Early Psychiatric Rehospitalization Prediction

Anh L Tran

Georgia Institute of Technology

April 2021

Author Note

Anh L Tran, Georgia Institute of Technology.

This study was my applied analytics practicum (CSE 6748) project. I would like to thank Amy E. Chadwick, Mark Metzger, Dr. Kate McDonald, Dr. Emily V. Trask, Dr. Gregory Aarons and other team members at Child and Adolescent Services Research Center (CASRC) and Health Services Research Center (HSRC) for their guidance, support and feedback to me during my master program and this project.

Correspondence should be addressed to: Anh L Tran. Email: alt041@health.ucsd.edu.

Table of Contents

Page

Abstract……………………………………………………………………………………….. 3

Background and Objective……………………………………………………………………. 4

Methods…..…………………………………………………………………………………… 5

Key Elements…………………………………………………………………………….. 5

Data Source…………………………………………………………………….… 5

Outcomes and Predictor Variables…………………………………………….…. 6

Machine Learning Algorithms…………………………………………………... 7

Predictive Model Pipeline……………………………………………………………….. 9

Training and Test………………………………………………………………… 9

Missing Data Imputation……………………………………………………….... 9

Feature Engineering……………………………………………………………... 9

Variable Selection……………………………………………………………… 10

Modelling………………………………………………………………………. 11

Performance Evaluation Metrics……………………………………………….... 12

Results……………………………………………………………………………………….... 13

Children, Youth, and Families System of Care (CYF SOC) Findings………………………… 13

Exploratory Data Analysis……………………………………………………… 13

Machine Learning Model Results..…………………………………………….. 16

Adult and Older Adult System of Care (AOA) SOC Findings………………………………… 22

Exploratory Data Analysis……………………………………………………… 22

Machine Learning Model Results……………………………………………… 25

Discussion……………………………………………………………………………………. 30

Summary of Findings…………………………………………………………………… 30

Strengths and Limitations………………………………………………………………. 32

Conclusions………………………………………………………………………………….... 33

References……………………………………………………………………………………. 34

Appendices…………………………………………………………………………………… 36

ABSTRACT

Objective: Early psychiatric rehospitalization is disruptive to the clients' lives and put a strain on the mental health care system. With a large sample size data and information, we explored state-of-art machine-learning (ML) algorithms to predict 7, 30, 60 and 90-day readmissions as well as identifying influencing factors for both the youth (CYF) and the adult (AOA) systems of care in San Diego County.

Methods: Our 12-year dataset covered 118,893 psychiatric hospitalization records of 44,605 clients who were discharged from inpatient services. We analyzed each system of care separately. Seven ML algorithms (Naïve Bayes, K-Nearest Neighbors or KNN, Logistic Regression, Neuro Networks, Decision Tree, Random Forest, and Gradient Boosting) and three sets of variables (all predictors, predictors selected from the ML models, and predictors selected by the Boruta algorithm) were used to predict the readmission outcomes and evaluated.

Results: The CYF and AOA samples had similar results. Naïve Bayes and KNN had the lowest performances (accuracy and AUC < 0.65). Logistic Regression, Neuro Networks, and Decision Tree performed moderately (0.65-0.75 accuracy and 0.68 - 0.72 AUC). The best ML algorithms were Random Forest and Gradient Boosting (0.70-0.92 accuracy and 0.70 – 0.79 AUC). Among the set of predictor variables, Boruta-selected variables were preferred because there were much fewer variables (12-17 variables) used but the performances of the ML algorithms improved. We selected the Gradient Boosting with Boruta-selected variables as our final prediction model and tuned its parameters. We identified that length in the psychiatric hospital, history of psychiatric hospitalization, number of services before the index hospitalization, and some demographics variables were strong predictors for our model.

Conclusions: The Gradient Boosting model with Boruta-selected predictor variables yielded acceptable results. Although its performances was considered modest, there are many potentials to improve this model when combining with other information such as the outcome assessment data. The prediction model could contribute valuable information to help prevent and reduce psychiatric rehospitalization.

Keywords: Psychiatric Rehospitalization, Machine Learning, System of Care.

1. Background and Objective

Early psychiatric rehospitalization (EPR) occurs when a client is readmitted to a psychiatric inpatient service within 90 days from the previous inpatient psychiatric hospital discharge (Zhao et al. 2020). Because hospitalization is a preferred care modality for clients with severe mental illnesses or experiencing acute psychiatric crisis and costs more than outpatient services, being rehospitalized, specifically EPR, would cause many negative consequences such as being disruptive for clients and their families, increasing the risk of complication, and representing a strain on limited health care resources (Vigod et al., 2013). Hence, being able to identify influencing factors and predict the likelihood of EPR can help health care providers to provide early intervention and appropriate post-discharge cares for the high-risk clients in order to reduce rehospitalization rate.

Previous studies have identified some associations between psychiatric rehospitalization and pre-discharge client-level characteristics such as age, comorbidities, diagnostics, length of stay in the hospital, clinical history, type of service history and many other factors (Yu et al. 2015, Hung et al. 2017, Donisi et al. 2016, Zhao et al. 2020). Specifically, previous admissions and the amount of previous psychiatric service (Hamilton et al. 2016, Hung et al. 2017), diagnosis of schizophrenia, bipolar disorder, or depression (Hamilton et al. 2016, Hung et al. 2017), homeless at admission and discharge (Laliberté et al. 2019), living alone (Hung et al. 2017, Webber et al. 2004, Zhao et al. 2020), and having substance use disorders (Morel et al. 2020) are reliable predictors of readmission.

With the abundance of administrative records of inpatient clients and high dimensionality of psychiatric data, many recent studies have applied machine learning (ML) approach to detect complex patterns among risk factors and to develop hospitalization/rehospitalization prediction models. For example, Zhao et al. 2020 used random forest, a tree-based classification algorithm, to identify top risk factors for EPR. Blankers et al. 2020 applied ten machine learning algorithms, including the traditional statistical technique - generalized linear model (GLM/logistic regression) to predict psychiatric hospitalization. Morel et al. 2020 used the ensemble model, a technique that combines individual models to boost the performance, such as extreme gradient boosting to predict psychiatric readmission.

Despite the considerable number of studies on readmission prediction, most EPR predictive models are designed for 90-day psychiatric rehospitalization and focus on the adult population. It is essential if we could predict much earlier than 90 days and identify the unique factors contributing to the likelihood of the readmission for both the children/ teenager and the adult populations.

In San Diego County, California, the Behavioral Health Services Department oversees mental health services via two main systems of care: the Children, Youth and Families System of Care (CYF SOC) and the Adults and Older Adults System of Care (AOA SOC). The CYF SOC serves about 15,000 clients annually and 3-4% of them received psychiatric inpatient services. The AOA SOC serves about 42,000 clients annually and 7-10% of them were hospitalized in a psychiatric hospital (County of San Diego Behavioral Health Services). By identifying risk/ protective factors and predicting the EPR likelihood, the County could provide better mental health services to the clients.

In this study, we build predictive models to predict 90, 60, 30 and 7-day psychiatric rehospitalization and identify influencing factors for both CYF and AOA systems.

2. Methods

2.1 Key Elements

2.1.A. Client data source

The County of San Diego Behavioral Health Services uses the Cerner Community Behavioral Health (CCBH) platform to collect client data. CCBH is an electronic health record system that tracks demographics, diagnosis, and episodes in different levels of care including acute care hospitalization, emergency screening unit, outpatient and jail mental health services. CCBH data are routinely downloaded, imported and stored in our Microsoft SQL Server.

In our study, we extracted 12 years (07/01/2008 to 06/30/2020) of inpatient service data. The dataset is organized by the inpatient record level. Each row in the dataset includes a hospital stay, also known as the “index hospitalization”, and the subsequent hospital stay called the “readmission.” Each client can have one or multiple indexes of hospitalization. Furthermore, we wrote SQL queries to join multiple tables (i.e. client table, subunit table, level of care table, service table and assignment table) to construct a main dataset that includes client demographics, all services received before the index hospitalization, level of care, and the 7, 30, 60 and 90-day readmissions.

The whole original dataset consists of 118,907 distinct psychiatric hospitalization records of 44,611 unique clients. For the purpose of analysis, we only included cases that have date admitted to psychiatric hospital since 07/01/2008 and have an inpatient discharge. The final study population covered 118,893 psychiatric hospitalization records of 44,605 clients. The dataset was split into two subdatasets (i.e. CYF and AOA) and were analyzed separately.

2.1.B. Outcomes and predictor variables

The primary outcome of interest was a dichotomized measure of client psychiatric rehospitalization within 7, 30, 60 or 90 days after being discharged from an inpatient service (the index hospitalization). Each of these timeframe outcomes were analyzed separately. Candidate predictor variables include the client demographics (which are slightly different between CYF and AOA), length and history of psychiatric hospitalization, number of total services received, number of each group of services received 7/30/60/90 days before the index hospitalization, the first five level of care (LOC) services received before the index hospitalization and discharge status. The candidate predictors were selected based on the prior knowledge and availability of information, in which each variable should have less than 30% missing data. Table 1 lists all variables, and table 2 lists the service groups and their LOCs.

Table 1. Predictor variables by themes and population systems

Themes	Candidate predictor variables	
	CYF	AOA
Demographics	Age (num), Age group (cat), Gender (cat), Race (cat), Primary language (cat), Psychiatric diagnosis (cat), Living situation (cat), Insurance status (cat), substance abuse (cat), co-occurring substance use (cat)	
	Receive Child Welfare Services/ CWS (cat), Receive Alcohol and Drug Services/ADS (cat), Probation or Juvenile Justice/JJ Involvement (cat), In Special Education (cat)	Employment status (cat), Education status (cat), Sexual orientation (cat) , Jail/Justice Involvement (cat)
Length and history of psychiatric hospitalization	First time being hospitalized (cat) Days in the psychiatric hospital in last 12 months (num) Number of psychiatric hospitalizations in lifetime (num) Type of psychiatric hospitalization (cat)	
Number of total services	Number of all services up to the index hospitalization (num) Number of services before the index hospitalization since previous hospitalization (num) Number of services before the index hospitalization since previous hospitalization at same subunit (num)	
Service group received up to 7/30/60/90 days before the index hospitalization	Number of Outpatient (OP) Services (num) Number of Emergency/Crisis Services (num) Number of Inpatient (IP) Services (num)	
	Days in Day (DT) Services (num)	Days in 24-hour Services (num)
Order of LOC services	1st LOC service received before the index hospitalization (cat) 2nd LOC service received before the index hospitalization (cat) 3rd LOC service received before the index hospitalization (cat) 4th LOC service received before the index hospitalization (cat) 5th LOC service received before the index hospitalization (cat)	
Discharge	Discharge Status (cat)	
Total Variables	185	258

Table 2. Service groups and LOCs

Population	Outpatient Services	Emergency Services	Inpatient Services	24-hour Services	Day Services
CYF	Outpatient (OP), OP-Fee for Service (OP-FFS), OP-Residential (OP-R), Juvenile Forensic Service (JFS), Wraparound (WRAP), Therapeutic Behavioral Services (TBS)	Crisis Stabilization (CS), Urgent Outpatient (UO), Emergency Screening Unit (ESU), Crisis Outpatient (CO)	Inpatient – CAPS (IP-CAPS), Inpatient – Fee for Service (IP-FFS), Psych Health Facility (PHF)	No services	Day Treatment-Community (DT-C), DT-Residential (DT-R), DT-Closed Treatment Facility (DT-CTF)
AOA	ACT, Case Management (CM), CM-Institutional (CMI), CM-Strengths (CMS), CM-Transitional (CM-TRAN), Fee for Service (FFS), OP, OP-Low Income Health Program (OP-LIHP), Prevention (PREV), Eating Disorder Program (ED), Jail	Crisis Stabilization (CS), Urgent Outpatient (UO), PERT, Crisis Residential (CR), Early Psychosis Unit (EPU)	IP-County (IP-CNTY), IP-FFS, State Hospital (IP-SH), IP-LIHP	Edgemoor, Long Term Care (LTC), LTC-Institutional (LTC-I), LTC-Residential (LTC-R), Residential	No services

2.1.C. Machine learning algorithms

The seven statistical/ML algorithms used in this project were Logistic Regression, Naïve Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, Gradient Boosting and Artificial Neural Network. These represent the most commonly used types of algorithms for classification problems, which are applicable to the outcomes of interest. All algorithms had implementations in Python 3.8 and the sklearn library.

Logistic Regression, also known as logit regression or logit model, is a statistical model used to estimate the probability of a binary (0-1) outcome. It uses the Sigmoid function, which is an S-shaped curve that takes input data and maps it into a value between 0 (event does not occur) and 1 (event occurs). This model is also considered a type of Generalized Linear Models (Nelder 1972). For our Logistic Regression model, “liblinear” solver was used to handle L1 penalty, which limits size of the coefficients.

Naïve Bayes is a technique used to classify cases into labels (in our case rehospitalization Yes or No). The model is based on Bayes’ Theorem with an assumption of independence among predictors. For example, an animal may be considered to be a bird if it has two wings, two legs, a beak, feathers and lays eggs. Although these features depend on each other, each of them independently contributes to the probability that the animal is a bird and that is why it is known as ‘Naive’ (Hand 2001). The default Naïve Bayes model was used.

K-nearest neighbors (KNN) is an algorithm that estimates how likely a new case is to be a member of one group or the other depending on the majority of k (k>1) closest matching neighbor data. KNN determines the closest matching neighbor by a similarity measure such as the Euclidean distance [formula = $(\sqrt{((X1-X2)^2+(Y1-Y2)^2)}$] between two data points (Altman 1992). We used k=3 in our KNN model.

Decision Tree, also known as Classification and Regression Tree, is a robust algorithm that repeatedly partitions the data into a number of smaller subgroups (called leaf nodes) with similar response values based on a set of decision rules (called decision nodes) like an upside-down tree. The model tries to minimize cross-entropy or Gini index, a measure of purity to ensure the subset is as pure or homogeneous as possible, which is also where the branches stop splitting (Song 2015). Because our outcome variable is categorical, the model predicts the class label (i.e. rehospitalization Yes of No) from the class that has majority presentation in the subgroup. We used the default Decision Tree model.

Random Forest is an ensemble learning method that produces many decision trees using the training data as an input. These tree are made by randomly selecting a number of predictors and a number of rows from the original dataset. Each individual tree in the random forest outputs a class prediction and the class with the most votes becomes the model’s prediction. Random forest is considered more reliable and stable than just a decision tree (Biau 2016). We used the default 100 trees for the Random Forest model.

Gradient Boosting is a ML technique that produces a strong prediction model in the form of an ensemble of weak prediction models, often in the form of decision trees. Gradient Boosting trains many models in a gradual, additive and sequential manner in order to optimize a loss function, a measure indicating how good model’s coefficients are at fitting the underlying data (Natekin 2013). In our case, the loss function is a measure of how good our predictive model is at classifying rehospitalization. We also used the default 100 trees for the Gradient Boosting model.

Artificial Neural Networks, or simply called Neural Networks, is a ML model inspired by the biological neural networks of the brain. It is modelled to learn tasks based on provided examples, without being explicitly programmed. A neural network can have any number of layers with any number of neurons in those layers. The inputs are fed forward through the neurons in the network to get the output(s) at the end. During the transmission of the "signal" through the layers of neurons, there are weights and thresholds to moderate the strength of the signal at the connection to improve the prediction accuracy (van Gerven 2018). We used 100 neurons in 50 hidden layers with the hyperbolic tan function for our Neural Networks model.

