1 Machine Learning to Predict Successful Discharge of Children from Behavioral Health Treatment

1.1 Why This Topic and Data were chosen:

- It has become increasingly important to achieve an accurate assessment of likelihood of successful discharge from treatment. There are multiple contingencies to consider: contractual and client care outcome goals, improved care, increased revenue, and business stability during times of crisis, such as COVID-19. Efficient and effective utilization of services enables client flow through our system and provides resources for new clients. Moreover, the field is shifting from fee for service to value-based cost reimbursement models that structure payment based on an expected time to discharge and an expected treatment outcome.
- Importance of ascertaining successful outcomes in naturalistic settings (Webb et al., 2020) when the use of Randomized Clinical Trials is not feasible or practical. Machine learning allows for the use of naturally occurring data to train new models and utilize them with future data.
- The use of statistical prediction models is highly recommended for their versatility at many levels of mental health care delivery: diagnosis, prediction, and treatment planning (Garb et al., 2019)
- It is important to accurately identify high-risk clients at onset of treatment to intervene early and at the right level of intensity to ensure successful outcomes.
- The use of machine learning models to predict client outcomes is a nascent field of study in behavioral health; however, there is growing body of literature that supports machine learning to augment traditional clinical judgment and to strengthen the prediction of outcomes (Webb et al., 2020).
- Previous studies have shown the utility of machine learning to predict depression, treatment of obsessive-compulsive disorder in children, pediatric bipolar disorder, foster care permanency and likelihood of remaining cancer-free over five years (Gao et al., 2018; Lenhard et al., 2018; Wolff et al., 2020; Xu et al., 2019; Youngstrom et al., 2018)

1.2 Data Source Description, Extraction (ETL) and Exploration

Data for this project was extracted from our agency's EMR system. There were challenges with the extraction of this data that necessitated additional steps:

- Our system is a legacy system using an Intersystems Cache database, with the requirement that data be extracted using 32- ODBC drivers, with Crystal Reports and 32-bit WinSQL to generate SQL statements to extract data from underlying tables.
- There are hundreds of tables within the system that need to be considered and selection of tables to join is complex. The vendors provide an older version of Crystal Reports to be used in crafting reports and data extractions. I first applied Crystal Reports to test my connections between tables to get the views I needed, then ported the Crystal Reports structure into SQL, which then was used with our agency SQL platform, 32-bit WinSQL. The SQL setups used to extract the data will be included as .txt files in the GitHub repository.
- Resultant data tables were exported as .csv files and stored in Postgres on my project computer. Schema to create tables in Postgres is included in the SQL subfolder.

- All data was collected on previously discharged clients in order to have completed outcome data on discharge reason. Only children and youth between the ages of 6 and 25 are included, as I am focusing on programs serving children and young adults. At the time of data extraction, there was one client currently 27 years old.
- Data was loaded into Python using Pandas and Psycopg2 library to directly connect to the Postgres database. An additional revised table with alternative CANS items was directly imported into Pandas after extraction from our company EMR (transfer and load process). This revised CANS measure table replaced the earlier table to test an alternative set of features. Copies of all tables are also kept in Power BI for later use.
- Tables were:
    - **<u>Demographics</u>**—de-identified client information: patient ID, birth date, gender, ethnicity, education level, employment status, country of birth and primary language. These features/variables were used in previous studies (cited above) as predictors of treatment outcome. Patient ID ("patid") is the Primary Key.
    - **<u>Diagnosis</u>**—Patient ID, episode of treatment, client diagnosis at admission, and trauma history and preexisting medical conditions. Only primary DSM-5 diagnoses were included and type of program that client had been enrolled in. The type of program is a proxy feature representing level of care.
    - **<u>Level of Care</u>**—client program will be categorized along a continuum of care from prevention/early intervention to intense crisis-based care.
    - **<u>Discharge data</u>**—patient ID, date of discharge, episode (selected to most recent treatment episode), length of stay, reason (successful vs. not successful), program from which discharged. This data will be thoroughly examined for outliers, especially in length of stay when preprocessing data prior to machine learning.
    - **<u>Assessment data</u>**—use of an assessment tool that has been adopted agency-wide for children and youth called the Children and Youth Needs and Strengths Assessment (CANS). Included are initial scores on the following dimensions: depression, anxiety, conduct/oppositional disorders, school performance, and on the strengths dimension, included resiliency and family supports. The CANS tool is much more extensive in its inclusion of needs and strengths, but the first machine learning models seek to use fewer features to minimize overfitting and to simplify interpretation in the first exploratory model. CANS data was joined to the other tables using Patient ID as the foreign key.
    - **<u>Outcome</u>**—from the discharge data, did client successfully achieve treatment goals or did client discharge unsuccessfully? Binary coding (1=successful; 0=not successful) was used to create the y-variable for all train-test splitting and ssubsequent analysis

