Early Psychiatric Rehospitalization Prediction

Anh Tran

CSE 6748 Applied Analytics Practicum presentation

Overview

- Background and Objective
- Methods
  - Key Elements
  - Predictive Model Pipeline
- Results
  - CYF findings
  - AOA findings
- Discussion

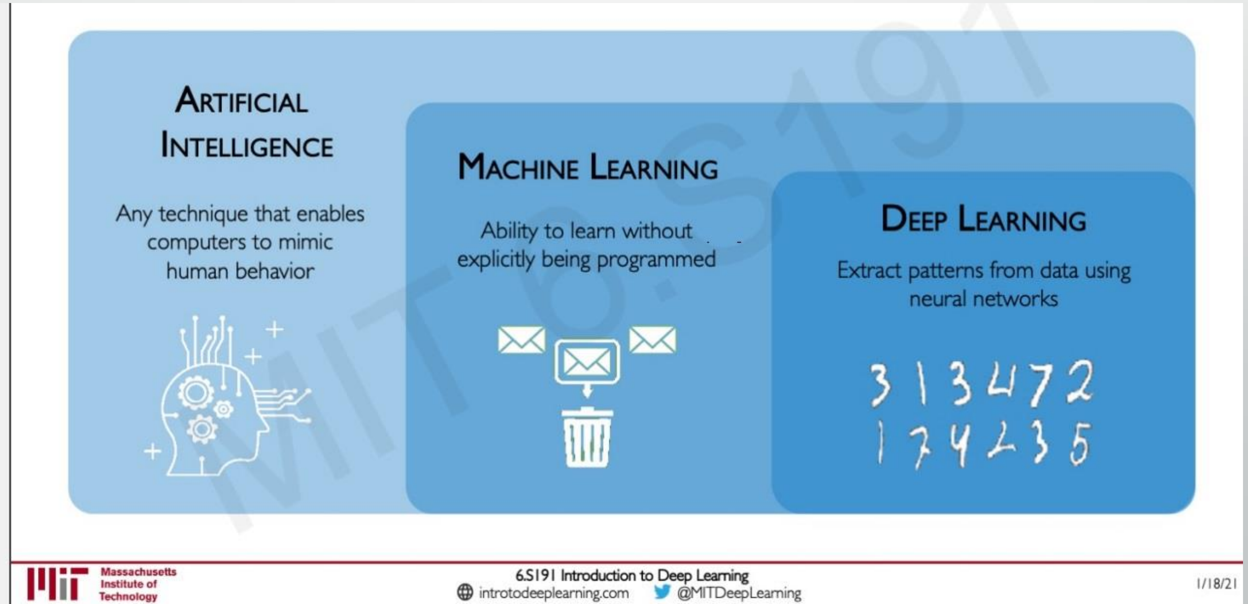
Background and Objective

Early psychiatric rehospitalization (EPR)

- 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.
- EPR is disruptive for clients and their families, increases the risk of complication, and represents a strain on limited health care resources (Vigod et al., 2013).
- Previous studies have identified some associations between psychiatric rehospitalization and 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).
- 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 care for the high-risk clients in order to reduce rehospitalization rate.

Machine learning approach

- With the abundance of data and input features, many recent studies have applied machine learning approach to develop hospitalization/rehospitalization prediction models such as:
    - Supervised learning models: Logistic Regression, Naïve Bayes, etc. (Li et al. 2020)
    - Random forest (Zhao et al., 2020)
    - Gradient boosted decision trees XGBoost (Morel et al. 2020)
    - Generalized linear model (GLM), ensemble methods and deep learning (Blankers et al. 2020)



Objectives

- 1) Predict 90/60/30/7-day rehospitalization since a previous inpatient psychiatric hospital discharge for CYF population and identify the influencing factors.
- 2) Predict 90/60/30/7-day rehospitalization since a previous inpatient psychiatric hospital discharge for AOA population and identify the influencing factors.

Methods

Key Elements

- Client data source
- Outputs and predictor variables
- Machine learning models

Client data source

- Data source: Cerner Community Behavioral Health (CCBH)
  - Data Warehouse database in Microsoft SQL Server at HSRC
- Dataset:
  - Inpatient assignment data was pulled from 07/01/2008 to 06/30/2020 (12 years)
    - Each row in the dataset includes a hospital stay, also known as the "**index hospitalization**", and the subsequent hospital stay called the "**readmission**."
    - The dataset includes client demographics, all services received before the index hospitalization, level of care, and the 7, 30, 60 and 90-day readmissions.
  - 118,893 psychiatric hospitalization records of 44,605 clients who have received at least one psychiatric IP service.
    - CYF sample includes 13,950 hospitalizations of 6,310 unique clients
    - AOA sample includes 104,920 cases of 38,272 unique clients

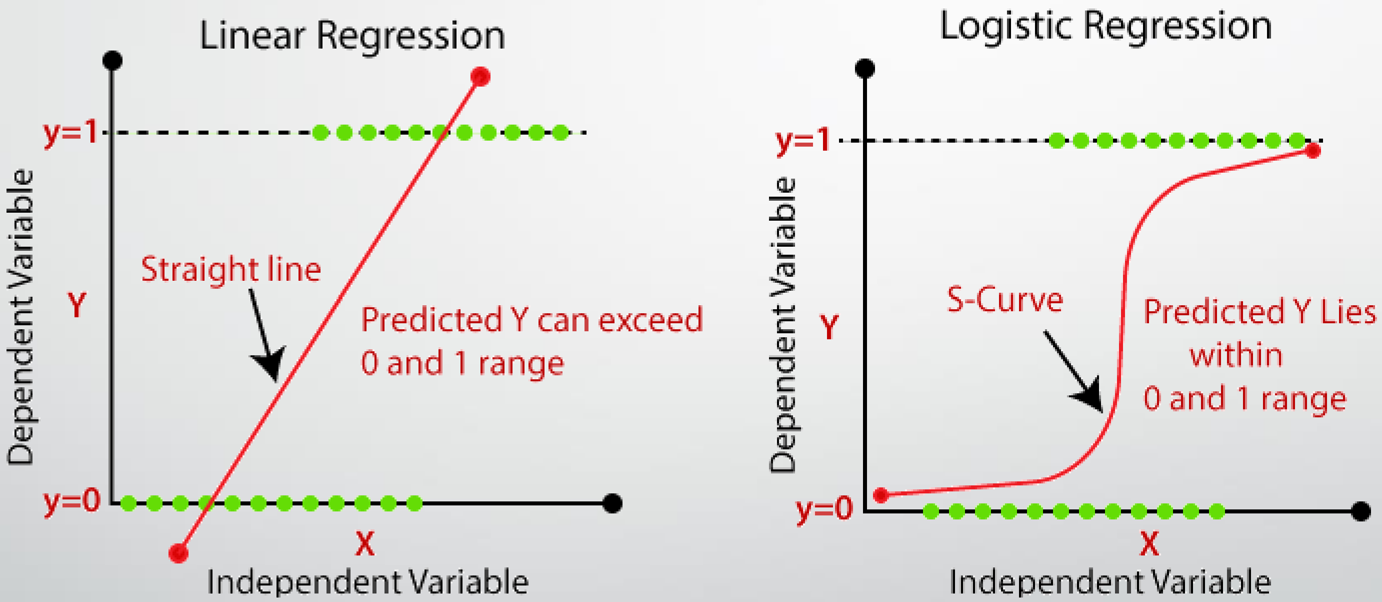
Outputs and predictor variables

- Output: a dichotomized measure of client psychiatric rehospitalization within 7, 30, 60 or 90 days after being discharged from an index hospitalization.(4 analyses)
- Predictor variables:
  - Excluded variables with 30% or more missing data

<table>
<tr><th></th><th colspan="2">Candidate predictor variables</th></tr>
<tr><th>Themes</th><th>CYF</th><th>AOA</th></tr>
<tr><td rowspan="2">Demographics</td><td colspan="2">Age (num), Age group (cat), Gender (cat), Race (cat), Primary language (cat), Psychiatric diagnosis (cat), Living situation (cat), Insurance status (cat), co-occurring substance use (cat)</td></tr>
<tr><td>Receive Child Welfare Services/ CWS (cat), Receive Alcohol and Drug Services/ADS (cat), Probation or Juvenile Justice/JJ Involvement (cat)</td><td>Employment status (cat), Education status (cat), Sexual orientation (cat), Jail/Justice Involvement (cat)</td></tr>
<tr><td>Length and history of psychiatric hospitalization</td><td colspan="2">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)</td></tr>
<tr><td>Number of total services</td><td colspan="2">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)</td></tr>
<tr><td rowspan="2">Service group received up to 7/30/60/90 days before the index hospitalization</td><td colspan="2">Number of Outpatient (OP) Services (num)
Number of Emergency/Crisis Services (num)
Number of Inpatient (IP) Services (num)</td></tr>
<tr><td>Days in Day (DT) Services (num)</td><td>Days in 24-hour Services (num)</td></tr>
<tr><td>Order of LOC services</td><td colspan="2">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)</td></tr>
<tr><td>Discharge</td><td colspan="2">Discharge Status (cat)</td></tr>
<tr><td>Total Variables</td><td>185</td><td>258</td></tr>
</table>

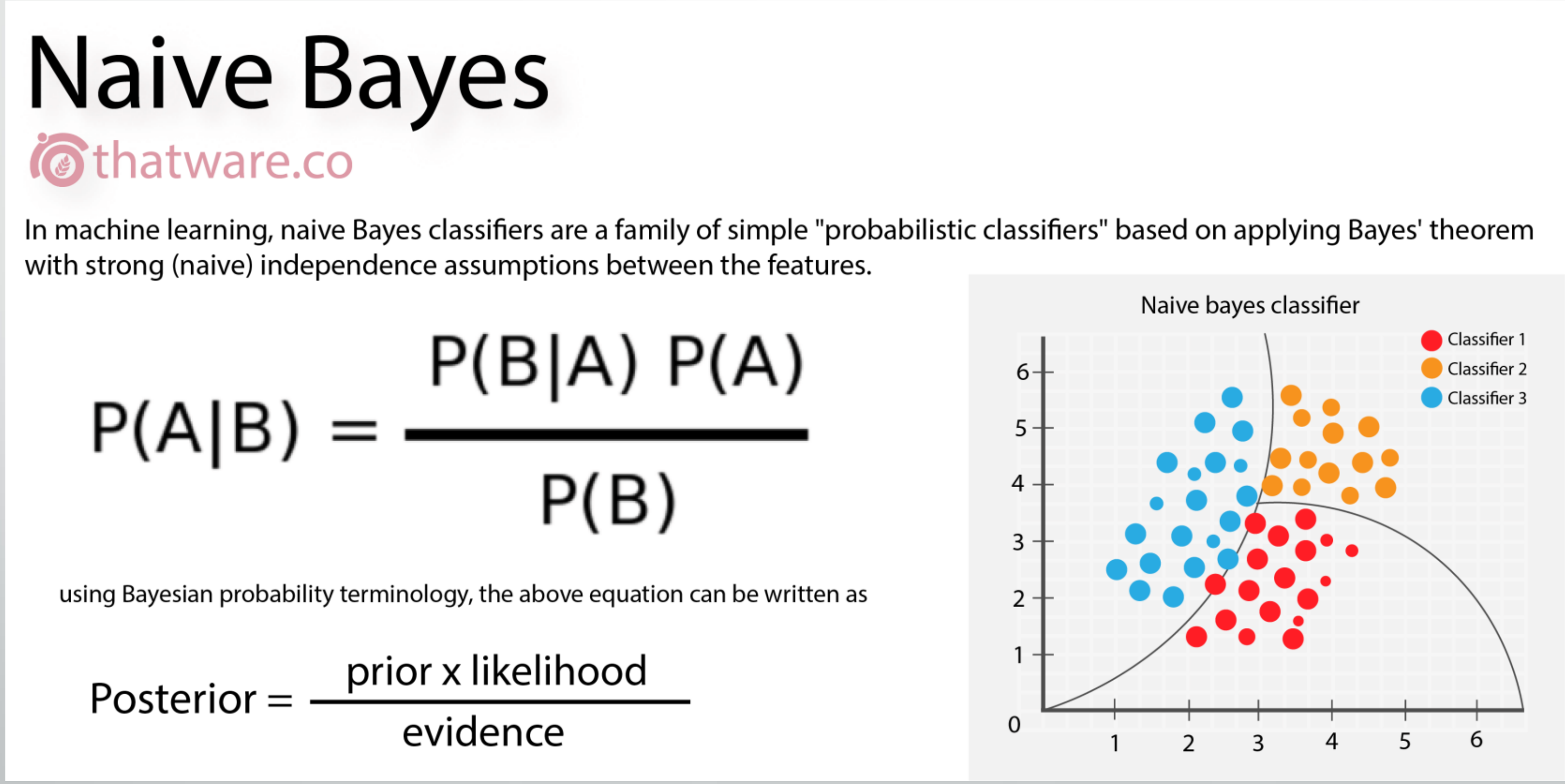
Machine learning algorithms

Algorithm	Descriptions
Logistic Regression	A statistical algorithm used to estimate the probability of a binary (0-1) outcome using the Sigmoid function



