Writing Assignment 3

Qi Mao

maoxx241@umn.edu

Haowen Luo

luoxx560@umn.edu

March 28, 2019

1 Background

As international students, the safety of us is always the most inportant problem concerned by our parents. So, we would like to build a model to help them understand what situation of crime in the major cities of the United State. The purpose of this project aims to discuss the relationship between the type of crime and the area of Chicago. Why is Chicago? As one of the biggest and most developed cities in Midwest of the United States, Chicago is representative. And it is a well-known city near Minneapolis which means more persuasion for parents of international students. For another important reason, Chicago Data Portal provides the most detailed and reliable data to analysis on the website. This model we build should not only provide a clear and visualized information to parents, but also help the police to predict what kinds of crime happened in this area after they receive a report. And According to the paper "Using machine learning algorithms to analyze crime data", we can see a very successful example of using machine learning to predict crime.[2] This gives us confidence in using machine learning to predict crime.

2 Research Problem

The problem we are researching is to predict the type of crime based on the location of the crime.

3 Approach

In this project, we will use the historical crime record of Chicago. In this dataset, crimes are documented in details with time, latitude, longitude, location description, and crime type. We will use Python and R to write the Project.

First, we will download the dataset from the Chicago Data Portal website. And we will use the Jupyter notebook and RStudio to analyze and clean up the data.

Then, we will visualize the data. In this part, we will conduct exploratory data analysis. we will explor two maps, one for the overlook the other one for details. and we will use plot to show the tend of the main type of crime.

Finally, we will choose four models to evaluate the dataset and find the relationship between the type of crime and the location of the crime. We will choose the k-NN model, Naive Bayes, random forest, and Logistic Regression. According to the accuracy, precision, and recall, we will show the performance of each model. and we will plot the ROC curve of each model. From the random forest, we can find the feature importance.

4   Preliminary result

We predict that there is relationship between the type of crime and the location of the crime exists. According to the paper "An Experimental Study of Classification Algorithms for Crime Prediction", we can make a predict that Naive Bayes will have a well performance.[1]

5   Timeframe

According to our schedule, we decide to meet every Tuesday and Thursday after the lecture. For the next week, we will collect the data from some reliable source and clean it up. This step aims to make the data usable. When we reform and clean-up it, we can find the outlier data which should be paid more attention. In the second week, we plan to make the data visualized which would be easier to understand, and it is important for a complete academic project. Then, we are going to build models for training and predict what kinds of crime happened with known location information. Finally, we will do the conclusion and summary, complete the final report of this project.

6   Note

Haowen write the timeframe part and the Background part, Qi write the remaining part.

References

[1] Rizwan Iqbal, Masrah Azrifah Azmi Murad, Aida Mustapha, Payam Hassany Shariat Panahy, and Nasim Khanahmadliravi. An experimental study of classification algorithms for crime prediction. *Indian Journal of Science and Technology*, 6(3):4219–4225, 2013.

[2] Lawrence McClendon and Natarajan Meghanathan. Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal (MLAIJ)*, 2(1):1–12, 2015.