9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Using Machine Learning to Predict Involuntary Substance Use Treatment Terminations

Keywords: substance abuse, machine learning, data science, bias detection, harm reduction

Extended Abstract

The United States National Institute on Drug Abuse estimates that substance abuse costs the nation over $600 billion annually. [1] In an effort to predict the likelihood of treatment completion, prior machine learning research on publicly-available addiction treatment datasets has focused almost exclusively on correlations between treatment outcomes and either patient demographics (e.g., race, ethnicity, income status) or disorder presentation features (polysubstance use, age of first use). [2, 3] By contrast, we tested machine learning algorithms to isolate and identify features that predicted a higher likelihood of early unilateral treatment termination by the provider. Our most significant finding was the demonstrably higher impact of socioeconomic variables - namely, income source, health insurance, and primary method of payment - on the termination outcomes predicted by our highest-performing classifier, in contrast to other demographic, behavioral, or treatment-related variables. The prevalence of missing values and imbalance in target classes, however, presented significant challenges to the development of a model that performed well with respect not only to accuracy but also precision and recall.

The publicly-available Treatment Episode Data Sets-Discharges (TEDS-D) contains de-identified episode-level intake, treatment, and discharge data for every patient treated from 2006-2020 at a facility receiving federal funding from the public agency charged with improving the quality and availability of treatment and rehabilitative services in the US for mental illness and addiction. Our dataset, the 2019 TEDS-D, was comprised of 1,722,503 rows capturing 76 individual patient-level features, including demographic data (e.g., race, ethnicity, Veteran status, sex), intake and discharge information (living arrangements prior to arrival, reason for discharge, discharge destination), and limited course-of-treatment data (number of self-help groups attended within 30 days prior to discharge, mental health diagnostics). Our target outcome, involuntary patient terminations, represented only 5.53% of all discharges.

Our approach was to train and test a broad range of machine learning algorithms in search of the highest-performing classifier best able to successfully predict a termination outcome based on a presented feature set. We would then perform permutation analysis on the selected classifier, an approach that randomizes values for each independent variable to determine its relative impact on classifier performance. Our initial efforts to develop an effective classifier included Support Vector Machines, Random Forest Classifiers (RFC), and multi-layer perceptron neural networks, with and without gradient boost and linear embedding. Area under the receiver operating characteristic curve (ROC AUC) measurements for all classifiers fell within the range $[0.80, 0.84]$, commensurate with the highest-performing classifiers documented in the literature. [4] After some minor dimensionality reduction, feature preprocessing, and hyperparameter tuning, we selected an RFC with an ROC AUC score of 0.86 for further analysis.

Applying permutation analysis, we identified that the top four features with relative weight scores at or above 0.055 contributing to a termination outcome were HLTHINS (health insurance), PRIMINC (income source), SERVICES (type of treatment setting), and PRIMPAY (primary method of payment). The next most heavily-weighted feature scores just above 0.025, less than half the weight of its immediate frontrunner. The relative weights of primary income source, primary method of payment, medical insurer, and EDUC (education, the next highest-ranking demographic variable) may indicate a socioeconomic bias; uninsured patients were more likely to face termination than all patients combined (7.53% versus 5.53%), and privately insured patients were somewhat less likely (4.67%). Other variables posited to be potential areas of provider bias - Veteran status, race, ethnicity, intravenous drug use - ranked below the upper twelve factors and represent decreasing feature weights contributing toward termination.

The next highest-scoring features, PSYPROB and DSMCRIT, both relate to mental illness and suggest that terminations may be a response to behaviors correlated with specific co-occurring psychiatric diagnoses. Indeed, the differences between rates for termination for patients with schizophrenia, psychosis, and bipolar disorder versus the entire population are the greatest of all the variables within the top seven weighted features, with 12.42% of schizophrenic/psychotic patients and 11.09% of bipolar patients being terminated. Flags for specific drug or alcohol use (e.g., METHFLG, an auto-calculated variable indicating whether any data in a patient's row indicated that the patient used methamphetamines) were ranked extremely low, suggesting that the type of drug or alcohol is not heavily weighted in a termination outcome.

