International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.049 ISSN: 2581-4621

Crime Rate Prediction Using Machine Learning

P.Poornima[1],L.Tharun kumar[2], S.Sumera[3], K.Deepika[4],K.Yoshitha[5]
Assistant professor, Student,Department of CSE, AITS, Tirupati

Abstract

The accurate prediction of crime is crucial for formulating effective policing strategies and implementing crime prevention and control measures. Machine learning is the prevailing method for crime prediction, but there has been limited systematic comparison of different machine learning techniques. This study utilizes historical data on public property crime between 2015 and 2018 from a section of a large coastal city in southeast China to evaluate the predictive power of several machine learning algorithms. The results demonstrate that the LSTM model surpasses KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks when using only historical crime data for prediction. Moreover, the inclusion of built environment data such as points of interest (POIs) and urban road network density as covariates in the LSTM model enhances its predictive ability compared to the original model that solely relies on historical crime data.

Keywords: Random Forest, Prediction, Social network, naïve bayes

Introduction

The incidence of crime is escalating rapidly, with each passing day. It is a major concern that is becoming increasingly intense and intricate. Crime is a dynamic phenomenon, with constantly evolving patterns that defy straightforward explanations of criminal behavior. The various categories of crime include kidnapping, theft, murder, rape, and more. Law enforcement agencies rely on information technologies (IT) to gather data on crime.

Crime is increasing considerably day by day. Crime is among the main issues which is growing continuously in intensity and complexity. Crime patterns are changing constantly because of which it is difficult to explain behavior in crime patterns. Crime is classified into various types like kidnapping, theft murder, rape etc. The law enforcement agencies collects the crime data information with the help of information technologies(IT).

1.1Related Work

Many researches have been done which address this problem of reducing crime and many crime-predictions algorithms has been proposed. The prediction accuracy depends upon on type of data used, type of attributes selected for prediction. In mobile network activity was used to obtain human behavioral data which was used to predict the crime hotspot in London with an accuracy of about 70% when predicting that whether a specific area in London city will be a hotspot for crime or not. Data collected from various websites, newsletter was used for prediction and classification of crime using Naive Bayes algorithm and decision trees and found that former performed better.

1.2 MACHINE LEARNING

Machine learning is the study of computer algorithms that may improve themselves automatically with the use of data and practice (ML). It is a part of artificial intelligence, theoretically. Technically, artificial intelligence is where it belongs. To create a model that can generate judgments and predictions without ever being explicitly programmed, machine learning algorithms utilize sample data, or "training data." ML algorithms are utilized in a variety of applications, such as computer vision, speech recognition, email filtering, and medicine, where it is difficult or impractical to design traditional algorithms that can complete the necessary tasks.

The theory, techniques, and application domains of machine learning are provided through learning about mathematical optimization. A similar area of research focuses on exploratory data mining for unsupervised data analysis. Several machine learning applications employ data and neural networks in a manner akin to that of the human brain. Machine learning, applied to solve issues in enterprises, is sometimes referred to as predictive analytics.

2. PROPOSED METHODOLOGY

2. 1 Overview

Predictive modeling was used for making predictions since it has the method which is able to build a model and has the capability to make predictions. This method consists of different algorithms of Machine Learning that can study properties from the data used for training which is used for producing predictions. It is split in two major classes one is Regression and other is classification of patterns. Regression models are based upon analysis of the relationship that are present between trends and variable in order to make predictions about the continuous variables. Whereas, the job of classification is to assign a particular class labels to a data value as output of the prediction. Division of pattern classification is in two ways i.e., Supervised and Unsupervised learning. It is already known in supervised learning that which class labels are to be used for building classification models. In unsupervised learning, these class labels are not known.

Data collection is a process in which information is gathered from many sources which is later used to develop the machine learning models. The data should be stored in a way that makes sense for problem. Data pre-processing basically involves methods to remove the infinite or null values from data which might affect the performance of the model. In this step the data set is converted into the understandable format which can be fed into machine learning models.

