A Machine Learning Approach to Identify Socio-economic Factors Responsible for Patients Dropping Out of Substance Abuse Treatment

Gautam, Prateek (School: Marquette High School)

This study will help determine whether certain socio-economic factors emerge as common threads between patients dropping out of substance abuse treatment, which would help healthcare providers administer more targeted and equitable treatments that meet the needs of all patients. This study uses patient level data from the Treatment Episode Data Set - Discharge (TEDS-D) 2017 because we can be sure it includes almost all public treatment facilities, which are the ones the government can control and affect if needed based on the results of this study. A multiple logistic regression was run in R coding. 70% of the data was used to train the model and 30% was used to test. The model was found to be accurate in predicting a patient dropping out of opioid abuse treatment 68% of the time. 22 out of the 30 independent variables chosen for the analysis were found to be significant predictors. Next, Random Forest Classification was run. For the Random Forest Classification, 200 trees were generated. Within each tree, at every single node 5 different variables were evaluated to determine what the best way to split the data was. The resulting prediction was 89% accurate, and a smaller number of variables were determined to be significant. Comparing the analyses indicates that the Random Forest Classification method is a better predictor of dropping out of opioid abuse treatment, with the Length of Stay, State/Region a person gets treated in, Age, and Employment Status being the most significant variables found. The MLRM was not as accurate as the Random Forest; however, the Analysis of Maximum Likelihood Estimates gives us the ability to generate an equation using the “significant” variables that can predict the likelihood of a patient dropping out of opioid abuse treatment.

Awards Won: