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

Predicting Crime in a particular region using Decision Trees

Kosireddi Sai Divya[1], Pandranki Shanmukha Rao[2], Chintagunti Dilip[3], Nali Ravi Teja[4], M. Chandra Sekhar[5], G. Rajendra Kumar[6]

Department of Information Technology, Vignan's Institute of Information Technology (VIIT)

Abstract:

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. In general, while we are going out to any place it was quite normal to check for weather and traffic through network. Similarly, we can also check for the crime rate before going. Using the concept of Machine Learning, we can extract previously unknown, useful information from a structured data i.e., crime data. Predicting a crime in a particular region can be done based on the data that was collected. Using this concept, we can predict the high probability of crime rate using Decision Trees. It also provides visualization techniques to understand easily. During prediction we will undergo various techniques like Data collection, Data cleaning, Data handling, Predictive modeling, Model selection, Prediction, Visualization using Random Forest classification.

With the results, we can provide identification of crime by predicting the crime rate with an improved accuracy than with other methods which will directly help police to practice and for better strategies.

Introduction:

Crimes now-a-days are increasing day by day and with different level of intensity and versatility. The result is great loss to society in terms of monitory loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. To overcome this problem the computing era can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. The crime rate prediction strategies can be applied on historical data available in the police records by examining the data at various angles like reason of crime, frequency of similar kind of crimes at specific location with other parameters to prepare the machine learning model for crime prediction. It is the major challenge to understand the versatile data available with us then model it to predict the crime incidence with acceptable accuracy and further to reduce the crime rate.

The main scope of the project is to predict the crime rate in a particular region based on the historic data and visualize it graphically using statistical tools so that, it is easy to look and understand the data to support public safety, financial success and better outcomes. Public safety and protection related to crime, and a better understanding of crime is beneficial in multiple ways: it can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime, and more concerted efforts by citizens and authorities to create healthy neighbourhood environments. With the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of

research. This project presents the visualization techniques and classification algorithms that can be used for predicting the crimes and helps the law agencies. In future, there is a plan for applying other classification algorithms on the crime data and improving the accuracy in prediction.

In recent time, many researchers have conducted various experiments to predict crimes using various machine learning methods and inputs. For crime prediction, KNN (K- Nearest Neighbour), Decision trees and some other algorithms are used. Data collected from various websites and newsletter were used for prediction and classification of crime using Naïve Bayes, Decision trees, Support Vector Machine (SVM) and Artificial neural networks (ANN) but there does not exist any particular method that can solve different crime datasets problems.

To predict the crime of a particular region some of the standards should be followed. In this model the standards we followed are mentioned as below:

- Collecting the dataset from Kaggle and exploring the dataset as per requirement. In this dataset 1994 areas are present. Size of the dataset is 1994 rows and 128 columns. The dataset features are: the occurrence month, the occurrence day of the week, the occurrence time and the crime location.
- Importing the required modules into the jupyter notebook and implementing data cleaning, analysing correlation (relationship between variables using heat map) on the data set are done.
- After that, we have used machine learning algorithms like regression and decision tree classifier to train the dataset and build machine learning model to get good accuracy.
- The accuracy rate is 85% with Decision tree classifier and 0.65 R2 score with Linear Regression. To increase the previous accuracy rate, Decision tree classifier algorithm is optimized.

After following these standards, the accuracy increased to 90%.

Literature Survey:

[1] prediction is done using ID3(Iterative Dichotomiser 3) algorithm with the help of WEKA mining software as tool at an accuracy of 72.7%[2] prediction and classification of crime is done based on the only feature that is location that is Indore City and it is done by using KNN, decision trees and random forests but it doesn't achieve that much of accuracy.[3] The author predicts the rate based on the features like sex, age and relationship and the algorithms used are KNN and Artificial Neural Networks(ANN) and got an accuracy around 85%.[4]prediction is done using the techniques like KNN, Decision Tree an Extra Tree Classifier with an accuracy of 88%.[5]For prediction the author used Deep learning techniques along with the some Machine learning techniques like XGBoost and KNN whereas deep learning techniques like LSTM(for analysis),RME and MAE are used for visualization. The project is main purpose is to predict and forecast the crime in Chicago and Los Angeles.[6] Based on the past crime data, the prediction is done using Linear regression, Naïve Bayes and KNN around with an accuracy of 73.6%,69.5% and 76.9% respectively.