2.2. *Predictive Model Pipeline*

2.2.A. Training and Test

For each population (CYF/AOA) x each outcome (7/30/60/90-day readmissions), we applied stratified sampling technique. The dataset was split at 80% as training dataset and the remaining 20% as test dataset with respect to the outcome variable. This technique ensures that the training and test sets have approximately the same percentage of samples of each outcome class as the complete set.

2.2.B. Missing Data Imputation

Missing data might cause some issues and affect many machine learning algorithms. Thus we had excluded predictors with 30% or more missing data at the beginning. For the candidate predictors, only a few demographics (i.e. primary language, insurance status, living situation, employment status, education and substance use) predictors have missing data less than 1%, except the diagnosis (about 2% for CYF and 5% for AOA) and sexual orientation (about 21%). For the CYF system, sexual orientation data are only collected for clients who are 13 years old or older; thus we did not include it in the CYF predictor set.

To impute the missing data, we applied the KNN algorithm on the training and test datasets. The number of k neighbors were determined by the formula: [take square root the number of predictors and then divided by 2]. Because the KNN only takes numerical values, each label in each categorical variable were encoded as a number based on the alphabetical order. After imputing, these numbers were converted back to their corresponding labels.

2.2.C. Feature Engineering

Categorical variables were one-hot encoded: each category in each categorical variable is represented by a dummy variable, where 1 indicates the category and 0 otherwise. Since one-hot-encoding directly induces multicollinearity, we dropped one of the columns from the encoded features.

Both numeric and categorical predictors were combined to create the input features. Because the ML models require data to be scaled and Gaussian distribution is preferred, we applied the Min-Max scaling technique to normalize the input data for both the training and test datasets. This technique shifts and rescale values so that they end up ranging between 0 and 1 (Patro 2015).

The Yes:No rehospitalization of each timeframe outcome is imbalanced; there are much fewer rehospitalized cases than the non-rehospitalized ones. Plus, the shorter the timeframe is, the wider the ratio gap is. This would impact the ML models because they would have poor performance on the minority class (i.e. Rehospitalization Yes), which is our outcome of interest. To resolve this problem, we applied a sampling technique called Synthetic Minority Oversampling Technique, or SMOTE. The algorithm is similar to the KNN. It selects samples in the minority class that are close to each other and then draws lines between them. New sample points are generated from the points on these lines (Chawla et al. 2002). Because we need to ensure the integrity of the test dataset for the evaluation purpose, only the training dataset was augmented.

2.2.D. Variable Selection

There are three sets of predictor variables used for the ML models: 1) all variables, 2) variables selected from the Correlation and Boruta algorithm and 3) variables selected from the feature importance produced by ML models.

Pearson correlation coefficients for numerical variables and Cramer's V (strength of association) coefficients for categorical variables were computed. Any feature with coefficient > 0.9 was considered redundant and excluded. After that, the Boruta variable selection was applied. The Boruta algorithm is a wrapper built around the random forest classification algorithm. It tries to capture all the important features in the dataset with respect to an outcome variable. The idea behind this algorithm is that the features compete with a randomized version of themselves, which are called shadow features created by duplicating the dataset and shuffling the values in each column. The importance of each original features is compared with a threshold, which is defined as the highest feature importance recorded among the shadow features. After a number of defined iterations, a feature is selected only if it’s capable of doing better than its best randomized version (Kursa 2010). The strongly and moderately predictive features were selected as the final second set for the ML models.

Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting an outcome after fitting the training data into the model. The higher the value the more important the feature. For the tree-based models (i.e. Decision Tree, Random Forest and Gradient Boosting), feature importance score is calculated as the reduction in node impurity weighted by the probability of reaching that node. That is, features that tend to split nodes closer to the root of a tree will result in a larger importance value. The feature importance score can be retrieved easily from the feature_importances_ property of the trained tree-based models. The third set of variables were selected from the top 15 of the average feature importance scores of the three tree-based models. This method follows a two-step Big Data analytic approach (Zhao and Castellanos, 2016).

2.2.E. Modelling

The ML algorithms were first applied to training data to parameterize and fit the model. Predicted classifications were then calculated for the test set. Three sets of selected features were used to fit the models, and their results were compared. The best set of features and the best model were then selected as the final model and we tuned the parameters to improve its performance. The tuning process used 5-fold cross-validation run for each parameter list. A grid search was used to identify the best hyper-parameters (e.g. the learning rate, the depth and the number of trees).

Figure 1. Predictive Model Pipeline – Feature Engineering

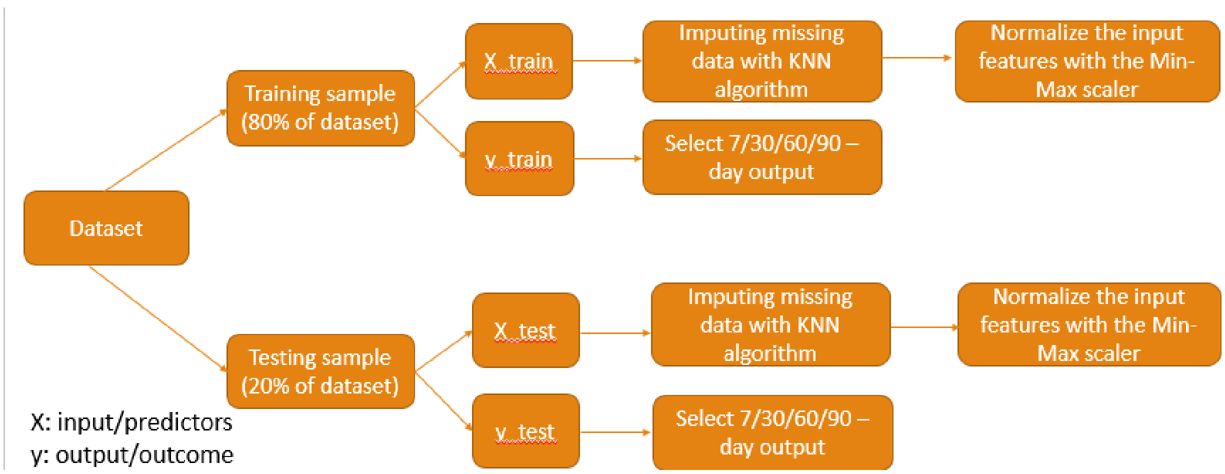
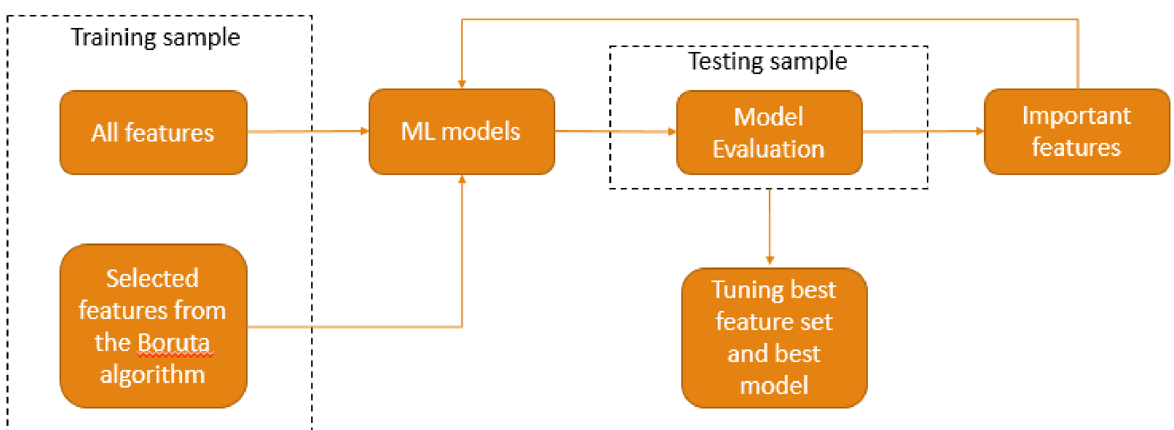


Figure 2. Predictive Model Pipeline – Modelling



2.2.F. Evaluation Metrics

There are seven main metrics to evaluate the performance of a ML model. They are: confusion matrix, accuracy, precision, recall, f1-score, precision-recall curve and AUC-ROC.

The confusion matrix is a straightforward table to view True and False Positive and Negative results.

The accuracy is the number of correct predictions over the total number of test dataset.

The precision is the ratio between the True Positives and all the Positives. It measures the model ability to not classify a case as positive, if it should be negative.

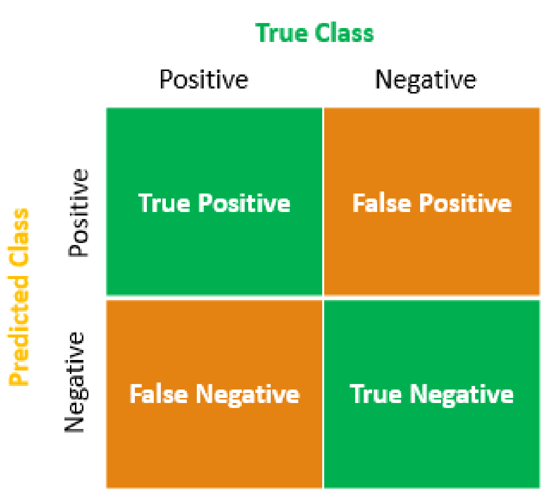
The recall is the measure of the model correctly identifying True Positives. It is the opposite of the precision.

The F1-score is the weighted harmonic mean of precision and recall (F1= 1 is the best).

The precision-recall curve is a metric for demonstrating the tradeoff between precision and recall for imbalanced datasets. The area under the precision-recall curve is the average precision score (AP = 1 is the best).

The AUC-ROC (area under the ROC curve), or AUC for short, is a performance measurement for the classification problems at various threshold settings. ROC (Receiver Operating Characteristics) is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Generally speaking, an AUC of 0.5 indicates that the model is no better than chance; an AUC of 0.7 to 0.8 indicates modest or acceptable discriminative ability, and a threshold greater than 0.8 indicates good discriminative ability (Hanley 1982).

Figure 3. Evaluation metrics and formulas



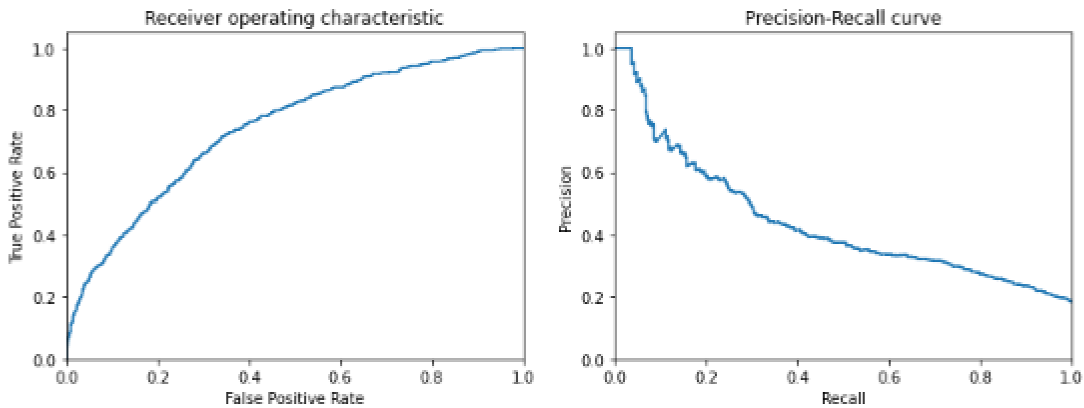
$$Accuracy = \frac{True\ Postive + True\ Negative}{Total\ (Positive + Negative)}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Figure 4. AUC-ROC and Precision-Recall Curve Examples



3. Results

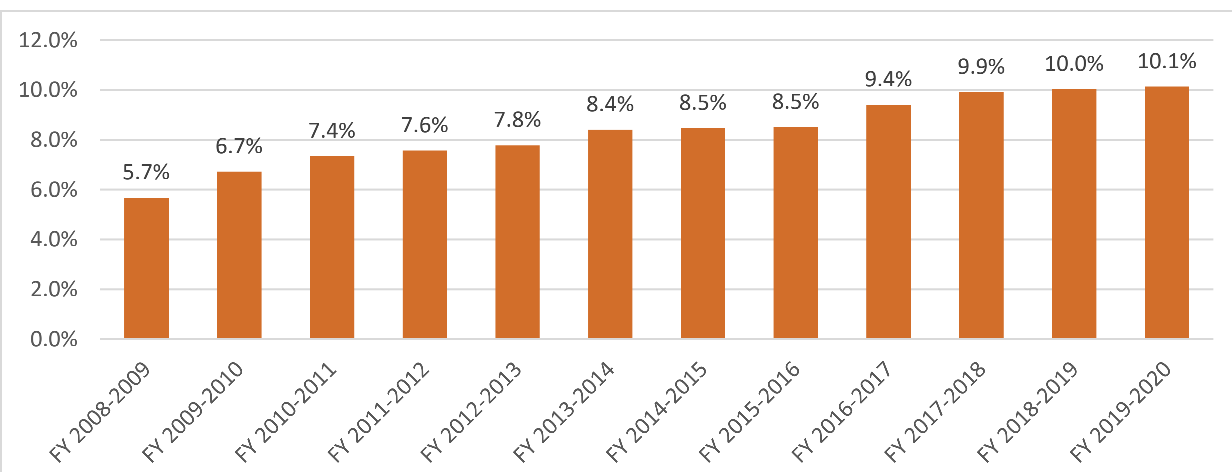
In this section, we present the exploratory data analysis summary and the machine learning model results for the CYF and AOA samples. Because the AUC is a standard measure of prediction accuracy that indicates discriminative ability**,** we used it along with the accuracy as the main metrics to evaluate the models. The full list of evaluation metric results are shown in the appendices.

3.1 Children, Youth and Family System of Care (CYF SOC) Results

3.1.A. Exploratory Data Analysis

The CYF sample includes 13,950 cases of 6,310 unique clients who have received at least one psychiatric IP service from FY 2008-2009 to FY 2019-2020 in San Diego County, California. The below figure 5 provides the distribution of the number of clients in psychiatric hospital by fiscal years.

Figure 5. Percent of clients received psychiatric IP services by fiscal years (N = 6,310)



The majority of CYF-sample clients receiving IP services are teenagers who have been diagnosed with depressive disorders. Most of the clients live with family, have English as their primary language and have Medi-Cal only. For reported gender and ethnicity, there are more female and Hispanic clients than other groups in the sample. About 1/6 of the sample used alcohol and drug or had substance abuse issues. A small portion of the sample received CWS, ADS and/or Probation/Juvenile Justice services. Almost 1/4 of the sample were in special education. About half of the sample were discharged with satisfactory outcome and achieved the discharge goals. Table 3 provides summary statistics of the study CYF sample.

Table 3. Summary statistics of the study CYF sample*

Variables (number of category)	Statistics
Age	Mean: 14.7 (Std : 2.35)
Age group (3)	16-17 years old: 44.8%
Gender (3)	Female: 12.8%
Race/Ethnicity (7)	Hispanic: 12.2%
Primary language (5)	English: 85.4%
Diagnosis (8)	Depressive disorders: 48.8%
Insurance status (4)	Medi-Cal only: 84.3%
Living situation (6)	House or Apartment: 82.0%
Substance abuse (2)	Yes: 13.8%
Co-occurring substance use (2)	Yes: 17.8%
Receive CWS (2)	Yes: 12.2%
Receive ADS (2)	Yes: 3.0%
Probation/JJ Involvement (2)	Yes: 4.7%
Receive Special Education (2)	Yes: 23.8%
Discharge status (16)	Satisfactorily achieved goals: 48.9%

**Full categories are displayed in the appendices.*

In average, the clients in the CYF sample received 100.3 services up to the index hospitalization. 71.5 services were received before the index hospitalization since previous hospitalization but only 0.2 services at the same subunit. There are four main groups of services that serve the CYF clients: Outpatient, Emergency, Inpatient and Day Services. Tables 4 and 5 summarize the group of services and the LOCs that CYF clients received before the index hospitalization.

