

Crime Data Analysis and Prediction

Ritisha D¹, Pooja R² and Dr. Jasmine K.S³

¹RV College of Engineering/MCA, Bangalore, India

Email: ritishad.mca20@rvce.edu.in

²⁻³RV College of Engineering/MCA, Bangalore, India

Email: poojar.mca20@rvce.edu.in, jasmineks.mca20@rvce.edu.in

Abstract—Crimes against women have created a negative impact on society both socially and economically. The Law is facing numerous challenges while trying to prevent crimes against women. Crimes against women are rapidly growing in India and it is a big threat to humanity. In every part of India, women face various issues such as kidnapping, domestic violence, acid attacks, rapes, etc. The crime data analytic system will be able to analyze and predict the rate of crimes happening in various states in India by using machine learning algorithms such as linear regression and random forest classifiers. By using this web application, the user will know the statistics and occurrence of crimes in one particular state with the data visualization done on the dataset available.

Index Terms— Crime Analysis, Flask Architecture, Linear Regression, Random Forest classifier.

I. INTRODUCTION

Crimes against women have created a negative impact on society both socially and economically. The Law is facing numerous challenges while trying to prevent crimes against women. Crimes against women are rapidly growing in India and is a biggest threat to humanity. In India, women face various issues such as kidnapping, domestic violence, acid attacks, rapes, etc. The proposed work is a web application that detects and predicts the crime rate against women using the features present in the dataset. The various tasks considered are Exploratory data analysis, data pre-processing, training the dataset, building a machine learning model, evaluating the model and classifying the type of crime with the help of Machine learning algorithms and data visualization.

The analysis and prediction of crimes against women using machine learning algorithms will help women to create awareness among themselves and avoid travelling to a certain area at a particular time. By using this web application, the user will get to know the statistics and occurrence of crimes in a particular area with visualization of different categories of crime rates against women. Also, police can increase their patrol force in a particular area where crime rates are high and can reduce the rate of crime against women. Gradually, the rate of crime against women can be reduced with help of police, hoping India will be a better living place for women.

II. PROPOSED SYSTEM

The system uses linear regression algorithm in major section i.e. predicting the Crime Patterns with minimum accuracy rates as factor. Linear Regression is the suitable algorithm to predict pattern of this data compared to some other algorithms for testing purpose such as random forest algorithm, and found linear regression as the

best fit for this proposed system with minimal accuracy in the results. The system is Scalable as the Dataset increases, the variation in the Visualization increases. As the increase in the system's workload, the system would be able to process and so the system is Scalable to an apex.

A.Unique Technical Features

- Develop a platform that can be used to analyse crime data using descriptive and predictive data analytics techniques.
- State-wise analysis of Crimes against Women.
- Predicting the crime hotspots in India
- Web application for analysis of crime against women in India.

B. Flask Architecture and Web Application

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or other components where pre-existing third-party libraries provide common functions. Flask supports extensions that will add application features implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several other common framework related tools.

C. Machine Learning Algorithms

The following are machine learning algorithms implemented for the work

Linear Regression:

Linear regression is one of the easy and most popular Machine Learning algorithms. It is a statistical procedure that is used for predictive analysis. Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. The linear regression model provides a sloped line representing the connection between the variables. Mathematically, we represent a linear regression as:

a. $y = a_0 + a_1x + \epsilon$ b. Y=Dependent Variable (Target Variable)

c. X= Independent Variable (predictor Variable)

a_0 = intercept of the road (Gives a further degree of freedom)

a_1 = rectilinear regression coefficient (scale factor to every input value).

ϵ = random error

The values for x and y variables are training datasets for Linear Regression model representation.[7].

Random Forest:

Random Forest is another classification method, also known as the improved CART method (Classification and Regression Trees). It is based on ensemble learning and using that it makes a lot of classification trees. Everyone can be built using a single deterministic algorithm or they can be built using different algorithms. There built depends on two factors:

- The best split is chosen at each node from a random subset.
- These trees are built using two-thirds to create the model and the rest is employed to predict the accuracy.

III. METHODOLOGY

The proposed work is a web application built which detects and predicts the crime rate against women using the features present in the dataset gathered together. The various tasks to be considered are Exploratory data analysis, data pre-processing, training the dataset, building a machine learning model, evaluating the model and classifying the type of crime with the help of Machine learning algorithms and data visualization.