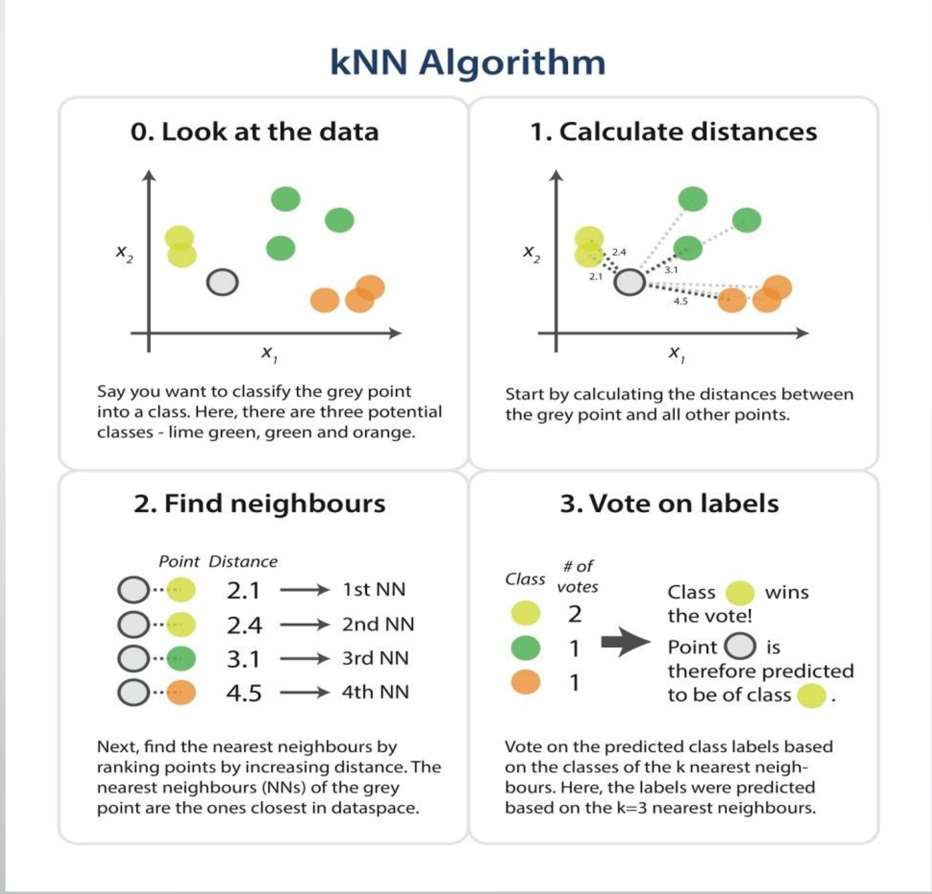
Machine learning algorithms

Algorithm	Descriptions
Naïve Bayes	A Bayes' Theorem algorithm with an assumption of independence among predictors to classify cases into labels



Machine learning algorithms

Algorithm	Descriptions
K-nearest neighbors (KNN)	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

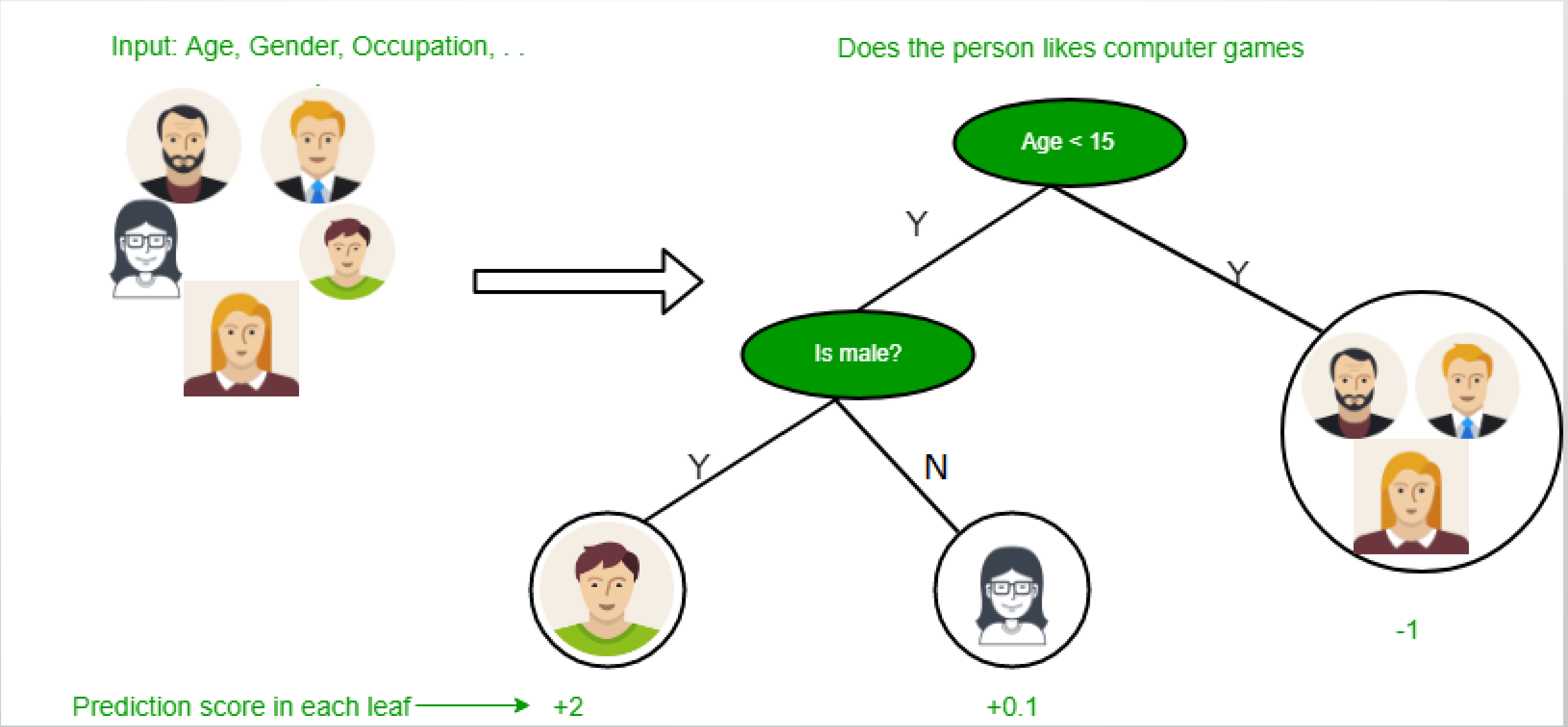


Machine learning algorithms

Algorithm	Descriptions
Decision Tree	An upside-down-tree-like model 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) in order to predict the class label

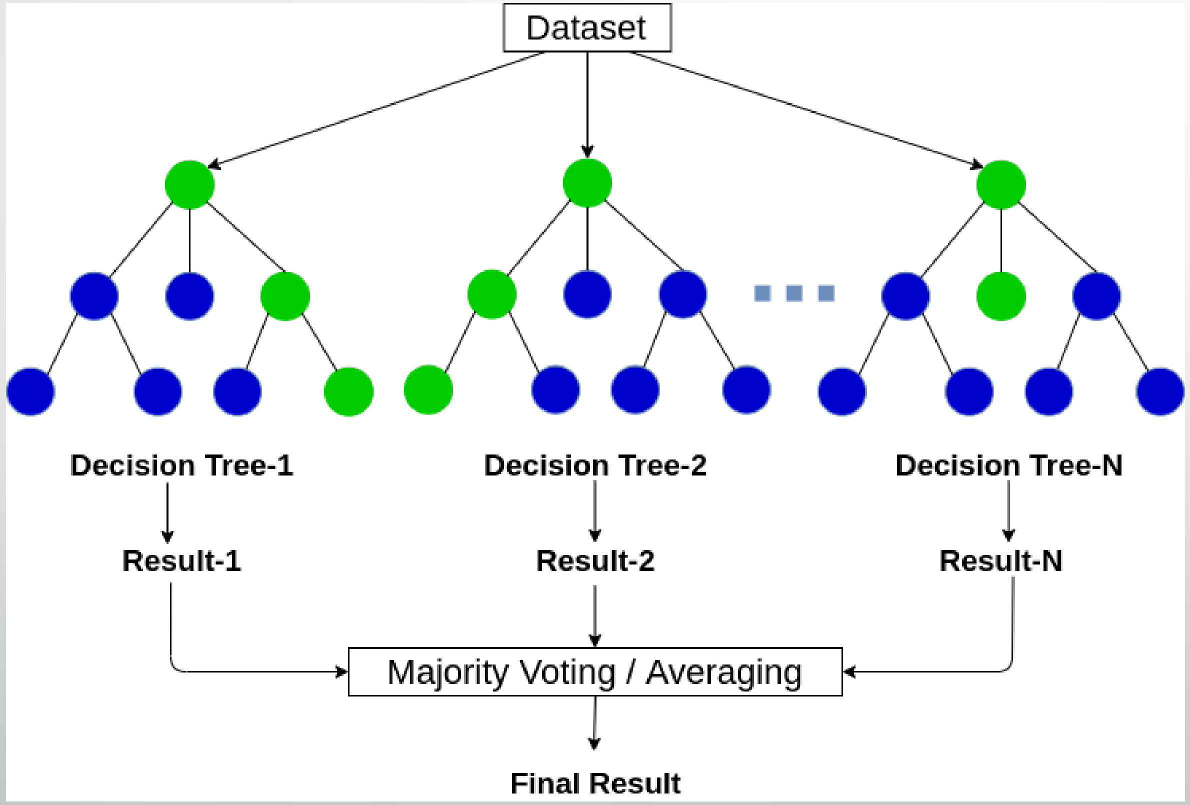
INPUT: S, where S = *set of classified instances*
OUTPUT: *Decision Tree*
Require: $S \neq \varnothing$, $num\_attributes > 0$

```
procedure BUILDTREE
    repeat
        maxGain ← 0
        splitA ← null
        e ← Entropy(Attributes)
        for all Attributes a in S do
            gain ← InformationGain(a, e)
            if gain > maxGain then
                maxGain ← gain
                splitA ← a
            end if
        end for
        Partition(S, splitA)
    until all partitions processed
end procedure
```



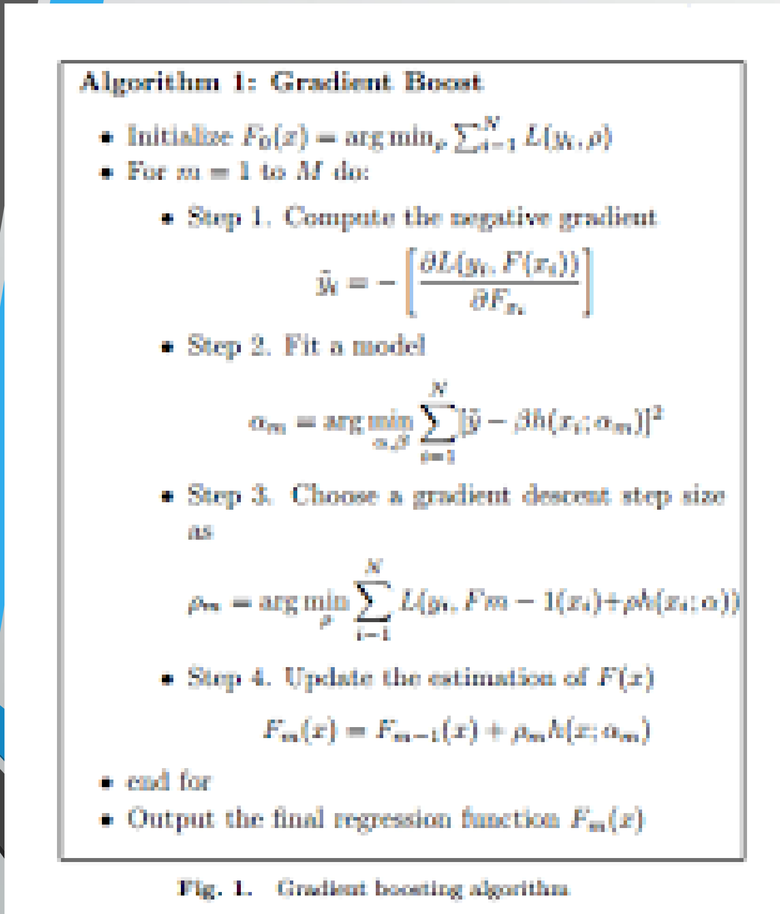
Machine learning algorithms

Algorithm	Descriptions
Random Forest	An algorithm that produces many decision trees. Each individual tree in the random forest makes a class prediction and the class with the most votes becomes the model's prediction



