JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CRIME TYPE AND OCCRANCE PREDICTION USING MACHINE LEARNING

Shylaja L N[1] , Monisha K M[2] , Shreya K Shetty[3] , Sushma A C[4]

1. Guide, 2,3&4. Student Computer Science Department, Bahubali College of Engineering.

ABSTRACT: Crime is one of the serious issues in our society. It is the most predominant aspect of our society. It is also predominant in society. So, the prevention of crime is one of the important tasks. The crime analysis should be in a systematic way. As the analysis makes it important in the detecting and prevention of crime. The analysis detects the investigating patterns and helps in the detection of trends in crime. The main of this paper is the analysis of the efficiency of the crime investigation. The model is designed for the detection of crime patterns from inferences. The inferences are collected from the crime scene and these inferences, the paper demonstrates the prediction of the perpetrator. The paper gives the research way for the prediction of perpetrator age and gender. This paper gives two major aspects of crime prediction. One is perpetrator gender and the other is perpetrator age. The parameters used are analysis of the various actors like the year, month, and weapon used in the unsolved crimes

Index Terms: Crime Type and Occurrence, Random Forest, Decision Tree

I. INTRODUCTION Crime has become a major thread imposed which is considered to grow relatively high in intensity. An action stated is said to be a crime, when it violates the rule, against the government laws and it is highly offensive. The crime pattern analysis requires a study in the different aspects of criminology and also in indicating patterns. The Government has to spend a lot of time and work to imply technology to govern some of these criminal activities. Hence, use of machine learning techniques and its records is required to predict the crime type and patterns. It imposes the uses of existing crime data and predicts the crime type and its occurrence bases on the location and time. Researchers undergone many studies that helps in analyzing the crime patterns along with their relations in a specific location. Some of the hotspots analyzed has become easier way of classifying the crime patterns.

This leads

to assist the officials to resolve them faster. This approach uses a dataset obtained from Kaggle open source based on various factors along with the time and space where it occurs over a certain period of time. We implied a classification algorithm that helps in locating the type of crime and hotspots of the criminal actions that takes place on the certain time and day. In this proposed one to impose a machine learning algorithms to find the matching criminal patterns along with the assist of its category with the given temporal and spatial data. Traditional machine learning models have proven to be effective for crime prediction. Various types of models such as decision trees, support vector machines, logistic regression, and random forests have been utilized to analyze crime data and identify patterns that can be used to predict criminal a cover larger geographic areas or accommodate increasing volumes of crime data. The safety and security of a country are a very important part of its growth and economic development, crime prediction will decrease economic loss and increase public safety in exchange.

II. RELATED WORKS

The main advance in programming improvement measure. Prior to building up the instrument it is important to decide the time factor, economy and friends' strength. When these things are fulfilled, at that point the subsequent stage is to figure out which working framework and language can be utilized for building up the device. When the developers begin assembling the apparatus the software engineers need parcel of outer help. This help can be gotten from senior developers, from book or from sites. The major part of the project development sector considers and fully survey all the required needs for developing the project. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Prior to building the framework the above thought is considered for building up the proposed framework. The significant piece of the undertaking advancement area considers and completely survey all the necessary requirements for building up the venture. For each undertaking Literature survey is the main area in programming improvement measure. Prior to building up the instruments and the related planning it is important to decide and survey the time factor, asset prerequisite, labor, economy, and friends' strength. When these things are fulfilled and completely surveyed, at that point the following stage is to decide about the product details in the separate framework, for example, what kind of working framework the venture would require and what are largely the important programming are expected to continue with the subsequent stage like building up the apparatuses, and the related activities. Here we have taken the general surveys of different creators and noted down the fundamental central issues with respect to their work.

Crime against women these days has become problem of every nation around the globe many countries are trying to curb this problem. Preventive are taken to reduce the increasing number of cases of crime against women. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analyzing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department

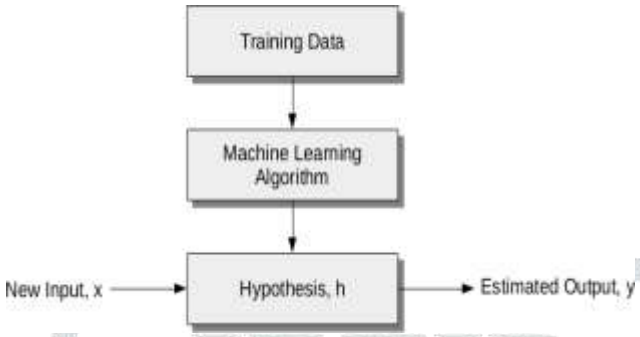

Figure 2: System architecture of crime type and occurance prediction

Crime is a foremost problem where the top priority has been concerned by individual,the community and government. This paper investigates a number of data mining algorithms and ensemble learning which are applied on crime data mining. This survey paper describes a summary of the methods and techniques which are implemented in crime data analysis and prediction. Crime forecasting is a way of trying to mining out anddecreasing the upcoming crimes by forecasting the future crime Using crime datasets requires different strategies for the varying types of data that describe illicit activity. Falade et al. (2019) provides a survey of crime prediction efforts wherein various machine learning methods have been applied to multiple types of datasets: criminal records, social media, news, and police reports.

