

For office use only	Team Control Number	For office use only
T1 _____	1906204	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	C	F4 _____

2019
MCM/ICM
Summary Sheet

Analysis of the opioid crisis and strategies

Summary

The United States is undergoing an unprecedented opioid crisis. Opioid drugs have been widely used as medicine in a variety of treatments. The fast pace of life and fierce social competition put a lot of pressure on people. People with mental or physical illness will probably be treated with many drugs and they are likely to get addicted to them, especially to opioid drugs. Thus, more and more drug identification cases have been confirmed in recent years.

The goal of our model is to find the spread and characteristics of the reported synthetic opioid and heroin incidents and do possible explanation for current situation and prediction for future case distribution, based on provided data. Specifically, our model is inspired from "Recommendation System". The first step of our model has the similar purpose as "Recommendation System" does that is to find similarity and correlation between different zones and drugs. It is rough and inaccurate to directly address the data regardless of plenty of factors behind this sophisticated problem. In order to find how graphical location, marital status, education level, age distribution and other factors contribute to the opioid crisis, one appropriate method is to first find similar zones based on social structure, then compare that how their drug spread and opioid drug identification cases distribution relates each other and then we extend our model to serve different purposes, such as tracing the drug source, predicting the drug spread.

For the first part, we construct a weighted directed graph based on the similarity(described above) and use "Walk Around" strategy to simulate the drug spread process to trace the drug start source. Still based on similarity, we use SVR regression to fit the data distribution over time and predict the following two years of synthetic opioid drug identification and heroin cases distribution situation. Then we use a SVM discriminator to predict whether a county will be in the risk of opioid crisis. A county in opioid crisis will face a continues increase in drug abuse.

For the second part, We first binarize all the data through the Kmeans algorithm, then we use association rule learning algorithms to look for factors that lead to opioids and drug addiction. We then introduced the time factor, further streamlined the factors through correlation analysis methods, and found people who abused opioids. After these steps we can find all the main factors. However, due to the large number of these factors, we also need to use the PCA algorithm to reduce the output factors to make the prediction model simpler. Our model finds that difference in population distribution has a huge impact on the abuse of opioids.

For the third part, we extract three main features from part two, reintegrate the data and invoke this into our previous model, making our model do regression with multiple dimensions. We design some strategy aimed for different groups and use our model to verify the effectiveness of our strategies. Our model finds that special attention should be payed to female householder without husband as well as householder who is 65 years old or older and improving overall education level can also decrease opioid addiction rate.

Keywords: Recommendation System, Regression, PCA, Association Rule Learning