Machine learning algorithms

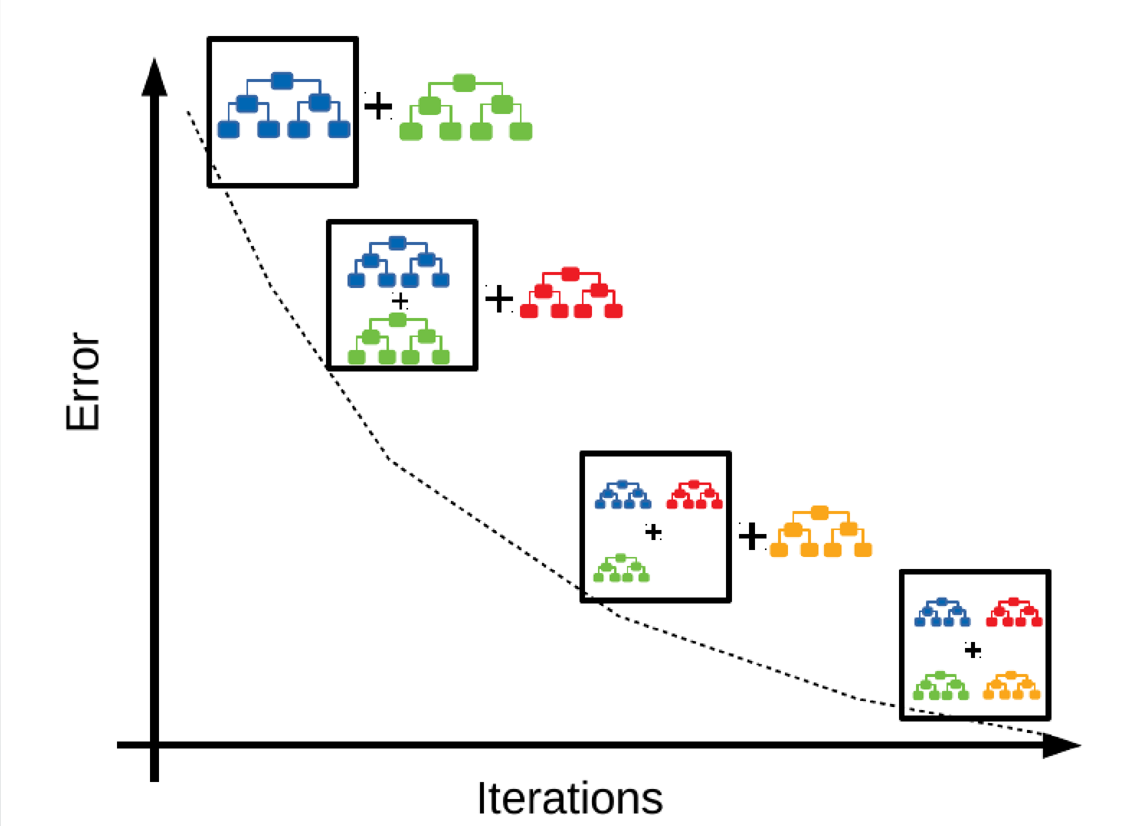
Algorithm	Descriptions
Gradient Boosting	An algorithm that produces a strong prediction model in the form of an ensemble of weak prediction models, often in the form of decision trees.



Algorithm 1: Gradient Boost

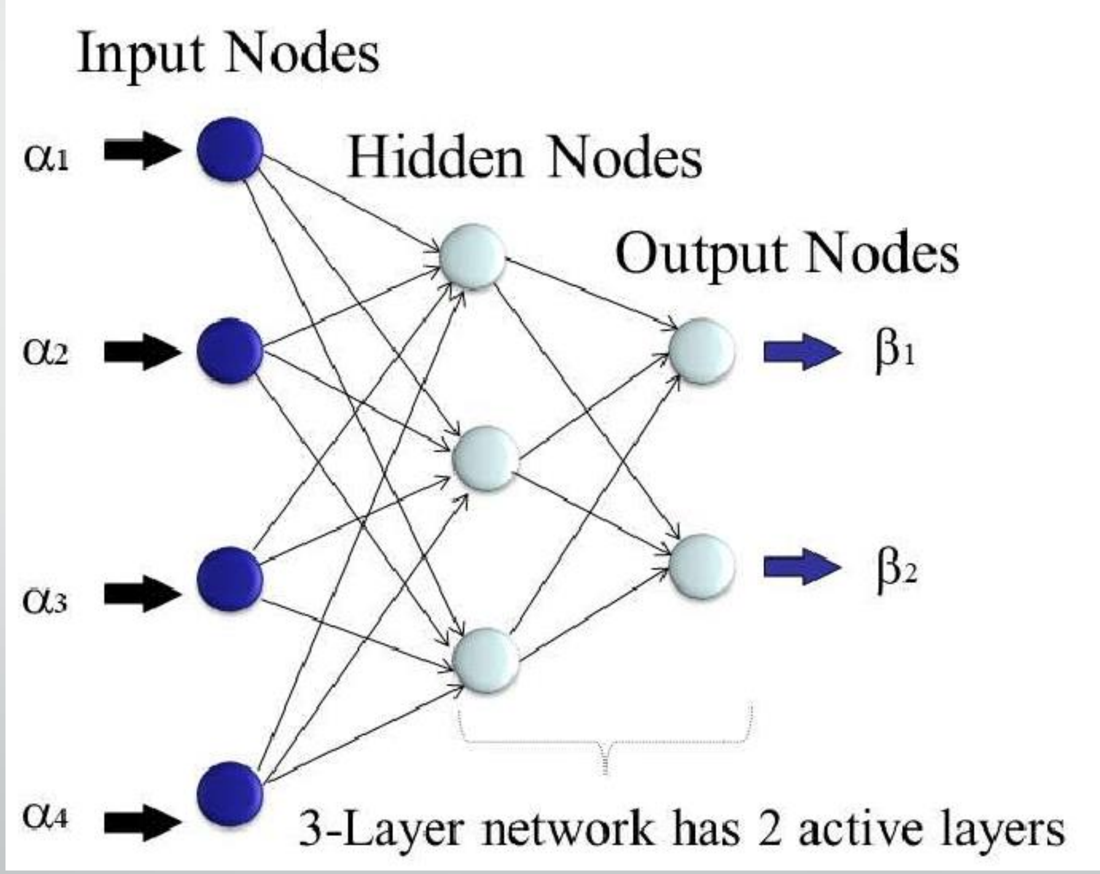
- Initialize $F_0(x) = \arg\min_\rho \sum_{i=1}^{N} L(y_i, \rho)$
- For $m = 1$ to M do:
  - Step 1. Compute the negative gradient
    $$\tilde{y}_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F_{x_i}}\right]$$
  - Step 2. Fit a model
    $$\alpha_m = \arg\min_{\alpha, \beta} \sum_{i=1}^{N} [\tilde{y} - \beta h(x_i; \alpha_m)]^2$$
  - Step 3. Choose a gradient descent step size as
    $$\rho_m = \arg\min_\rho \sum_{i=1}^{N} L(y_i, Fm-1(x_i) + \rho h(x_i; \alpha))$$
  - Step 4. Update the estimation of $F(x)$
    $$F_m(x) = F_{m-1}(x) + \rho_m h(x; \alpha_m)$$
- end for
- Output the final regression function $F_m(x)$

