Key Visuals In Pre-Processing Phase: the Preliminary Story

237
Average LOS in Days

725
Number Successful

1237
Number of Clients

64%
% Trauma_HX

7/15/2016
Beginning Discharge Date

05/21/2020
Ending Discharge Date

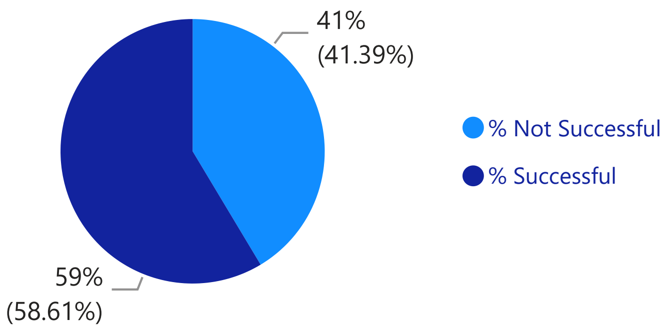
While the majority of clients had evidence of previous trauma at admission, the trauma feature did not contribute greatly to prediction of successful discharge in any of the Machine Learning Models.

Ethnicity
- [ ] Mexican/Mexican Ame...
- [ ] No Entry
- [ ] Not Hispanic
- [ ] Other Hispanic/Latino
- [ ] Unknown

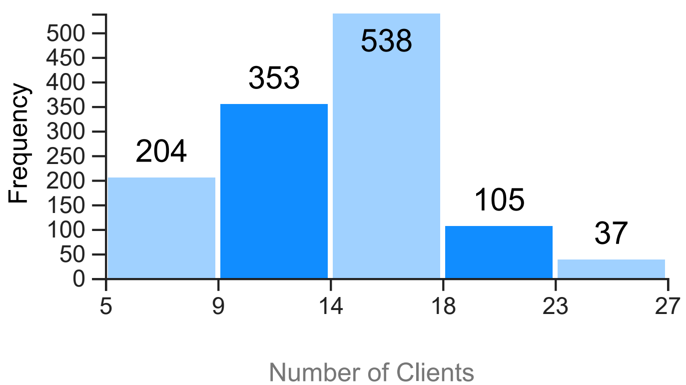
Original Program
All

Gender
- [ ] F
- [ ] M

Percent of Clients Successful vs, Not Successful Discharges



Distribution of Age at time of Discharge



Logistic Regression Analysis Final Model

Training Data Set N = 927
Test/Validation Data Set N =310

Use of Elastic Net Regularization in the Logistic Regression Model slightly improved classification

(Webb, et.al (2020) supports use of ENR)

Logistic Regression Final Model

Feature	Coefficient	P
_Constant (y-intercept)	-0.1072	0.75
Age at Discharge	-0.0368	0.06
Defiant oppositional behavior (CANS)	-0.2544	0.00
DSM 5 Oppositional defiant disorder	-0.6632	0.02
DSM5 ADHD	-0.7246	0.02
Family strength (CANS)	-0.2955	0.00
Length of Stay in Days	0.0044	0.00
Level of Care	0.2949	0.00
Recreational functioning (CANS score)	-0.2222	0.02
Social functioning (CANS)	-0.1095	0.22

$P < 0.05$ is statistically significant

Machine Learning Process:

1. Split into train (75%) and test (25%)- did not use defaults because of relatively small data set and needing sufficient training data.
2. Feature engineering based on previous research into factors that relate to successful discharge. Used many features and then reduced based on several models tested in sequence.
3. Model choice also based on research that pointed to logistic regression as the easiest to interpret for stakeholders as a benefit and applicable to binary outcomes.
4. Also used Random Forest, ensemble, ADABoost and Gradient Boosted Tree algorithms for comparison.
5. Future research will entail selection of other features with the goal of improving model fit and generalizability.

Elastic Net Regularization

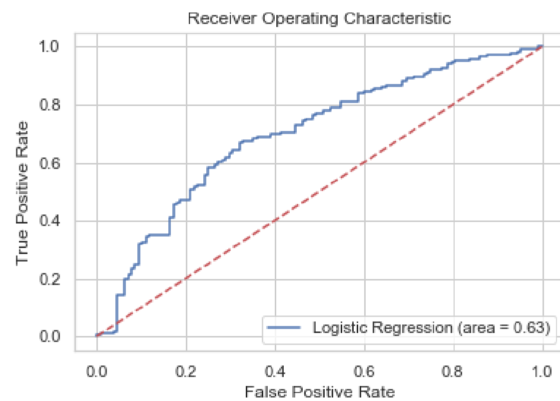
Confusion Matrix for Elastic Net Regularization Model

Category	Predicted Not Successful	Predicted Successful
Actual Not Discharged Successfully	50	78
Actual Discharged Successfully	21	161