III. PROPOSED METHOD

we will be using the technique of machine learning and data science for crime prediction of crime data set. The crime data is extracted from the official portal of police. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtain will be high. The Logistic Regression classification and various other algorithms (Decision Tree and Random Forest) will be tested for crime prediction and one with better accuracy will be used for training. Visualization of dataset will be done in terms of graphical representation of many cases for example at which time the criminal rates are high or at which month the criminal activities are high. The whole purpose of this project is to givea just idea of how machine learning can be used by the law enforcement agencies to detect, predict and solve crimes at a much faster rate and thus reduces the crime rate. This can be used in other states or countries depending upon the availability of the dataset.

Input Segmentation:

In this step we give the input for x and y values in the numerical form.

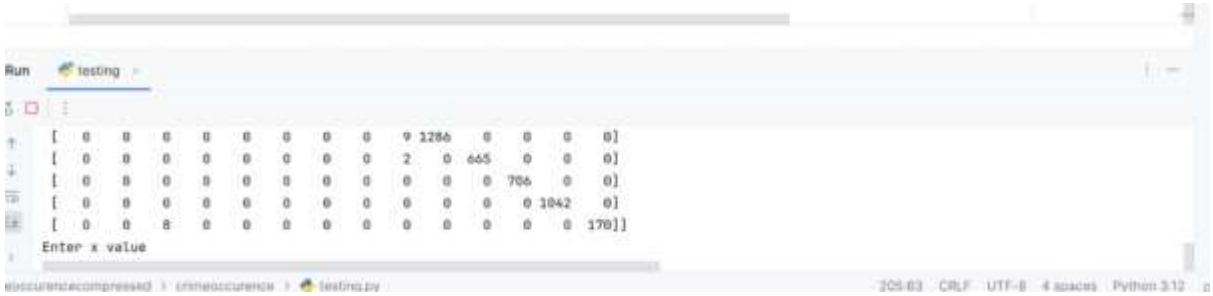

Figure 4: input for x and y value

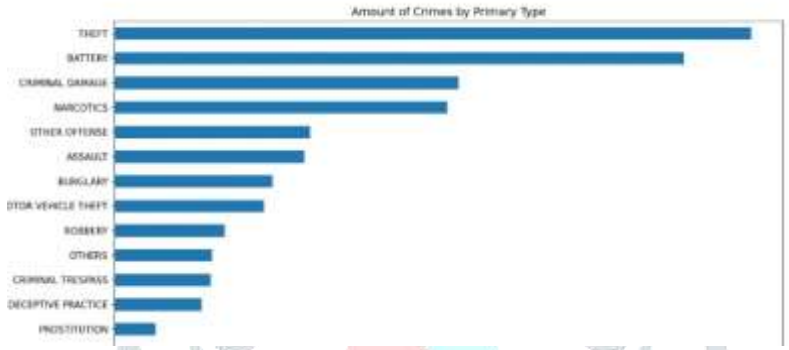

Figure 5: Graph of Primary Crime Type

E. Classification

Random Forest is a popular machine learning algorithm used for both classification and regression tasks. It's an ensemble learning method, meaning it builds multiple decision trees and merges them together to get a more accurate and stable prediction.

Here's a breakdown of how Random Forest works:

1. **Decision Trees**: Random Forest is made up of multiple decision trees. Each decision tree is trained on a random subset of the training data and selects the best feature to split the data at each node based on certain criteria (usually Gini impurity or information gain for classification, and mean squared error for regression).
2. **Random Subsets**: When building each decision tree, Random Forest randomly selects a subset of the training data (with replacement) and a subset of the features to consider for each split. This randomness helps to introduce diversity among the trees.
3. **Voting or Averaging**: Once all the trees are built, Random Forest combines their predictions through either voting (for classification) or averaging (for regression). For classification tasks, the class with the most votes from the individual trees is selected as the final prediction. For regression tasks, the average of the individual tree predictions is taken.

4. **Bootstrapping**: Random Forest often uses a technique called bootstrapping. This involves randomly sampling the training data with replacement to create multiple bootstrap samples. Each decision tree is then trained on one of these bootstrap samples.

5. **Bagging**: Random Forest is a specific type of ensemble method known as bagging (bootstrap aggregating). Bagging helps to reduce overfitting by training each tree on a slightly different subset of the data.

