**Capstone proposal Alissa Mezzacappa**

**Drug Consumption Risk Data Science Career Track**

Psychoactive drug abuse creates risks for users and those around them. Drug consumption risks include poor health, early mortality, and societal consequences. For example, the opioid epidemic not only claims the lives of users, but leaves children without mothers and creates criminals out of law abiding citizens because of the addictive nature of these drugs. However, it is possible that some fraction of this destruction is preventable. In the case of opiods, if a pain management doctor was given the risk profile of patient for opioid addiction, this insight could inform his/her choice of prescription pain medication, potentially saving a life and a lot of societal suffering.

Fehrman et al. 2015[[1]](#footnote-1) created a database[[2]](#footnote-2) available on the UCI Machine Learning Repository that can enable just this type of modeling of drug risk behavior. Using an online survey methodology, usable data from 1885 persons was collected on the Big Five personality traits, impulsivity, sensation seeking, demographic information, and drug use of 18 different legal and illegal drugs. With this dataset, one can identify the association of personality profiles with drug consumption and predict the risk of consumption of individual drugs existing for specific personality type and demographic background.

Understanding the profile of a drug user could create an alternative to failed anti-drug education programs like DARE. Individuals are not born using drugs. An individual has risk factors which are an attribute, characteristic, or event which catalyze to drug use. In this sense, drug use is a symptom of the underlying profile. If drug prevention attempted to mitigate underlying personality attributes which contribute to drug use rather than devoting curriculum time on purporting the negative attributes of drug use, educators could engage students in positive ways that decrease neuroticism and increase conscientiousness, two known attributes of drug users.

To enable educators and doctors prevent drug consumption, this data will be used to identify the associations of personality profiles with drug consumption and predict the risk of consumption per drug for any given profile. Here correlation will be used to create profiles of drug users. To approach the problem of predicting the risk of consumption per drug, decision tree and random forest will be tested identify the most accurate method by first, training the model with known drug user types in accordance with their personality data, and then cross-validating the model using leave one out method.

Deliverables of the project will include code, a prediction model for drug user type for unknown test subjects given inputs of Big Five personality traits, impulsivity, sensation seeking, and demographic information with known accuracy, a paper summarizing results and conclusions and a slide deck.

1. E. Fehrman, A. K. Muhammad, E. M. Mirkes, V. Egan and A. N. Gorban, "The Five Factor Model  of personality and evaluation of drug consumption risk.," arXiv, 2015 [↑](#footnote-ref-1)
2. https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29 [↑](#footnote-ref-2)