Elastic Net Regularization Classification Report

#	Element	precision	recall	f1-score	support
1	Not Successful	0.70	0.39	0.50	128
2	Successful	0.67	0.88	0.76	182
3	accuracy			0.68	310
4	macro avg	0.69	0.64	0.63	310
5	weighted avg	0.69	0.68	0.66	310

Default Regularization



Random Forest, Ensemble, ADABoost and Gradient Boosted Tree Models

To be included next commit for Segment 3

Level of Care

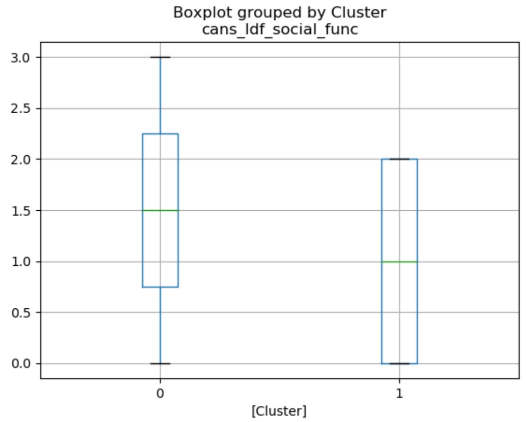
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4

Examination of Cluster Differences in CANS Scores

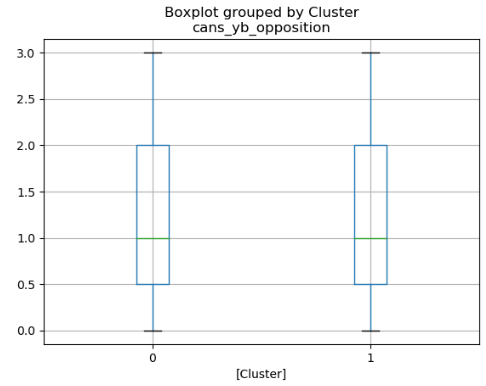
(Lower Score is Better in Terms of Behavioral Issues)

Delving into differences by cluster provides an understanding of the criteria by which clusters were derived. Higher risk clients with higher initial CANS scores tend to be in the first cluster (0) compared to somewhat better functioning clients in cluster (1). Such an understanding, along with more research, should help us to create a profile of clients likely to need more intervention along the trajectory of treatment.

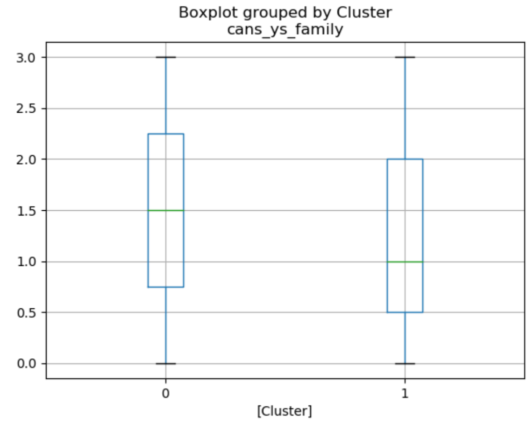
CANS Social Functioning by Cluster



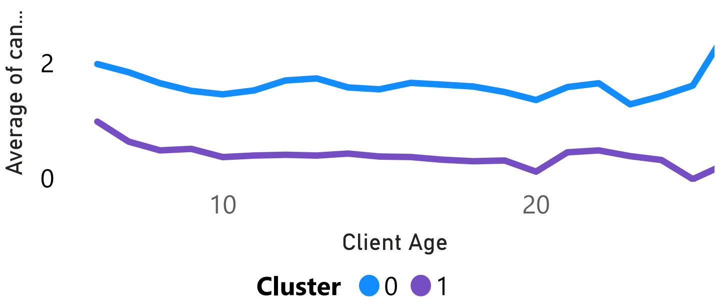
CANS Oppositional Behavior Scores by Cluster



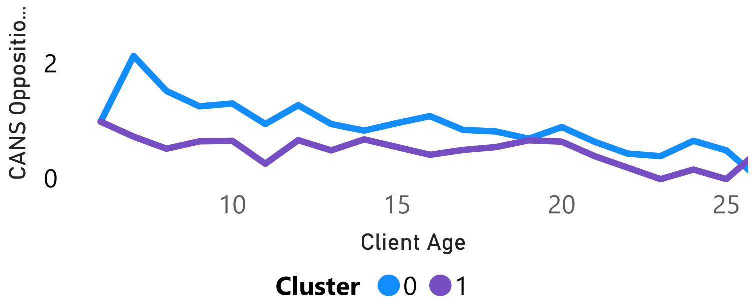
Child Family Strengths by Cluster



CANS Child Social Functioning Scores by Age and Cluster



CANS Oppositional Behavior Score by Client Age and Cluster



CANS Family Strengths by Client Age and Cluster