1.3 Questions we hope to answer

- What factors lead to successful completion of treatment for youth and children who receive mental health services at our agency?
- Conversely, what types of client characteristics are indicative of high risk and subsequent failure to complete treatment and achieve treatment goals? I am looking for a profile of high-risk youth who may need early and vigorous intervention to succeed.

- Can a simpler, more parsimonious model suffice to fit the data and thus have a replicable model for future data sets? Thus, we start simply and then add or remove features and vary the learning algorithm if our accuracy, precision, and sensitivity are lower than optimal. I am aiming to avoid overfitting and thus lose the ability to generalize to other datasets.

2 Methodology Used in Machine Learning

The overriding purpose of this study require the following elements:

- Prediction of outcome, using several supervised machine learning algorithms (logistic regression, random forest, ensemble, ADABoost and Gradient Boosted Tree methods)
- I also aim to classify children into groups based on their initial characteristics, thus we will also implement an unsupervised machine learning mode using PCA. This classification model will enable visualization of groupings of clients that have similar characteristics and thus similar risk profiles. Two visualizations that I may employ are: a boxplot to examine how each client cluster is distributed and an interactive plot using Hvplot method.

2.1.1 Pre-processing and data exploration

The following steps were taken to explore the data and prepare it for the machine learning phase. All code is included in the ML subfolder in the form of jupyter notebooks. Key findings will be in the Power BI dashboard and the Google slide presentation.

- Import from Postgres and direct import of revised CANS dataset (.csv format) using Pandas
- Merging the individual datasets into one DataFrame for analysis using Pandas merge function
- Listing categorical variables and using value count to determine need for binning. Variables were binned to reduce the categories.
- Descriptive statistics on key predictors individually
- Grouping by the key categorical features to examine how they related to successful discharge proportions
- Graphical representation of the data for each key feature and for the outcome of successful discharge. Examples of these are included in the dashboard and Google slides.
- Removal of categorical features with limited variability, i.e., majority of clients in one category, such as country of origin.
- Creation of dummy variables for categorical predictors using the Pandas get_dummies function.

2.1.2 Analysis—Machine Learning Model Selection

The decision to employ Logistic Regression was based on a search of the extant peer-reviewed literature that is directed at treatment outcomes in the fields of medicine and behavioral health. The preliminary feature selection followed researching the machine learning peer-reviewed literature that delineated factors such as diagnosis, age, ethnicity, gender, length of stay, level of care and family factors are features that have shown to contribute to models with high precision and accuracy (Elgin, 2018; Webb et al., 2020) in predicting outcomes.

Logistic regression was chosen first as ease of interpretation of results provides a level of transparency and understanding that contributes to application in direct service settings. Providing internal and external stakeholders with detailed understanding of the factors that weigh into prediction of successful discharge will boost the subsequent use of the models in data driven leadership to improve policy and practice. Other methods, such as Random Forest, are less prone to interpretability, but nevertheless, are useful. Therefore, for the purposes of this project and to provide a comparative analysis that might boost precision, I also used Random Forest modeling. Finally, I employed Principal Component Analysis (PCA) in unsupervised machine learning to classify clients along several dimensions, such as length of stay, level of care, age, social functioning, family strength, and oppositional behavior to create a preliminary classification of risk. These factors were the factors that were significant predictors of outcome in the final logistic regression model. I plan to conduct further research in later projects to create a classification scheme to ascertain risk prior to treatment.