Data Pre-processing plays a vital role before implementing the machine learning algorithms. From the Dataset acquired from the government portal there were undefined and unnamed features that needed to be dropped as they were not related and had no meaning to adding these features to the algorithms. The system uses a linear regression Algorithm in a major section i.e. Predicting the Crime Patterns with minimum accuracy rates as a factor. Linear Regression is the suitable Algorithm to predict the pattern of this data compared to other algorithms used for testing purposes such as Random Forest Algorithm, and found linear regression is the best fit for this proposed system with maximum accuracy in the results. The system is scalable as the dataset increases,

the variation in the visualization of analysis and the accuracy of the prediction increases. With the increase in the system's workload, the system would be able to process and so the system is scalable to an apex.

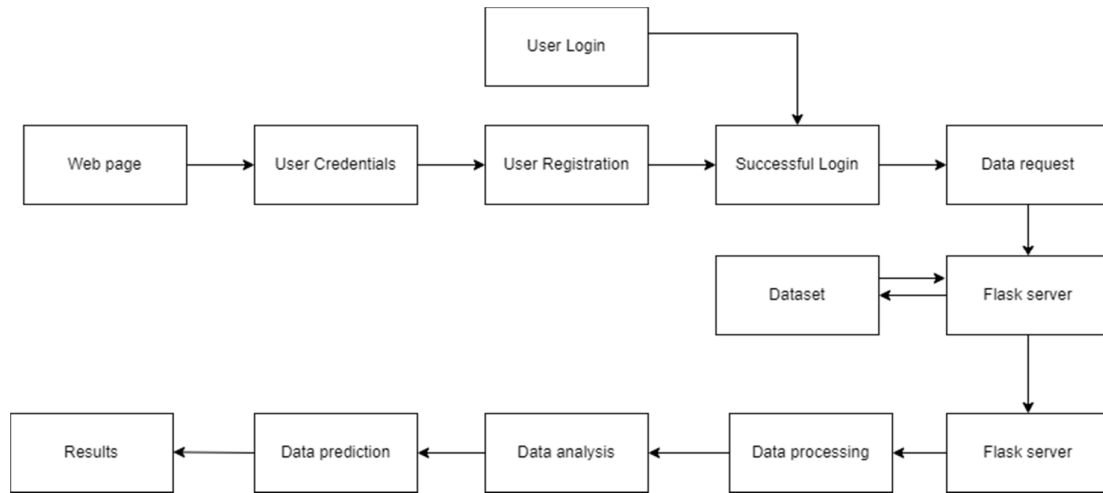


Figure.1 Block diagram of the Application

Figure. 1 describes the proposed system of the application. Python programming language is used for implementation. The user will request the analysis or prediction of crime against women in India using three inputs: year, the state of India in which the user wants to visualize and the type of crime category. After fetching the inputs from the web page, the query values will be fetched to the server which will request for the crime prediction or analysis visual from the model which in turn would be fetched back to the user.

IV. IMPLEMENTATION

In the proposed work the data is first pre-processed by removing the missing value and columns which are not needed for the analysis and prediction model. Then we are splitting the data into X and Y attributes, where Y contains column class and its corresponding row entities and X attribute contains all other column and their corresponding row entries with the help of sklearn split data library. The test and train data is split into a 3:1 ratio in which the training data set size would be 0.80 of the total dataset size and the testing dataset size would be 0.20 of the total data set. After dividing the training and testing dataset, the Machine learning algorithms Random Forest algorithm and Linear Regression that are compatible and relevant for our dataset. Random Forest is implemented to the train data which is 80% of the total dataset. Then, we tested the model using the test data that consists of 20% of the dataset. The accuracy score of the predictive model is observed to be 0.7692307692307693 that has 76.923% accuracy. Linear Regression was implemented on the prediction model and has observed accuracy of 83% percentage. Linear Regression gave better accuracy results when compared with Random Forest Algorithm.

A. Modules of the Application

User Registration/Login: This module enables the data administrator to register as an authorized person and enables him/her to register via an OTP verification sent to the email provided and then login into the dashboard where he/she can make changes to the dataset present.

