CRIME ANALYSIS , PREDICTION AND EMBEDDING USING MACHINE LEARNING TECNIQUES

LSTM and ARIMA based crime prediction, analysis and forecast

Crime incidents embedding using restricted Boltzmann machines

Abstract:

Crime is a socio-economical problem which affects both the quality of life and economic growth . The specifics of how crime is conducted varies depending on the type of society and community. Earlier researches in crime prediction have found that numerous factors like education, poverty, employment, and climate affect the crime rate.

Machine-learning-based crime analysis predominantly involves data collection, classification, pattern identification, prediction, and visualization. The main objective of this work is to create a prediction model that can accurately predict crime. In our project, we have planned to implement classification algorithms, K-Nearest Neighbour (KNN), decision tree , Naïve Bayes to analyze the crime dataset compiled between a particular time period for Banglore and Chennai.

This study also applies different machine learning algorithms, namely, the logistic regression, support vector machine (SVM),multilayer perceptron (MLP), random forest, and extreme Gradient Boosting (XGBoost), and time series analysis by long-short term memory (LSTM) and autoregressive integrated moving average (ARIMA) model to better fit the crime data.

Exploratory data analysis predicts more than 35 types of crimes and suggests how the crime rates increase or decline on a particular year. It addition it also shows that how crime rates vary from month to month within a partuclar year. Now once analysis of the data is done it also predicts the rate of crime in upcoming years by using ARIMA model. Moreover, crime forecasting results were further identified in the main regions for both cities. Overall, these results provide early identification of crime, hot spots with higher crime rate, and future trends with improved predictive accuracy than with other methods and are useful for directing police practice and strategies.

Time series analysis is needed to generate visual patterns along with a deep learning algorithm specifically LSTM, which provides the better classification of crimes. Additionally, forecasting the crime trends through ARIMA model is a highly recommended and promising in forecasting future crimes.

Another module planned for this project is detecting related crime series, by unsupervised learning of the latent feature embeddings from narratives of crime record via the Gaussian-Bernoulli Restricted Boltzmann Machine (GBRBM). This is a novel approach from earlier works on crime analysis, which predominantly considers only time and location and at most category information. Once the embedding is done, crime cases which are related,are closer to each other in the Euclidean feature space, and the unrelated cases are far apart.can enable subsequent analysis such as detection and clustering of related cases.

The description of the crime in the police report would be processed directly, and then mapped into a feature vector space that automatically captures the similarity of incidents.

WHAT’S THE EXPECTED OUTCOME?

The objective is to analyze crime prediction using the data.

Theo overall planned outcome is as follows:

1. improving the predictive accuracy by implementing the Logistic Regression, SVM, Naïve Bayes, KNN, Decision Tree, MLP, Random Forest, XGBoost algorithms
2. time-series analysis by using LSTM
3. creating a visual summary by using exploratory data analysis
4. crime forecasting for the crime rate and high intensity crime areas for subsequent years by using an ARIMA model.
5. detection and clustering of related cases using the Gaussian-Bernoulli Restricted Boltzmann Machine (GBRBM)

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.