To: Community Leaders and Government Officials

From: H2Li

Subject: Final Project Update

Date: December 4, 2019

To Whom It May Concern,

Over the past several months, we have conducted an investigation into predictors of violent crime in communities. Given a dataset from the UCI Machine Learning Repository with 128 attributes and data from approximately 2000 communities, our team aimed to discover what attributes, if any, serve as the main predictors of violent crime through employing various machine learning models. The attributes included in the dataset involve the community, such as the percent of the population considered urban, and the median family income, and involving law enforcement, such as per capita number of police officers, and percent of officers assigned to drug units. Our objective was two-fold: to see how accurate a machine learning model could predict violent crime in a community given these attributes and to see what attributes carried the most weight in the model; that is, what attributes have the most predictive power in estimating violent crimes in one's community.

After testing a variety of different machine learning models, we found that we were able to achieve the highest accuracy and the lowest error using deep learning models (i.e. feed forward neural networks). We found that we were able to achieve comparable accuracy and error scores using the top 5 "important" features: Percent of kids born to parents who have never been married, percent of children in family housing with 2 parents, percent of community members who are white, percent of the population under the poverty line, and percent of vacant housing that is boarded up. However, we noticed that all of these features had relatively low feature importance, and all of these attributes correlated to poverty metrics. That is, the top features in predicting violent crime also showed a strong relationship to poverty. Thus, we hypothesized that perhaps just poverty alone would be a good predictor of violent crime. Using only the feature that measures the percent of community members below the poverty line, we were able to create several models with comparable accuracy and error metrics compared to the models created with the full dataset and the reduced data. This indicates to us as a team that poverty is one of the most significant predictors of violent crime.

This does not necessarily indicate that poverty, and all of the other factors that correlate strongly to poverty, is the root cause of violent crime. Crime is an incredibly complex socio-economic issue whose cause cannot be boiled down to just one single attribute. However, our findings do indicate that the strong correlation between poverty and violent crime cannot be ignored. That is, violent crime rates are higher in communities that experience higher rates of poverty. While unsurprising, this result also suggests that tackling the issue of poverty should be a priority in communities looking to potentially lower violent crime. While there is no easy or obvious way to do this, community leaders, such as yourself, should consider government programs that help individuals in your community find jobs, stable housing, and other resources so that crime is not a necessity in their everyday lives. Again, while this will not guarantee a reduction in rates of violent crime, our findings do indicate that there is a proportional relationship between rates of violent crime and poverty in communities across the United States.

To help community members better understand violent crime in their own community and the relationship between poverty and violent crime, we developed a website that uses a neural network to predict rates of violent crime in a user's community given the poverty rate. To use the website, the user simply just has to input the percent of individuals below the poverty line, and the model will output if the rate of violent crime in their community is low, medium, or high compared to other communities in the database. While this product is not super useful in its current form, we envision incorporating geographic data in the future, so the user simply would have to input their zip code instead of searching for their community's poverty rate. However, this was outside of the scope of the project. In addition, it is important that we make very clear that people in poverty are not inherently dangerous, and that the predictive power of poverty does not indicate that poor people are criminals. Crime is an incredibly complex social and economic issue that is not due to poverty alone and is not just committed by those in poverty. There are complex reasons why there is such a strong relationship between poverty and violent crime.

In sum, during the course of this study, we were able to show that rates of violent crime can be predicted using community statistics. That is, we found that rates of violent crime can be predicted just as accurately using only a few community statistics, and just poverty alone. This suggests that there is a strong relationship between rates of violent crime and poverty, as the variables found to have the most predictive power in predicting rates of violent crime were also strongly correlated to poverty in the community. Though this does not necessarily imply that these factors directly lead to violent crime, there is an underlying correlation. We apologize that our findings were not more conclusive but we appreciate the opportunity to apply machine learning to this complex issue. Feel free to contact us with any further questions.

Sincerely,

H2Li
Heidi Hufford, Hannah Levy, and Lindsey Nield
CSCI 470 Final Project
Colorado School of Mines