Time Series Analysis and Trend Forecasting of Crime Data

ITCS5156-Applied Machine Learning Project Proposal

Project proposal by ***Ramasri Saladi***

(Niner Id: 801254656)

Journal Details:

- ***Title:*** *Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data.*
- ***Author:*** *Mingchen Feng, Jinchang Ren*
- ***Year:*** *July 2019*
- ***Conference/Journal:*** *IEEE*
- *M. Feng et al., Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data, in IEEE Access, vol. 7, pp. 106111-106123, 2019, doi: 10.1109/ACCESS.2019.2930410.*

Problem Statement:

Crime prevention has always been a top priority for governments to ensure a safe living environment for their citizens. Accurately forecasting crimes will help in reducing the crime rate. As this is an active research area, many researchers applied several machine learning, deep learning, and time series forecasting algorithms on real-world crime data sets from major cities like Chicago, Los Angeles, etc. To date, Prophet, LSTM, and Ensemble of classifiers are some of the most successful models. Prediction results from these studies illustrate varying degrees of accuracy. My study focuses on leveraging the effective strategies of big data analytics in present crime prevention research as well as applying improvements to the current methods.

DataSet:

- *https://www.kaggle.com/currie32/crimes-in-chicago*
- *The dataset consists of data from the year 2001 to 2021.*
- *In Chicago dataset there are 21 columns. The columns are ID, Case Number, Date, Block, IUCR, Primary Type, Description, Location Description, Arrest, Domestic, Beat, District, Ward, Community Area, FBI Code, X Coordinate, Y Coordinate, Year, Updated On, Latitude, Longitude, Location, Neighborhood, Municipality and County. While most of the column names and description are obvious, some are not, including IUCR (Illinois Uniform Crime Reporting Code), Primary type (description of IUCR code), description (secondary description of IUCR code), Domestic (if the crime comes under domestic violence or not), and Beat (the smallest police geographic area).*

Motivation:

Till date all the papers were focussed on predicting crimes on a yearly, monthly, daily basis, but my idea is to predict crimes based on time of the day(i.e. Morning, Afternoon, evening, night). This prediction helps in distribution of troops based on the crime type and severity. For this purpose new columns were needed like Day of the week,Day of month, Day of year.

Related Papers:

- ***Paper 1 Title:*** *A Comparative Study on Crime in Denver City Based on Machine Learning and Data Mining.*
- ***Author:*** *Md. Aminur Rab Ratul*
- ***Year:*** *January 2020*
- ***Conference/Journal:*** *Researchgate*
- ***Why:***

  Md. Aminur Rab Ratul analyzed the Denver County crime dataset and applied various classification algorithms like Random Forest, Decision Tree, AdaBoost classifier, Extra tree classifier, K-Neighbors classifier, 4 ensemble models to classify 15 different classes of crimes and concluded that Ensemble models produce high accuracy when compared to other classification models.

Related Papers (Contd):

- ***Paper 2 Title:*** *Modeling Daily Crime Events Prediction Using Seq2Seq Architecture*
- ***Author:*** *Mingchen Feng, Jinchang Ren*
- ***Year:*** *July 2019*
- ***Conference/Journal:*** *Researchgate*
- ***Why:***

  Jawaher Alghamdi built ARIMA, RNN, Conv1D, (Seq2Seq) based LSTM models to predict crimes week ahead of occurrence. By comparing the results of these models, we concluded that Seq2Seq model is highly effective.

Summary:

- *Long Short-Term Memory (LSTM) is an artificial neural network architecture used in the field of deep learning. Developed to deal with the vanishing gradient problem. Information will be passed through series of layers i.e., forget gate layer, input gate layer, tanh function layer, sigmoid function layer.*
- *For performance evaluation, the Root Mean Square Error (RMSE) and spear-man correlation are used in terms of different parameters and different sizes of training samples.*
- *The optimal time period for crime trends forecasting is 3 years where the RMSE is the minimum and the spearman correlation is the highest.*

Timelines:

Task	***Start Date***	***End Date***
Data Selection	*01/15/2022*	*01/22/2022*
Project Proposal	*01/22/2022*	*01/29/2022*
Initial Analysis	*01/29/2022*	*02/05/2022*
Data Pre-Processing	*02/05/2022*	*02/12/2022*
Predictive Modeling	*02/12/2022*	*02/19/2022*
Validation and evaluation	*02/19/2022*	*02/26/2022*
Final Report	*02/26/2022*	*03/01/2022*

Thank you