Table 4. CYF group of services received within 90, 60, 30 and 7 days before the index hospitalization

Group of Services	Metrics	Within 90 days	Within 60 days	Within 30 days	Within 7 days
Outpatient Services	Total visits:	219,817	157,980	88,538	26,674
	Total clients:	257	278	371	622
	Percent user:	4.1	4.4	5.9	9.9
Emergency Services	Total visits:	56,466	48,457	38,954	29,094
	Total clients:	322	334	344	351
	Percent user:	5.1	5.3	5.5	5.6
Inpatient Services	Total visits:	4,108	2,966	1,672	387
	Total clients:	766	738	676	288
	Percent user:	12.1	11.7	10.7	4.6
Day Services	Total days:	67,343	45,527	22,445	4,988
	Total clients:	8	7	11	21
	Percent user:	0.1	0.1	0.2	0.3

Table 5. Top 10 common LOCs of each of the five CYF LOC events before the index hospitalization

5^{th} LOC	4^{th} LOC	3^{rd} LOC	2^{nd} LOC	1^{st} LOC
IPFFS	OP	OP	OP	OP
DTR	IPFFS	CS	CS	CS
OP	TBS	FFS	IP	FFS
WRAP	CS	IPFFS	IPFFS	IP
TBS	DTR	DTR	FFS	WRAP
PERT	WRAP	IPCAPS	WRAP	IPFFS
DTC	FFS	TBS	TBS	TBS
CS	OPR	WRAP	IPCAPS	DTR
FFS	PERT	IP	DTR	DTC
OPR	IPCAPS	OPR	OPR	OPR

The current CYF client sample has 1.3 psychiatric hospitalization (3.2 standard deviation) in average. The average length of stay in the psychiatric hospital is 6.6 days (18.4 days standard deviation). The gap between hospitalizations is 505.2 days or about 1.4 years. When we limited the gap to within a year, we found that the clients were readmitted to the psychiatric hospital in 60.8 days (86.4 days standard deviation) since a previous IP discharge. Figure 6 shows the distribution of readmission time in days; and figure 7 shows the percentages of 7, 30, 60 and 90-day readmission cases and clients.

Figure 6. Box plot of readmission time in days since IP discharge

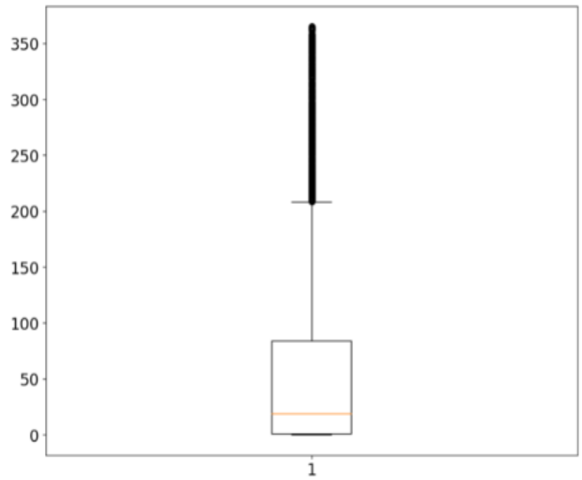
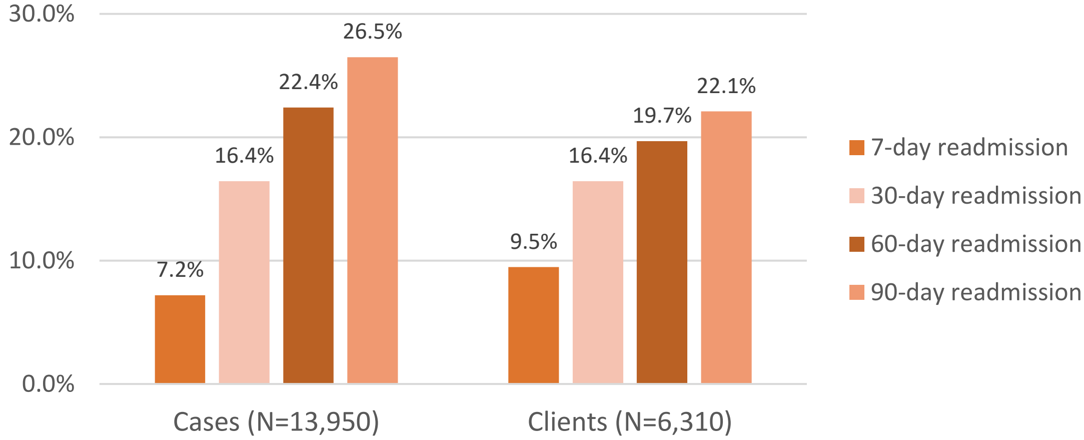


Figure 7. Percent of CYF cases and clients being readmitted to psychiatric hospital 7, 30, 60 and 90-days after being discharged



3.1.B. Machine Learning Model Results

All predictor variables

Figure 8 presents the accuracy for the models using all 185 predictor variables. Six out of seven models predicted the readmission in the shorter timeframe with higher accuracy than the longer timeframe. The Neural Networks and the tree-based models (i.e. Decision Tree, Random Forest and Gradient Boosting) had better performance (> 0.7 accuracy overall) compared to the statistical model Logistic Regression. The Naïve Bayes model had the lowest accuracy while the Gradient Boosting model performed the best across all timeframes (0.75 to 0.92 accuracy).

Figure 8. Model accuracy of ML algorithms by timeframes – all predictors - CYF

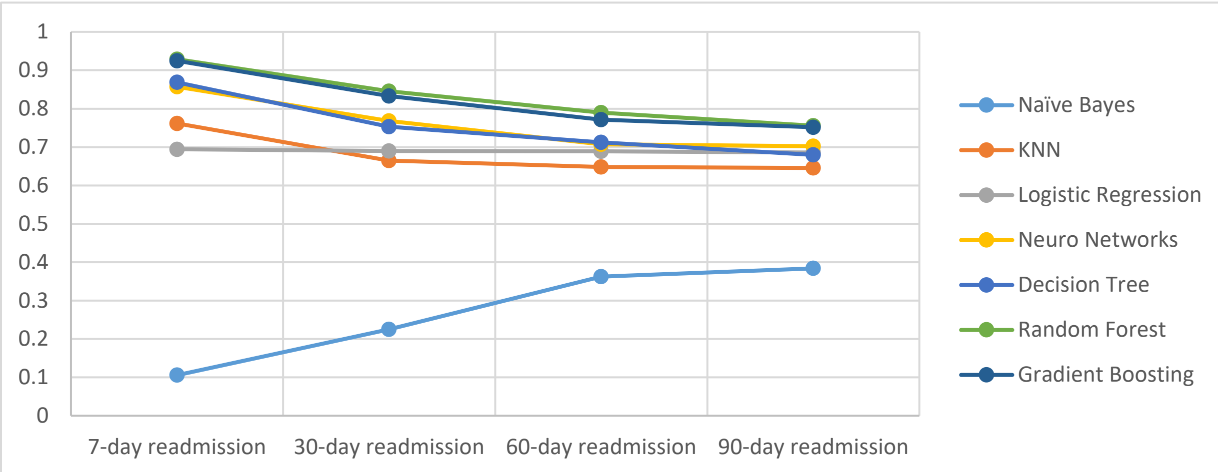
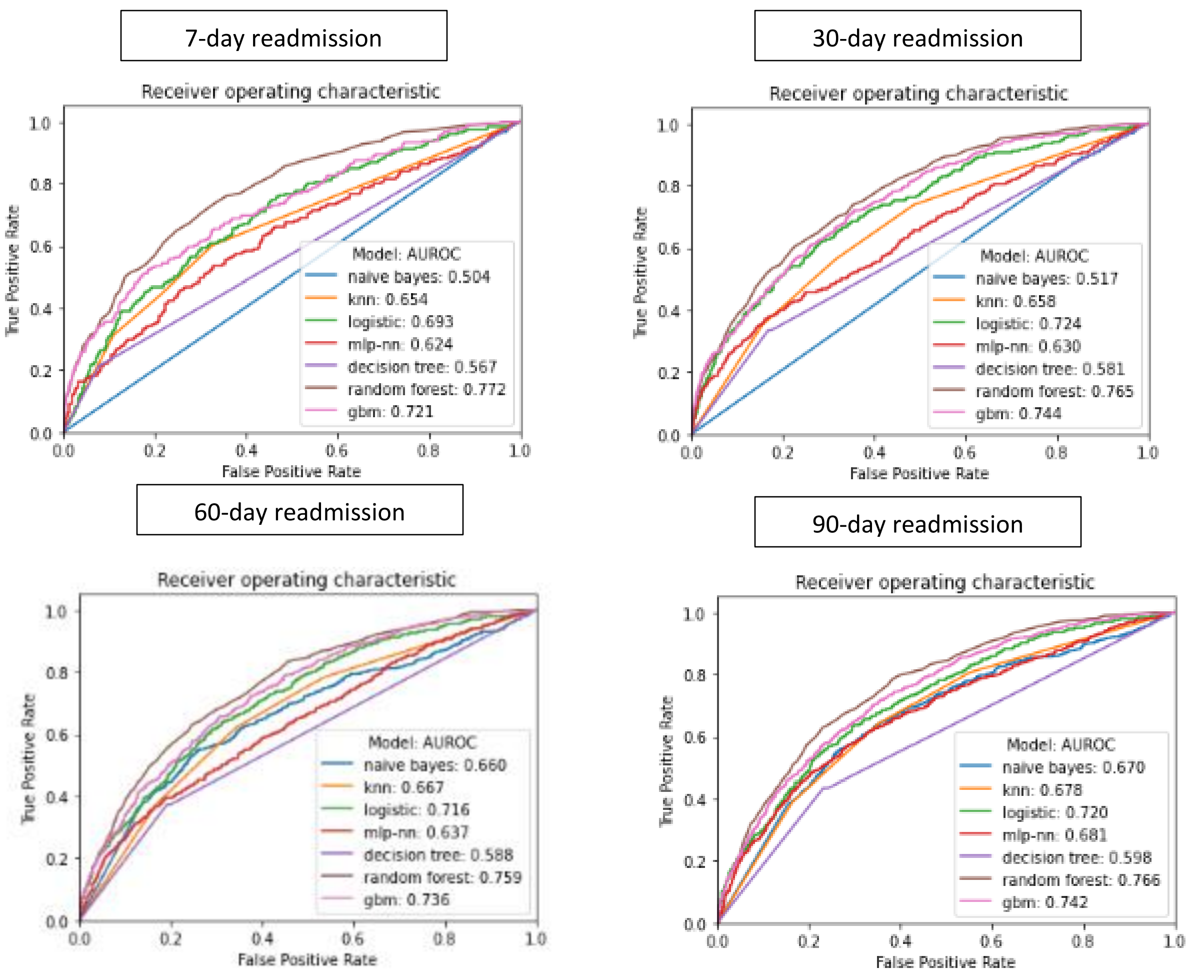


Figure 9 shows the AUC-ROC. The Random Forest and the Gradient Boosting models had acceptable discriminative ability scores (0.72-0.77) across four timeframes. The Logistic Regression model also performed moderately well (0.69-0.72).

Figure 9. Model AUC of ML algorithms by timeframes - all predictors - CYF



Variables selected from the ML models

The top 15 predictor variables were selected from the three tree-based models (i.e. Decision Tree, Random Forest and Gradient Boosting). They are OP services 7/30/60/90 days before the index hospitalization, IP services 30/60 days before the index hospitalization, ES services 90 days before the index hospitalization, number of services before the index hospitalization since previous hospitalization at same/ other subunit, number of all services up to the index hospitalization, IP-FFS, history of hospitalization, length of hospitalization, race and age. These predictor variables are common across the four timeframes.

Unexpectedly, the performance of all models got worse. The accuracy and AUC scores were lower than when using all predictors. Figures 10 and 11 present the accuracy and the AUC for the models.

Figure 10. Model accuracy of ML algorithms by timeframes – ML-selected predictors - CYF

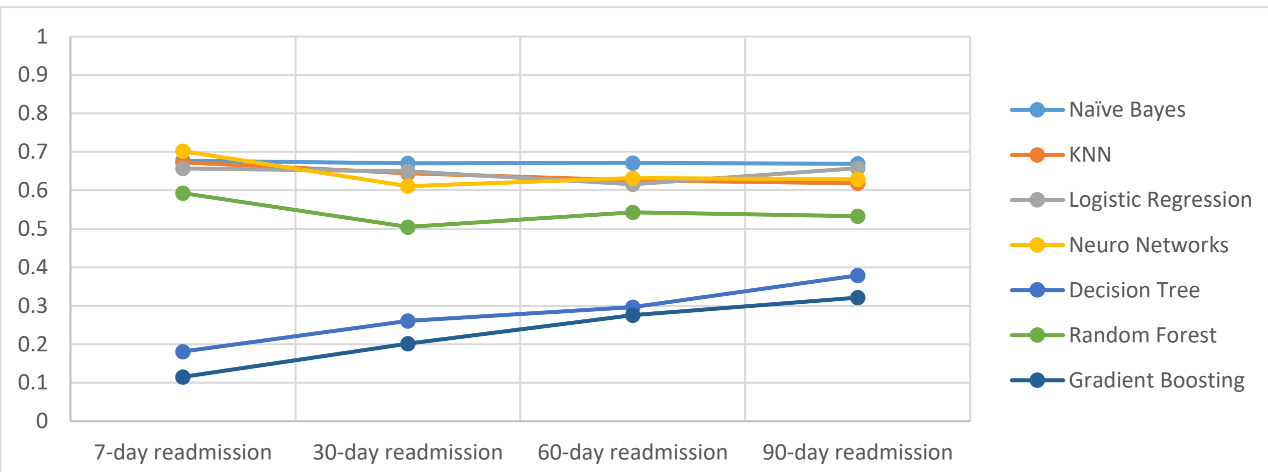
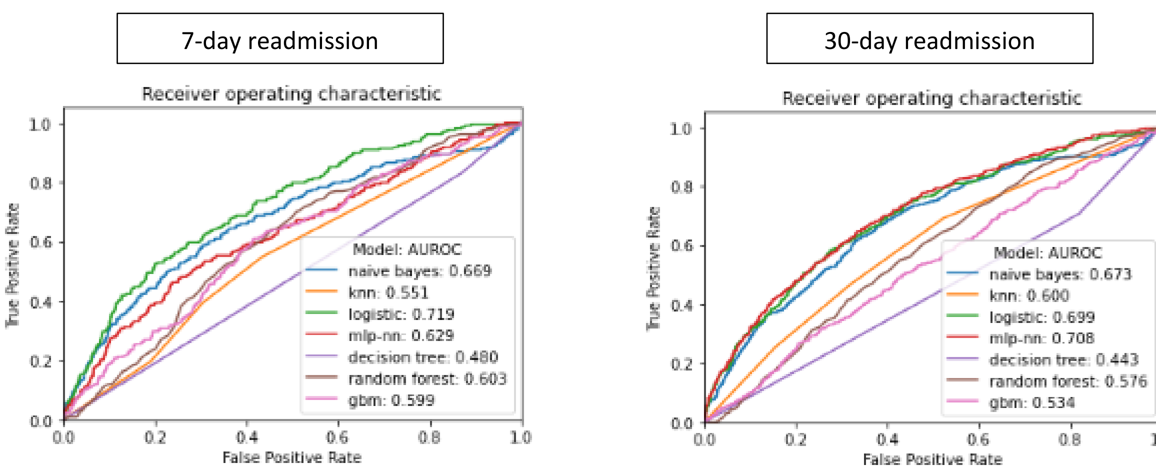
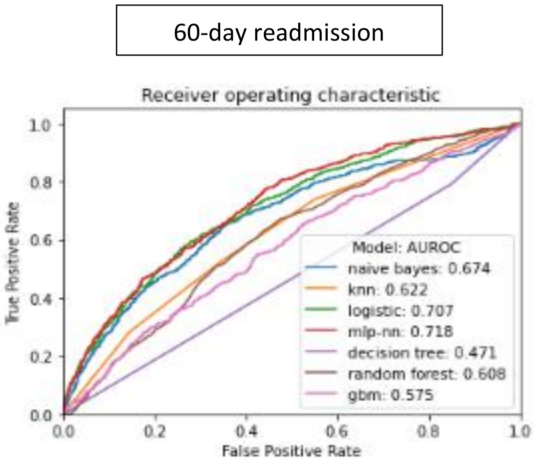
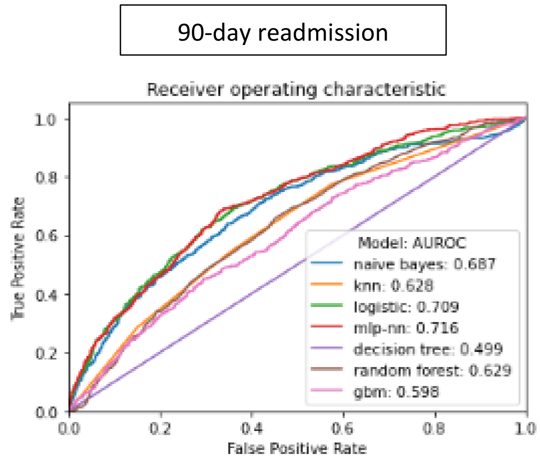