Support Vector Machine per-forms well for regression, time prediction series and classification problems. Support vector machine performance can be measured against Recurrent Neural Network. Thus, SVM had been applied in predicting hotspots of crime and predicting diseases like diabetic and pre-diabetic. Since it can make prototype of nonlinear relations in a coherent way. It performs well for anticipation of time series. For a predetermined degree of crime and data set it has to select a subset using Kclustering algorithm of crime data set and will determine a label for each data point in the set that is selected. Point where the crime rate is below given rate are called hotspots and where it is above given rate are called cold spots.

Decision trees are one of the most popular and powerful tool for classification and prediction. It has a structure like a tree, where all of the intermediate node represents a test on a peculiarity and the end product of test is denoted by every branch, and label of class are held by every leaf node. The target variable is generally categorical. Decision trees are used either for calculating the probability that a given record belongs to each category or to classify records (which is done by assigning records to the most similar class).

It comprises of huge number of constituents which work cooperatively to process and resolve problems. It is based on prediction by analyzing trends in an already existing large amount of historical data. It has more general and flexible functions forms and can effectively deal with than traditional statistical methods. These were used to estimate the relation between inputs and outputs by adjusting the weights in every iteration. ANN can realize and study patterns for obtaining knowledge. It displays a link between an input neuron and an output neuron. Neurons have some specified weights. Output is calculated by multiplying the input with the specific neuron weight and then comparing it with the threshold value. If its above given threshold then it is contemplated as the output.

Training and Testing:

In this step, after validating the assumptions of the algorithm that we have chosen. Model is trained on the basis of given training Sample. After training, the performance of the model is checked on the basis of error and accuracy. At last, the trained model is tested with some unseen data and the model performance is checked on the basis of various performance parameters depending on the problem.

2.2 Proposed framework

Random Forest

Random Forest technique is an ensemble learning method for classification, regression, and other tasks, operated by constructing a multitude of Decision Trees at training time and outputting the class, that is, the mode of the classes (classification) or means prediction (regression) of the individual trees. Random Decision Forests correct for Decision Trees' habit of overfitting to their training set. In this experiment, Random Forest was selected as a technique to estimate the predictors. Random forests are frequently used as black box models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

In the proposed system, we are introducing an application that will analyze the crime that criminals did in the past in a particular region or area. This prediction is based on attribute like a criminal record, time, place, which could represent the whole crime in form of bar graphs. We could further take this idea ahead and even make predictions of the crime in particular places or regions.

3. IMPLEMENTATION

Dataset

Data mining has been frequently used in crime prediction models for the last couple of years, considering different features. Used variables such as longitude (X), latitude (Y), address, day of week, date (YYYY-mm-dd--hh : MM : ss), district, resolution, and category to analyze and predict San Francisco crime data. The study used different techniques and principal component analysis to classify the accuracy and avoid overfitting. He also used four different classifiers: K-NN, Decision Tree, Bayesian, and Random Forest, applied them to the task, and obtained the log-loss of 2.39031 by the Random Forest classifier.

Crimes are being analyzed based on some of the parameters such as: 4.2 Project Objectives

The main objective of our project is to classify crime in a particular place. To analyze and interpret crimes at all places uniquely. To provide clear information about its environment to intelligent agents. Predictive modeling is a way of building a model that can make predictions. This includes the process in which a machine learning algorithm learns certain properties from various training Data Sets to make accurate predictions.

4.3 Project Outcomes Benefits:

The police can use the system in two ways: The system will be alert that a criminal offense is imminent based on any new weather event. The police can run the system once every day and support the predictions, deciding how to deploy resources(policemen) in each community/district. Longitude–X coordinates on the map where the crime has occurred.

- Latitude–Y coordinates on the map where the crime occurred.
- Address: – The place where the crime incident has taken place.
- Day of Week: – The day of the week (i.e. Monday)
- Date: – on which date the crime has taken place.

Crimes are being analyzed based on some of the parameters such as

- District: – Police district to which the crime is assigned.
- Resolution: – The resolution which needs to be taken to address the crime in a given area.
- Category: – The type of the crime. This is the label that we need to predict. We have many important files used in this project: MATPLOTLIB: It is a multiplatform visualization library in Python for 2D plots of NumPy arrays and is designed to work with the broader SciPy stack.

Sklearn: it is used in statistical modeling including classification, regression, and clustering and dimensionality reduction.

SEABORN: it is data visualization library based on matplotlib which provides the best interface for drawing informative and attractive statistical graphical designs.

