PREDICTING MANHATTAN CRIME USING NAIVE BAYES CLASSIFIER

Capstone Project

INTRODUCTION

A crime can happen anytime and anywhere without any warning. Sometimes, it can happen at the same place and at the same recurring time. I think it would be great if we can predict where and when the crime is going to be and we can prevent it before it happened. By using artificial intelligence and machine learning from historical data, we could build a classifier to make a prediction out of it. So in this project I will try to train a model using naive bayes classification algorithm to predict a crime that is going to happen around Manhattan on January 2020. Naive bayes is a classification algorithm that using a concept of likelihood a posterior event may happened after a prior event. Hopefully, someone will be interested on this matter of research and conduct any further research using more powerful algorithm and state of the art method. It will be very helpful for the law enforcer to handle the crime in the area.

DATA DESCRIPTION

- nyu-geojson data to retrieve the neighborhood latitude and longitude

- Foursquare API to retrieve the venues around the neighborhood

- NYPD complaint data on year range from 2015 – 2019

METHODOLOGY

- Retrieve venues data from Foursquare API

- Preprocess NYPD complaints data (clean data type, resample data, split train test data)

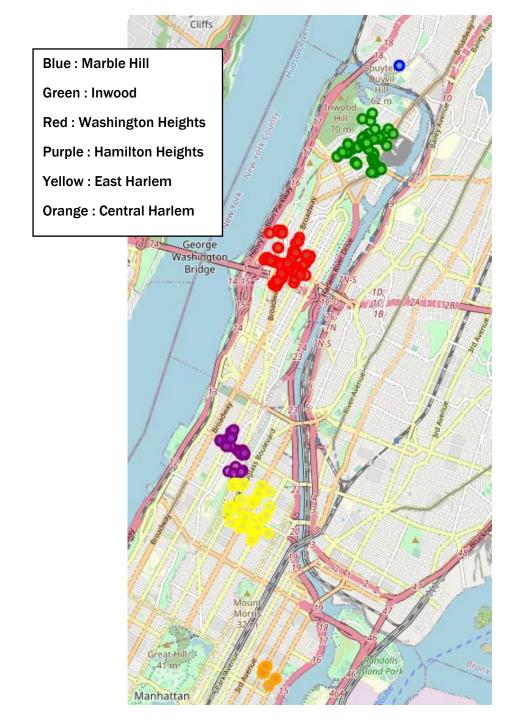- Training model

- Evaluate model

- Make a prediction

RESULT

ACCURACY : 0.4958 %
F1 SCORE : 0.3333 %

225 total crime of
INTOXICATED/IMPAIRED DRIVING is going
to happen on January 5th around these
neighborhood:

```
Washington Heights     85
Inwood                 57
Central Harlem         40
Hamilton Heights       32
East Harlem            10
Marble Hill             1
```

PREDICTION

PREDICTED CRIME MAP

Blue : Marble Hill

Green : Inwood

Red : Washington Heights

Purple : Hamilton Heights

Yellow : East Harlem

Orange : Central Harlem

CONCLUSION

- Naive bayes classifier can make a model that can predict a crime in Manhattan but the model itself is still need a lot of more improvement to increase the accuracy of model.

- Further research : improve the model performance, use another algorithm, add more features that correspond the happening of a crime from another source of data