Perfecting the Crime Machine

Yigit Alparslan[1], Ioanna Panagiotou[2] Willow Livengood[3], Dr. Robert Kane[4],Andrew Cohen[5]
[1,2,3,5]Department of Electrical and Computer Engineering, Drexel University
[4]Department of Criminology and Justice Studies, Drexel University
Philadelphia, PA
Email: {ya332[1], ip68[2], wgl28[3], rjk28[4], arc334[5] }@drexel.edu

***Abstract*—This study explores using different machine learning techniques and workflows to predict crime related statistics, specifically crime type in Philadelphia. We use crime location and time as main features, extract different features from the two features that our raw data has, and build models that would work with large number of class labels. We use different techniques to extract various features including combining unsupervised learning techniques and try to predict the crime type. Some of the models that we use are Support Vector Machines, Decision Trees, Random Forest, K-Nearest Neighbors. We report that the Random Forest as the best performing model to predict crime type with an error log loss of 2.3120.**

***Index Terms*—Crime Prediction, Crime Category, Algorithm, Machine Learning, Supervised Models, Unsupervised Models, Cluster, Urban Computing**

I. Introduction

Crime is a problem that we face every day in our society. Even though there are various reasons behind it, most of the reasons of crimes can be attributed to social-economical reasons. It is also shown that urban areas and cities show higher density of crime[1]. Crime also depends on different factors such as education, culture, economy level of neighbours and unemployment. There is a huge push towards using machine learning models to get statistics regarding crime predictions, to attest why they occur, when they would occur, and to whom it would occur[2][3][4][5][6][7]. One of the reasons we wanted to work with crime was because of individual incidents that we have seen on Drexel University campus, a rape incident dating around September 2019 that caused widespread backlash around Drexel University community and Philadelphia community regarding why Public Safety didn't take enough precautions. Philadelphia, being at the top 6 cities in the United States for population, and being our home appeals to us as a city that we can study, because we wanted to see if we could find any underlying reasons regarding crime by building predictive models, and see if we can systematically find those reasons with robust workflows. Some of the workflows that we adhere by in this study is to feature extraction, model selection, parameter tuning for those models, and feature selection.

II. Related Work

There is a huge push towards building predictive models and fight against crime. Studies show that one of the techniques used widely in this crime field is to look at how dense the crime points are on a map. It has been shown that the existence of crime dense areas can be used as an indicator of the future crime areas since crime changes depend on several different reasons on a multidimensional layer, this has been widely accepted as an indicator of future crime. In this study, we wanted to differentiate ourselves by following approaches.

1) Work with very large number of classes (30 labels)
2) Create features that doesn't depend on the city
3) Find optimal number of clusters in a data set
4) Cluster centers and use the distance as feature in our predictive models.
5) Work with different supervised learning models that incorporate the aforementioned aspects hoping that it would increase our model accuracies.

Researchers have focused on studying crime both from a time and location perspective[8]. The time perspective is the predictive aspect of crime as one might imagine. More specifically, one can create a grid on a city, and count the crime points on a grid and pose this problem as a regression over time series[9]. Other perspective is to use the location. Location might sound similar to the first time perspective but this is different and the difference lie on the fact that crime locations barely change over short amounts of time. So, if one were to study the crime dense neighbors of Philadelphia over a decade, and then guess the crime dense neighbors for the next year, month etc, one potential solution would be to flag the already existing crime dense areas and predict those neighbors as the future potential crime dense areas. We have to realize that the literature uses a special word for this, that is crime hot spot. There are mathematical models that labels an area as crime hot spot or not based on a Euclidean distance, that is a linear kernel functions.

Even though current literature is built on top of these approaches, we wanted to remove the assumption that current literature has, even this meant deviating from the current literature approaches. For this reason, there is a narrow common ground between our findings and the common ground where we can compare our findings. This meant creating models that would not depend on the city. For example, as we have seen with the time perspective, a predictive model that poses this crime problem as a regression is a model that would need crime counts over time. To get the crime counts, most researches had to create grids on a city and count the crime counts for each grid and sum them over different periods of time[10][11]. This way, one can do regression single grid, and