journal of Innovative Technology and Exploring Engineering (IJITEE)	2019	Regression
3	Crime Analysis and Prediction Using Fuzzy C-	B. Sivanagaleela , S. Rajesh.	Proceedings of the Third International Conference on	2019	Fuzzy c mean, Fuzzy clustering.

	Means Algorithm.		Trends in Electronics and Informatics (ICOEI 2019) by IEEE Xplore		
4	Crime Detection Technique Using Data Mining and K-Means.	Khushabu A. Bokde, Tiksha P. Kakade, Dnyaneshwari S. Tumsare, Chetan G. Wadha, Prof. Deepa Bhattacharya.	International Journal of Engineering Research & Technology (IJERT)	2018	K-Means clustering, Genetic Algorithm
5	Behavioural Analysis of Crime against Women using a Graph Based Clustering Approach.	Priyanka Das, Asit Kumar Das.	IEEE	2017	Graph based clustering Approach, Info map community detection technique.
6	Crime against women in India: unveiling spatial	patterns and temporal trends of dowry deaths in the	districts of Uttar Pradesh.	G. Vicente, T. Goicoa and P. Fernandez-Rasines, M. D. Ugarte.	J. R. Statist. Soc. A
7	Diagnosis of Crime Rate against Women using k fold Cross Validation through Machine Learning Algorithms.	P. Tamilarasi, Dr.R.Uma Rani .	IEEE	2020	KNN and decision trees, Naïve Bayes, Linear
8	Crime against Women (CAW) Analysis and Prediction in Tamil Nadu Police Using	S. Lavanyaa, D. Akila.	IJRTE	2019	Classification and Prediction Methods, Decision Tree.

	Data Mining Techniques.				
9	Crime Analysis and Prediction Using Data Mining.	Shiju Sathyadevan, Devan M.S, Surya Gangadharan. S	IEEE	2014	Naive Bayes, Apriori algorithm.
10	Approach of Predictive Modeling on Crime Against Women Problem.	Priya Gandhi, Shayog Sharma.	IJRAA	2018	Linear Regression
11	Prediction of crime occurrence from multimodal data using deep learning.	Hyeon-Woo Kang, Hang-Bong Kang.	PLOS ONE	2017	DNN-based prediction model with feature-level data fusion.
12	Mapping Crime against Women in India: Spatio-Temporal Analysis, 2001-2012.	Ritvik Chauhan, Vijay Kumar Baraik.	International Journal of Law and Political Sciences.	2016	Spatio-Temporal Analysis.
13	Factors Affecting Crime Against Women Using Regression and K-Means Clustering Techniques.	Bhajneet Kaur, Laxmi Ahuja and Vinay Kumar.	Springer Nature Singapore.	2018	K-Means Clustering Technique, Regression Model

repression, theft, burglary, arms act, explosive, drugs, and smuggling) have been compiled from the Bangladesh Police website to form the dataset. The input data utilised in these studies [2,5,7,12] comes from the NCRB's crime data, which is made available to the public along with other informational

III. METHODOLOGY

A. Datasets

Records of numerous crimes (such as dacoity, robbery, abduction, murder, women & child

resources including reports and software. Data World maintains the district-wise crime data and includes numerous crimes including kidnapping, murder, theft, robbery, etc., which will be utilised for [3] I crime data. Clustering and classification [4] applied to a real crime dataset collected by the police in England and Wales between 1990 and 2011 are examples of this. [9] assembling information from several online sources, including online newspapers, blogs, social media, RSS feeds, etc. Criminal Offenses Against Women in India, 2001–2013, Open Government Data (OGD) Platform, data.gov.in. [11] Their data collection is comprised of information culled from a wide range of publicly available resources, including Chicago crime statistics, population and weather data, and photographic imagery.

B. Pre-Processing

During data cleaning, issues with the actual values of the variables are addressed, such as those with inaccurate or miscoded values, outliers, and missing values. Repairing such issues is often crucial for developing accurate prediction models and botching the variable cleaning process might render the updated variables useless. We should have previously discovered the dirty variables during Data Understanding. Methods for fixing certain data issues are outlined in this section. [10]

C. Forecasting Models

(1) **Linear Regression:** Linear regression is a supervised learning algorithm that performs a regression task. Linear regression performs prediction of the value of the dependent variable (y) from the independent variable (x). Hence, this model determines a linear relation between x and y . [1,2,7,10].

(2) **Polynomial Regression Model:** Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as n th degree polynomial. It is also called the special case of Multiple Linear Regression in ML. Because we add some polynomial terms to the

Multiple Linear regression equation to convert it into Polynomial Regression. It is a linear model with some modification to increase the accuracy. The dataset used in Polynomial regression for training is of non-linear nature. Hence, "In Polynomial regression, the original features are converted into Polynomial features of required degree ($2,3,\dots,n$) and then modelled using a linear model.[1]

(3) **Random Forest Regression Model:** Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual. Ensemble learning method combines various base models in order to produce one predictive model which is optimal and more powerful.[1]

(5) **SVM:** A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labelled training data for each category, they're able to categorize new text.[7]

(6) **Decision Tree:** It is an information delineation structure comprising of hubs and branches composed inside the sort of a tree indicated, each inner non-leaf hub is labelled with estimations of the traits. The branches setting out from an encased hub are labelled with estimations of the characteristics in the hub. Each hub is labelled with a class (characteristic value). For the most part models that grasp grouping are the primary procedure of acceptance displaying, choice fitted to information handling. There are in profitable to build, simple to decipher, simple to incorporate with Database framework and they have practically identical or better precision in numerous applications.[8]