These are the steps I performed:

- Logistic regression using Statsmodels to show output of beta weights used to calculate odds and probability of successful discharge for selected key features as they individually and jointly contribute to the outcome. The first logistic regression trial involved the entire dataset, but the final model fitting involved splitting the data into training and testing samples before proceeding.
- Data was split into training and test sets using Train-Test-Split function, using 75% training and 25% percentage testing. This was modified from the first trials to provide a larger training sample, as the total number of clients was only 1,237. I wanted a larger training sample to possibly improve classification.
- Logistic model fitting using Scikit after running the logistic regression model using Statsmodels.
- Metrics to evaluate model:
    - Classification report: 1) Precision, 2) Sensitivity/recall, 3) F1 Score, 4) Accuracy
    - Confusion matrix
    - ROC curve to show plot of false positive rate vs. true positive rate at varying threshold levels (different points in 0,1) using the AUC-ROC library in Scikit.
- Random Forest with classification report and rankings of feature importance using Scikit.
- PCA in unsupervised learning to derive groups of youth based on age, level of care, length of stay, scores on key CANS measures and diagnosis at entry to begin exploration of a grouping method to classify youth at greater risk for unsuccessful discharge.
- Data from PCA was exported with the classifier added as an additional variable and joined to the original xvalue subset used for PCA. This was exported to .csv files to be used in Power BI for the dashboards, joining on the Pandas DataFrame index.
- All code is in the ML subfolder. Note books that reference “ML Model” are the preprocessing notebooks. Those referencing “analysis” are the analysis portion.
- For Random Forest, the first preprocessed data in used and the notebook is entitled: “Random_Forest_Successful_dc and the second version is referred to as “Random_forest with Gender Dummy var”. The PCA analysis notebook is called “PCA Analysis Client Data”.

All visualizations produced during pre-processing and analysis are included as .png files. The ROC curve produced during Logistic Regression analysis is called "LogisticReg_ROC.png".

- During Segment 3 of the final project, further analysis was conducted and will be added to the final presentation. This was an attempt to expand the variety of random forest approaches and improve model fit and prediction. Coding is in the process of finalization and will be included in the final segment, with a draft in this segment. Methods explored were:
  - Gradient boosted tree to examine various learning rates
  - Ensemble
  - ADABoost

2.2 Preliminary Results

At this point, the accuracy of the logistic regression model using Elastic Net Regularization is 0.68. Given that this is not optimal, I used the aforementioned methods to train the model, with the goal of improvement. The precision is 0.67, and the recall/sensitivity for successful discharge is 0.88. Analysis using Gradient boosted tree found that a learning rate of 0.75 yielded the best model in terms of accuracy (68%), with recall/sensitivity for successful discharge at 0.80.

2.3 Conclusion (next segment)

References Cited:

Elgin, D. J. (2018). Utilizing predictive modeling to enhance policy and practice through improved identification of at-risk clients: Predicting permanency for foster children. *Children and Youth Services Review*, *91*, 156–167. https://doi.org/10.1016/j.childyouth.2018.05.030

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neuroscience and Therapeutics*, *24*(11), 1037–1052. https://doi.org/10.1111/cns.13048

Garb, H. N., Wood, J. M., Garb, H. N., Base, J., Antonio, S., Antonio, S., & Wood, J. M. (2019). Methodological Advances in Statistical Prediction. *Psychological Assessment*, *March 11*, *2019*, 1–12.

Lenhard, F., Sauer, S., Andersson, E., Månsson, K. N. T., Mataix-Cols, D., Rück, C., & Serlachius, E. (2018). Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *International Journal of Methods in Psychiatric Research*, *27*(1), 1–12. https://doi.org/10.1002/mpr.1576

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, *88*(1), 25–38. https://doi.org/10.1037/ccp0000451

Wolff, J., Gary, A., Jung, D., Normann, C., Kaier, K., Binder, H., Domschke, K., Klimke, A., & Franz, M. (2020). Predicting patient outcomes in psychiatric hospitals with routine data: A machine learning approach. *BMC Medical Informatics and Decision Making*, *20*(1), 1–10. https://doi.org/10.1186/s12911-020-1042-2

Xu, Y., Ju, L., Tong, J., Zhou, C., & Yang, J. (2019). Supervised machine learning predictive analytics for triple-negative breast cancer death outcomes. *OncoTargets and Therapy*, *12*, 9059–9067. https://doi.org/10.2147/OTT.S223603

Youngstrom, E. A., Halverson, T. F., Youngstrom, J. K., Lindhiem, O., & Findling, R. L. (2018). Evidence-Based Assessment From Simple Clinical Judgments to Statistical Learning: Evaluating a Range of Options Using Pediatric Bipolar Disorder as a Diagnostic Challenge. *Clinical Psychological Science*, *6*(2), 243–265. https://doi.org/10.1177/2167702617741845