Figure 11. Model AUC of ML algorithms by timeframes – ML-selected predictors – CYF*







**knn: K-Nearest Neighbors, mlp-nn: Neuro Networks, gbm: Gradient Boosting*

Predictor variables selected from the Boruta algorithm

The Boruta algorithm selected the following predictors across the four timeframes: IP services 60/90 days before the index hospitalization, number of all services up to the index hospitalization, IP-CAPS, history of hospitalization, length of hospitalization, race and age.

There were increases in accuracy and AUC in the Naïve Bayes, KNN and Logistic Regression models across four timeframes. The performances of the Neuro Networks and Decision Tree were slightly lower than its version of all predictors. The Random Forest and Gradient Boosting models had similar performances (0.70 - 0.89 accuracy and 0.71-0.75 AUC) compared to their versions of all predictors. Figures 12 and 13 present the accuracy and the AUC for the models using the predictor variables selected from the Boruta algorithm.

Figure 12. Model accuracy of ML algorithms by timeframes – Boruta-selected predictors - CYF

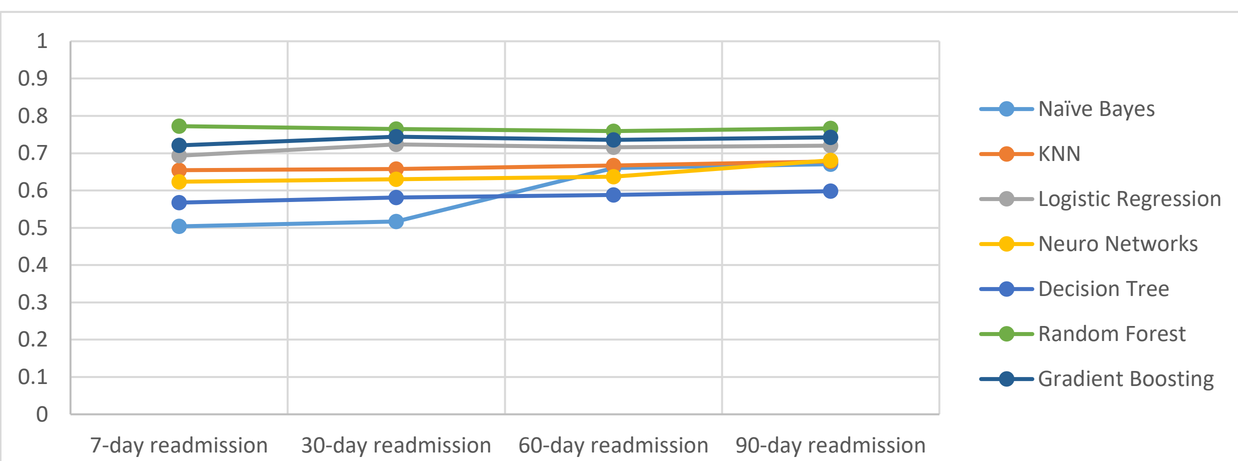
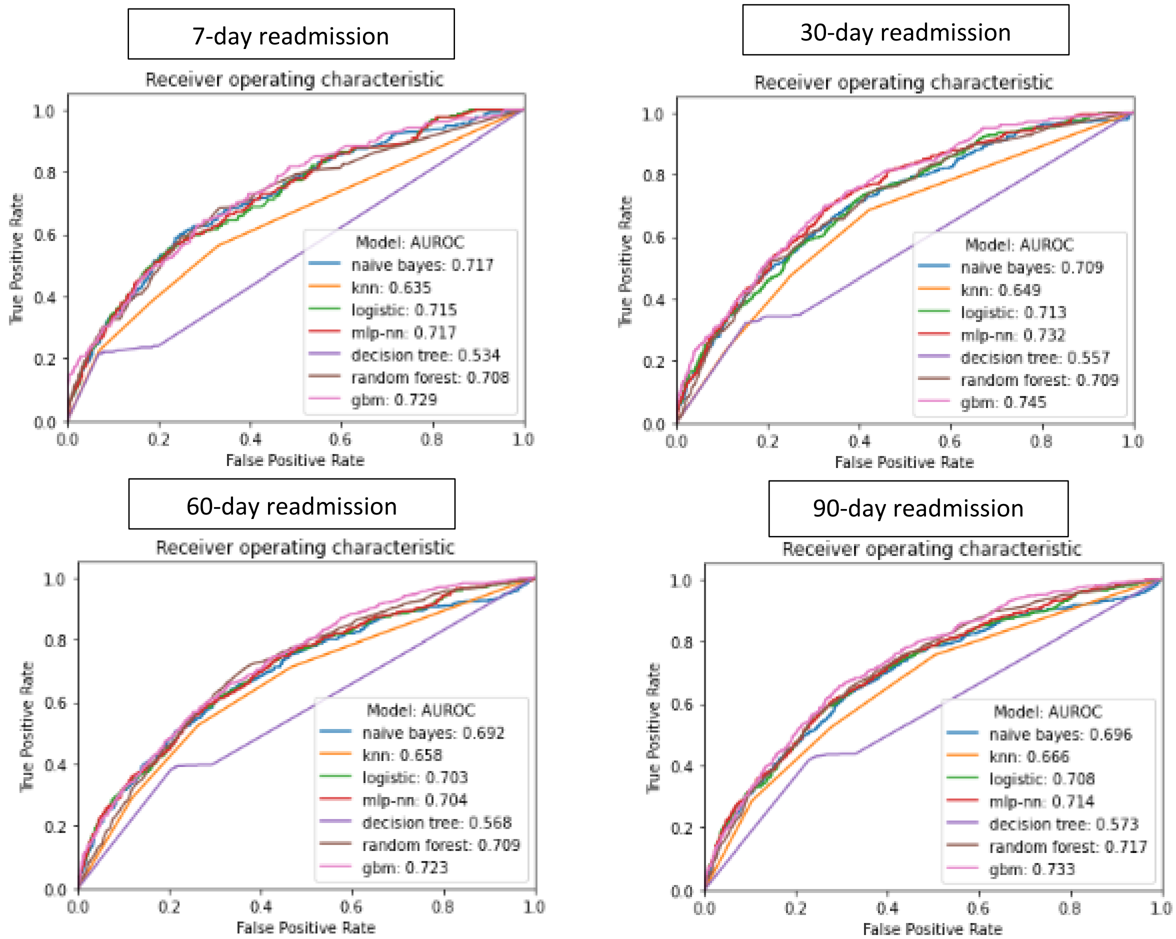


Figure 13. Model AUC of ML algorithms by timeframes – Boruta-selected predictors – CYF*



**knn: K-Nearest Neighbors, mlp-nn: Neuro Networks, gbm: Gradient Boosting*

Final set of variables and model

Among the models, Gradient Boosting consistently performed better than the other ML models across all timeframes. Thus, we chose it as our final model. Although using all 185 predictor variables with the Gradient Boosting model yielded the best accuracy and AUC, it consumed lots of computational power and took a long time to run (about 20-30 minutes). On the other hands, there were only 10 or fewer variables selected from the Boruta algorithms and the results were similar. The final model that we decided to pick and tune its parameters was the Gradient Boosting with Boruta-selected predictors.

There were three parameters that we tuned for the Gradient Boosting model: learning rate (moderate the contribution of each tree), n_estimators (the number of trees in the forest) and max_dept (how deep the built tree can be). Table 6 lists the best parameters, accuracy and AUC for each timeframe model. There were moderate improvements in the accuracy and AUC of this tuned final model compared to the previous models. The highest accuracy and AUC were 0.91 and 0.78, respectively, for 7-day readmission. The accuracy and AUC decreased as the readmission timeframe increased but the values were still larger than 0.72.

Table 6 Best parameters, accuracy and AUC of the final model - CYF

	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Learning rate	0.1	0.1	0.1	0.01
N_estimators	1500	1000	500	1500
Max_depth	7	7	3	3
Accuracy	0.9125	0.8401	0.7710	0.7441
AUC	0.7816	0.7426	0.7244	0.7335

Regarding the important predictor variables, length in the psychiatric hospital, history of psychiatric hospitalization, number of services up to the index hospitalization, and number of services before the index hospitalization since previous hospitalization are the strongest predictors for the model. Having IP (60 and 90 days) and ES (30 days) services before the index moderately contribute to the final model. IP-CAPS, IP-FFS, discharge status and demographics variables (i.e. age and race) also influence the model prediction. Table 7 shows the average important feature scores and coefficients of the Boruta predictor variables for all four readmission timeframes. The details of individual timeframes are listed in the appendices.

Table 7. Average important feature scores and coefficient of Boruta predictors across four timeframes - CYF

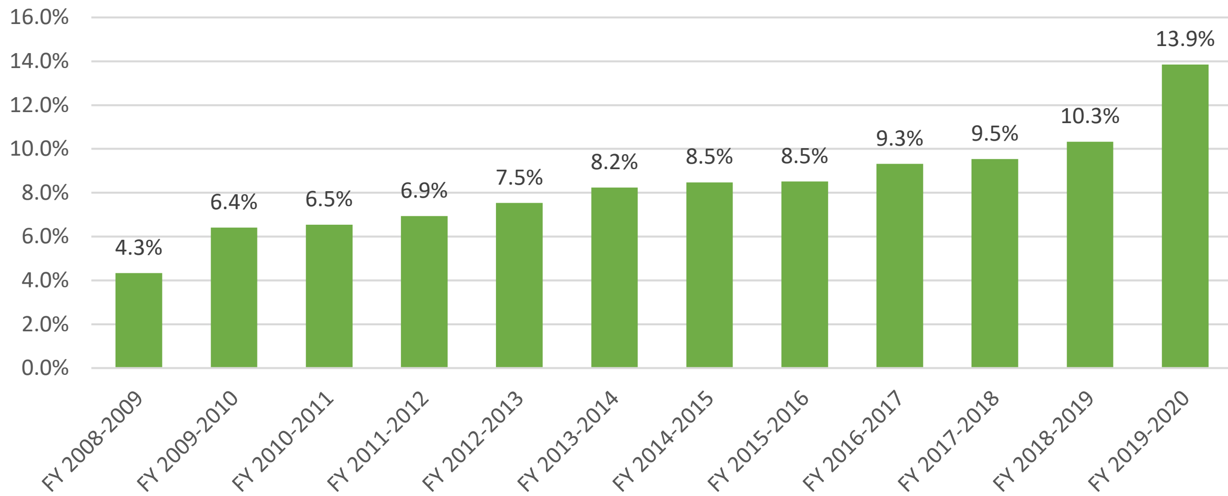
Boruta predictors	Important feature score	Coefficient
Length in hospital	0.41852	0.198852
History of hospitalization	0.205235	2.712986
Number of services up to index hospitalization	0.114051	1.24282
Number of services before index hospitalization since previous hospitalization	0.079563	1.749609
IP services 60 days before index hospitalization	0.032213	0.649447
IP services 90 days before index hospitalization	0.032076	0.211267
ES services 30 days before index hospitalization	0.072166	-0.98324
IP CAPS	0.017635	0.645254
IP FFS	0.023644	0.966863
Discharge status: other reasons	0.028372	0.821287
Race: unknown	0.056583	1.492222
Age	0.050025	1.102105

3.2 Adults and Older Adults System of Care (AOA SOC) Results

3.2.A. Exploratory Data Analysis

The AOA sample includes 104,920 cases of 38,272 unique clients who have received at least one psychiatric IP service from FY 2008-2009 to FY 2019-2020 in San Diego County, California. The below figure 14 provides the distribution of the number of clients in psychiatric hospital by fiscal year.

Figure 14. Percent of clients received psychiatric IP services by fiscal years (N = 38,272)



The majority of AOA clients receiving IP services are 26-59 years old, males, white, have English as the primary language and have been diagnosed with schizophrenia and other psychotic disorders. About 4% are identified as LGBTQ+. More than 1/4 of the sample completed high school diploma/GED. About 67% of the sample have Medi-Cal/ Medicare. 1/3 of the sample reported being not in labor force and more than half of the sample reported living independently. More than half of the sample used alcohol and drug and had substance abuse issues. Almost half of the sample were discharged to home or their shelters. Table 8 provides summary statistics of the study AOA sample.

Table 8. Summary statistics of the study AOA population*

Variables (number of category)	Statistics
Age	Mean: 38.3 (Std : 14.09)
Age group (3)	26-59 years old: 63.9%
Gender (3)	Male: 55.1%
Race/Ethnicity (7)	White: 46.4%
Primary language (5)	English: 83.8%
Education level (8)	High School Diploma/GED: 26.4%
Diagnosis (6)	Schizophrenia & Other Psychotic Disorders: 45.0%
Insurance status (4)	Medi-Cal and Medicare: 66.6%
Employment status (7)	Not in labor force: 30.3%
Living situation (6)	Living independently: 62.3%
Sexual orientation (3)	LGBTQ+: 4.4%
Substance abuse (2)	Yes: 60.3%
Co-occurring substance use (2)	Yes: 57.3%
Discharge status (24)	Discharged to Home/Shelter: 46.1%

**Full categories are displayed in the appendices.*

In average, the clients in the AOA sample received 55.6 services up to the index hospitalization. 32.6 services were received before the index hospitalization since previous hospitalization but only 0.3 services at the same subunit. There are four main groups of services that serve the AOA clients: Outpatient, Emergency, Inpatient and 24-hour Services. Tables 9 and 10 summarize the group of services and LOCs that AOA clients received before the index hospitalization.

