

Capstone Project 1 Proposal

According to National Survey on Drug Use and Health (NSDUH), 19.7 million americans(12 years or older) suffered a substance abuse disorder in 2017. In addition to that there are millions of people who are indirectly affected by this problem, such as families of abusers and those who were injured or killed by intoxicated drivers. Besides, the monetary cost associated with substance abuse to society and the economy because of reduced productivity, property damage, accidents, and health care expenses are astounding.

There are over 14,500 specialized substance abuse treatment facilities in the United States providing a variety of care options, including counseling, behavioral therapy, medication, case management, and other forms of care. National Institute on Drug Abuse (2017) states that an estimated 20.7 million people age 12 and older needed treatment for a substance use disorder. Unfortunately, addiction treatment completion rates remain relatively low, looming around 50%. Addiction is considered a highly treatable disease, but successful completion of the treatment program is an important factor leading to recovery.

Understanding the risk factors that may lead to an individual's drop out from the treatment program and predicting his\her likelihood of dropping out can help to make the programs more effective.

Problem

The problem is to predict the likelihood of a patient dropping out of the treatment before the actual completion of the program. Considering the demographic and clinical characteristics, which are the factors that affect a patient's successful completion of an addiction treatment program?

Client

Addiction treatment facilities and healthcare providers: With the understanding of risk factors, they may be able to target patients at high risk for dropout, and tailor treatment programs in order to maximize completion of treatment, influencing lower relapse rates.

Data

The Treatment Episode Data Set -- Discharges (TEDS-D) is a national data system of annual discharges from substance abuse treatment facilities. These data provide information on the demographic and substance abuse characteristics of substance abuse treatment discharges and their corresponding admissions aged 12 and older in facilities that report to individual state administrative data systems.

<https://datafiles.samhsa.gov/study-dataset/teds-d-2016-ds0001-teds-d-2016-ds0001-nid1845>

Method

1. Exploratory data analysis and statistical inference
2. Usage of Tree based methods and gradient boosting methods for drop out prediction.

Deliverables

Code on Github, project report and powerpoint presentation.