It is important to note, however, that the extreme imbalance in our target classes calls into question the validity of even the highest-performing classifier. [5] Despite our attempts to address this concern through sample rebalancing using scikit-learn's built-in subsample-weighting hyperparameter, our precision-recall AUC scores fell within the range of only [0.18, 0.24]. Additionally, we attempted to improve classifier performance through excluding rows of data with a high number of missing values. In doing so, our classifier's performance decreased from an ROC AUC score of 0.86 to 0.77 as we set progressively lower thresholds for allowable row-wise counts of missing values. In other words, terminated patients were more likely to have more complete data and fewer missing values than non-terminated patients. We measured predictor performance at varying thresholds of missing values tolerance and landed on a maximum allowable number of 3 missing values for a given row in the training dataset.

In sum, the most significant finding of our project was the demonstrably higher impact of socioeconomic variables, including income source, health insurance, and primary method of payment, on a termination outcome than any other demographic or use-related variable. Under our Random Forest Classifier and optimized, preprocessed feature set, demographic variables and flags for specific drug use or intravenous administration scored far below socioeconomic factors. However, further research is needed to better mitigate the impact of missingness and class imbalance on classifier performance to ensure that predictive power correlates to models performing well with respect not only to accuracy but also to precision and recall.

References

1. National Institute on Drug Abuse (2014). Principles of Drug Addiction Treatment: A Research-Based Guide (Third Edition). https://nida.nih.gov/sites/default/files/podat-3rdEd-508.pdf

2. Gautam, Prateek, and Pradeep Singh. "A Machine Learning Approach to Identify Socio-Economic Factors Responsible for Patients Dropping out of Substance Abuse Treatment." American Journal of Public Health Research, vol. 8, no. 5, 2020, pp. 140–146., https://doi.org/10.12691/ajphr-8-5-2.

3. Stafford, Celia, et al. "Predictors of PREMATURE Discontinuation of Opioid Use Disorder Treatment in the United States." 2021, https://doi.org/10.1101/2021.07.26.21261080.

4. Acion, Laura, et al. "Use of a Machine Learning Framework to Predict Substance Use Disorder Treatment Success." PLOS ONE, vol. 12, no. 4, 2017, https://doi.org/10.1371/journal.pone.0175383

5. Branco, Paula, Luís Torgo, and Rita P. Ribeiro (2016). "A survey of predictive modeling on imbalanced domains." ACM Computing Surveys (CSUR) 49.2 : 1-50.