Table 9. AOA group of services received within 90, 60, 30 and 7 days before the index hospitalization

Group of Services	Metrics	Within 90 days	Within 60 days	Within 30 days	Within 7 days
Outpatient Services	Total visits:	491,466	346,263	192,847	62,783
	Total clients:	2,076	2,289	2,676	3,219
	Percent user:	4.8	5.2	6.1	7.4
Emergency Services	Total visits:	312,409	258,965	193,694	123,911
	Total clients:	1,462	1,466	1,494	1,687
	Percent user:	3.3	3.4	3.4	3.9
Inpatient Services	Total visits:	62,068	45,394	24,902	4,125
	Total clients:	5,726	5,651	5,359	2,363
	Percent user:	13.1	12.9	12.3	5.4
24-hour Services	Total days:	15,481	7,514	3,309	1,179
	Total clients:	11	13	16	30
	Percent user:	<0.1	<0.1	<0.1	0.1

Table 10. Top 10 common LOCs of each of the five LOC events before the index hospitalization

5th LOC	4th LOC	3rd LOC	2nd LOC	1st LOC
IP-FFS	IP-FFS	IP-FFS	IP-FFS	IP-FFS
CS	CS	CS	CS	CS
CR	CR	JAIL	OP	UO
JAIL	JAIL	OP	UO	OP
OP	OP	IP-CNTY	CR	OP-FFS
IP-CNTY	IP-CNTY	CR	OP-FFS	ACT
OP-FFS	UO	UO	JAIL	JAIL
UO	OP-FFS	OP-FFS	IP-CNTY	CR
CM-TRAN	CM-TRAN	CM-TRAN	CM-TRAN	CM-TRAN
PERT	PERT	ACT	ACT	IPCNTY

The current AOA client sample has 1.8 psychiatric hospitalization (4.79 standard deviation) in average. The average length of stay in the psychiatric hospital is 11.3 days (38.8 days standard deviation). The gap between hospitalizations is 619.2 days or about 1.7 years. When we limited the gap to within a year, we found that the clients were readmitted to the psychiatric hospital in 71.1 days (88.0 days standard deviation) since a previous IP discharge. Figure 15 shows the distribution of readmission time in days; and figure 16 shows the percentages of 7, 30, 60 and 90-day readmission cases and clients.

Figure 15: Box plot of readmission time in days since IP discharge

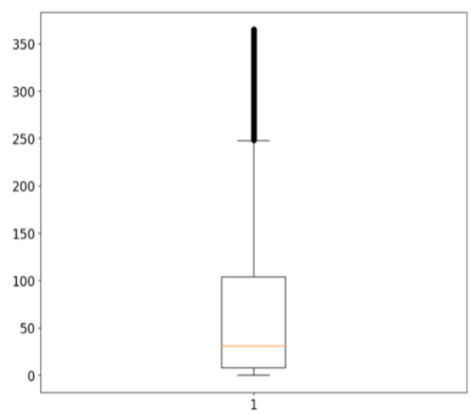
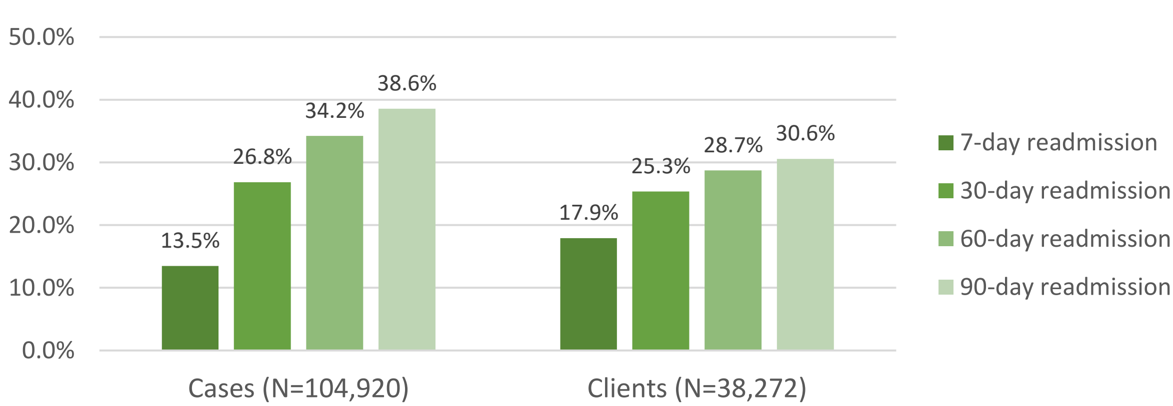


Figure 16. Percent of AOA cases and clients being readmitted to psychiatric hospital 7, 30, 60 and 90-days after being discharged



3.2.B. Machine Learning Model Results

All predictor variables

Figure 17 presents the accuracy for the models using all 258 predictor variables. Except the Naïve Bayes model with the lowest performance, all other models had the prediction accuracy over 60% and two models (i.e. Random Forest and Gradient Boosting) had 70% or more accuracy. These performances were consistent across the four readmission timeframes.

Regarding the AUC, four out of six models (i.e. Logistic Regression, Neural Networks, Random Forest and Gradient Boosting) had the performance over 0.70 in overall. Gradient Boosting had the best AUC score (0.77-0.79) compared to the other models.

Figure 17. Model accuracy of ML algorithms by timeframes – all predictors – AOA

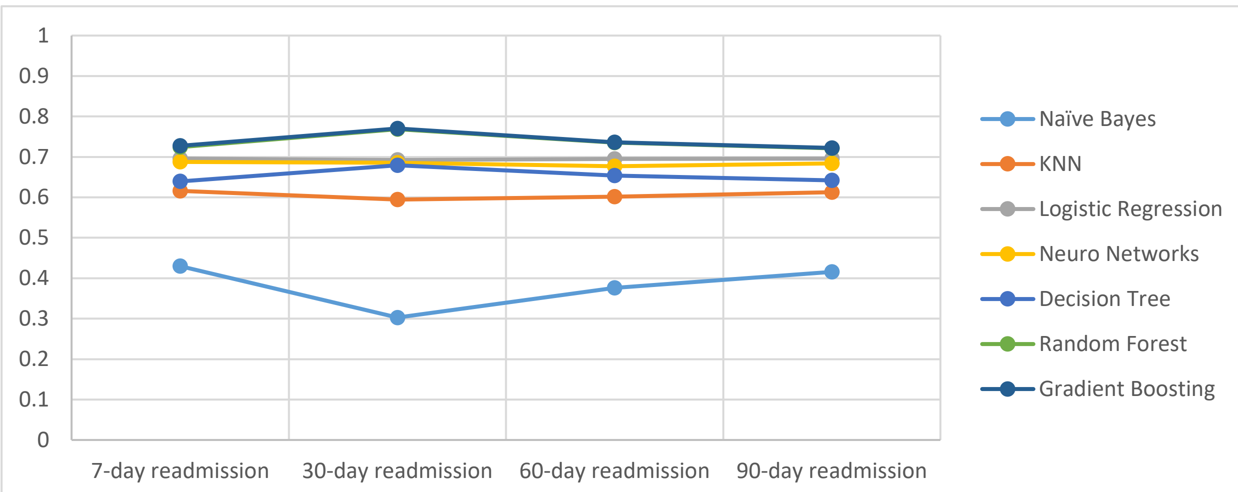
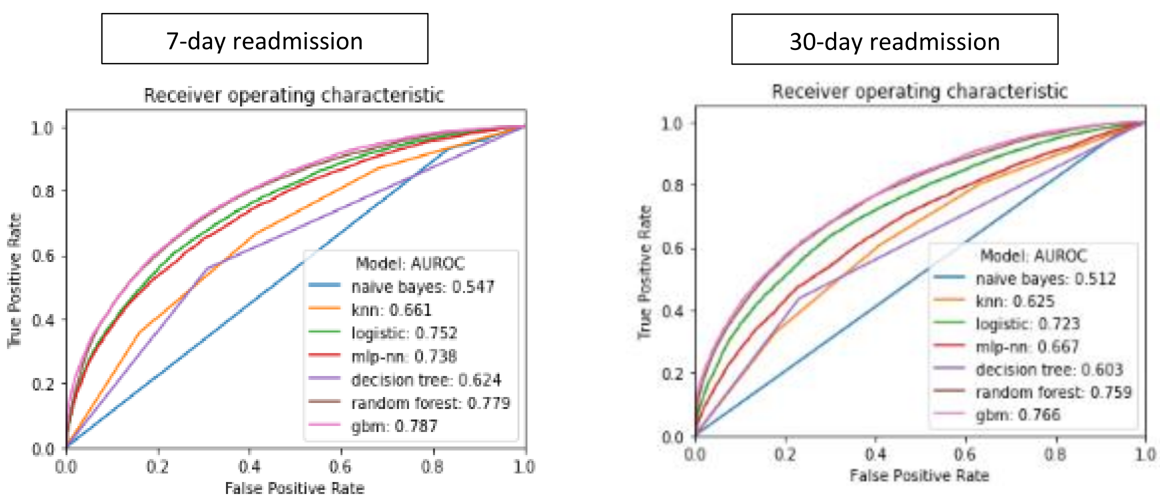
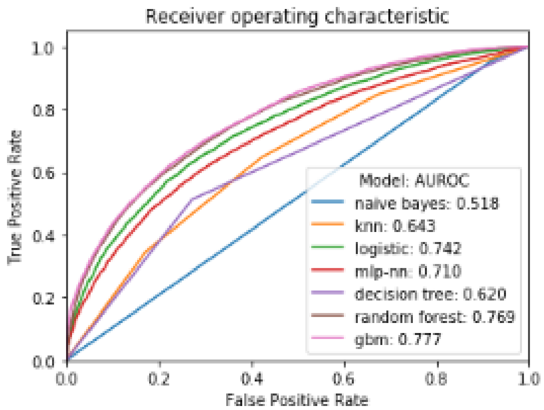


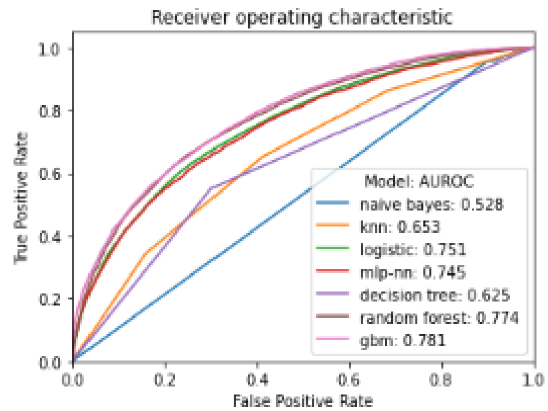
Figure 18. Model AUC of ML algorithms by timeframes – all predictors – AOA*



60-day readmission



90-day readmission



**knn: K-Nearest Neighbors, mlp-nn: Neuro Networks, gbm: Gradient Boosting*

Variables selected from the ML models

The top 15 predictor variables were selected from the three tree-based models (i.e. Decision Tree, Random Forest and Gradient Boosting). They are IP services 30/60/90 days before the index hospitalization, ES services 90 days before the index hospitalization, number of services before the index hospitalization since previous hospitalization at same/ other subunit, number of all services up to the index hospitalization, history of hospitalization, length of hospitalization, sexual orientation, discharge status, living situation, gender and age. These predictor variables are common across the four timeframes.

Interestingly, the accuracy performances of the tree-based models decreased significantly (below 0.60) while the Naïve Bayes increased (0.67-0.70) compared to when using all predictors. The accuracy of the Logistic Regression, KNN and Neuro Networks were similar to the previous their corresponding all-predictors models (0.61-070). The AUC of the models were also lower than the previous all-predictor models.

Figure 19. Model accuracy of ML algorithms by timeframes – ML-selected predictors – AOA

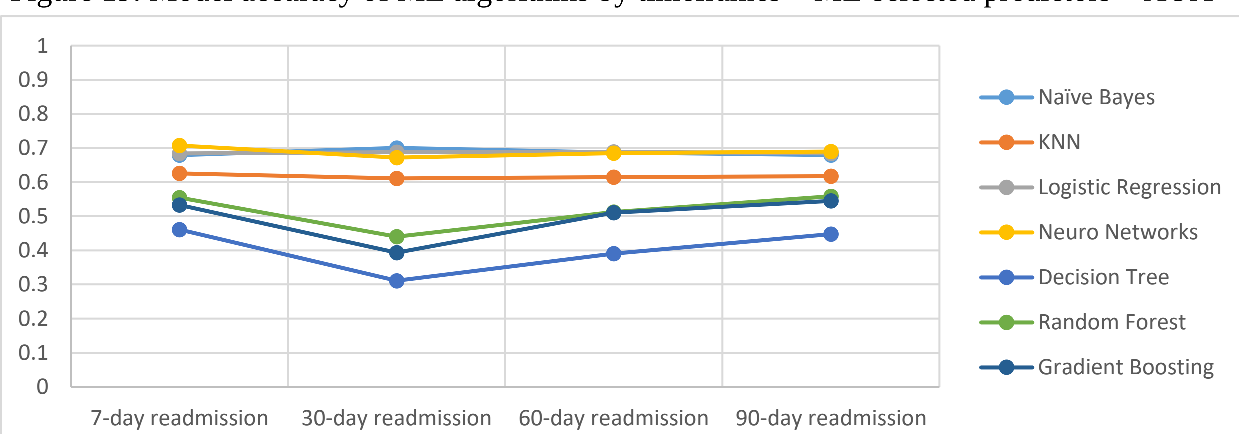
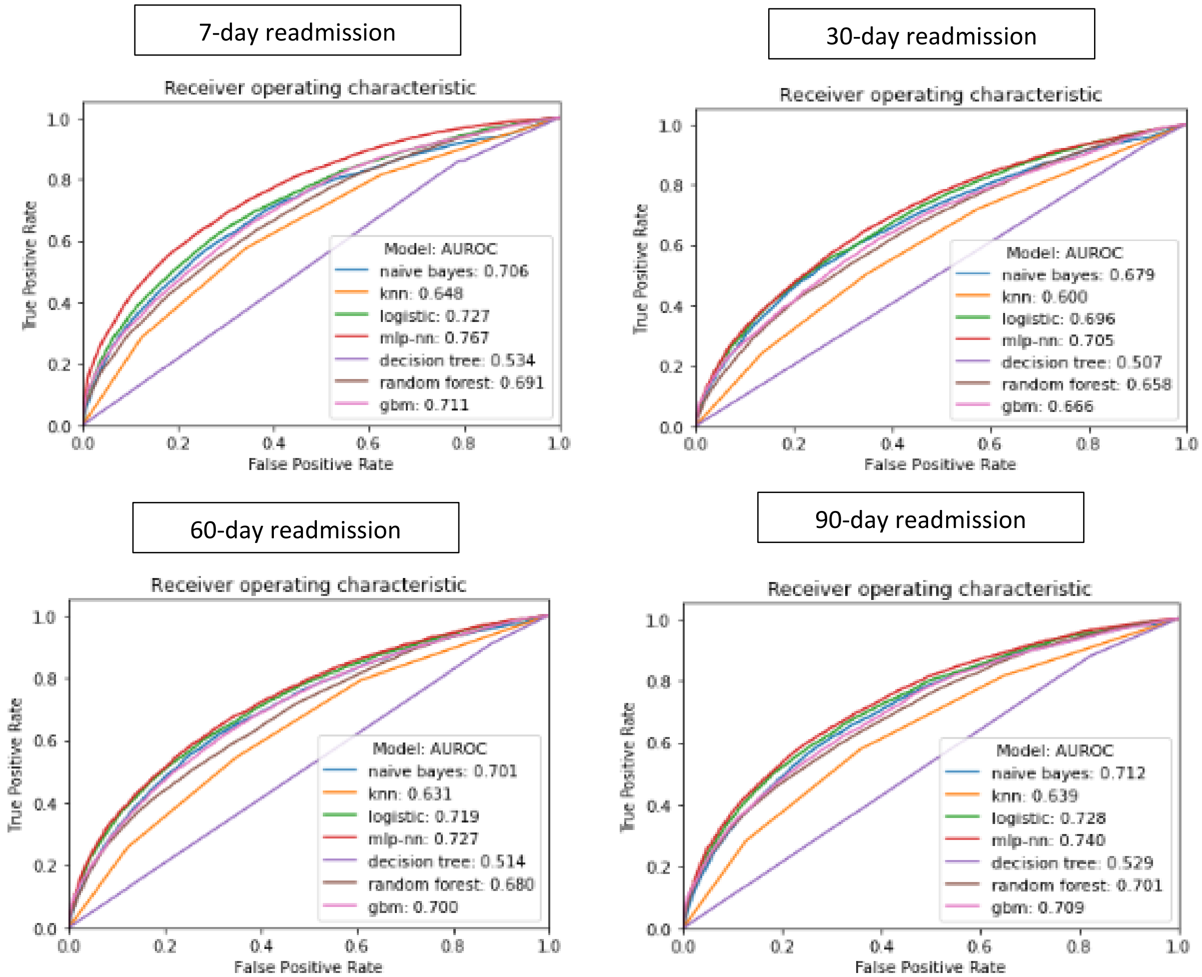


Figure 20. Model AUC of ML algorithms by timeframes – ML-selected predictors – AOA*



**knn: K-Nearest Neighbors, mlp-nn: Neuro Networks, gbm: Gradient Boosting*

Predictor variables selected from the Boruta algorithm

The Boruta algorithm selected the following predictors across the four timeframes are: IP services 30/60/90 days before the index hospitalization, ES services 60/90 days before the index hospitalization, number of all services up to the index hospitalization, history of hospitalization, length of hospitalization, living situation, diagnosis, race, gender and age.

