EMPIRICAL ANALYSIS FOR CRIME PREDICTION AND FORCASTING USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Project Reference No.: *45S_BE_0778*

College ***:*** *Sri Venkateshwara College of Engineering, Bengaluru*
Branch ***:*** *Department of Computer Science Engineering*
Guide(s) ***:*** *Dr. Suma T*
Student(S) ***:*** *Ms. Megha C*
Mr. Mahesh
Mr. Mittal. Savan Kumar

Keywords:
Crime Prediction, Logistic Regression, XG Boost, Decision Tree and Multilayer Perceptron (MLP).

Introduction:

Criminality is a negative phenomenon, which occurs worldwide in both developed and underdeveloped countries. Crimes take place due to various circumstances including specific motives, human nature and behavior, critical situations and poverty. Furthermore, multiple factors such as unemployment, gender inequality, high population density, child labour, and illiteracy, can cause an increase in violent crimes.

This study aims to analyze crime prediction in the Chicago and Los Angeles datasets. The algorithms used in this project are Logistic Regression, XG Boost, Decision Tree and Multilayer Perceptron (MLP). Crime and violation are the threat to justice and meant to be controlled. Crime is an offence and such acts are punishable by Law.

The Chicago and Los Angeles datasets have been collected throughout the years; it is no surprise that machine learning and deep learning methods may be useful in the prediction of crime types and forecasting future benefit. The overall crime rate forecasting results would benefit the police by using identified alleged crime areas to allocate additional resources and protective measures against criminals.

Objectives:

The main objective of this project would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy.
Visualization of dataset is done to analyze the crimes which may have occurred in the country.

The aim of our project is to make crime prediction using the features present in the dataset. With the help of machine learning algorithm, using python as core we can predict the type of crime which will occur in the particular area. This work helps the law enforcement

agencies to predict and detect crimes in Chicago and Los Angeles with improved accuracy and thus reduces the crime rate.

Methodology:

A. PREDICTIVE ACCURACY

This study used different parameters to assess the performance of multiple algorithms, which better react the real dataset application. Four different algorithms were applied to the Chicago and Los Angeles datasets to investigate the detailed predictive accuracy of the trained models.

B. TIME SERIES ANALYSIS THROUGH LSTM

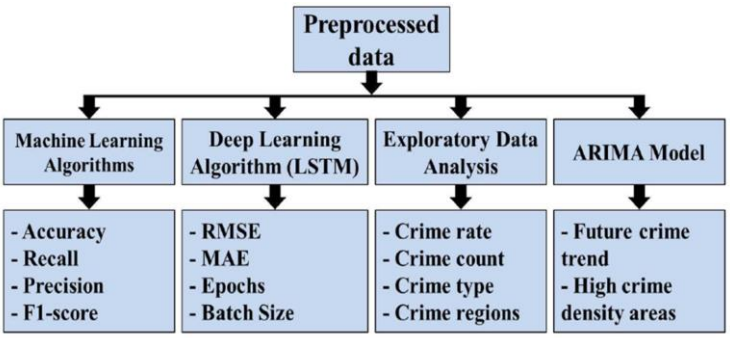
LSTM is an elegant variation in the RNN architecture, which is an approach that can be applied to model sequential data. The performance of LSTM seems to be adequate for time series analysis, especially for RMSE and MAE, where it can classify the data focusing on their variations.

C. EXPLORATORY DATA ANALYSIS

This section discusses the detailed periodic insights of the Chicago and Los Angeles statistics. The term crime count refers to the number of crime incidents, while high-intensity crime areas are the hot spot crime regions referring to the district level location.

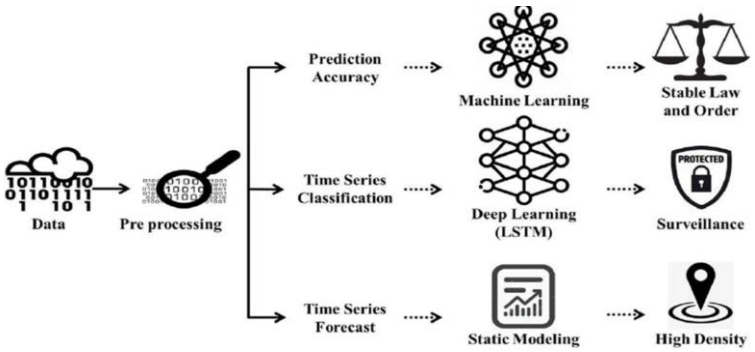
D. FORECASTING WITH AN ARIMA MODEL

An ARIMA model is a composite model for time series data combining a traditional autoregressive moving average (ARMA) model and autoregressive (AR) moving average (MA) processes.



Supervised Learning

Supervised machine learning builds a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.

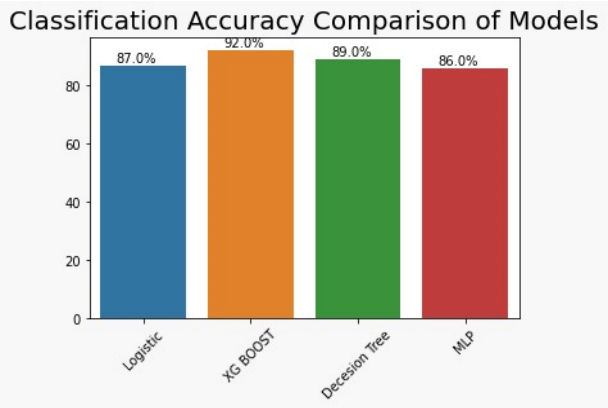


The performance of LSTM for time series analysis was reasonably adequate in order of magnitude of root mean square error (RMSE) and mean absolute error (MAE), on both data sets.

Our Current work applied different machine learning algorithms, namely, the logistic regression, decision tree, multilayer perceptron (MLP), and XG Boost, and time series analysis by long-short term memory (LSTM) and autoregressive integrated moving average (ARIMA) model to better the crime data.

Results and Conclusions:

We get the better accuracy in XG Boost as compared to the Logistic Regression, Decision Tree and MLP.



Crimes are serious threats to human society, safety, and sustainable development and are thus meant to be controlled. Investigation authorities often demand computational predictions and predictive systems that improve crime analytics to further enhance the safety and security of cities and help to prevent crimes. We achieved an improved predictive accuracy for crimes by implementing different machine learning algorithms on Chicago and Los Angeles crime datasets.

Using the concept of machine learning we have built a model using training data set that has undergone data cleaning. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding Chicago crimes datasets and Los Angeles crimes datasets that can help in capturing the factors that might help in keeping society safe.

Scope for future work:

The scope of using different methods for crime prediction and prevention can change the scenario of law enforcement agencies. Although the current systems have a large impact on crime prevention, this could be the next big approach and bring about a revolutionary change in the crime rate, prediction, detection, and prevention. Investigation authorities often demand computational predictions and predictive systems that improve crime analytics to further enhance the safety and security of cities and help to prevent crimes. The sparsity of crime in many areas complicates the application of the prediction rate area-specific modeling.

The idea behind this project is that crimes are relatively predictable; it just requires being able to sort through a massive volume of data to find patterns that are useful to law enforcement. This kind of data analysis was technologically impossible a few decades ago but the hope is that recent developments in machine learning are up to the task. Crime has been growing all around the world over the past few decades. As technology has advanced and more data has become available, data-driven approaches to predict crime can really help in preventing and deploying the police force more effectively. Adaptation of data-driven approaches would help the police force and make cities safer and more habitable.