Fig. 1. Gradient boosting algorithm

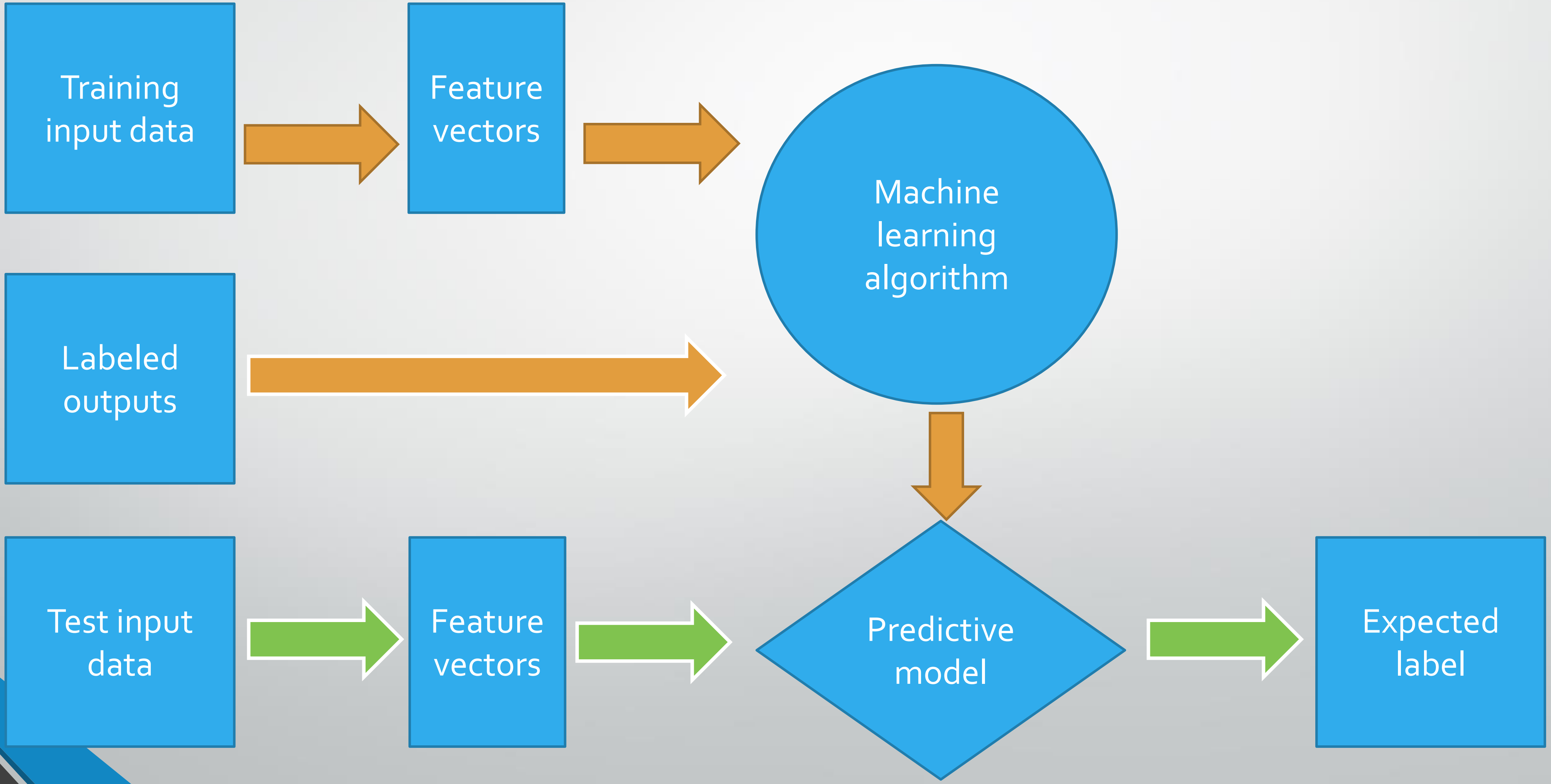


Machine learning algorithms

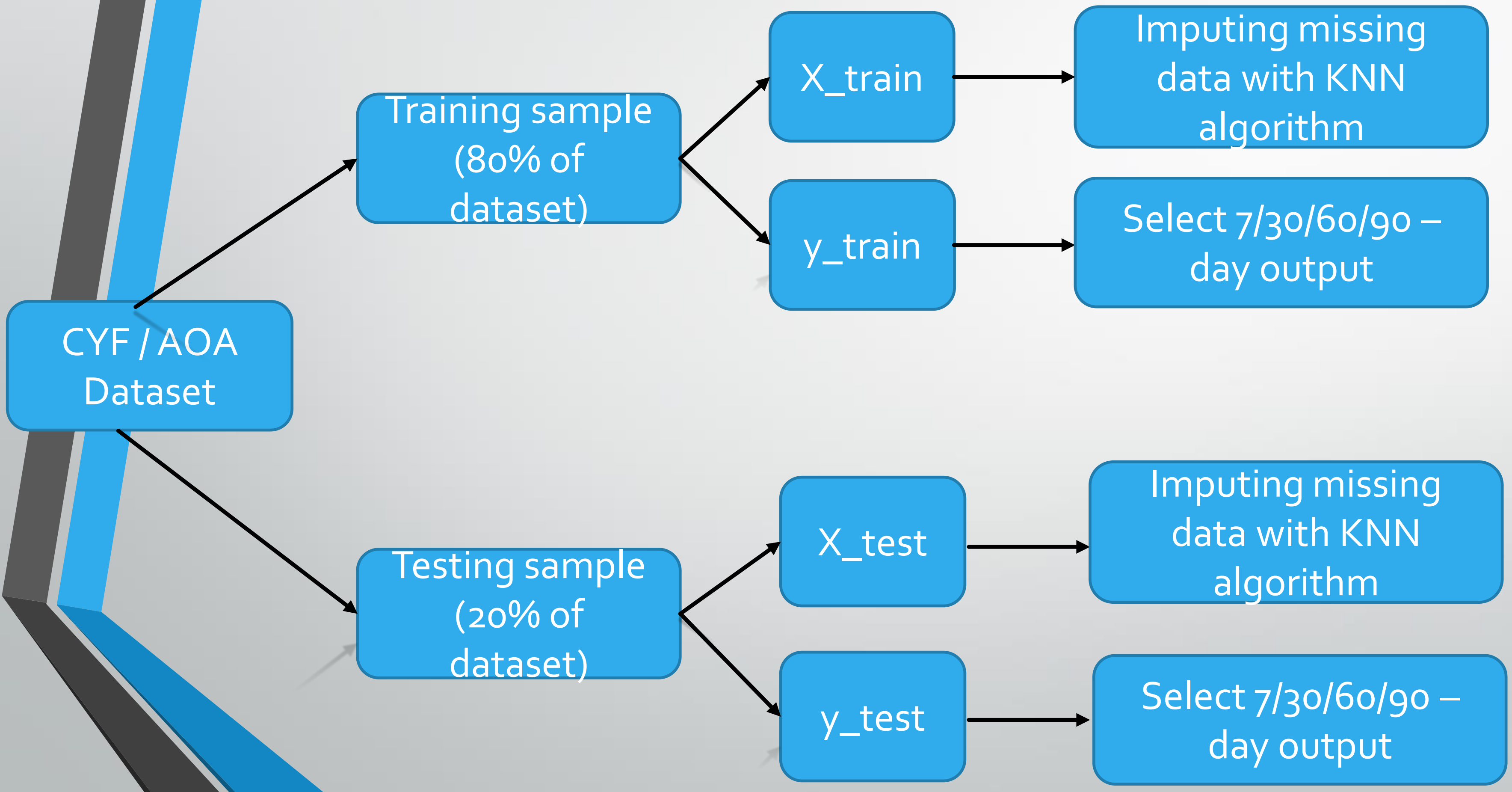
Algorithm	Descriptions
Artificial Neural Networks	A model inspired by the biological neural networks of the brain to learn tasks based on provided examples, without being explicitly programmed



Predictive model pipeline - Supervised Learning Models

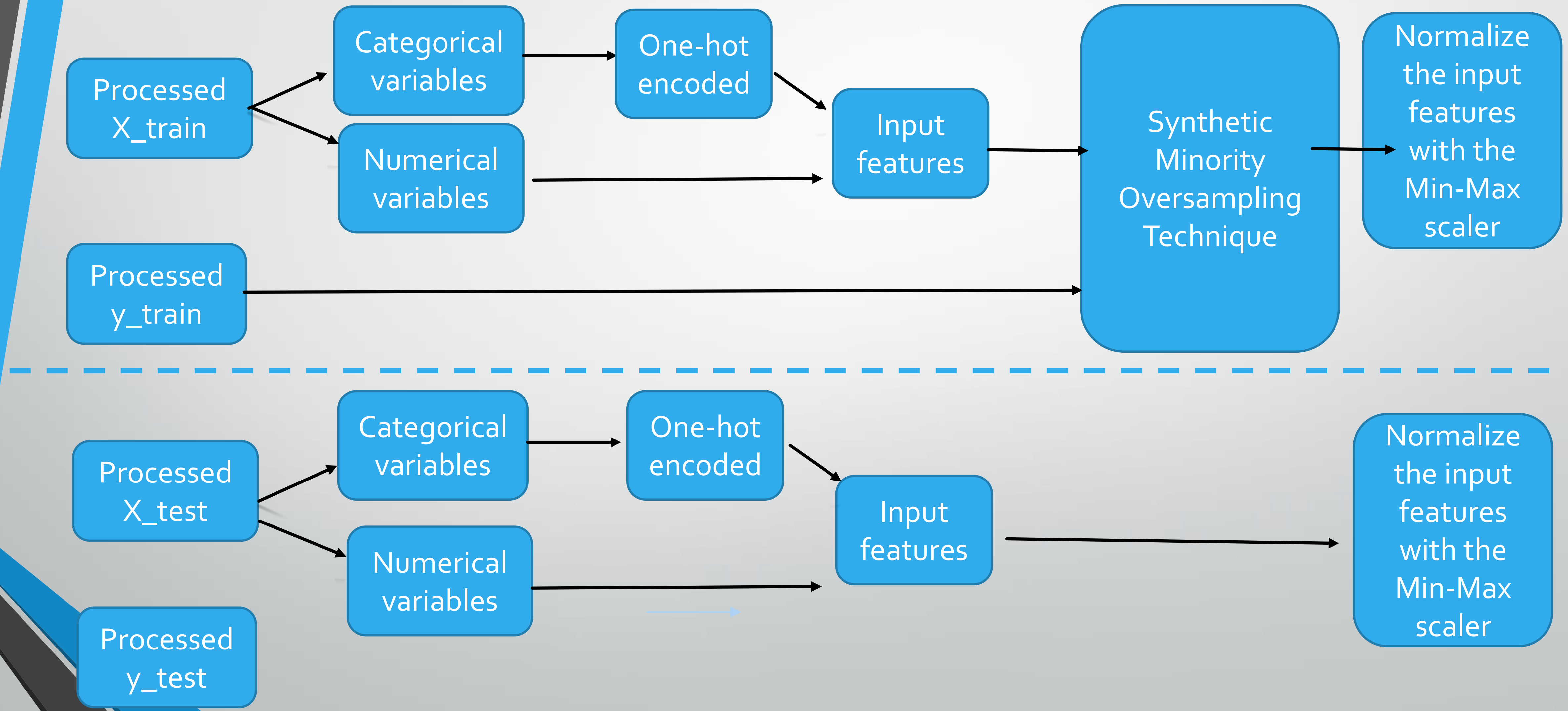


Predictive model pipeline – Feature Engineering



X: input/predictors
y: output

Predictive model pipeline – Feature Engineering (continued)