All models had the accuracy higher than 0.60. Among the models, Gradient Boosting consistently performed the best (0.71 – 0.74 accuracy) across all readmission timeframes. Regarding the AUC scores, all other models had the AUC over 0.70 except KNN and Decision Tree. Notably, the Gradient Boosting had AUC ranging from 0.75-0.78.

Figure 21. Model accuracy of ML algorithms by timeframes – Boruta predictors – AOA

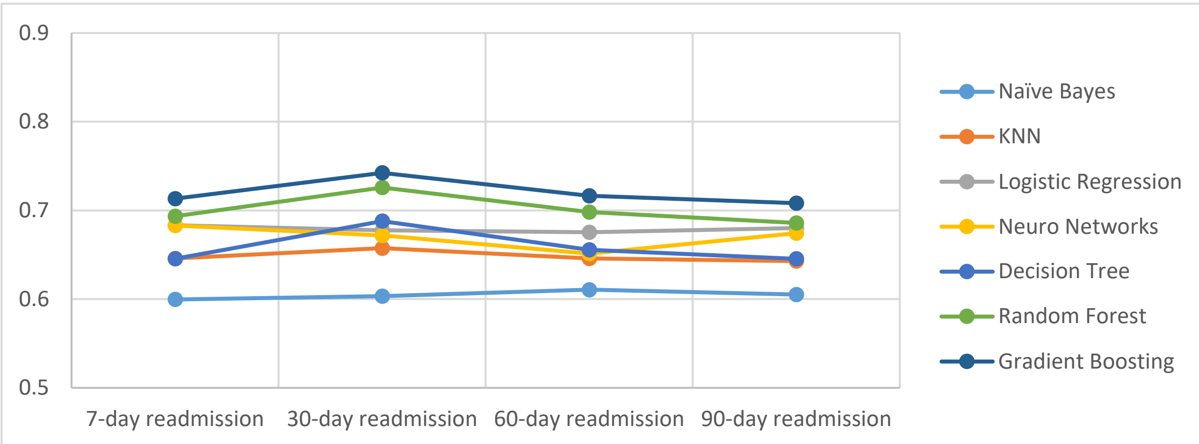
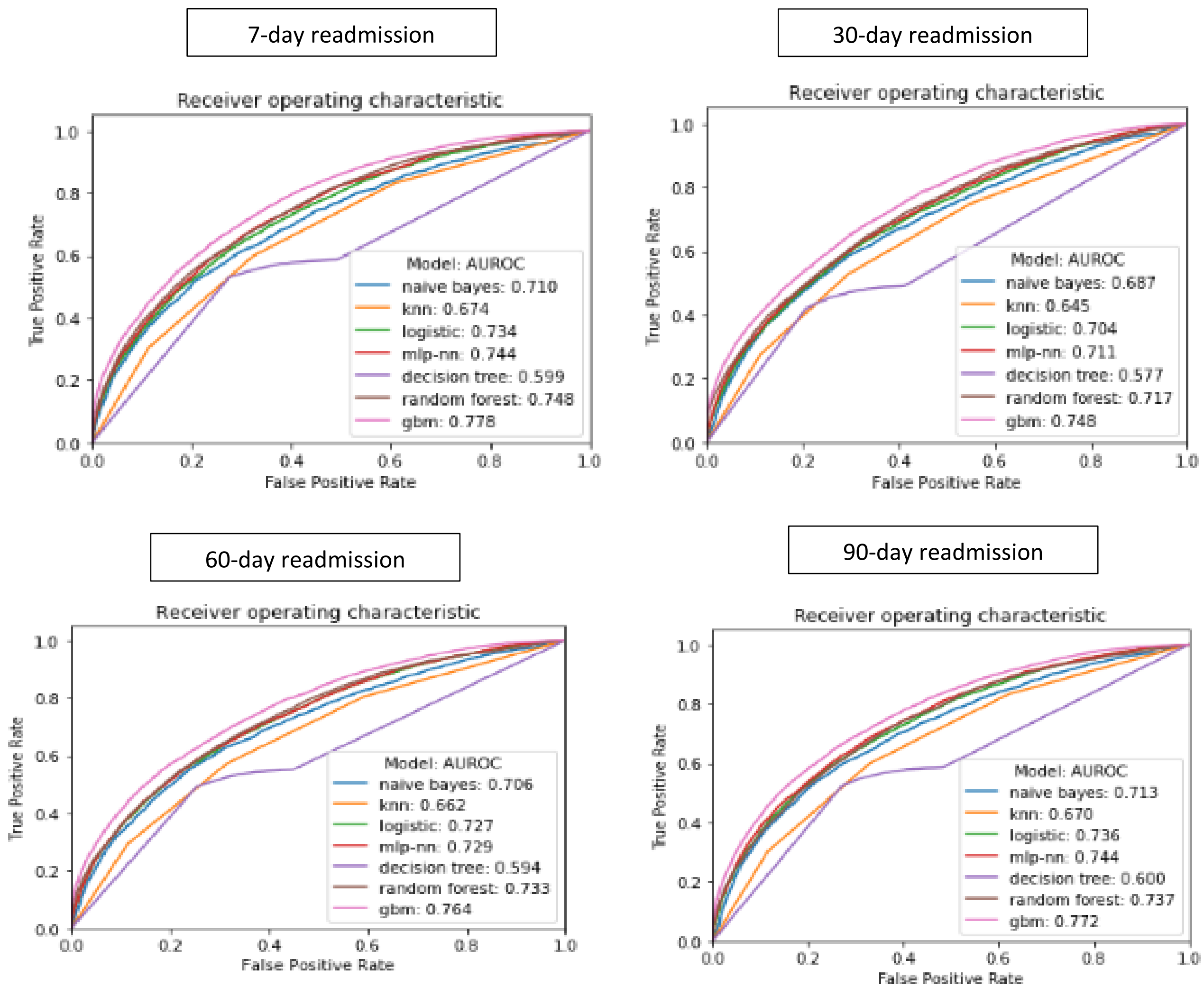


Figure 22. Model AUC of ML algorithms by timeframes – Boruta predictors – AOA*



**knn: K-Nearest Neighbors, mlp-nn: Neuro Networks, gbm: Gradient Boosting*

Final set of variables and model

Among the models, Gradient Boosting consistently performed better than the other ML models across all timeframes. Thus, we chose it as our final model. Although using all 258 predictor variables with the Gradient Boosting model yielded the best accuracy and AUC, it consumed lots of computational power and took a long time to run (about 20-30 minutes). On the other hands, there are only 10 or fewer variables selected from the Boruta algorithms and the results were similar. The final model that decided to pick and tune its parameters was the Gradient Boosting with Boruta-selected predictors.

There were three parameters that we tuned for the Gradient Boosting model: learning rate (moderate the contribution of each tree), n_estimators (the number of trees in the forest) and max_dept (how deep the built tree can be). Table 11 lists the best parameters, accuracy and AUC for each timeframe model. There were some improvements in the accuracy and AUC of this tuned final model compared to the other models. The highest accuracy and AUC were 0.87 and 0.78, respectively, for 7-day readmission. The accuracy and AUC decreased as the readmission timeframe increased but the values were still larger than 0.72. These results are similar to the final model of the AOA sample.

Table 11. Best parameters, accuracy and AUC of the final model - AOA

	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Learning rate	1	0.1	0.1	0.01
N_estimators	1500	1500	1500	1000
Max_depth	4	6	4	5
Accuracy	0.8734	0.7712	0.7342	0.7218
AUC	0.7789	0.7785	0.7639	0.7769

Regarding the important predictor variables, length in the psychiatric hospital, history of psychiatric hospitalization, number of services up to the index hospitalization, and number of services before the index hospitalization since previous hospitalization are the strongest predictors for the model. Having IP and ES services before the index hospitalization and discharge status moderately contribute to the final model. The demographics variables influencing the models include diagnosis, education level, living situation, sexual orientation, and co-occurring substance use. Table 12 shows the average important feature scores and coefficients of the Boruta predictor variables for all four readmission timeframes. The details of individual timeframes are listed in the appendices.

Table 12. Average important feature scores and coefficient of Boruta predictors across four timeframes - AOA

Boruta predictors	Important feature score	Coefficient
Length in hospital	0.3906	-5.1703
History of hospitalization	0.2143	5.3748
Number of services up to index hospitalization	0.1077	-1.8395
Number of services before index hospitalization since previous hospitalization	0.0871	-3.2599
Number of services before index hospitalization since previous hospitalization at the same subunits	0.0577	-0.3956
IP services 30 days before index hospitalization	0.0486	1.9231
ES services 7 days before index hospitalization	0.0470	-0.0244
IP services 60 days before index hospitalization	0.0378	3.1198
ES services 30 days before index hospitalization	0.0300	1.4333
ES services 60 days before index hospitalization	0.0254	0.4098
Diagnosed with Depressive disorders	0.0218	-0.2804
Diagnosed with Schizophrenia/Psychotic disorders	0.0171	0.1644
Unknown Education level	0.0148	-0.3531
Lives independently	0.0143	-0.3210
Missing sexual orientation	0.0127	-0.3717
Received IP-FFS	0.0114	0.1969
Received - IP LIHP	0.0098	-0.3392
Discharge status: to IMD/MHRC	0.0085	-3.1615
Discharge status: to Medical Hospital	0.0080	1.9624
Discharge status: to Psychiatric Hospital	0.0077	1.9732
No first LOC before index hospitalization	0.0065	-0.1842
Co-occurring substance use	0.0059	0.2581
Discharge status: to lower LOC	0.0051	0.9226

4. Discussion

4.1 Summary of Findings

Psychiatric rehospitalization, especially when it occurs early after the discharge from a previous psychiatric inpatient service, can be costly and burdensome to the clients, their families and the healthcare system. Using the administrative data from the County of San Diego Behavioral Health Services via CCBH system, we analyzed and applied the machine learning approach to predict the inpatient readmission within 7, 30, 60 and 90 days after a discharge as well as to identify the influencing predictor variables for both the CYF and AOA samples. To find the optimal solution, we used three different sets of variables to train the models: all variables, variables selected from the ML models and variables selected from the Boruta algorithm. By doing this process, we were able to find the "sweet point" to reduce a large number of predictors to the finest set while preserving the best performance of the models. This analytical procedure is widely applied in psychiatric research with high dimensional data (Zhao and Castellano, 2016).

There are not many differences in the overall model performance between the CYF and AOA samples. Among the models, we found that the Naïve Bayes and KNN models had poor performances (accuracy and AUC < 0.65) compared to other models. The Logistic Regression, which is a popular statistical model used predominately in psychiatric rehospitalization research (Grinshpoon et al., 2007), had a moderate performance (0.65 – 0.75 accuracy and 0.68 - 0.72 AUC). Its performance improved when using a smaller set of predictor variables. This shows that this model is more suitable for low dimensional data. The Neural Networks model had similar accuracy to the Logistic Regression but better AUC (0.70-0.74). Among the three tree-based models, Decision Tree had the lowest performance (accuracy 0.64-0.68 and AUC 0.62 – 0.66). The Random Forest and Gradient Boosting had the best performance (accuracy 0.70-0.92 and AUC 0.70 – 0.79). Because Gradient Boosting performed better than the random forest, we selected it as our final model and fine-tuned its parameters.

Regarding the predictor variables, we found that fitting all variables (185 for CYF and 258 for AOA) yielded the best performance, specifically with Gradient Boosting (accuracy 0.70-0.92). However, due to the high dimensionality, it took a long time to run all models (about 40 minutes). This duration decreased significantly (5-10 minutes) when reducing the number of variables to 15 or less. Thus, we selected the top 15 predictor variables based on the important feature scores from the three tree-based models for the second round. However, the results were unexpected. The performances of the Logistic Regression, Neuro Networks and the tree-based models got worse while the Naïve Bayes and KNN improved; yet their accuracy and AUC were still lower than 0.71. This pattern were more obvious for the CYF sample than the AOA sample. The third set of predictor variables were selected by the Boruta algorithm. The average number of selected variables were 12 for the CYF sample and 17 for the AOA sample. When fitting these variables, the performances of weak models (Naïve Bayes and KNN) improved by 0.1 or 0.2 points compared to all-variable versions. The performance of other models did not change much. Because the Boruta set of predictor variables are much smaller than all-variable set and the performance of the models were similar and some even better, we selected this set to train the final Gradient Boosting model.

The influencing predictor variables that the Boruta algorithm chose include length in the psychiatric hospital, history of psychiatric hospitalization, number of services up to the index hospitalization, number of services before the index hospitalization since previous hospitalization, having IP and ES services before the index hospitalization, and discharge status. They are the same for both CYF and AOA samples. For demographics variables, the CYF sample had age and race while the AOA-selected predictors also include diagnosis, sexual orientation, education level, living situation, and co-occurring substance use.

After tuning the Gradient Boosting model with the Boruta predictors, the average accuracy scores across four timeframes were 0.82 for the CYF sample and 0.78 for the AOA sample; the average AUC scores were 0.75 for the CYF sample and 0.77 for the AOA sample. The model performed very well for the 7-day readmission (CYF: 0.92 accuracy and 0.78 AUC, AOA: 0.87 accuracy and 0.78 AUC). However, the model performance decreased in 30, 60 and 90 days. This makes sense because the further the timeframe in the future, the less accuracy the prediction could be. Although the AUC scores are not high like the accuracy scores, the 0.7 or higher AUC scores could be considered moderately high for hospital readmission. Most of the hospital readmission (medical and psychiatric) prediction models in previous studies usually had an AUC score below 0.75 (Kansagara et al. 2011, Artetxe et al. 2018).

