Predicting Opioid Use Disorder (OUD) Using Machine Learning

Wadekar, Adway (School: Saint John's High School)

Opioid Use Disorder (OUD), defined as a physical or psychological reliance on opioids, is quickly becoming a public health epidemic. This project develops a supervised machine learning approach to predict adults at risk for OUD by considering interactions between various demographic, socioeconomic, physical, and psychological features in an integrated manner. A labeled data set was built by pooling the responses from the 2016 and 2017 editions of the National Survey on Drug Use and Health (NSDUH). This labeled data was split and used to train decision tree and random forest classifiers while downsampling to account for class imbalance. Both classifiers can predict adults at risk for OUD accurately with the average area under the ROC curve (AUC) over 0.80, even though the prevalence of OUD in the test data set is only about 1%. First use of marijuana (prior to 18 years) emerges as the dominant predictor for developing OUD in adult life. This is surprising because early initiation of marijuana ranks higher than both mental illness and disability, two conditions that are often comorbid with substance use disorders. The classifiers also reveal that early initiation affects some socioeconomic and demographic groups more than the others. These include young adults (18-34 years), or those with incomes less than $40,000, or of Hispanic and White heritage. The key takeaway is that curbing early initiation of marijuana will be an effective prevention strategy. This highlights the crucial role that educators, counselors, and parents can play in alleviating the opioid overdose crisis.

Awards Won:
American Statistical Association: Certificate of Honorable Mention