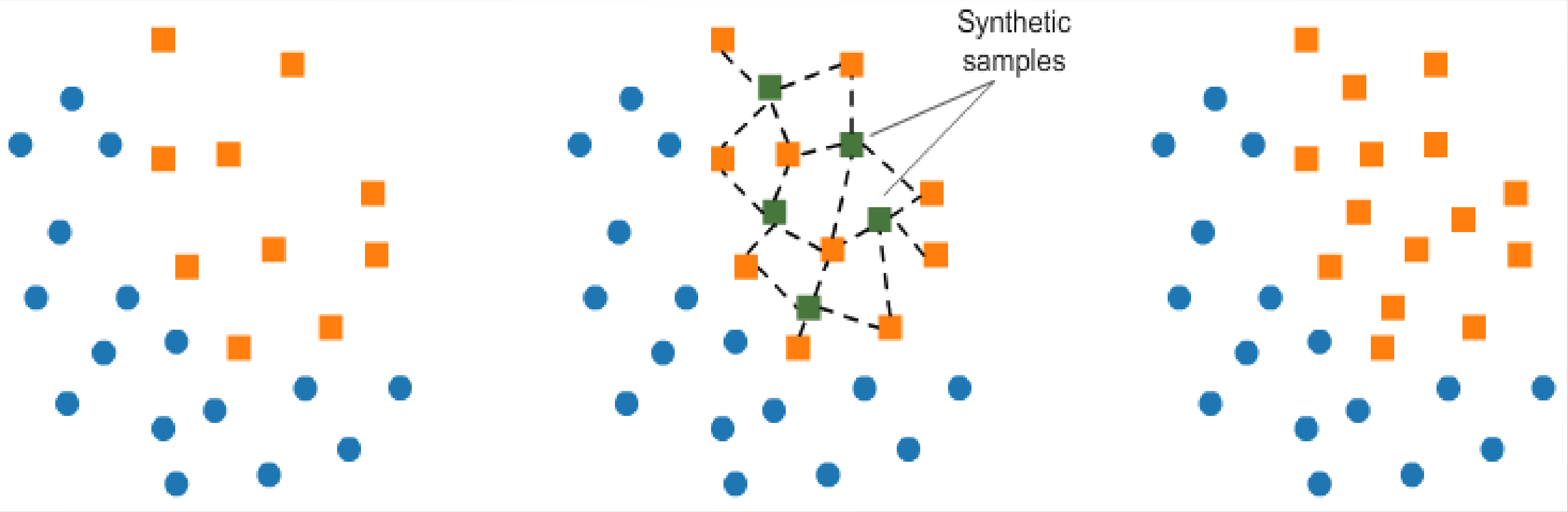


One-hot encoded example

Color
Red
Red
Yellow
Green
Yellow

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1

SMOTE example

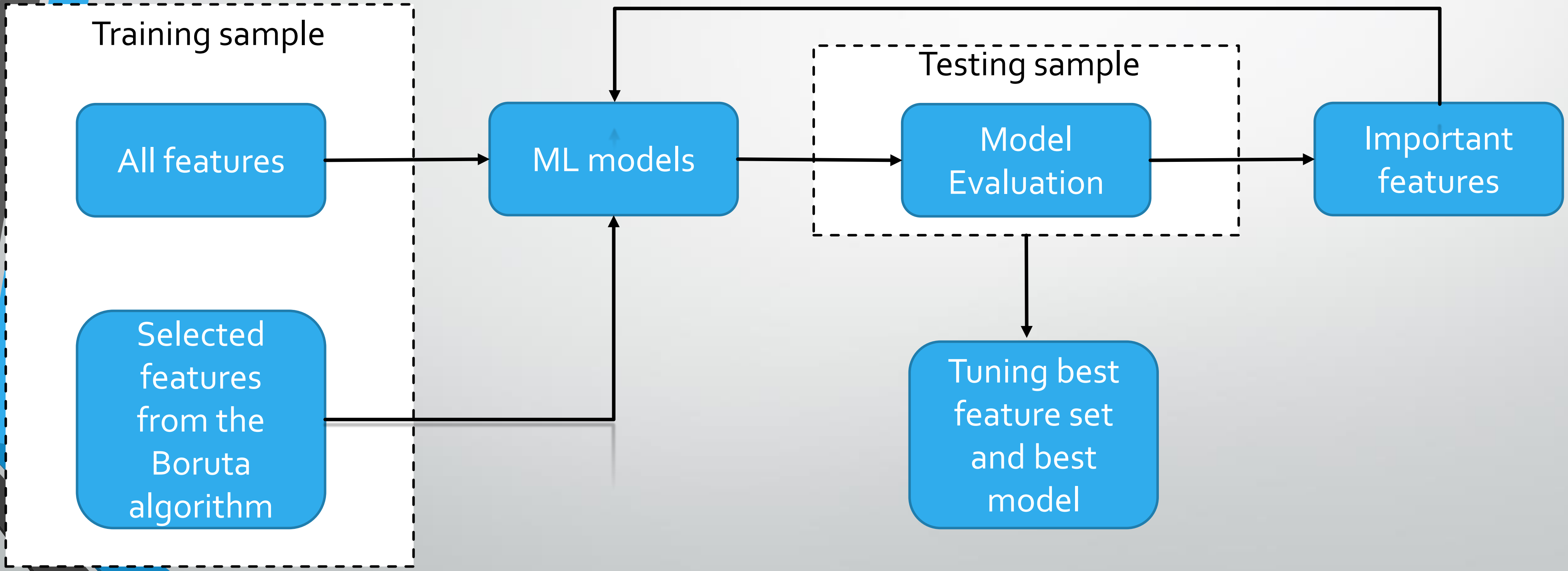


Min-Max scaler formula

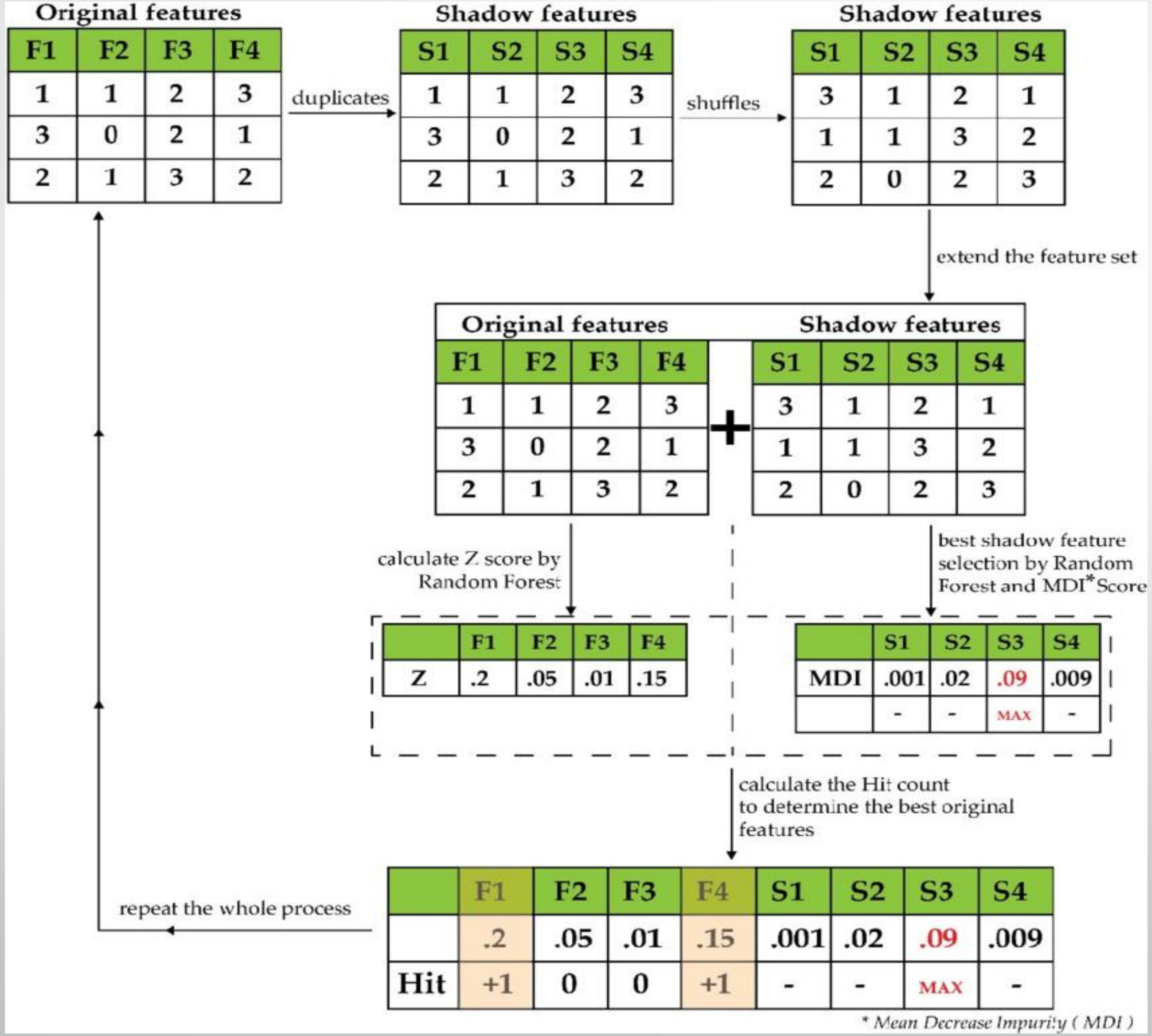
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Predictive model pipeline – Modelling

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.



Boruta Algorithm example



		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

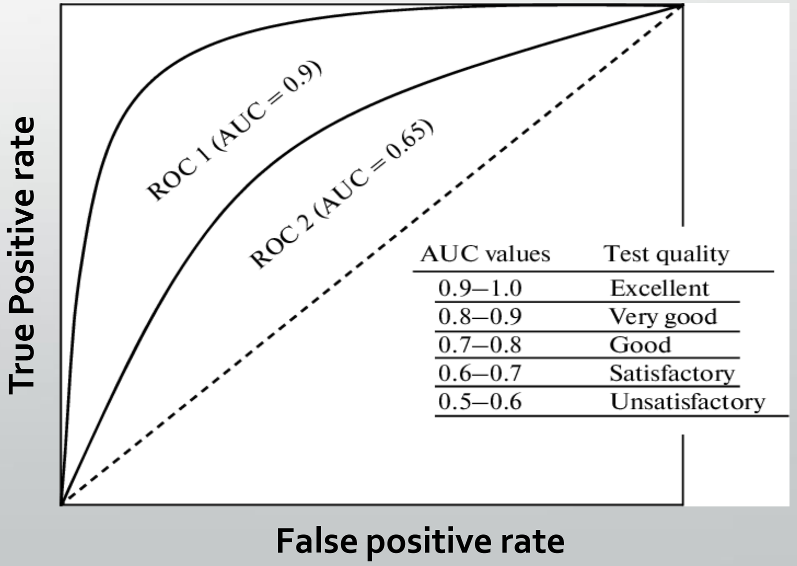
$$Accuracy = \frac{True\ Postive + True\ Negative}{Total\ (Positive + Negative)}$$

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

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

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

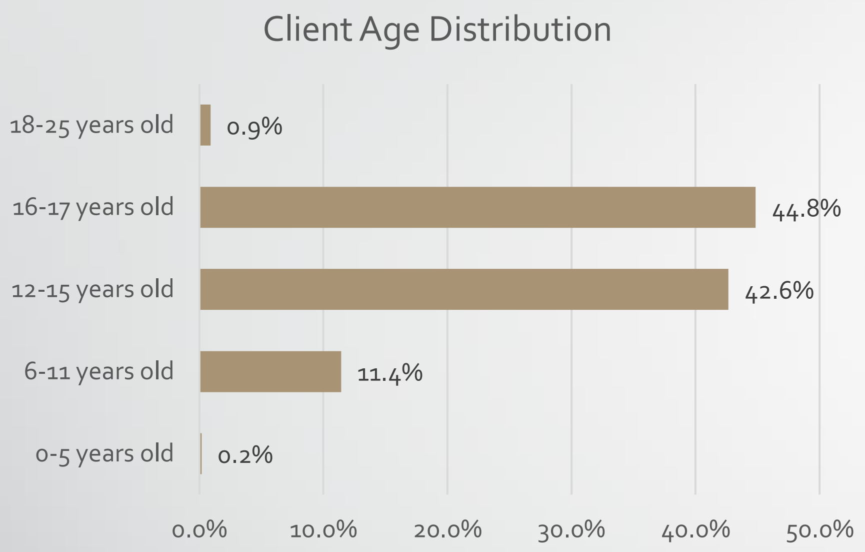
The AUC-ROC (area under the ROC curve), or **AUC** for short, is a performance measurement for the classification problems at various threshold settings. ROC (Receiver Operating Characteristics) is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes.

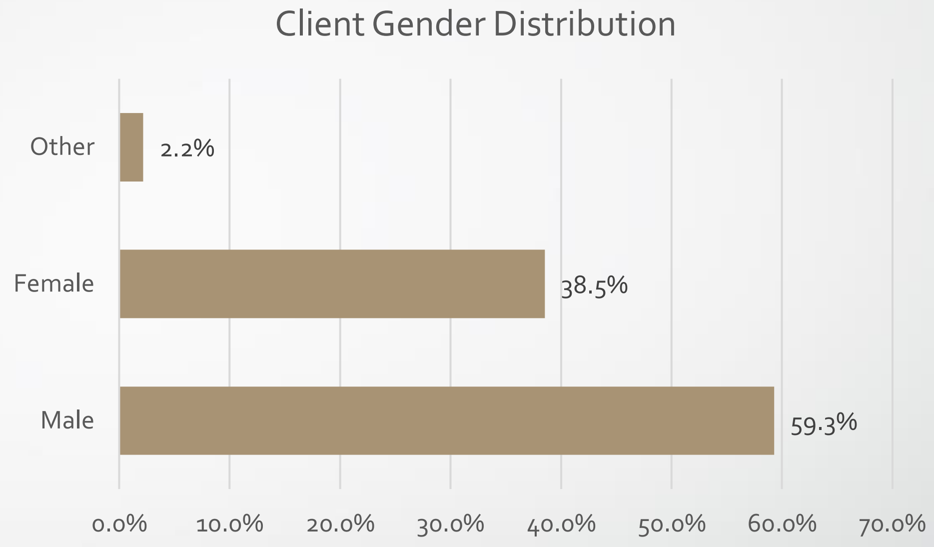


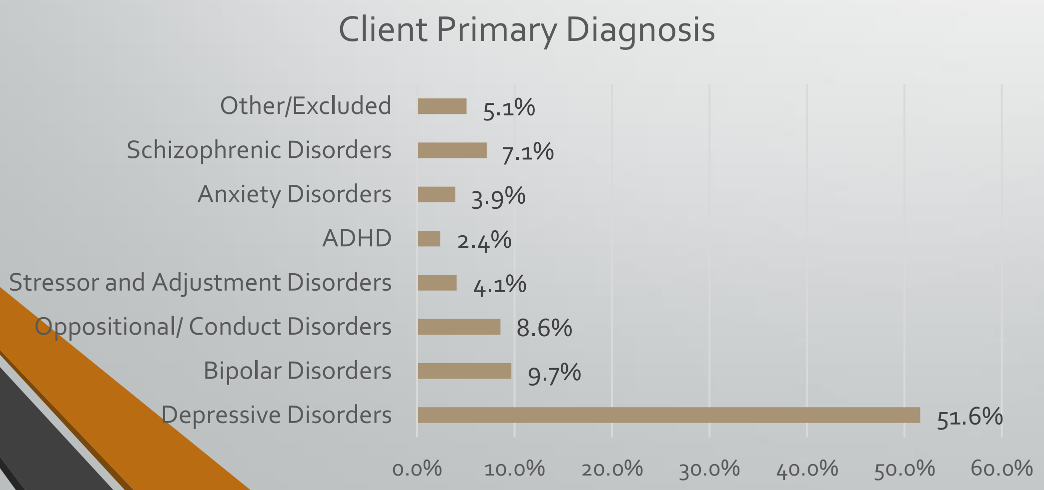
Results

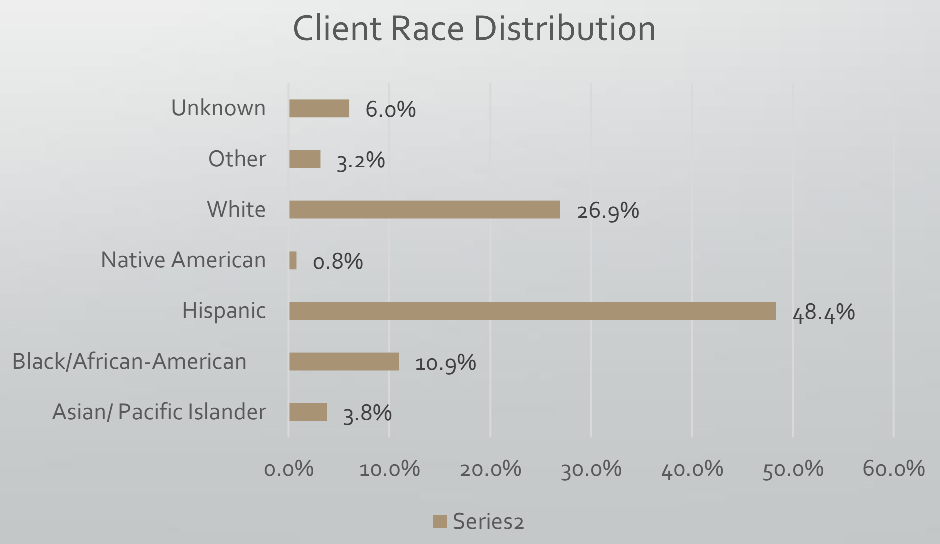
CYF Findings

CYF demographics
(Total unique clients: 6,310)