4.2 Strengths and limitations

The findings of our study should be considered in light of its strengths and limitations. A strength in this study is that we analyzed two relatively big datasets for both the CYF and AOA systems at four readmission timeframes (7, 30, 60 and 90 days). The datasets include 12 years of data and have more than 100 variables. This increased the complexity of the project but it also provided much helpful information. The second strength is the application of multiple ML models with different sets of number of variables. This method helped us to find the best combination to achieve the optimal solution.

Limitations of our study include the fact that we could not use all variables in the original dataset, especially demographics variables, due to a large portion of missing data. Because some of the variables (e.g. veteran status, history of trauma, domestic violence, sexual orientation for CYF, etc.) were not collected in the earlier years, these variables had more than 30% of missing data and was discarded by our selection threshold initially. For variables with a low percent of missing data, we had to impute them because most of ML models are not capable of working with missing observations. The second limitation is that not all ML models were well equipped. We used the default version for each model and only fine-tuned the Gradient Boosting. Some models might perform better than the other if its parameters are modified. Third limitation is the class imbalance; there are much fewer rehospitalized cases than non-rehospitalized ones. Although we applied the SMOTE technique to generate synthetic “Yes” examples to balance the Yes-No readmission classes, this technique did not take into consideration neighboring examples can be from the “No” class. This can increase the overlapping of classes and introduce additional noise. The fourth limitation is that we only used administrative data to build the model. The initial plan was to use both the administrative data and the outcome assessment data, which measures the clients’ emotional and behavioral symptoms at outpatient visits by clinicians. However, because there were some changes in the outcome assessment tools last four years and combining versions were difficult, we decided to use the administrative data only for this study project.

4.3 Conclusion

Our study analyzed the administrative data of behavioral health services in San Diego County, CA and attempted to build multiple ML models to predict the EPR at four timeframes for both the CYF and AOA samples. Although we found that the Gradient Boosting model with Boruta-selected predictor variables yielded acceptable accuracy and AUC, we believed that there are more potentials to improve this model. Our next steps are to find a way to bring in the outcome assessment data and potentially apply some deep learning models in addition to upgrading the Gradient Boosting. This study adds to the current literature on predicting psychiatric rehospitalization and machine learning. Effectively predicting the EPR and understanding its contributors could help inform the County stakeholders, mental health providers and researchers to improve mental health services planning, policies and programs to reduce hospitalizations.

References

Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 1992;46(3):175–85.

Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: a systematic review of methods. Comput Methods Programs Biomed. 2018;164:49–64.

Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25(2), 197-227.

Blankers, M., van der Post, L. F., & Dekker, J. J. (2020). Predicting hospitalization following psychiatric crisis care using machine learning. BMC medical informatics and decision making, 20(1), 1-11.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

County of San Diego Behavioral Health Services. Technical Resource Library.

https://www.sandiegocounty.gov/content/sdc/hhsa/programs/bhs/technical_resource_library.html

Donisi, V., Tedeschi, F., Wahlbeck, K., et al., 2016. Pre-discharge factors predicting readmissions of psychiatric patients: a systematic review of the literature. BMC Psychiatry 16, 449. https://doi.org/10.1186/s12888-016-1114-0

Grinshpoon, A., Abramowitz, M.Z., Lerner, Y., et al., 2007. Re-hospitalization of first-inlife admitted schizophrenic patients before and after rehabilitation legislation: a comparison of two national cohorts. Soc. Psychiatry Psychiatr. Epidemiol. 23, 355–359. https://doi.org/10.1007/s00127-007-0167-2.

Hamilton JE, Passos IC, de Azevedo CT, et al. Predictors of psychiatric readmission among patients with bipolar disorder at an academic safety net hospital. Aust N Z J Psychiatry. 2016;50(6):584–93.

Hand DJ, Yu K. Idiot’s Bayes—not so stupid after all? Int Stat Rev. 2001;69(3):385–98.

Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiol. 1982;143(1):29–36.

Hung YY, Chan HY, Pan YJ. Risk factors for readmission in schizophrenia patients following involuntary admission. PLoS ONE. 2017;12(10):e0186768.

Kallert TW, Glöckner M, Schützwohl M. Involuntary vs. voluntary hospital admission. A systematic literature review on outcome diversity. Eur Arch Psychiatry Clin Neurosci. 2008;258(4):195–209.

Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. JAMA. 2011;306(15):1688–1698

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. J StatSoftw, 36(11), 1-13.

Laliberté V, Stergiopoulos V, Jacob B, Kurdyak P. Homelessness at discharge and its impact on psychiatric readmission and physician follow-up: a population-based cohort study. Epidemiol Psychiatr Sci.2019;29:1–8.

Morel, D., Kalvin, C. Y., Liu-Ferrara, A., Caceres-Suriel, A. J., Kurtz, S. G., & Tabak, Y. P. (2020). Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. International journal of medical informatics, 139, 104136.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

Nelder J, Wedderburn R. Generalized linear models. J R Stat Soc Ser A (General). 1972;135(3):370–84.

Patro, S., & Sahu, K. K. (2015). Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462.

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.

van Gerven M, Bohte S, editors. Artificial neural networks as models of neural information processing. Lausanne: Frontiers Media; 2018.

Vigod, S.N., Kurdyak, P.A., Dennis, C.L., et al., 2013. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. BJPsych 202, 187–194. https://doi.org/10.1192/bjp.bp.112.115030

Webber M, Huxley P. Social exclusion and risk of emergency compulsory admission. A case-control study. Soc Psychiatry Psychiatr Epidemiol.2004;39(12):1000–9.

Zhao, Y., Castellanos, F.X, 2016. Annual research review: discovery science strategies in studies of the pathophysiology of child and adolescent psychiatric disorders – promises and limitations. J. Child Psychol. Psychiatry 57, 421–439. https://doi.org/10. 1111/jcpp.12503.

Zhao, Y., Hoenig, J. M., Protacio, A., Lim, S., & Norman, C. C. (2020). Identification of risk factors for early psychiatric rehospitalization. Psychiatry research, 285, 112803.

Appendices

A. CYF demographics (Total unique clients: 6,310)

Variables	%
Age group	
0-5 years old	
6-11 years old	
12-15 years old	
16-17 years old	
18-25 years old	
Gender	
Female	
Male	
Other	
Unknown	
Race/Ethnicity	
Asian/ Pacific Islander	
Black/African-American	
Hispanic	
Native American	
White	
Other	
Unknown	
Primary spoken language	
English	
Spanish	
Asian languages	
Middle Eastern languages	
Other/Unknown	
Primary diagnosis	
Depressive Disorders	
Bipolar Disorders	
Oppositional/ Conduct Disorders	
Stressor and Adjustment Disorders	
ADHD	
Anxiety Disorders	
Schizophrenic Disorders	
Other/Excluded	
Insurance status	
Medi-Cal only	
Private	
Other insurance	
Uninsured/Unknown	

These data are for internal use only.

Variables	%
Living situation	
House or Apartment	
Group Home	
Foster Home	
Correctional Facility	
Children’s Shelter	
Homeless	
Lives Independently	
Other	
Unknown	
Discharge status	
Satisfactorily achieved discharge goals	
To home/ shelter	
To Crisis Residence	
To Correctional Facility	
To Lower Level of Care	
To Higher Level of Care	
Left the crisis mental health treatment against medical recommendation	
Detained in a correctional facility	
To residential treatment facility	
Homeless	
To medical facility	
Released from 24-hour care by a court order	
Transfer to medical hospital	
Had substance abuse	
Yes	
No	
Had co-occurring substance use	
Yes	
No	
Received CWS	
Yes	
No	
Received ADS	
Yes	
No	
Juvenile Justice Involvement	
Yes	
No	
Received Special Education	
Yes	
No	

These data are for internal use only.

B1. CYF Performance Metrics – All predictors

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Naïve Bayes	Accuracy	0.1057	0.2247	0.3624	0.3839
	Precision	0.0747	0.1604	0.2392	0.2873
	Recall	0.9663	0.9271	0.8639	0.8958
	F1 score	0.1388	0.2734	0.3747	0.4351
	AUC score	0.5038	0.5166	0.6598	0.6705
	AP score	0.0751	0.1622	0.3571	0.4066
KNN	Accuracy	0.7616	0.6645	0.6480	0.6455
	Precision	0.1435	0.2492	0.3389	0.3957
	Recall	0.4423	0.5626	0.6224	0.6414
	F1 score	0.2167	0.3455	0.4389	0.4894
	AUC score	0.6544	0.6576	0.6673	0.6779
	AP score	0.1272	0.2370	0.3238	0.3872
Logistic Regression	Accuracy	0.6943	0.6900	0.6889	0.6857
	Precision	0.1364	0.2822	0.3739	0.4340
	Recall	0.5817	0.6287	0.6029	0.6143
	F1 score	0.2210	0.3896	0.4615	0.5087
	AUC score	0.6933	0.7235	0.7163	0.7199
	AP score	0.1572	0.3545	0.4346	0.4958
Neural Networks	Accuracy	0.8570	0.7677	0.7072	0.7022
	Precision	0.1625	0.2994	0.3551	0.4442
	Recall	0.2212	0.3554	0.3971	0.4953
	F1 score	0.1874	0.3250	0.3749	0.4683
	AUC score	0.6235	0.6300	0.6366	0.6810
	AP score	0.1467	0.2913	0.3508	0.4421

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Decision tree	Accuracy	0.8685	0.7534	0.7122	0.6796
	Precision	0.1807	0.2707	0.3556	0.4028
	Recall	0.2163	0.3349	0.3712	0.4344
	F1 score	0.1969	0.2994	0.3632	0.4180
	AUC score	0.5675	0.5809	0.5880	0.5981
	AP score	0.0975	0.1953	0.2713	0.3248
Random forest	Accuracy	0.9287	0.8452	0.7896	0.7556
	Precision	0.6552	0.5181	0.5377	0.5510
	Recall	0.0913	0.2278	0.3468	0.4168
	F1 score	0.1603	0.3165	0.4217	0.4746
	AUC score	0.7724	0.7650	0.7592	0.7662
	AP score	0.2819	0.3974	0.4712	0.5266
Gradient Boosting	Accuracy	0.9247	0.8330	0.7713	0.7516
	Precision	0.4853	0.4483	0.4776	0.5420
	Recall	0.1587	0.2665	0.3630	0.4019
	F1 score	0.2391	0.3343	0.4125	0.4615
	AUC score	0.7209	0.7442	0.7357	0.7424
	AP score	0.2689	0.4056	0.4644	0.5164

B2. CYF Performance Metrics – Predictors selected from ML models

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Naïve Bayes	Accuracy	0.6771	0.6706	0.6713	0.6692
	Precision	0.1310	0.2556	0.3509	0.4100
	Recall	0.5913	0.5718	0.5721	0.5670
	F1 score	0.2145	0.3533	0.4350	0.4759
	AUC score	0.6694	0.6734	0.6736	0.6866
	AP score	0.1654	0.2947	0.3793	0.4408
KNN	Accuracy	0.6728	0.6448	0.6258	0.6183
	Precision	0.0944	0.2137	0.3014	0.3558
	Recall	0.3942	0.4692	0.5251	0.5440
	F1 score	0.1523	0.2937	0.3830	0.4302
	AUC score	0.5515	0.5997	0.6216	0.6281
	AP score	0.0817	0.1978	0.2904	0.3428
Logistic Regression	Accuracy	0.6573	0.6495	0.6158	0.6577
	Precision	0.1326	0.2557	0.3296	0.4107
	Recall	0.6490	0.6424	0.7131	0.6725
	F1 score	0.2202	0.3658	0.4508	0.5100
	AUC score	0.7189	0.6993	0.7072	0.7094
	AP score	0.1756	0.3288	0.4243	0.4753
Neural Networks	Accuracy	0.7018	0.611	0.6319	0.6283
	Precision	0.1250	0.2445	0.3381	0.3904
	Recall	0.5000	0.7039	0.6937	0.7158
	F1 score	0.2000	0.3629	0.4546	0.5060
	AUC score	0.6287	0.7079	0.7176	0.7158
	AP score	0.1360	0.3391	0.4332	0.4818

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Decision tree	Accuracy	0.1810	0.2609	0.2964	0.3789
	Precision	0.0714	0.1385	0.2092	0.2647
	Recall	0.8317	0.7084	0.7844	0.7564
	F1 score	0.1315	0.2317	0.3303	0.3921
	AUC score	0.4802	0.4431	0.4708	0.4995
	AP score	0.0719	0.1440	0.2117	0.2647
Random forest	Accuracy	0.5928	0.505	0.5427	0.5326
	Precision	0.1027	0.1877	0.2802	0.3272
	Recall	0.5769	0.6446	0.6807	0.7240
	F1 score	0.1744	0.2907	0.3970	0.4507
	AUC score	0.6030	0.5757	0.6085	0.6290
	AP score	0.0929	0.1769	0.2737	0.3395
Gradient Boosting	Accuracy	0.1151	0.2018	0.2760	0.3211
	Precision	0.0764	0.1580	0.2257	0.2736
	Recall	0.9808	0.9408	0.9352	0.9445
	F1 score	0.1418	0.2706	0.3636	0.4243
	AUC score	0.5994	0.5336	0.5745	0.5978
	AP score	0.1184	0.1721	0.2742	0.3377

B3. CYF Performance Metrics – Predictors selected from Boruta algorithm

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Naïve Bayes	Accuracy	0.8043	0.7573	0.7362	0.7208
	Precision	0.1762	0.3180	0.4086	0.4715
	Recall	0.4423	0.4738	0.4311	0.4479
	F1 score	0.2521	0.3806	0.4196	0.4594
	AUC score	0.7166	0.7090	0.6920	0.6955
	AP score	0.2041	0.3312	0.4132	0.4637
KNN	Accuracy	0.7828	0.7061	0.6896	0.6699
	Precision	0.1434	0.2610	0.3612	0.4046
	Recall	0.3846	0.4738	0.5251	0.5223
	F1 score	0.2089	0.3366	0.4280	0.4560
	AUC score	0.6348	0.6495	0.6585	0.6655
	AP score	0.1234	0.2286	0.3239	0.3844
Logistic Regression	Accuracy	0.7312	0.6581	0.6659	0.6839
	Precision	0.1543	0.2596	0.3546	0.4304
	Recall	0.5817	0.6333	0.6224	0.5981
	F1 score	0.2440	0.3682	0.4518	0.5006
	AUC score	0.7154	0.7131	0.7025	0.7080
	AP score	0.2017	0.3382	0.4256	0.4798
Neural Networks	Accuracy	0.6828	0.6416	0.6975	0.6724
	Precision	0.1372	0.2672	0.3762	0.4210
	Recall	0.6154	0.7335	0.5592	0.6306
	F1 score	0.2244	0.3917	0.4498	0.5049
	AUC score	0.7169	0.7318	0.7038	0.7138
	AP score	0.2000	0.3427	0.4257	0.4859

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Decision tree	Accuracy	0.8749	0.7620	0.7036	0.6749
	Precision	0.1948	0.2790	0.3465	0.3937
	Recall	0.2163	0.3235	0.3841	0.4208
	F1 score	0.2050	0.2996	0.3643	0.4068
	AUC score	0.5338	0.5569	0.5679	0.5729
	AP score	0.1032	0.2005	0.2710	0.3222
Random forest	Accuracy	0.8885	0.7914	0.7441	0.7211
	Precision	0.2512	0.333	0.4120	0.4714
	Recall	0.2500	0.3257	0.3679	0.4344
	F1 score	0.2506	0.3295	0.3887	0.4521
	AUC score	0.7084	0.7094	0.7089	0.7174
	AP score	0.2057	0.2966	0.3815	0.4605
Gradient Boosting	Accuracy	0.8925	0.8047	0.7541	0.7319
	Precision	0.2604	0.3695	0.4353	0.4932
	Recall	0.2404	0.3417	0.3760	0.4438
	F1 score	0.2500	0.3550	0.4035	0.4672
	AUC score	0.7291	0.7452	0.7234	0.7332
	AP score	0.2909	0.3954	0.4349	0.4939

