**Fighting Crime with Predictive Analytics**

Nicole Lin & Renel Chesak

A predictive model was built from demographic records in the 1990s order to determine (1) what factors most influence crime rate and (2) how we can predict the crime rate in a given county.

Several model selection techniques were used to identify the highest-performing model, being sure that all of the statistical science behind the analysis was sound, model assumptions were met and the model predicts adequately.  In order to boost the predictive accuracy of this model in the future, we could incorporate more factors that might have an influence on crime, such as income inequality, technology, globalization, criminal justice responses, and even simple things such as whether it’s a payday, a holiday, or a game day.

It is important to note that this model was built with data from the 1990s in the US, and due to changes in the economy, society, culture, and public policy, it can only be used to make predictions about this specific era and region. This model could easily be updated with current data to predict crime happening now. It could then be used by law enforcement to better allocate resources, by policy makers to drive decisions and regulations, and by non-profits to develop community outreach. This is a trend that is happening in cities all across the country, and has been linked to a significant drop in crime rate (Siegel, 2013).

The two final models discussed here have some important differences. When it comes to methodology, Renel’s main goal was to make sure model assumptions were met, whereas Nicole preferred to match what has been done in the literature. Nicole’s model includes a new variable, population density, which is largely responsible for its ability to explain 10% more of the variation in crime rate. The two models also differ in the number of predictors in the model. Depending on the user’s available resources for collecting, storing, processing, and updating the model, they may choose to have a variable with only 5 predictors (Renel’s model) as opposed to 7 (Nicole’s model). Or, they may have bountiful resources and decide that the 10% boost in explanatory ability (Nicole’s model) is more important. However, it is worth noting that both models comparable prediction accuracy (70% for Renel’s model and 73% for Nicole’s model).

Our models indicate that population density has the largest influence on crime, followed by per capita income, location in the North West region of the US (as opposed to the West), and poverty rate. These models can predict with about 73% accuracy, and can explain about 62% of the variation seen in crime rate. This model is quite proficient in predicting crime for a county in the 1990s, especially given that it was only built on simple demographics. Given these findings, we recommend police forces increase patrols in highly populated areas and areas with a high poverty rate. We also recommend that policy makers look for ways to decrease the poverty rate in their counties, and that community outreach groups do the same.

Siegel, E. (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken, New Jersey: John Wiley & Sons, Inc.