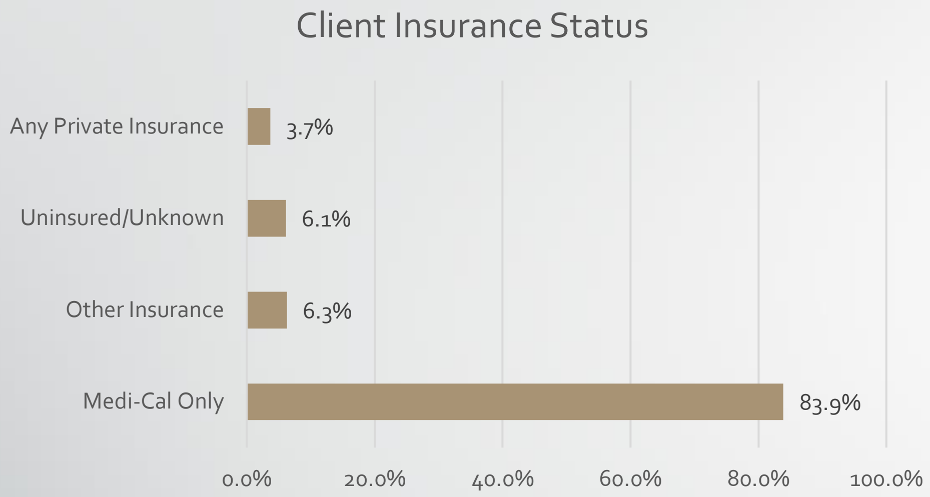


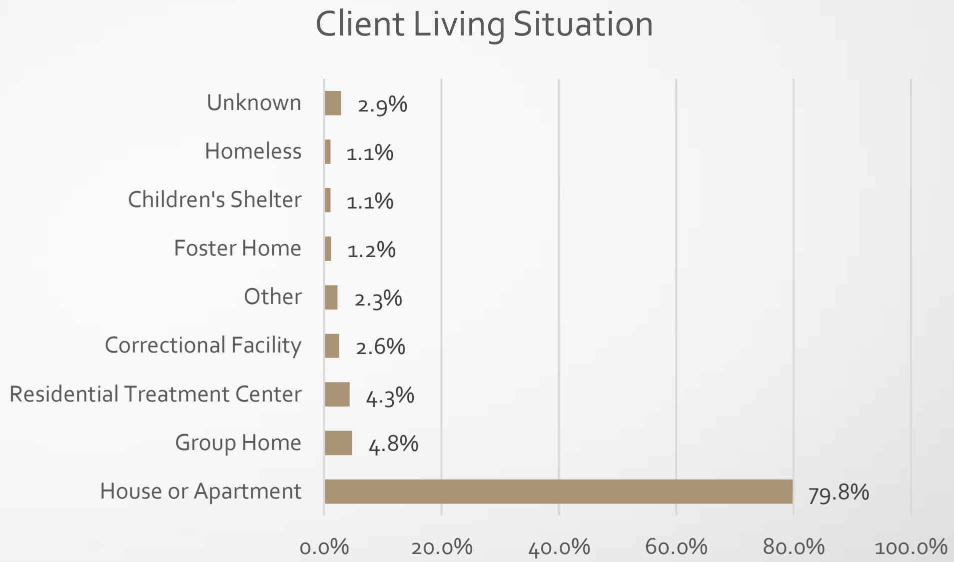


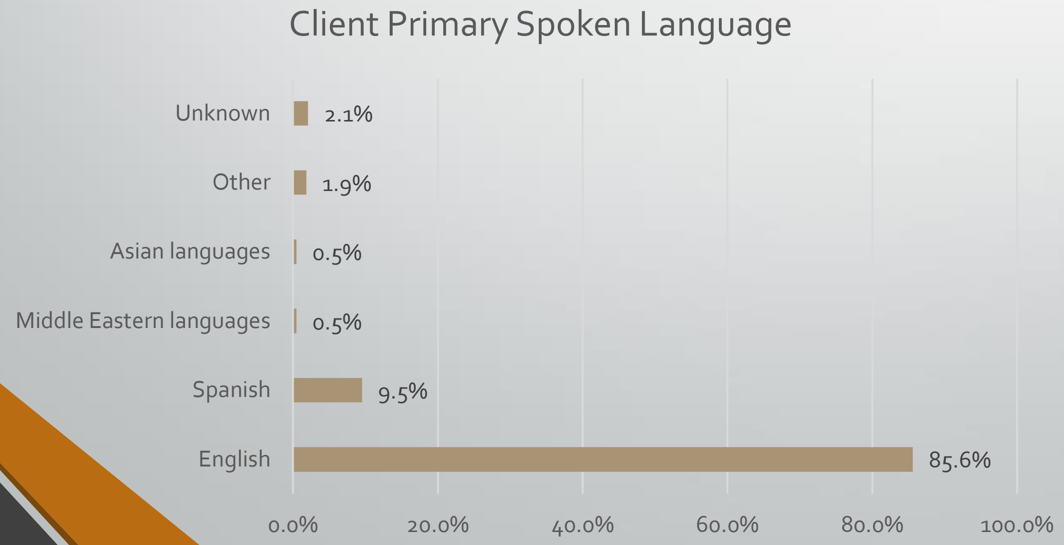


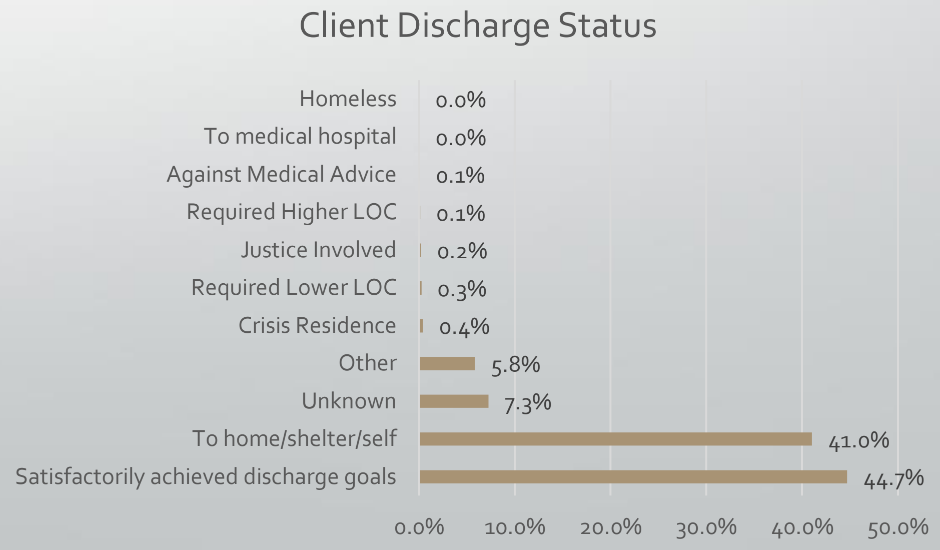


CYF demographics
(Total unique clients: 6,310)

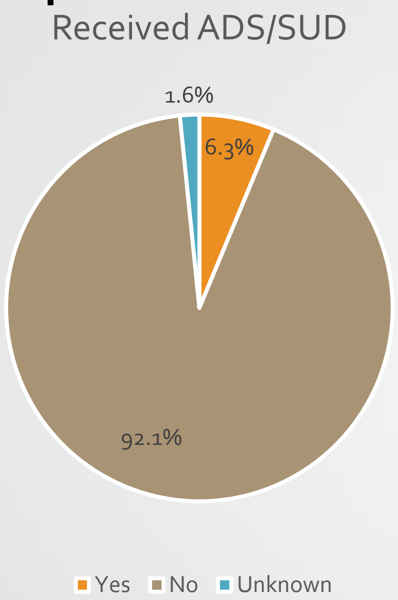


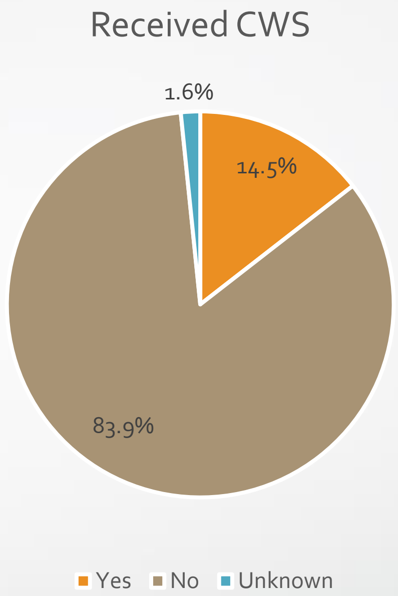


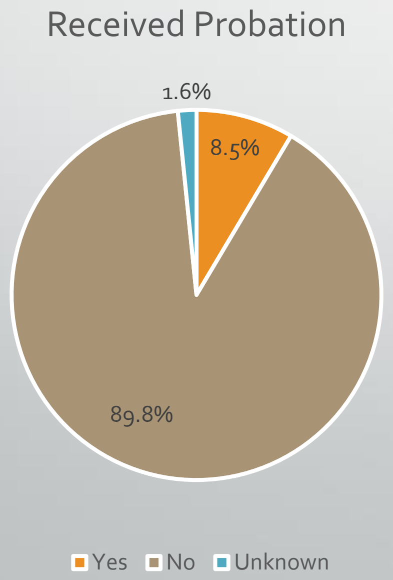


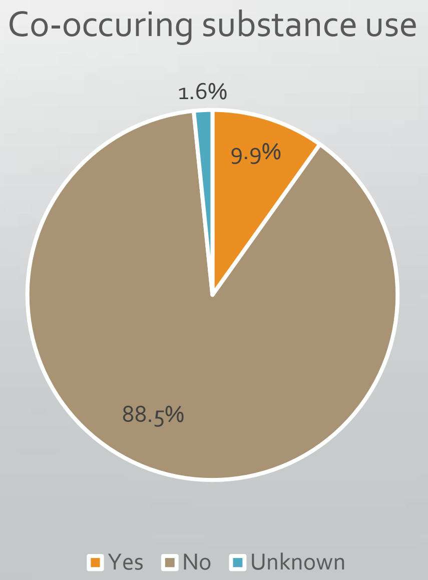


CYF demographics
(Total unique clients: 6,310)









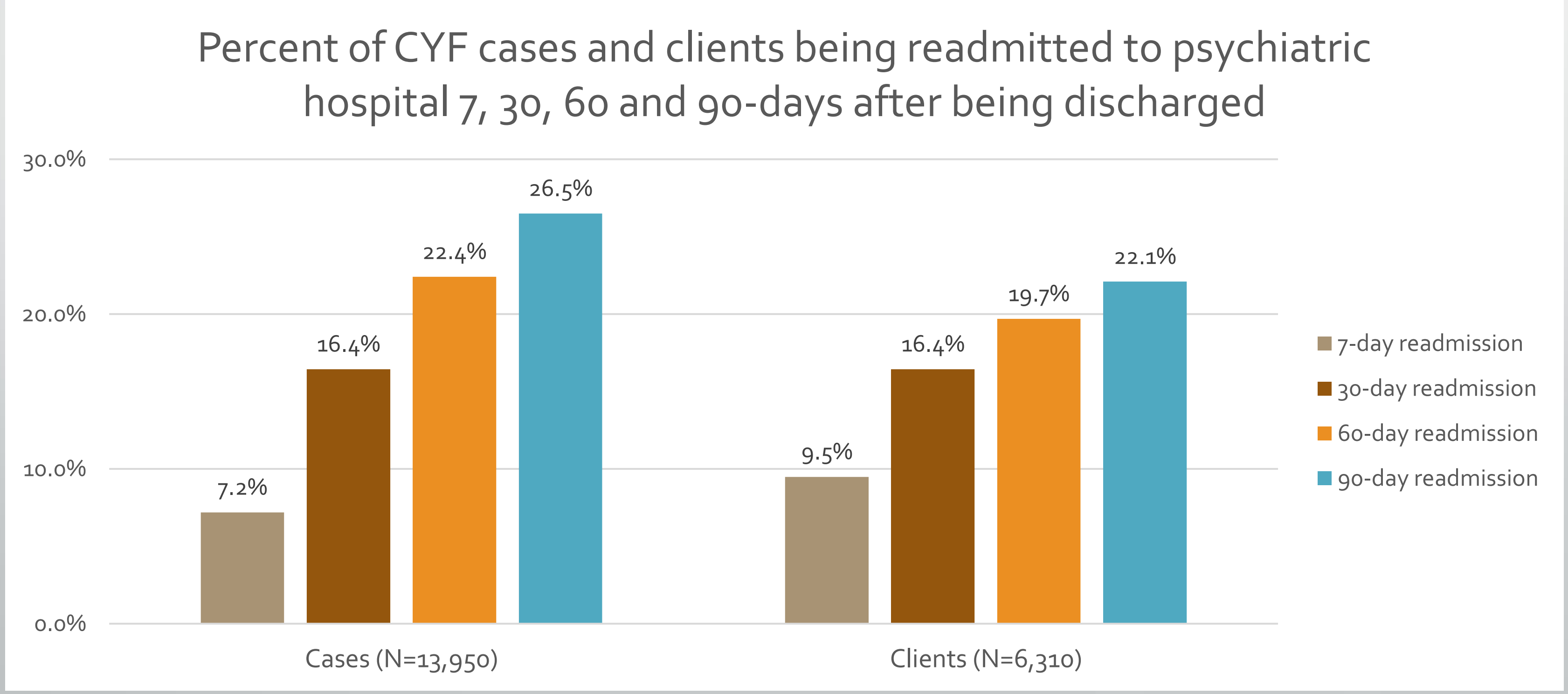
CYF services

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/ program.