Existing System:

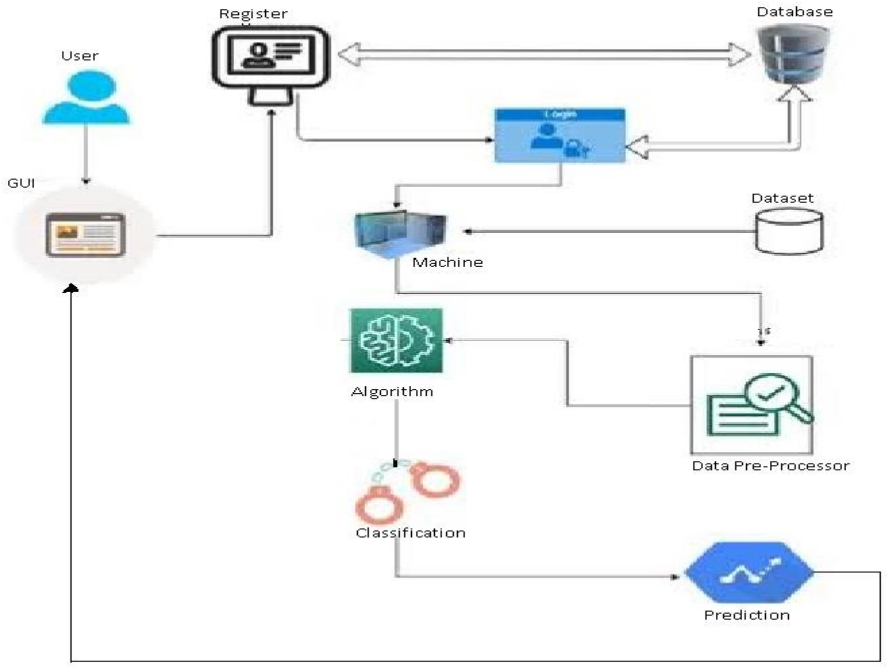
In recent times, many researches have conducted various experiments to predict crimes using various machine learning methods and inputs. For crime prediction, KNN (K- Nearest Neighbour), Decision trees, Neural Networks and some other algorithms are used. Data collected is from various websites and newsletter were used for prediction and classification of crime using Naïve Bayes, Decision trees, Support Vector Machine (SVM) and Artificial Neural Networks (ANN) but there does not exist any particular method that can solve different crime datasets problems and predict crime accurately.

Proposed System:

In the proposed system, the Machine Learning Model is built in such a way that it can predict the crime based on the data collected from various features like the type of the crime. The model

classifies the data based on the location given as input using decision trees and random forests and then it performs prediction using prediction functions. This system provides the accurate result than the existing systems.

Architecture:



The above figure depicts the architecture of the proposed system. It can be accessed by interacting the user through web application. Through that the user can enter the location of the place that he wants to know the crime rate. After the location is entered the respective area's data is automatically inserted to the machine/model from the dataset and the machine will undergo all the steps of the algorithm, thus undergo for classification. At last, with the help of some predict functions the machine will predict the output and displays the result to the user's application.

Methodology:

The methodology includes Machine Learning approach.

Data Collection:

Data is collected from digitalized information of crime reports which is of 1994 areas and the size of the dataset is 1994 rows and 128 columns.

Data Pre-processing:

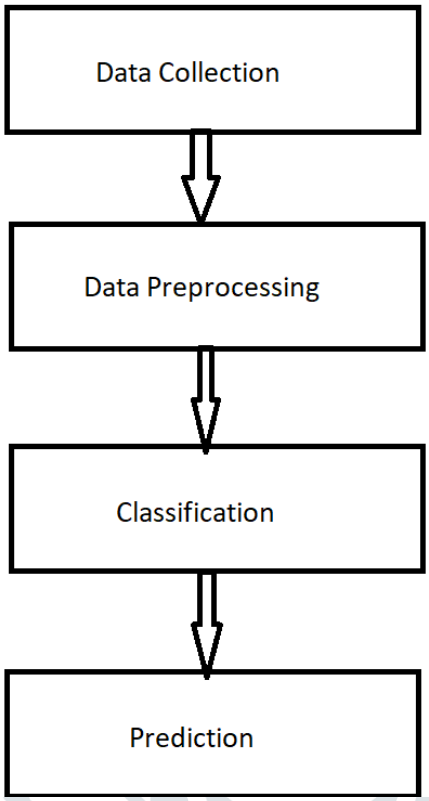
Data pre-processing includes data cleaning which removes missing values and noisy data and resolve inconsistencies. Further data integration and reduction also takes place.

Analysis:

The analysis includes graphical representation of data which provides the keen view towards the prediction.

Training and Testing:

The splitting of dataset for training and testing. For training, 80% of the dataset is used and for testing, 20% of the data is used.