Figure 1 allows the user to create an account and then register with name and password so that they can add the database in future.

Periodical Analysis of Crime Against Women: This module would present the comparative observations for all type of crimes against women, using bar graphs for making the study easier and interesting which will grab user's interest.

Predicting Crimes against Women: This module asks for the user choice of a particular category and redirects to the visualization of the crime rates accordingly.

Visualization of Crime Hotspots: This module displays the statistics of the crimes that happened previously the predicts the crime rates for the upcoming years with the respect to the input provided by the user.

The interface is divided into two main sections. The left section, titled 'Create An Account', contains input fields for 'Name', 'Email', and 'Password', followed by a 'Signup' button and a link that says 'Already Have An Account?'. The right section, titled 'Admin Signup Section', features a stylized icon of a person's head and shoulders with a plus sign, indicating an admin role.

Figure. 1: This page allows the user to create an account

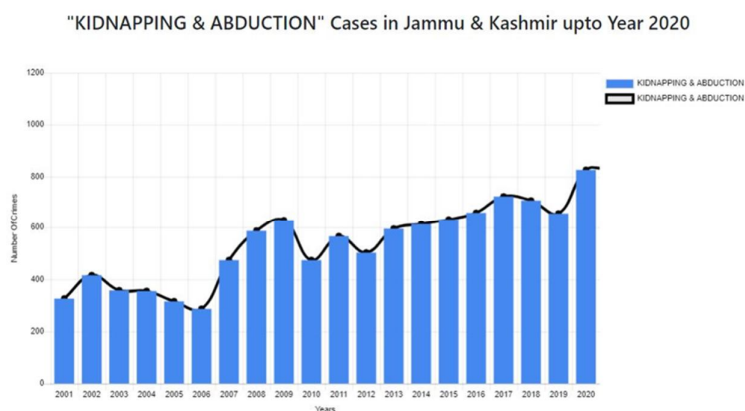


Figure 2: Kidnapping and abduction cases in Jammu and Kashmir up to the year 2020 on selecting crime type, state and year we can predict analysis

V. CONCLUSIONS

This application was successful in analysing and predicting various crimes against women in India. Random Forest and Linear Regression predictive models were implemented on the data set and accuracy of 76.923% and 83% of accuracy was obtained respectively. Linear Regression Algorithm showed better performance in terms of accuracy for the prediction of crime rate against women in India. The Police Department of a particular state can visualize the statistics and prediction results of various crime types and increase their Police force so that crime in that particular state is reduced and women can travel without any hesitation to that part of India. Gradually crimes rates will be decreased with help of the police force and Government officials implementing strict rules, hoping India to be a better and safe place for women. As part of future work, analysis and prediction can be done district wise for each state in India and user statistics using the application that can be implemented.

REFERENCES

- [1] Asit Kumar Das, Danilo Pelsui, Janmenjoy Nayak & Priyanka Das. A Graph-Based Clustering Approach for Relation Extraction From Crime Data. IEEE ,2019.
- [2] Asit Kumar Das, Danilo Pelsui, Janmenjoy Nayak & Priyanka Das. Behavioral Analysis of Crime against Women using a Graph-Based Clustering Approach. International Conference on Computer Communication and Informatics, 2017.

- [3] Aman Kumar, Nikhil Tiwari, Prakhar Gupta & Dr S.N.Rajan .Women Crime Prediction”, International Research Journal of Engineering and Technology,2017.
- [4] B. Sivanagaleela, S. Rajesh.Crime Analysis and Prediction Using Fuzzy C-Means Algorithm. International Conference on Trends in Electronics and Informatics,2019.
- [5] Keerthi. R, Karthik. B, Pavithraa.S & Dr V.Gowri. Prediction of Crime rate Analysis using Machine Learning Approach”, International Research Journal of Engineering and Technology,2020
- [6] Lavanya, D. Akila. Crime against Women (CAW) Analysis and Prediction in Tamilnadu Police Using Data Mining Technique. International Journal of Recent Technology and Engineering,2019
- [7] Manisha Khandelwal ,Muskaan, Teena Khandelwal & Purnendu Shekhar,. Women Safety Device Designed using IoT and Machine Learning. 2018 IEEE Smart World,2019.