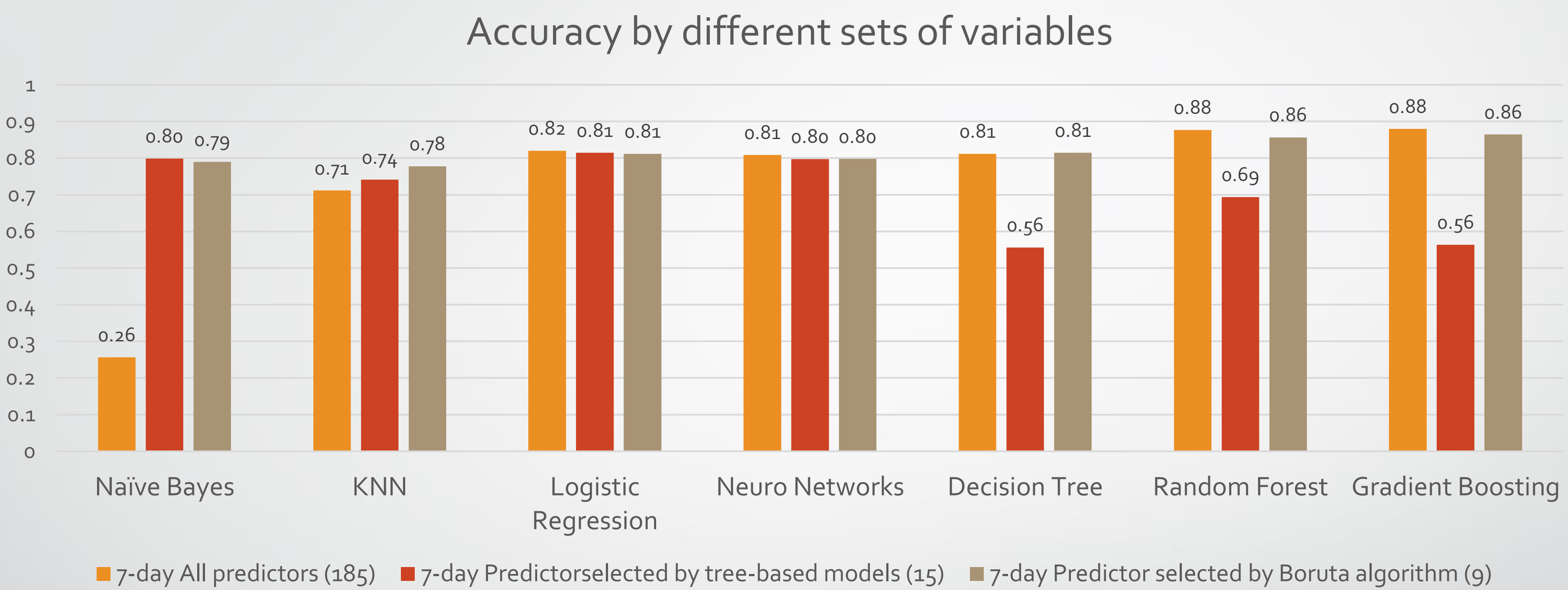
Group of Services	Metrics	Within 90 days	Within 60 days	Within 30 days	Within 7 days
Outpatient Services	Total visits:	219,817	157,980	88,538	26,674
	Total clients:	257	278	371	622
	Percent user:	4.1	4.4	5.9	9.9
Emergency Services	Total visits:	56,466	48,457	38,954	29,094
	Total clients:	322	334	344	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

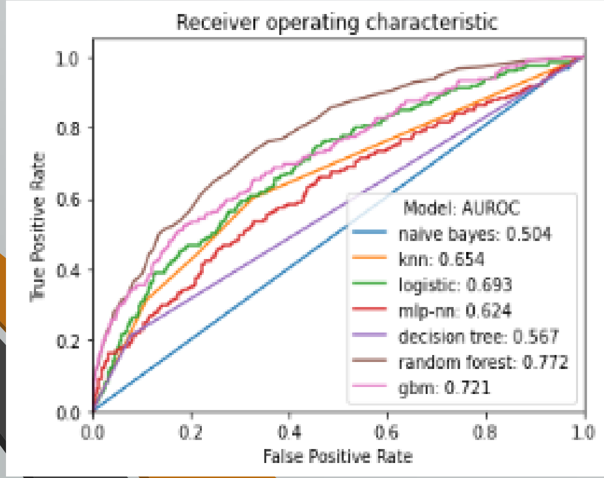
CYF hospitalization and readmission rates

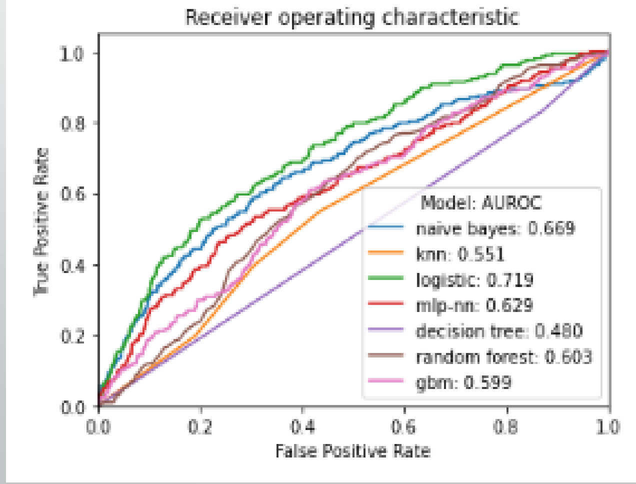
The current CYF client sample has 1.3 psychiatric hospitalization (3.2 standard deviation) in average. The average length of stay in the psychiatric hospital is 6.6 days (18.4 days standard deviation). The gap between hospitalizations is 505.2 days or about 1.4 years.

