



Crime Analysis and Prediction System

Akhand Prakash Singh¹ Aman Singh² Gourav³ Lokraj⁴ Mr. Sachin Kumar⁵

^{1,2,3,4}Student ⁵Project Mentor

^{1,2,3,4,5}Department of Information Technology

^{1,2,3,4,5}NIET, Gr. Noida, U.P, India

Abstract- Prevention is better than cure. Preventing crime from happening is better than investigating crime. Just as vaccines are given to a child to prevent disease, in today's world with such high crime rates and violent crime, it is important to have a crime prevention program in place. By vaccinating the community in crime, it refers to various methods such as public education, public awareness, skills development and police vigilance and other prevention measures. Predicting crime incidents relies heavily on crime history and a variety of geographical and demographic knowledge.

Key Words: Crime, Data mining, Prevention, Dataset, K-Means, Analysis.

1. INTRODUCTION

The crime rate in India is rising day by day, becoming a major concern, preventing good governance in India. Due to the rapid growth in crime, it is not possible to analyze such crime-related data and detect criminal patterns or predict future crimes by intelligence agencies or local law enforcement agencies. This project presents a detailed analysis of various relevant crime patterns and statistical analysis of crime data. This research will assist law enforcement agencies in developing strategies and strategies to deal with crime and disruption.

We have introduced a data mining system to predict crime. The K-means algorithm plays an important role in crime analysis and reporting. The K-means algorithm will integrate partners, collaborations and disintegration of organized crime groups, identify various key crime patterns, hidden links, link predictions and statistical analysis of crime data. In

this paper, the k-means algorithm is used to predict crime.

The goal could be to train a predictive model. Training will be conducted using a set of training data that will be validated using a test database. Modeling will be done using the best algorithm depending on the accuracy. K-Means will be used to predict crime. Database observations were made to analyze potential crimes in the country.

2. LITERATURE REVIEW

According to our research, crime analysis was done from past records. The Crime Analysis Tool (CIA) is also used to solve violent crimes and accuracy is limited. In our research we have read many different papers on this and the different strategies used by different people.

[i] This paper deals with crime in the city of Saint-Francois. The study proposes a crime prediction model by analyzing and comparing three prediction algorithms: Naive Bayes, Random Forest, and the Gradient Boosting Decision Tree. The model analyzes the top ten cases to make predictions about different categories, accounting for 97% of incident sites. The accuracy of the Naive Bayes, Informal Forest, strategies 65.82%, 63.43%, and 98.5% respectively, and the proposed model is re-evaluated for accuracy and matrix memory. The results show that the forecast model is the Gradient Boosting Decision Tree better than the other two techniques for predicting crime, based on historical data from a city. The analysis and prediction model can help the security agencies utilize the resources efficiently,

anticipate the crime at a specific time, and serve society well.

[ii] This paper is designed to prevent and respond to crime by predicting crime based on prudence. Using crime data from the city of Baltimore between 2011 and 2016, they analysed and predicted the different types of crimes and tools used in crime. Among the various machine learning algorithms, crime tools and crime types are predicted using the Multiclass Logic Regulation, which can increase model stability during faster learning and unnecessary flexibility, and the Multiclass Neural Network, which can create more accurate and complex models. In predicting the types of crimes and tools used in crime, the Multiclass Logistic Regression has shown greater accuracy, precision, and memory than the Multiclass Neural Network.

[iii] In this case, the purpose is to train a model for predicting future robberies. In this case, the data set contains the date and the level of crime that occurred in the corresponding years. In this work the crime rate is based solely on crime. Algorithm retrospective algorithm predicts future crime rate using previous data information. The date is given as an algorithm and the result is a percentage of the crime rate for that year. In mathematics, a logic (or logic) model is used to model a probability of a particular class or event such as passing / failing, winning / losing, alive / dead or healthy / sick. Moving backwards is a mathematical model that in its basic form uses a planning function to model a binary-dependent variant.

The model predicts the type of crime and the data detection helps to analyze the data set and predict crime.

3. PROPOSED METHODOLOGY

There is list of modules this project contains:

4. ALGORITHM

The domain contains many integration algorithms. But we used the K-Means merging algorithm because it is easier to use and has better accuracy than other merging algorithm. The algorithm takes a label-free database as included, divides the database into k-number of clusters, and repeats the process until it finds the best collections. The value of k should be determined in advance in this algorithm. The algorithm performs mainly two functions:

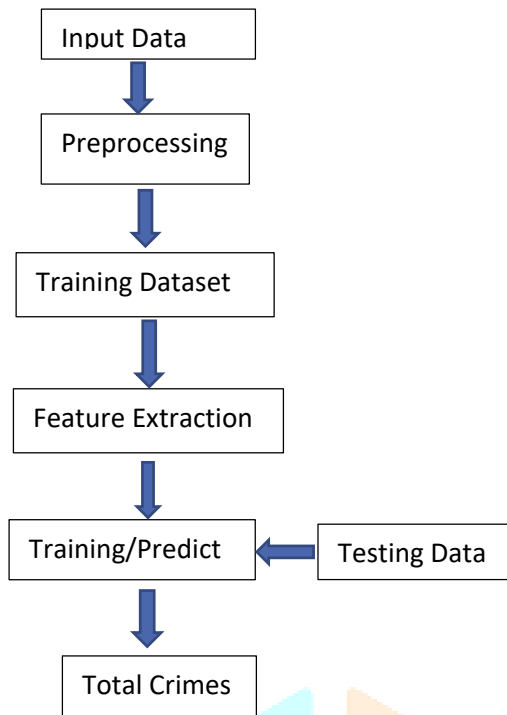
- ✓ Data Collection
- ✓ Dataset
- ✓ Data Preparation
- ✓ Model Selection
- ✓ Analyze and Prediction
- ✓ Accuracy on test set
- ✓ Saving the Trained Model

Data collection is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions and etc. Comparison of Machine Learning Algorithms for Predicting Crime Hotspots taken from Kaggle and some other source.

The dataset consists of 821 individual data. There are 27 columns in the dataset, which are described below as state, district, year of crime, type of crime etc. Then we will transform the data by getting rid of missing data and removing some columns. First, we will create a list of column names that we want to keep or retain. Next, we drop or remove all columns except for the columns that we want to retain. Finally, we drop or remove the rows that have missing values from the data set.

While creating a machine learning model, we need two datasets, one for training and other for testing. But now we have only one. So, let's split this in two with a ratio of 80:20. We will also divide the data frame into feature column and label column. Here we imported train_test_split function of sklearn. Then use it to split the dataset. Also, test_size = 0.2, it makes the split with 80% as train dataset and 20% as test dataset. Once the model is trained, we need to Test the model. For that we will pass test to the predict method. In the actual dataset, we chose only 3 features these are State, District and Year. We will predict total number of rapes, theft, molestation, murder etc.

- Determines the best value of K-center or centroid points by a repetitive process. Provides data point to its nearest k center.
- Those data points that are close to a certain k point, form a cluster.



4.1 Advantages:

- The purpose of this work is to improve our previously proposed prediction framework through alternative crime mapping and feature engineering approaches, and provide an open-source implementation that police analysts can use to deploy more effective predictive policing.
- This work helps the law enforcement agencies to predict and detect crimes in India with improved accuracy and thus reduces the crime rate.

4.2 Main Features:

- **Login Page:** You can simply login to the website by setting the password.
- **Upload Dataset:** You can simply upload the dataset for which you want to train the model.
- **Train and Test:** After uploading the dataset to the website you can train it by simply clicking on train.
- **Predict:** You can predict the future crimes by just adding the year and the name of the state you want to know about. You can also compare the crimes in different years

5. CONCLUSION

With the help of machine learning technology, it is easy to find correlations and patterns between different data. The work of this project is very much focused on predicting the type of crime that can occur if we know where it happened. We apply the concept of learning his machine model using a set of trained data sets and data conversion. The model predicts the type of crime with good accuracy. Viewing data helps to analyze the data set. Graphs include bar, pie, line and scatter graphs each has its own characteristics. We did a lot of graphs and found interesting statistics that helped to understand Indian crime data sets that can help capture the features that can help keep the community safe.

6. ACKNOWLEDGMENT

We would like to express our sincere gratitude towards our mentor Prof. Mr. Sachin Kumar for his constant help, encouragement, guidance and inspiration throughout the project work. Also, we would like to thank the Head of Information Technology Department for their valuable guidance, ability to motivate us and even willingness to solve difficulty made it possible to make our project unique and made task easier.

7. REFERENCES

- [1] Muzammil Khan and Azmat Ali, "A crime prediction model using San Francisco crime data", 2022, pp. 4830411
- [2] Kyoung-Sook KIM and Yeong-Hoon JEONG, "A study on crime prediction to reduce crime rate based on artificial intelligence.", 2021

[3] G. Nani Babu, G. Balu Nitin Chandra, A. Avinash, Y. Aditya, "Crime rate prediction using KNN", 5 May, 2020

[4] U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," in Proc. 2nd Asian Conf. Defence Technol. (ACDT), Jan. 2016, pp. 123128.

[5] J. M. Caplan, L. W. Kennedy, and J. Miller, "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting," Justice Quart., vol. 28, no. 2, pp. 360381, Apr. 2011.

[6] M. Cahill and G. Mulligan, "Using geographically weighted regression to explore local crime patterns," Social Sci. Comput. Rev., vol. 25, no. 2, pp. 174193, May 2007.

[7] A. Almeahmadi, Z. Joudaki, and R. Jalali, "Language usage on Twitter predicts crime rates," in Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN), 2017, pp. 307310.

[8] H. Berestycki and J.-P. Nadal, "Self-organised critical hot spots of criminal activity," Eur. J. Appl. Math., vol. 21, nos. 45, pp. 371399, Oct. 2010.

[9] K. C. Baumgartner, S. Ferrari, and C. G. Salfati, "Bayesian network modeling of offender behavior for criminal proling," in Proc. 44th IEEE Conf. Decis. Control, Eur. Control Conf. (CDC-ECC), Dec. 2005, pp. 27022709

