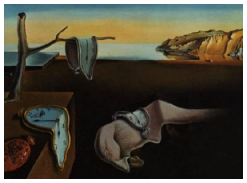

Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *doi: 10.18637/jss.v000.i00*

Team 15 The Scientists: Crime Prediction Proposal

Xiangyu Jin (xjin13)
UIUC

Nicholas Wong (nwong27)
UIUC

Abstract

We are trying to predict crime type based on arrestee's information

Keywords: R, group project.

1. Overview

To predict the type of crime an arrestee may have committed based on their background information such as race, sex, age, and etc, is a interesting topic to us. We want to see if there is any relationships between crime types and arrestee's background information, and to see crime distribution in Chicago.

There are numerical and categorical data in these arrestee's attributes. Thus, when we try to find fit models, we need to be careful. Moreover, since it's a large data set, we need to some data cleaning and feature engineering. There is a paper about crime prediction, we want to do the similar prediction on crime type based on arrestee's information. https://arxiv.org/pdf/1508.02050.pdf

In Chicago Data Portal, there is record that has crime record in Chicago since 2001. The data set contains information about every arrest and arrestee, and is updated daily. This data set is provided by Chicago Police Department. The link to the data set website is: https://data.cityofchicago.org/Public-Safety/Arrests/dpt3-jri9

The intuition is we can use Classification models or random forests model to predict the crime type based on information about arrestees. We can use data before 2018 as training data and data after 2018 as testing data to test our model accuracy. In this course, we learned how to use R to perform data manipulation, data cleaning, data visualization, and data analysis to finish this project.We will probably use tidyr to clean the data and ggplot2 to draw visualization graph. package caret will be used for training the prediction models.

2. Contribution

Xiangyu wrote the proposal and Nick participated in decision making.

Xiangyu 90 Nick 10

References

Affiliation:

Xiangyu Jin (xjin13)
Department of Statistics
E-mail: xjin13@illinois.edu

Nicholas Wong (nwong27)
Department of Statistics
E-mail: nwong27@illinois.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.18637/jss.v000.i00 *Accepted:* yyyy-mm-dd