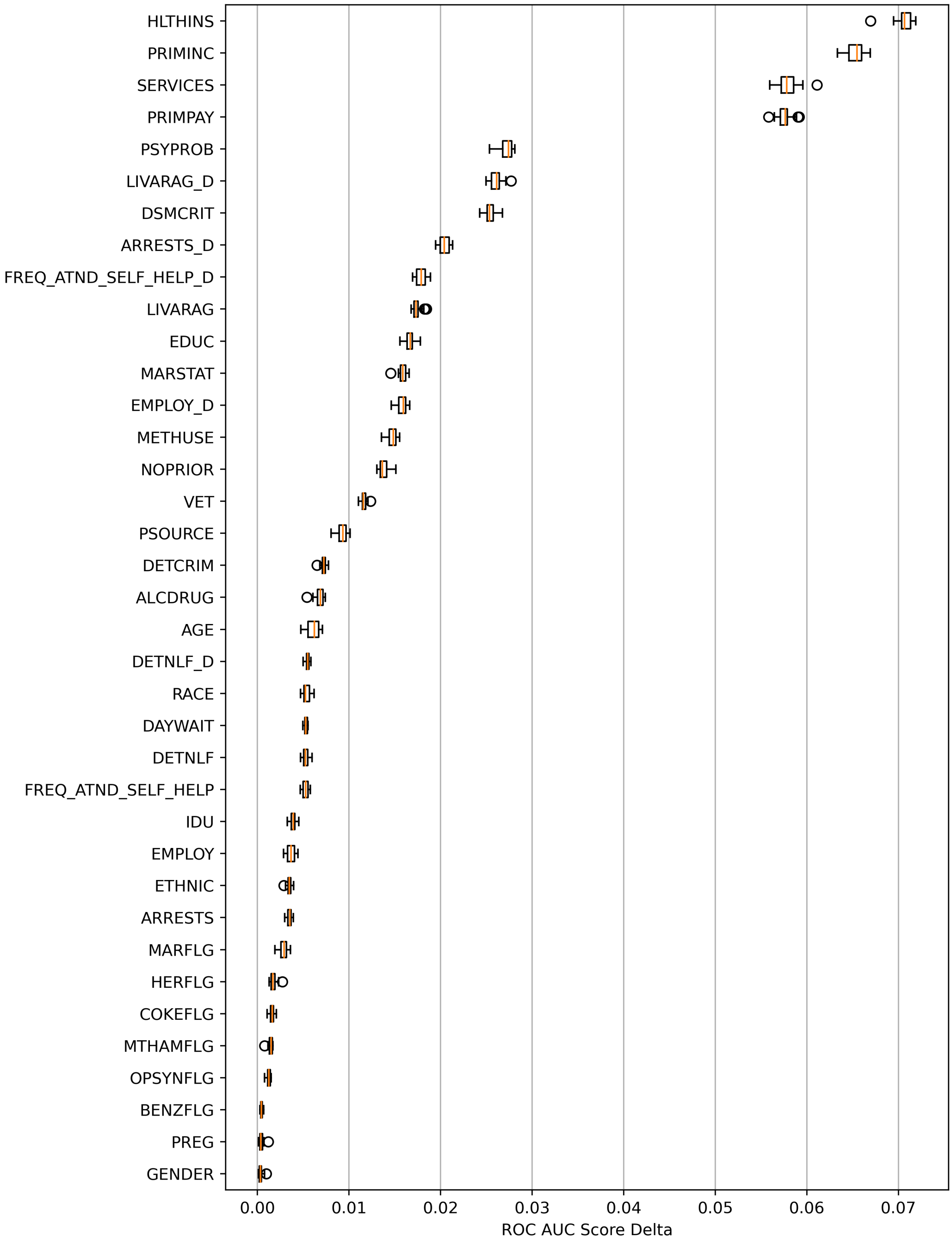

Figure 1: Change to ROC AUC scores represented by randomization of values for each variable in the TEDS-D 2019 dataset.

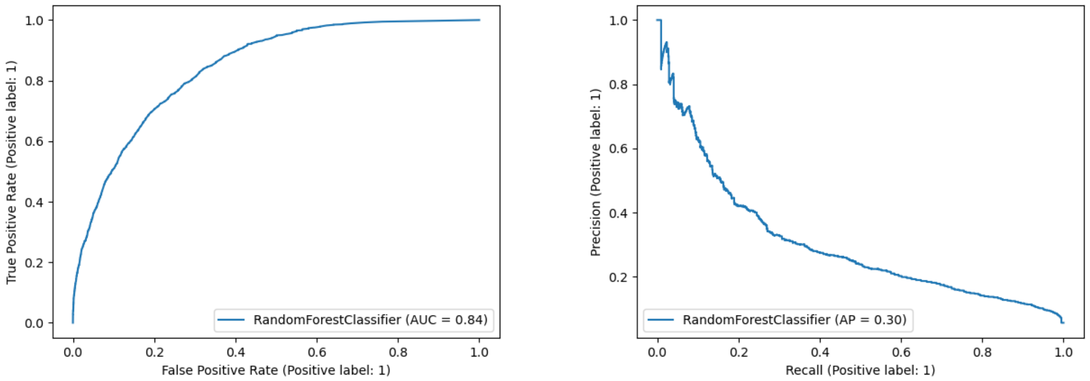

Figure 2: ROC AUC scores (left) and precision-recall curve (right) for the top-performing Random Forest Classifier

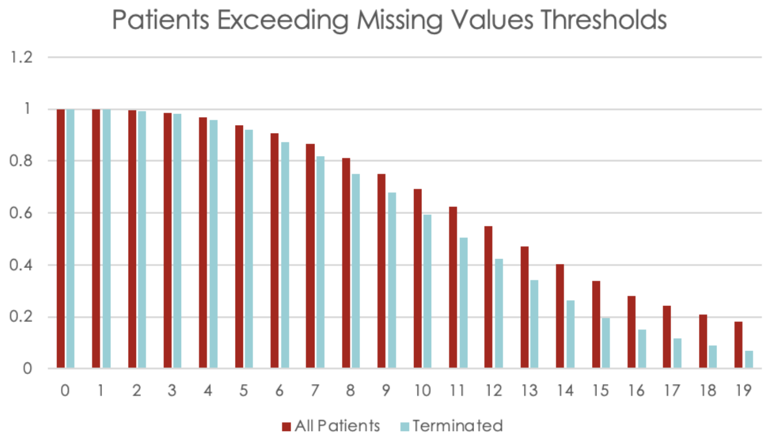

Figure 3: Proportion of episode data (rows) exceeding increasing missing values thresholds among all patients (red) and terminated patients(blue)