Prediction of Crime Rate in Chicago

Project Description:

Crime in Chicago is a very interesting topic for exploration for all kinds of reasons. Personally, I have been living in Chicago for more than 2 years now and crime here is always a topic of conversation with friends and family. Another reason is the availability of huge amounts of high-quality crime datasets open for data scientists to investigate.

In this notebook, I will explore crimes in Chicago from a perspective of a Chicago resident who wants to know more about the subject so he can better navigate his way through the city. Objective is to build an intelligent crime prediction model, where the model predicts different types: crime type, time of the crime, the place of crime.

Problem Statement:

- How has crime in Chicago changed across years? Was 2016 really the bloodiest year in two decades?
- Are some types of crimes more likely to happen in specific locations or specific time of the day or specific day of the week than other types of crimes?

Dataset Details:

- The Chicago Crime dataset contains a summary of the reported crimes occurred in the City of Chicago from 2005 onwards.
- Dataset has been obtained from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.
- Number of Attributes: 23

Attribute Information:

Here are some of the important fields used in the dataset:

- There are columns like ID and Case Number which helps us uniquely identify our crime record.
- Columns like Date and Year tells us when this crime happened.
- X/Y Coordinate, Latitude/Longitude, Location tells where exactly did the crime happened.
- Beat, District, Ward, Community Area tells us in which area of Chicago this crime happened.

Data source: https://www.kaggle.com/currie32/crimes-in-chicago

Method

The information presented in this data set is quite comprehensive, including information about the date and time of the crime, location of the crime, type of crime, etc. We will focus on the time of the crime

and the type of crime. The type of crime is given a standardized set of codes called the Illinois Uniform Crime Reporting (IUCR) codes. Thus, each IUCR corresponds to a specific type of crime. The list of crime codes and corresponding crimes can also be found through the city of Chicago's website.

K-means clustering provides a way to group data points together in a way that minimizes differences between the data points in the same group. By applying these methods, we can take n data points and partition them into k clusters.

Ethical Problems

Predictive policing relies on a large database of previous crime data and forecasts where crime is likely to occur. Since the program relies on old data, those previous arrests need to be unbiased to generate unbiased forecasts. One way crime data can be biased is if it lacks necessary information and context. Many cities including Los Angeles, New York, and Chicago have stopped using predictive policing due to violation of the ethical framework of justice and fairness because they perpetuate systemic racism through the use of biased data.

Challenges/Issues

Crime is one of the most dangerous phenomena for any country. To reduce and nonproliferation of crime, it requires new techniques that can deal with the vast amount of data, where the data cannot be analyzed with traditional analysis techniques. Therefore, Deep Learning algorithms techniques need to be employed through using repeated neural network (RNN) and the employment of one of its types (LSTM).

References

- https://www.kaggle.com/currie32/crimes-in-chicago
- https://www.kaggle.com/femiogunbode/eda-of-crime-in-chicago-from-2012-2016/discussion
- https://www.kaggle.com/djonafegnem/chicago-crime-data-analysis
- http://pandas.pydata.org/
- http://pandas.pydata.org/pandas-docs/stable/tutorials.html