6. **Tuning Parameters**: Random Forest has several hyperparameters that can be tuned to optimize its performance, such as the number of trees in the forest, the maximum depth of each tree, and the number of features to consider for each split.

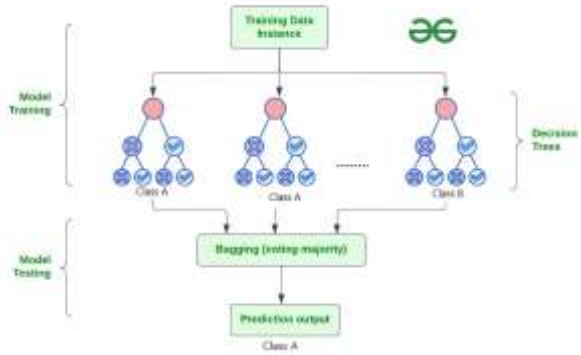

Figure 6 : General architecture of Rndom forest algorithm

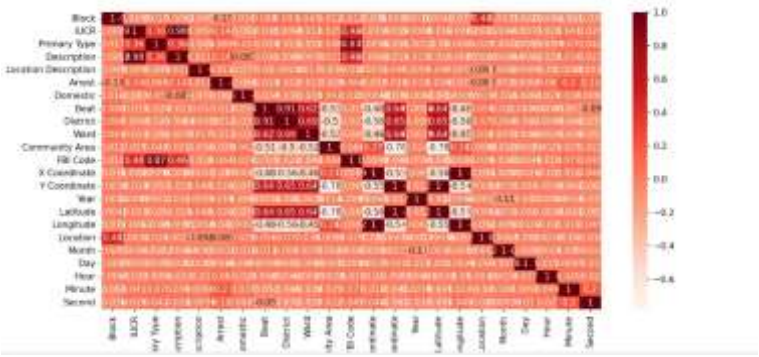

Figure 7: Confusion Matrix

IV. RESULTS AND DISCUSSIONS

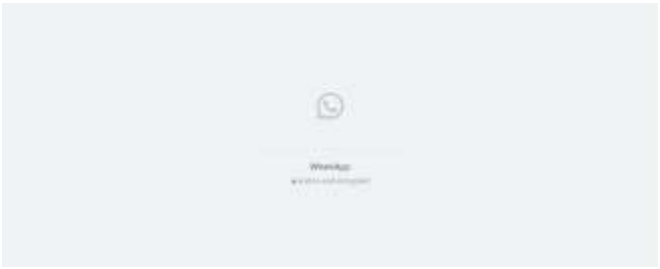

Figure 8 : Normal Images Results

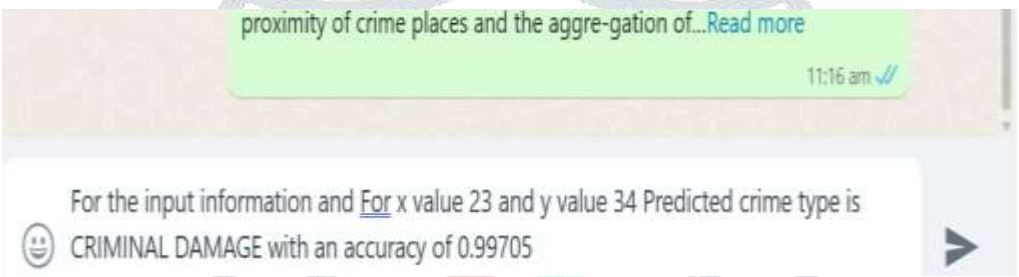

Figure 9: Online message sent

V. CONCLUSION AND FUTURE SCOPE

In this paper focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps the crime analysis to analysis these crime networks by means of various interactive visualization. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

REFERENCES

[1]. “Crime Type and Occurrence Prediction Using Machine Learning Algorithm”, Journal of Engineering Science(JES), 2022.

[2]. “Crime Type and Occurrence Prediction Using Machine Learning Algorithm”, Juni Khyat ((UGC Care Group I Listed Journal), 2022.

[3]. “Crime Type and Occurrence Prediction Using Machine Learning Algorithm”, IJPREMS, 2023.

[5]. “An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach”, IEEE Access, 2021.

[6]. “Crime Prediction Using Machine Learning: A Comparative Analysis”, Journal of Computer Science, 2023.

[7]. “A Survey of Crime Prediction Using Machine Learning and Recurrent Neural Networks”, JETIR, 2019.

[8]. “Crime Prediction Using Machine Learning Algorithms”, IJARSCT, 2022

[9]. “Crime Prediction and Analysis Using Machine Learning Algorithms”, IRJET, 2018.

[10]. “Crime Analysis Using Machine Learning YMER, 2023.