TensorFlow: TensorFlow is of a Python library mainly used for fast numerical Computations for better accuracy.

FOLIUM: it is used for visualizing geospatial data of a given area.

SQUARIFY: Squarify is a Processing library that implements the squarify Tree map layout algorithm.

The crime data set of San Francisco city is taken fromkaggle.com in csv format which was derived from incidents derived from SFPD Crime Incident Reporting system. the at-tributes of the data set are Date, category of crime, description, Dayo Week, Police Department District, Address, Latitude and longitude and contained 884k data points.

4. Results

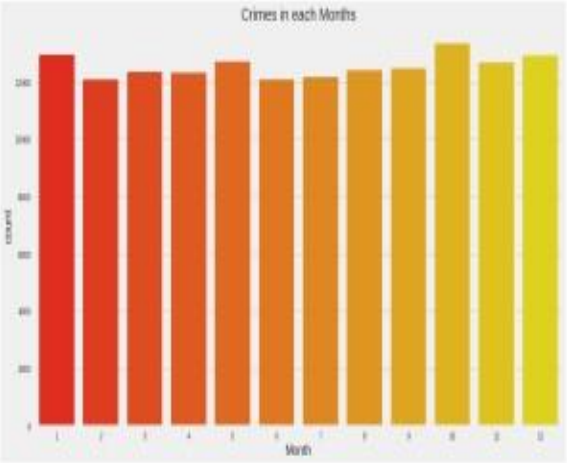

Figure: Crimes in each month

➢ Most occurring crimes in the city
➢Percentage of different crimes happening in the city per day

Figure: Representation of pie-chart

Crime solving is very difficult work which requires experience and intelligence of human along with Artificial intelligence approaches which assist them in problems of crime detection. The efficiency of neural networks in classifying data is 95%. We have taken 80% of the data for training neural networks and 20% for classification. The accuracy for predicting crime is basically depends upon on the crime data set used. If used training data set is very large, then model will be trained with very good accuracy while if the dataset used for training purpose is having less size, then small degree of training is attained**.**

5.Conclusion

Crime prediction is one the current trends in the society. Crime prediction intends to reduce crime occurrences. It does this by predicting which type of crime may occur in future. Here, analysis of crime and prediction are performed with the help of various approaches some of which are random forest, Artificial Neural network, Decision trees, Extra trees and Support vector machine. From the results obtained we saw that the training time of SVM is very high thus it should be avoided for this dataset. For MLP we saw that its accuracy is very low hence MLP is not working good for this dataset. Here we can see that for this data set Decision tree, random forest and Extra tree classifier are working best with optimal training and good accuracy.

6.References

[1]A. Bogomolov, B. Lepri, J. Staiano, N.
Oliver, F. Pianesi and A. Pent-land, "Once upon a crime: towards crime prediction from demographics and mobile data", IEEE, Proceedings of the 16th international conference on multimodal interaction, 2014, pp. 427-434.

[2]Ubon Thansatapornwatana,"A Survey of Data Mining Techniques for-Analyzing Crime Patterns", Second Asian Conference on Defense Technology ACDT, IEEE, Jan 2016, pp. 123–128.

[3]H. Adel, M. Salheen, and R. Mahmoud, "Crime in relation to urban design. Case study: the greater Cairo region," Ain Shams Eng. J., vol.7, no. 3, pp. 925-938, 2016.

[4]J. L. LeBeau, "The Methods and Measure of Centrography and the spatial Dynamics of Rape" Journal of Quantitative Criminology, Vol.3,No.2, pp.125-141, 1987.

[5]Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, Alex Pentland." Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data", in ACM International Conference on Multimodal Interaction (ICMI 2014).

[6]Shiju Sathyadevan, Devan M. S.,Surya S Gangadharan, First," Crime Analysis and Prediction Using Data Mining" International Conference on Networks Soft Computing (ICNSC), 2014.

[7]Sunil Yadav, Meet Timbalier, Ajith Yadav, Rohit Vishwakarma and Nikhilesh Yadav," Crime pattern detection, analysis and prediction, International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2017.

[8]Amanpreet Singh,Narina Thakur, Aakanksha Sharma," A review of supervised machine learning algorithms",3rd International Conference on Computing for Sustainable Global Development,2016.