Introduction

When it comes to examining behaviors of drug use, the relation between one's personality and drug tendencies is often questioned. The Five Factor Model of Personality is an overarching scale that divides an individual's personality into five traits: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. One can rank at or between the extremes of each trait. From these ratings, there may be attributes of a trait ranking that correlate with an individual's tendencies to use drugs or their likelihood of an addiction. There may also be social and environmental factors that contribute, which we hope to identify by looking at the relations between personality, demographics, and type and intensity of drug usage.

Team Members

Madelyn Redick
redick.m@northeastern.edu
Second Year
Intermediate/Proficient data science experience

Sahana Dhar
dhar.s@northeastern.edu
Second Year
Intermediate/Proficient data science experience

Glen Lim
lim.gl@northeastern.edu
Fourth Year
Intermediate/Proficient data science experience

Timeline

We intend to dedicate the first two weeks to cleaning and organizing the dataset so that its layout is best fit for the ML tools we intend to use. During this, we will begin creating visualizations to reveal any missing data and to start examining signs of correlations or patterns in our dataset. Once basic visualizations and reformatting of the dataset is complete, we will spend the rest of October and the beginning of November experimenting with different machine learning tools in order to find the approach that best fits our dataset. We will refine our work and look for areas of improvement until we are sure that our prediction model is robust and reliable, or decide that it is not a strong model and not enough of a correlation to analyze. By the end of November, we plan to have determined the performance and effectiveness of our prediction models, by employing visualization techniques to explore their trade-offs and enhance our understanding of the overall problem.

Project Description

The main objective of our project in the field of data science is to identify any correlation between people's demographics and personality types with their drug habits. Then, with the use of advanced machine learning tools, we hope to build the most fitting machine learning model that will read data of people's scores of the Five Factor Model of Personality and their demographics, and best predict how frequently they may partake and which drug they might be most susceptible to use. This can only be achieved if there is a strong correlation between the two aspects, which we will identify through a thorough analysis, which consists of creating visualizations, and the testing of different classifiers and regressors. If successful, and the predictive power of our model is strong, then this model can join the development of the use of data science and predictive models in healthcare.

As drug use increases, a larger problem in society is successfully identifying individuals who need preventative measures or rehabilitation. Through this project, we are aiming to identify potential connections between personality type and the frequency of drug use. Given this dataset of over 1500 people's personality scores in the Five Factor Model and their frequency of use for 18 different drugs, ranging from recreational to addictive substances, we hope to identify any correlation and generate a predictive model between levels of personality type and an at-risk level of drug use. We will be employing feature selection techniques, as well as cross-validating our model in order to ensure that we can discover the best fit model for our data with the strongest correlation and highest predictive power. We will be exploring different ML algorithms, starting from baseline classification problems such as logistic regression or decision trees, to more advanced techniques such as neural networks. If successful, this model could potentially help identify more at-risk human beings, which would greatly assist preventative efforts and tailoring rehabilitation programs to best support personalities that may be at greatest risk.