Prediction of Successful Discharge

Children and Youth in Behavioral Health Treatment: A Machine Learning Study

Judith Calvo, PhD

Successful Discharge

- Hypothesized Predictors: age, diagnosis, length of stay, initial scores on standard instrument, ethnicity, language and level of care
- Outcome of interest: child successfully completes treatment program

Why Project Was Chosen

Pull off proposal

Data Source

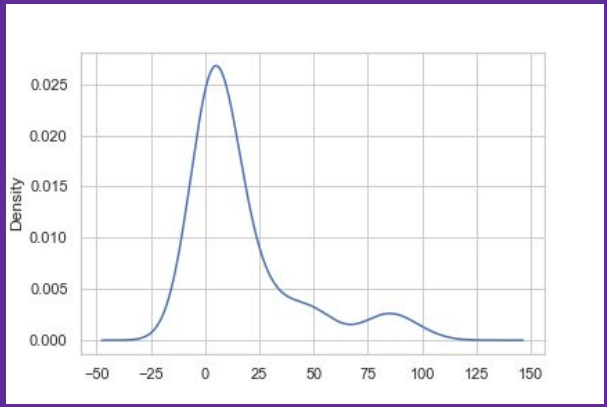
Pull off proposal

Questions We Hope to Answer

Pull off proposal

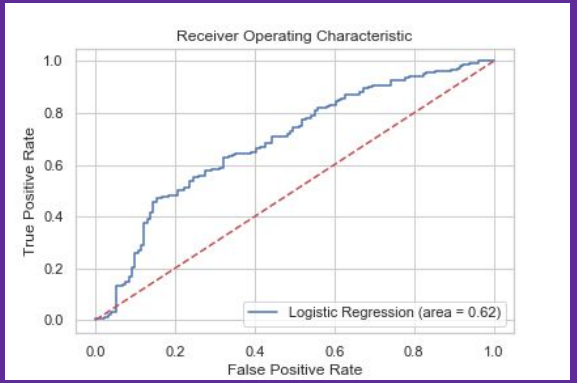
Data Exploration

- Describe with summary statistics
- List categorical variables
- Encoding of categorical variables
  - Scikit Encoder
  - Density plotting to ascertain possible binning
  - Custom encoding of level of care based on youth's program
  - Create dummy variables
- Graphs
  - Visuals of data characteristics
- SMOTE to correct imbalance



Data Analysis

- Logistic regression with feature output of beta weights
- Calculation of odds and probability for key features
- Logistic model fitting
- ROC curve
- Random Forest to compare results
- PCA in unsupervised learning to create youth groupings/profiles based on entering characteristics
- Drop and add new features with next run to improve model fit and prediction, repeat analysis



Tools Used

- Data preprocessing in Pandas
  - Encoding
  - Dummy variable creation
  - Matplotlib and Seaborn for visualizations
  - SMOTE to correct imbalance as needed
- Statsmodel to create full logistic model with betas and statistical significance
- Calculation of odds and probability for key features using Numpy exponentiation formulas
- Logistic model fitting using Scikit Logistic Regression Library
- ROC curve using Scikit AUC library
- Random Forest to compare results
- PCA and hvplot for interactive graphing
- Power BI for interactive dashboard using Python Scripting

Dashboard Tools and Interactive Elements

- Power BI for interactive dashboard using embedded Python Scripting
- Interactive Elements
  - Slicers for exploratory data by discharge status, Level of Care and levels on initial CANS scores for key features (i.e., levels of oppositional behavior at intake, child resilience, parental support) that were predictive of outcome
  - Python script embedded in Power BI using hvplot to allow for hover, rotation of plot, etc. for some of the graphs, and especially the final phase using PCA.

Storyboard--Data exploration phase

- Successful discharge by original program (1= successful)
- Successful discharge by ethnicity (no strong difference in proportions based on ethnicity