C. CYF Important feature scores (Imp) and coefficients (Coef) of the final model by timeframes

Timeframe	7-day readmission		30-day readmission		60-day readmission		90-day readmission	
Selected predictors	Imp	Coef	Imp	Coef	Imp	Coef	Imp	Coef
Length in hospital	0.4495	0.6835	0.4718	1.9172	0.4269	-0.7968	0.3259	-1.0085
History of hospitalization	0.2104	4.2345	0.1993	2.9647	0.2134	2.1867	0.1979	1.4661
Number of services up to index hospitalization	0.1882	1.1732	0.1033	0.8195	0.0519	0.6727	0.1128	2.3060
Number of services before index hospitalization since previous hospitalization	-	-	-	-	0.0688	1.9419	0.0903	1.5573
IP services 60 days before index hospitalization	0.0297	0.9954	0.0337	0.8710	0.0265	-0.8541	0.0390	1.5854
IP services 90 days before index hospitalization	0.0400	0.8623	0.0269	-0.6620	0.0265	1.4798	0.0348	-0.8351
ES services 30 days before index hospitalization	-	-	0.0846	0.4864	0.0637	-1.7551	0.0683	-1.6810
IP CAPS	0.0240	0.8571	0.0179	0.8485	0.0093	0.3789	0.0194	0.4965
IP FFS	0.0298	1.3453	0.0175	0.5884	-	-	-	-
Discharge status: other reasons	0.0284	0.8213	-	-	-	-	-	-
Race: unknown	-	-	0.0451	1.4426	0.0688	1.4180	0.0558	1.6161
Age	-	-	-	-	0.0441	1.5176	0.0559	0.6866

D. AOA demographics (Total unique clients: 38,272)

Variables	%
Age group	
<18-25 years old	
26-59 years old	
60+ years old	
Gender	
Female	
Male	
Other/ Unknown	
Sexual orientation	
Heterosexual	
LGBTQ+	
Unknown/Not reported	
Race/Ethnicity	
Asian/ Pacific Islander	
Black/African-American	
Hispanic	
Native American	
White	
Other	
Unknown	
Primary spoken language	These data are for internal use only.
English	
Spanish	
Asian languages	
Middle Eastern languages	
Other/Unknown	
Primary diagnosis	
Schizophrenia and other psychotic disorders	
Depressive disorders	
Bipolar disorders	
Stressor and Adjustment disorders	
Anxiety disorders	
Other/ Excluded	
Insurance status	
Medi-Cal + Medicare	
Private	
Uninsured/Unknown	
Employment status	
Not in labor force	
Not seeking work	
Seeking work	
Competitive job	
Resident/ Inmate of institution	
Other	
Unknown	

Variables	%
Education level	
High school not completed	
High school diploma/ GED	
Some college/ vocational training	
Associate's Degree	
Bachelor's Degree	
Master's Degree	
Doctoral Degree	
Unknown/ Not reported	
Living situation	
Lives independently	
Homeless	
Board and Care	
Institutional	
Justice Related	
Other/Unknown	
Discharge status	
Satisfactorily achieved discharge goals	
To home/ shelter	
Unplanned discharge	
To Crisis Residence	
To Psychiatric Hospital	These data are for internal use only.
To Medical Hospital	
Homeless	
To Nursing Home	
To Institute for the Mentally Disordered	
To residential treatment facility	
Left the crisis mental health treatment against medical recommendation	
Detained in a correctional facility	
To State Hospital	
Transfer to medical hospital	
To Higher Level of Care	
To Same Level of Care	
To Lower Level of Care	
Death	
Incarcerated	
Moved	
Released from 24-hour care by a court order	
Dropped out the treatment	
Other/ Unknown	
Had substance abuse	
Yes	
No	
Had co-occurring substance use	
Yes	
No	

E1. AOA Performance Metrics – All predictors

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Naïve Bayes	Accuracy	0.4298	0.3028	0.3763	0.4157
	Precision	0.3997	0.2734	0.3508	0.3948
	Recall	0.9539	0.9601	0.9637	0.9671
	F1 score	0.5634	0.4257	0.5144	0.5607
	AUC score	0.5472	0.5118	0.5182	0.5281
	AP score	0.4105	0.2739	0.3513	0.3997
KNN	Accuracy	0.6160	0.5945	0.6014	0.6124
	Precision	0.5016	0.3526	0.4442	0.4979
	Recall	0.6658	0.6066	0.6476	0.6518
	F1 score	0.5722	0.4460	0.5269	0.5646
	AUC score	0.6607	0.6252	0.6433	0.6535
	AP score	0.5036	0.3460	0.4426	0.4974
Logistic Regression	Accuracy	0.6954	0.6923	0.6946	0.6957
	Precision	0.5984	0.4471	0.5474	0.5981
	Recall	0.6393	0.6075	0.6298	0.6419
	F1 score	0.6182	0.5151	0.5857	0.6192
	AUC score	0.7521	0.7227	0.7421	0.7509
	AP score	0.6610	0.5092	0.6128	0.6596
Neural Networks	Accuracy	0.6876	0.6857	0.6769	0.6839
	Precision	0.5928	0.4257	0.5255	0.5806
	Recall	0.6071	0.4816	0.5929	0.6483
	F1 score	0.5999	0.4519	0.5572	0.6126
	AUC score	0.7381	0.666	0.7100	0.7447
	AP score	0.6499	0.4361	0.5764	0.6572

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Decision tree	Accuracy	0.6394	0.6793	0.6541	0.6416
	Precision	0.5311	0.4100	0.4956	0.5341
	Recall	0.5561	0.4369	0.5134	0.5515
	F1 score	0.5433	0.4230	0.5044	0.5427
	AUC score	0.6240	0.6027	0.6204	0.6247
	AP score	0.4665	0.3306	0.4213	0.4675
Random forest	Accuracy	0.7243	0.7685	0.7355	0.7212
	Precision	0.6651	0.6135	0.6462	0.6613
	Recall	0.5744	0.3776	0.5044	0.5674
	F1 score	0.6164	0.4675	0.5666	0.6107
	AUC score	0.7792	0.7587	0.7689	0.7738
	AP score	0.6882	0.5634	0.6427	0.6809
Gradient Boosting	Accuracy	0.7271	0.7703	0.7359	0.7222
	Precision	0.6674	0.6058	0.6432	0.6594
	Recall	0.5829	0.4194	0.5159	0.5781
	F1 score	0.6223	0.4957	0.5726	0.6161
	AUC score	0.7865	0.7663	0.7765	0.7814
	AP score	0.7145	0.5950	0.6717	0.7088

E2. AOA Performance Metrics – Predictors selected from ML models

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Naïve Bayes	Accuracy	0.6789	0.7002	0.6862	0.6796
	Precision	0.5999	0.4475	0.5456	0.5991
	Recall	0.5027	0.4857	0.5056	0.5105
	F1 score	0.5470	0.4658	0.5249	0.5513
	AUC score	0.7058	0.6790	0.7010	0.7123
	AP score	0.6089	0.4541	0.5523	0.6050
KNN	Accuracy	0.6254	0.6108	0.6148	0.6169
	Precision	0.5127	0.3468	0.4491	0.5028
	Recall	0.5782	0.5055	0.5447	0.5794
	F1 score	0.5435	0.4114	0.4923	0.5384
	AUC score	0.6481	0.5995	0.6310	0.6389
	AP score	0.4967	0.3301	0.4344	0.4889
Logistic Regression	Accuracy	0.6840	0.6872	0.6889	0.6840
	Precision	0.5916	0.4346	0.5456	0.5920
	Recall	0.5837	0.5399	0.5519	0.5805
	F1 score	0.5876	0.4816	0.5488	0.5862
	AUC score	0.7269	0.6960	0.7187	0.7281
	AP score	0.6381	0.4767	0.5865	0.6359
Neural Networks	Accuracy	0.7071	0.6718	0.6848	0.6891
	Precision	0.6181	0.4225	0.5361	0.5942
	Recall	0.6294	0.5983	0.5984	0.6105
	F1 score	0.6237	0.4952	0.5655	0.6022
	AUC score	0.7667	0.7050	0.7267	0.7400
	AP score	0.6917	0.4873	0.5967	0.6504

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Decision tree	Accuracy	0.4609	0.3110	0.3907	0.4475
	Precision	0.4061	0.2720	0.3501	0.4015
	Recall	0.8600	0.9311	0.9075	0.8827
	F1 score	0.5517	0.4210	0.5052	0.5520
	AUC score	0.5345	0.5068	0.5144	0.5286
	AP score	0.4032	0.2718	0.3494	0.3997
Random forest	Accuracy	0.5543	0.4400	0.5125	0.5584
	Precision	0.4578	0.3085	0.4009	0.4605
	Recall	0.8453	0.8712	0.8536	0.8478
	F1 score	0.5939	0.4557	0.5456	0.5968
	AUC score	0.6910	0.6577	0.6804	0.7006
	AP score	0.5909	0.4196	0.5358	0.6033
Gradient Boosting	Accuracy	0.5327	0.3932	0.5105	0.5448
	Precision	0.4475	0.2946	0.4017	0.4535
	Recall	0.9017	0.9005	0.8745	0.8813
	F1 score	0.5981	0.4440	0.5505	0.5989
	AUC score	0.7106	0.6664	0.7000	0.7093
	AP score	0.6237	0.4536	0.5668	0.6241

E3. AOA Performance Metrics – Predictors selected from Boruta algorithm

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Naïve Bayes	Accuracy	0.5996	0.6033	0.6107	0.6051
	Precision	0.4882	0.3716	0.4577	0.4923
	Recall	0.7903	0.6865	0.7356	0.7782
	F1 score	0.6035	0.4822	0.5643	0.6031
	AUC score	0.7097	0.6874	0.7059	0.7134
	AP score	0.6146	0.4543	0.5579	0.6064
KNN	Accuracy	0.6459	0.6575	0.6460	0.6429
	Precision	0.5367	0.3971	0.4861	0.5328
	Recall	0.5976	0.5267	0.5728	0.5984
	F1 score	0.5655	0.4528	0.5259	0.5637
	AUC score	0.6741	0.6452	0.6621	0.6696
	AP score	0.5210	0.3723	0.4666	0.5168
Logistic Regression	Accuracy	0.6833	0.6775	0.6754	0.6801
	Precision	0.5858	0.4273	0.5224	0.5782
	Recall	0.6097	0.5841	0.6199	0.6295
	F1 score	0.5976	0.4936	0.5670	0.6028
	AUC score	0.7342	0.7043	0.7269	0.7362
	AP score	0.6404	0.4855	0.5927	0.6419
Neural Networks	Accuracy	0.6830	0.6720	0.6513	0.6745
	Precision	0.5777	0.4236	0.4939	0.5637
	Recall	0.6618	0.6073	0.6939	0.6885
	F1 score	0.6169	0.4991	0.5770	0.6199
	AUC score	0.7438	0.7110	0.7286	0.7445
	AP score	0.6512	0.4979	0.5946	0.6558

Models	Metrics	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Decision tree	Accuracy	0.6457	0.6879	0.6555	0.6454
	Precision	0.5411	0.4210	0.4975	0.5410
	Recall	0.5347	0.4269	0.4984	0.5295
	F1 score	0.5379	0.4239	0.4980	0.5332
	AUC score	0.5985	0.5775	0.5935	0.6005
	AP score	0.4778	0.3435	0.4285	0.4777
Random forest	Accuracy	0.6933	0.7258	0.6983	0.6859
	Precision	0.6020	0.4897	0.5610	0.5932
	Recall	0.6042	0.4559	0.5508	0.5897
	F1 score	0.6031	0.4722	0.5559	0.5915
	AUC score	0.7477	0.7166	0.7328	0.7371
	AP score	0.6591	0.5269	0.6083	0.6491
Gradient Boosting	Accuracy	0.7134	0.7423	0.7165	0.7082
	Precision	0.6265	0.5228	0.5868	0.6203
	Recall	0.6362	0.4834	0.5852	0.6265
	F1 score	0.6313	0.5023	0.5860	0.6234
	AUC score	0.7781	0.7479	0.7639	0.7716
	AP score	0.7034	0.5707	0.6558	0.6977

F. AOA Important feature scores (Imp) and coefficients (Coef) of the final model by timeframes

Timeframe	7-day readmission		30-day readmission		60-day readmission		90-day readmission	
Selected predictors	Imp	Coef	Imp	Coef	Imp	Coef	Imp	Coef
Length in hospital	0.4741	-9.3431	0.4182	-3.2049	0.3580	-4.2073	0.3121	-3.9259
History of hospitalization	0.2007	3.5219	0.1828	4.7507	0.2413	6.0961	0.2321	7.1304
Number of services up to index hospitalization	0.0770	-2.7077	0.1419	-1.9454	0.1046	-1.4649	0.1074	-1.2399
Number of services before index hospitalization since previous hospitalization	-	-	0.0794	-3.5652	0.0844	-3.2682	0.0974	-2.9465
Number of services before index hospitalization since previous hospitalization at the same subunits	0.0577	-0.3956	-	-	-	-	-	-
IP services 30 days before index hospitalization	0.0229	2.8407	0.0509	1.6389	0.0720	1.2896	-	-
IP services 60 days before index hospitalization	0.0180	1.7616	0.0246	3.0897	0.0229	3.3327	0.0859	4.2952
ES services 7 days before index hospitalization	0.0470	-0.0244	-	-	-	-	-	-
ES services 30 days before index hospitalization	0.0300	1.4333	-	-	-	-	-	-
ES services 60 days before index hospitalization	0.0254	0.4098	-	-	-	-	-	-
Received IP-FFS	0.0114	0.1535	0.0105	0.1466	0.0094	0.2186	0.0144	0.2690
Received - IP LIHP	0.0100	-0.2312	0.0097	-0.3836	0.0085	-0.3403	0.0110	-0.4017
No first LOC before index hospitalization	0.0049	-0.1436	0.0071	-0.1349	0.0061	-0.2185	0.0080	-0.2397

Timeframe	7-day readmission		30-day readmission		60-day readmission		90-day readmission	
Selected predictors	Imp	Coef	Imp	Coef	Imp	Coef	Imp	Coef
Discharge status: to IMD/MHRC	-	-	0.0079	-2.9706	0.0078	-3.2424	0.0099	-3.2714
Discharge status: to Psychiatric Hospital	0.0075	2.7332	0.0071	1.9394	0.0075	1.6636	0.0088	1.5566
Discharge status: to lower LOC	0.0054	1.1293	0.0048	0.7159	-	-	-	-
Discharge status: to Medical Hospital	0.0080	1.9624	-	-	-	-	-	-
Diagnosed with Depressive disorders	-	-	0.0146	-0.2635	0.0217	-0.2740	0.0292	-0.3036
Diagnosed with Schizophrenia/Psychotic disorders	-	-	0.0122	0.1582	0.0135	0.1618	0.0257	0.1733
Unknown Education level	-	-	0.0115	-0.2998	0.0131	-0.3793	0.0198	-0.3801
Lives independently	-	-	-	-	0.0120	-0.3183	0.0167	-0.3236
Unknown sexual orientation	-	-	0.0112	-0.3463	0.0119	-0.3702	0.0152	-0.3986
Co-occurring substance use	-	-	0.0057	0.2721	0.0054	0.2509	0.0065	0.2514