Classification:

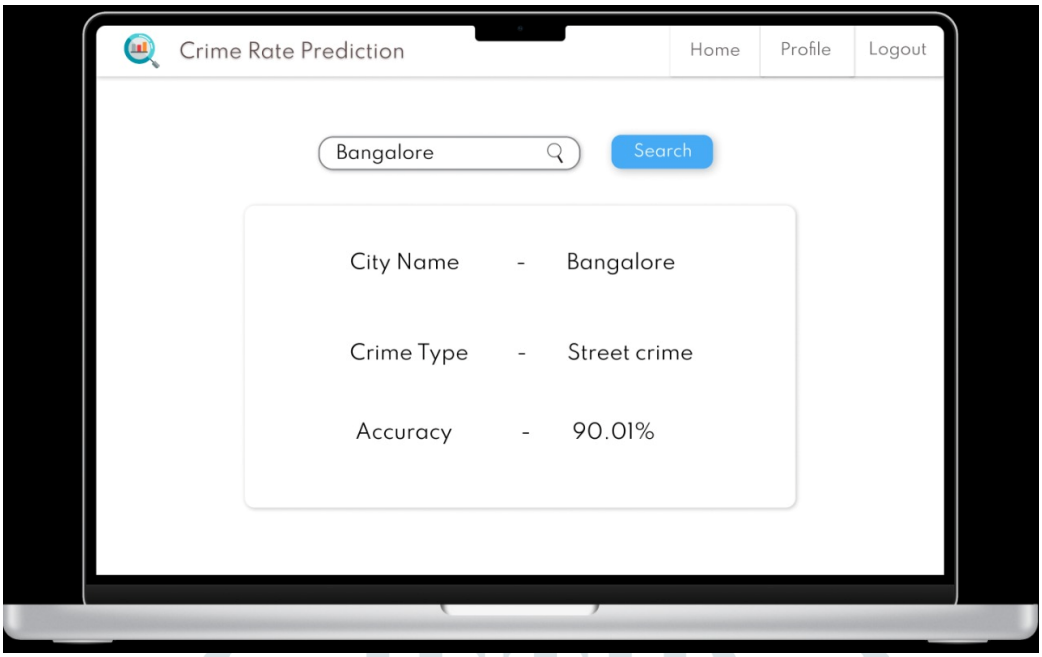
Classification is the process of separating or analysing the data based on the features. Decision Tree classifier is used for classification along with that Random tree classifier is used for bagging concept.

Prediction:

Prediction is performed based on the trained data. Whenever the data is inserted by the user it will undergo prediction functions and predicts the crime.

Results:

Our model gives an accuracy of 90% which was greater than the previous models to predict the crime in a particular area.



Conclusion:

With the help of machine learning technology, it has become easy to find out relation and patterns among various data and provides awareness of the crime. The work in this project mainly revolves around predicting the crime of region where it has occurred. Using the concept of machine learning we have built a model using data sets that have undergone data cleaning. The model is classified using decision trees and prediction is done using random forest classifier.

Future work:

In future this model can be build for predicting crime in any area by inserting the data to the model based on the track of user's location and also, they can find the crime rate for other places rather than the areas in their location.

References:

[1] Adewale Opeoluwa Ogunde, Gabriel Opeyemi Ogunleye, Oluwaleke Oreoluwa. A Decision Tree Algorithm Based System for Predicting Crime in the University. Machine Learning Research. Vol. 2, No. 1, 2017, pp

[2] Kunal Diwan, Sahitya Nigan, Sourabh Tiwari, Vikram Aditya Singh Bhatti "PAASBAAN"-Crime Prediction and classification in Indore city

[3] Prof. ShivPrasadMore, Sakshi Mench, Saloni Kuge, Hafsa Bagwan "Crime Prediction using ML approach" by International Journal of Advanced Research in Computer and Communication Engineering Vol. 10, Issue 5, May 2021.

[4] Pratibha Kumari, Lokesh Chouhan, Akanksha Gotalot "Crime prediction and analysis" Conference Paper by Research Gate in February 2020

[5] Wajiha Safat, Sohail Asghar, Saira Andleeb GillAni "Empirical Analysis for crime prediction and forecasting using Machine Learning and Deep Learning techniques" by IEEE date of publication May 6, 2021

[6] Sakib Mahmud, Musfika Nuha, Abdur Sattar "Crime rate prediction using Machine Learning and Data Mining" by Research Gate in January 2021

[7] Prajakta Yerpude and Vaishnavi Gudur "Predictive Modelling of Crime Dataset using Data Mining" by International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.7, No.4, July 2017

[8] Sirivanth Paladugu, Tarun Sai Yakkala, Neeraj Boggarapu, Sri Krishna Kumar Modekurty "Crime rate Prediction using Machine Learning" by International Journal of Research in Engineering, Science and Management Volume 4, Issue 9, September 2021

[9] Abhishek Yadav, Rakshit adgaonkar, Rohit Ingale, Omkar Pathak "Crime Prediction Using Machine Learning Algorithms" by International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 4, Issue 2, April 2021

[10] Anish Krishnan, Aditya Sarguru, A.C. Shantha Sheela "Predictive Analysis of Crime Data using Deep Learning" by International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 4023-4031