(7) **Naive Bayes:** Naive Bayes which is a supervised learning method as well as a statistical method for classification. Naive Bayes classifier is a probabilistic

classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output. The algorithm classifies a news article into a crime type to which it fits the best. The advantage of using Naive Bayes Classifier is that it is simple and converges quicker than logistic regression. Compared to other algorithms like SVM (Support Vector Machine) which takes lot of memory the easiness for implementation and high performance makes it different from other algorithms.[9]

IV.Comparative Analysis

TABLE I
COMPARATIVE ANALYSIS

Forecasting Models	Pros.	Cons.
Linear Regression [1],[2],[10]	Works well irrespective of other data size. Gives the information about the relevance of features.	The assumption of Linear Regression.
Polynomial Regression [7,8]	Broad range of functions that can be fit. Very flexible on nonlinear problems.	Strong sensitivity to outliers.
Random Forest [9,12]	Powerful Accurate Good performance on many problems including nonlinear.	No interpretability. Overfitting can easily occur. Need to choose the number of trees.

Decision Tree Regression [10]	Interpretability Works well on both linear and nonlinear problems. No need to apply feature scaling.	Poor result on small datasets. Overfitting can easily occur.
Naïve Bayes [9]	Performs better for independent predictors. Requires small amount of training data.	Assumption of independent Predictors.

V. CONCLUSION

In this study, we give an analysis on newly accessible algorithms for analysis and prediction of crime against women in India and in some other nation. Specifically, we focus on the situation in India. The goal of the analysis is to provide effective and understandable solutions, and this will be accomplished by applying various machine learning models, each of which produces accurate and effective results. This will demonstrate that the data is trustworthy enough to be used in the creation of new laws, the prevention of new crimes, and the implementation of new strategies to combat these activities. Several types of regression models are educated using the crime statistics from past years. This analytical effort may be expanded by using many years as testing data. Additionally, it is feasible to apply many alternative regression models into the provided information to forecast the crime patterns for several different time periods.

VI.REFERENCES

[1] A.A.Biswas and S.Basak "Forecasting the Trends and Patterns of Crime in Bangladesh using

- Machine Learning Model," pp. 114-118, doi:10.1109/ICCT46177.2019.8969031
- [2] S. Lavanyaa and D. Akila "Crime against Women (CAW) Analysis and Prediction in Tamilnadu Police Using Data Mining Techniques", pp 261-265, doi:10.31838/jcr.07.03.98
- [3] B. Sivanagaleela and S. Rajesh "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm", pp 595-599, doi:10.1109/ICOEI.2019.8862691
- [4] Khushabu A. Bokde, Tiksha P. Kakade, Dnyaneshwari S. Tumsare , Chetan G. Wadhai and Deepa Bhattacharya "Crime Detection Technique Using Data Mining and K-Means", pp 223-226, doi:10.17577/IJERTV7IS020110
- [5] Priyanka Das and Asit Kumar Das "CBehavioural analysis of crime against women using a graph-based clustering approach", doi:10.1109/ICCGI.2017.8117714
- [6] G. Vicente, T. Goicoa and P. Fernandez-Rasines, M. D. Ugarte "Crime against women in India: unveiling spatial patterns and temporal trends of dowry deaths in the districts of Uttar Pradesh.", pp 655-679, doi: https://doi.org/10.1111/rssa.12545
- [7] P. Tamilarasi, Dr.R.Uma Rani "Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning Algorithms", pp 1034-1038, doi:10.1109/ICCMC48092.2020.ICCMC-000193
- [8] S. Lavanyaa , D. Akila "Crime against Women (CAW) Analysis and Prediction in Tamil Nadu Police Using Data Mining Techniques", pp 595-599, doi:10.1109/ICOEI.2019.8862691
- [9] Shiju Sathyadevan, Devan M.S, Surya Gangadharan. S. "Crime Analysis and Prediction Using Data Mining ", pp 406-412, doi:10.1109/CNSC.2014.6906719
- [10] Priya Gandhi, Shayog Sharma "Approach of Predictive Modeling on Crime Against Women Problem", pp 284-288, https://www.ijrra.net/Vol5issue1/IJRRRA-05-01-65.pdf
- [11] Hyeon-Woo Kang, Hang-Bong Kang "Prediction of crime occurrence from multimodal data using deep learning", pp 1-19, doi: https://doi.org/10.1371/journal.pone.0176244
- [12] Ritvik Chauhan, Vijay Kumar Baraik "Mapping Crime against Women in India: Spatio-Temporal Analysis, 2001-2012 ", pp 2243-2254, doi: doi.org/10.5281/zenodo.1127980
- [13] Bhajneet Kaur, Laxmi Ahuja and Vinay Kumar "Factors Affecting Crime Against Women Using Regression and K-Means Clustering Techniques", pp 149-162, doi: https://doi.org/10.1007/978-981-10-3953-9_15

Cite this article as :

Shivani Surve, Dr. Rocky Upadhyay, Dr. Sheshang Degadwala, Dhairya Vyas, "A Review on Forecasting Crime against women in India using Machine Learning Approaches", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 436-442, November-December 2022. Available at doi : https://doi.org/10.32628/CSEIT228666
Journal URL : https://ijsrcseit.com/CSEIT228666