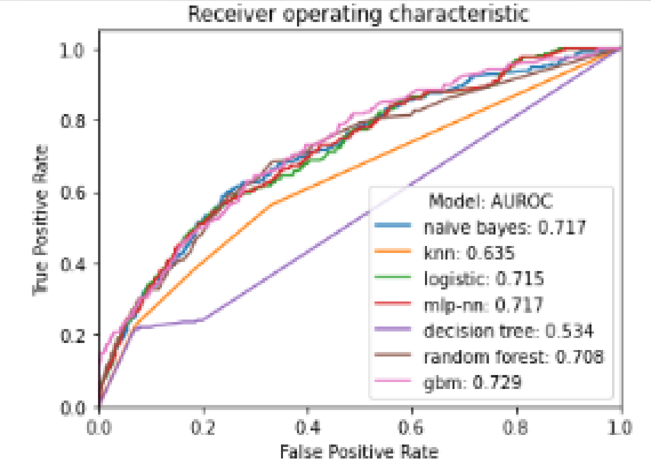


7-day readmission prediction

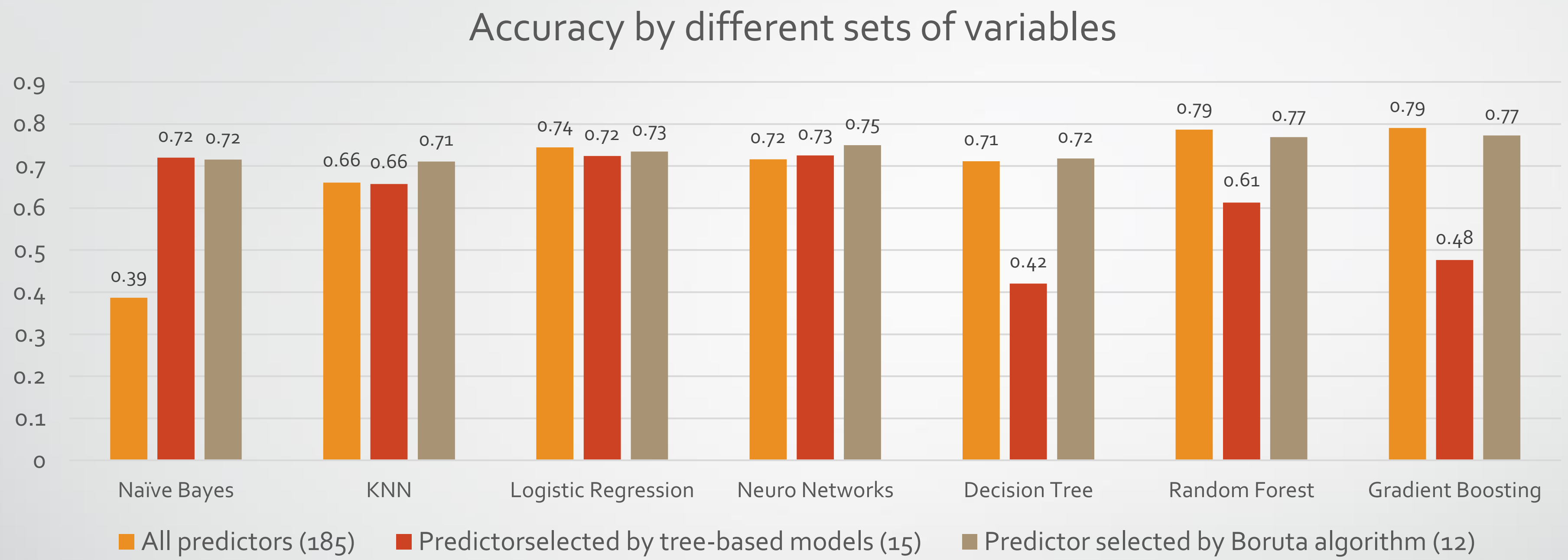


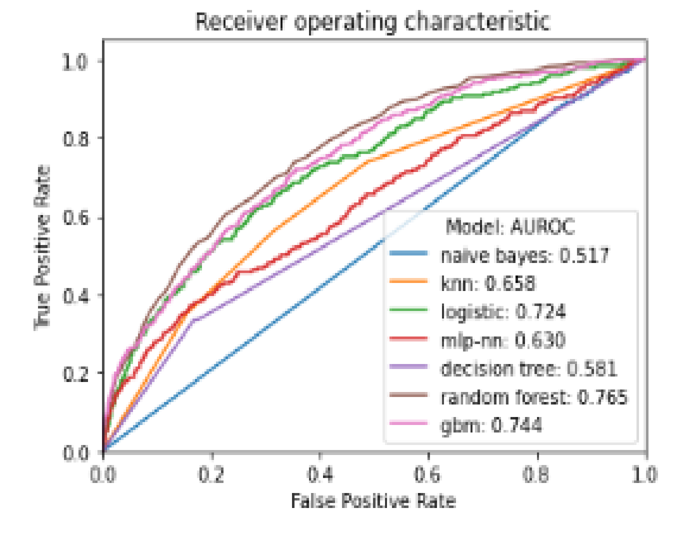


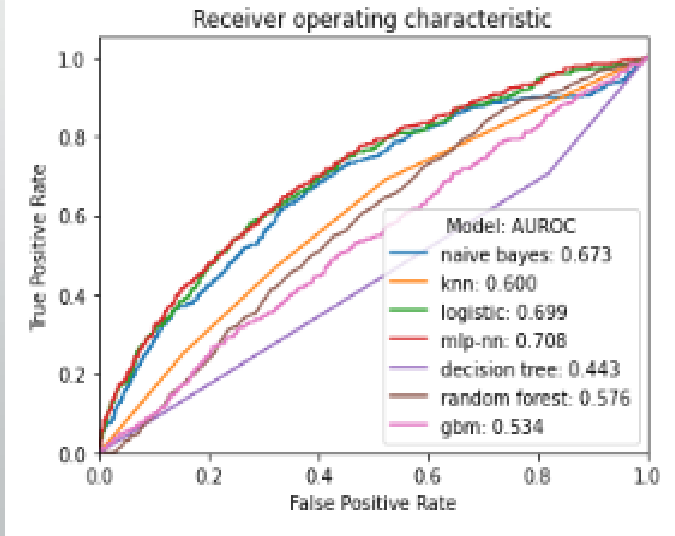


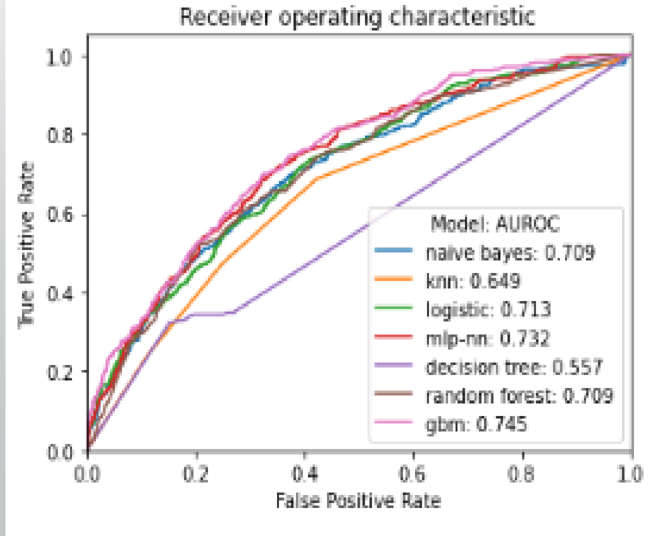


30-day readmission prediction

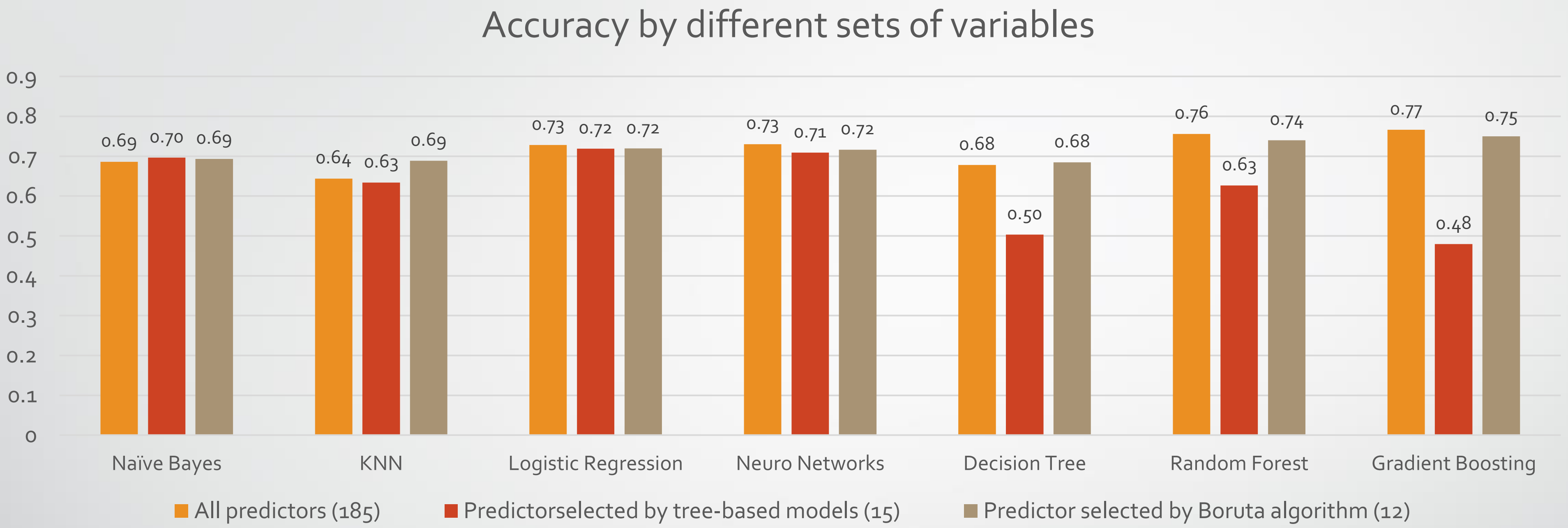


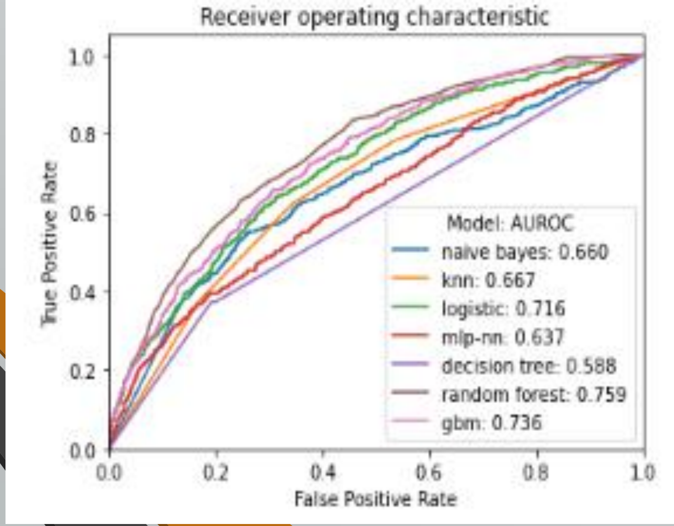


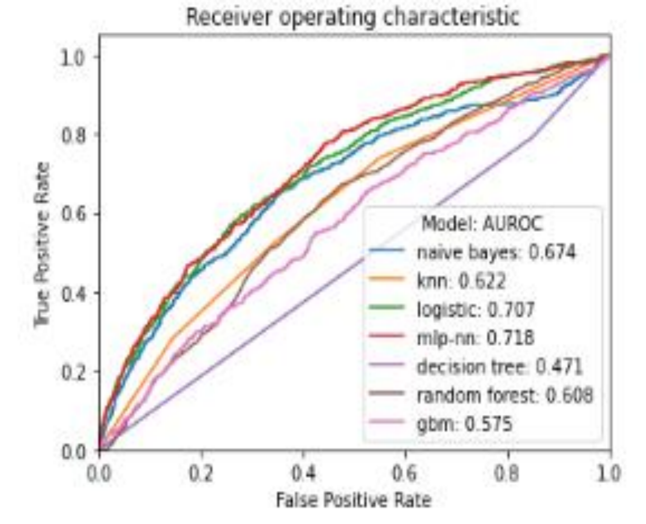


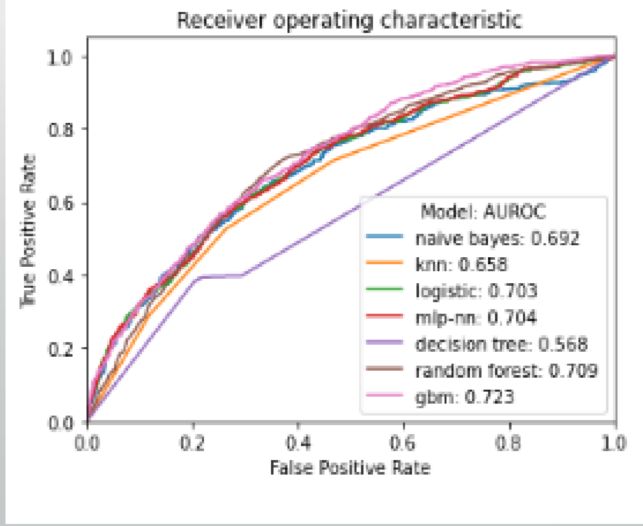


6o-day readmission prediction

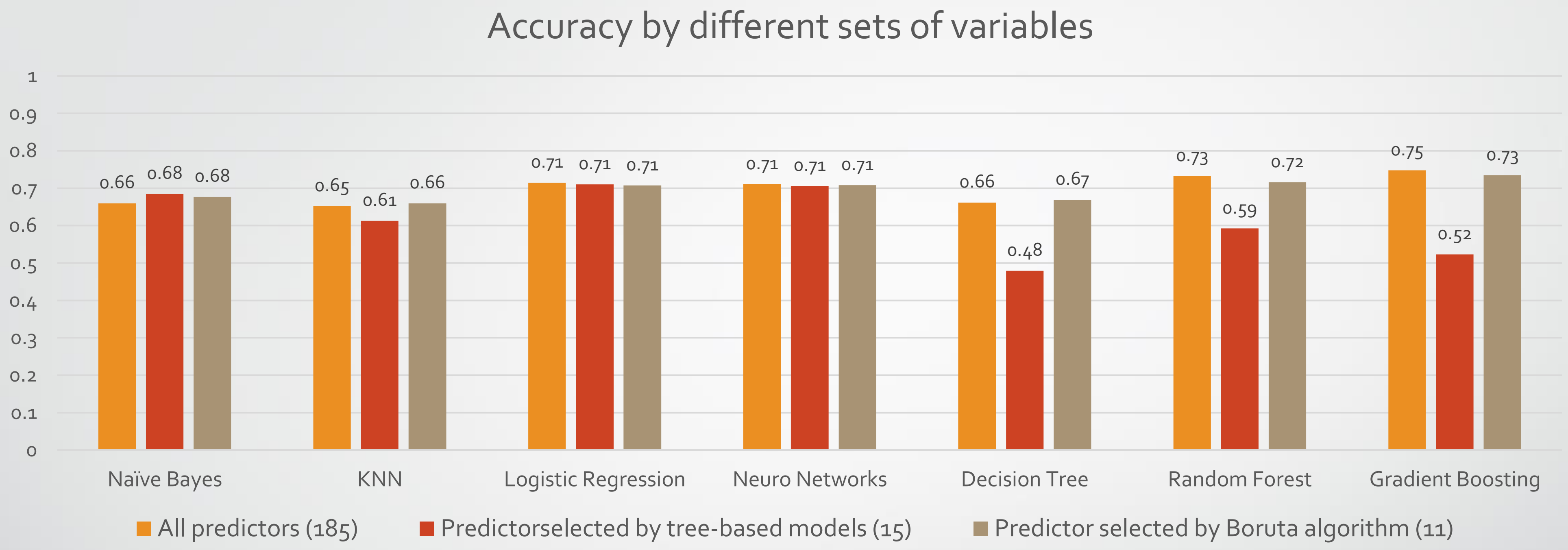


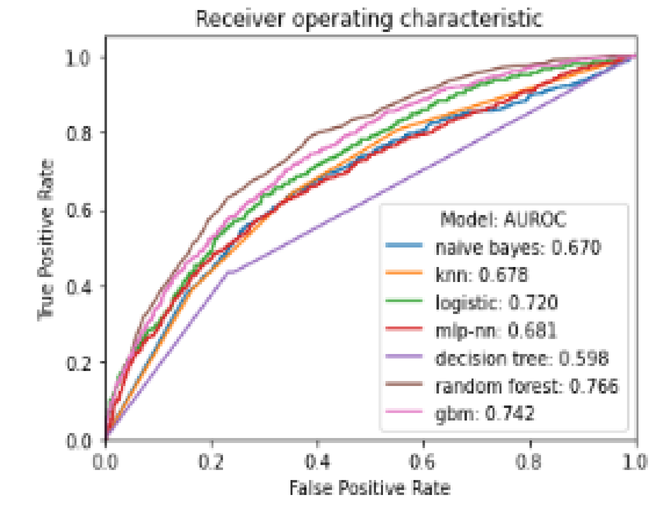


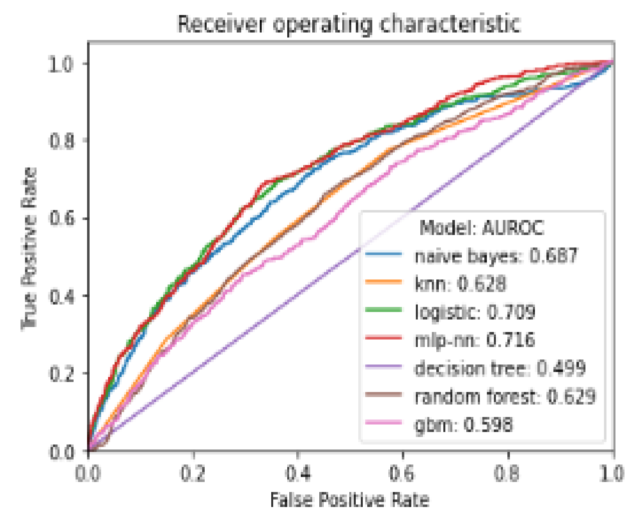


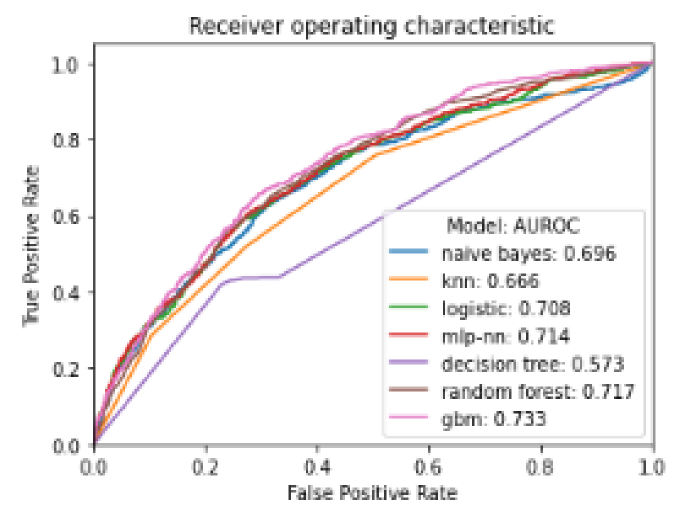


90-day readmission prediction







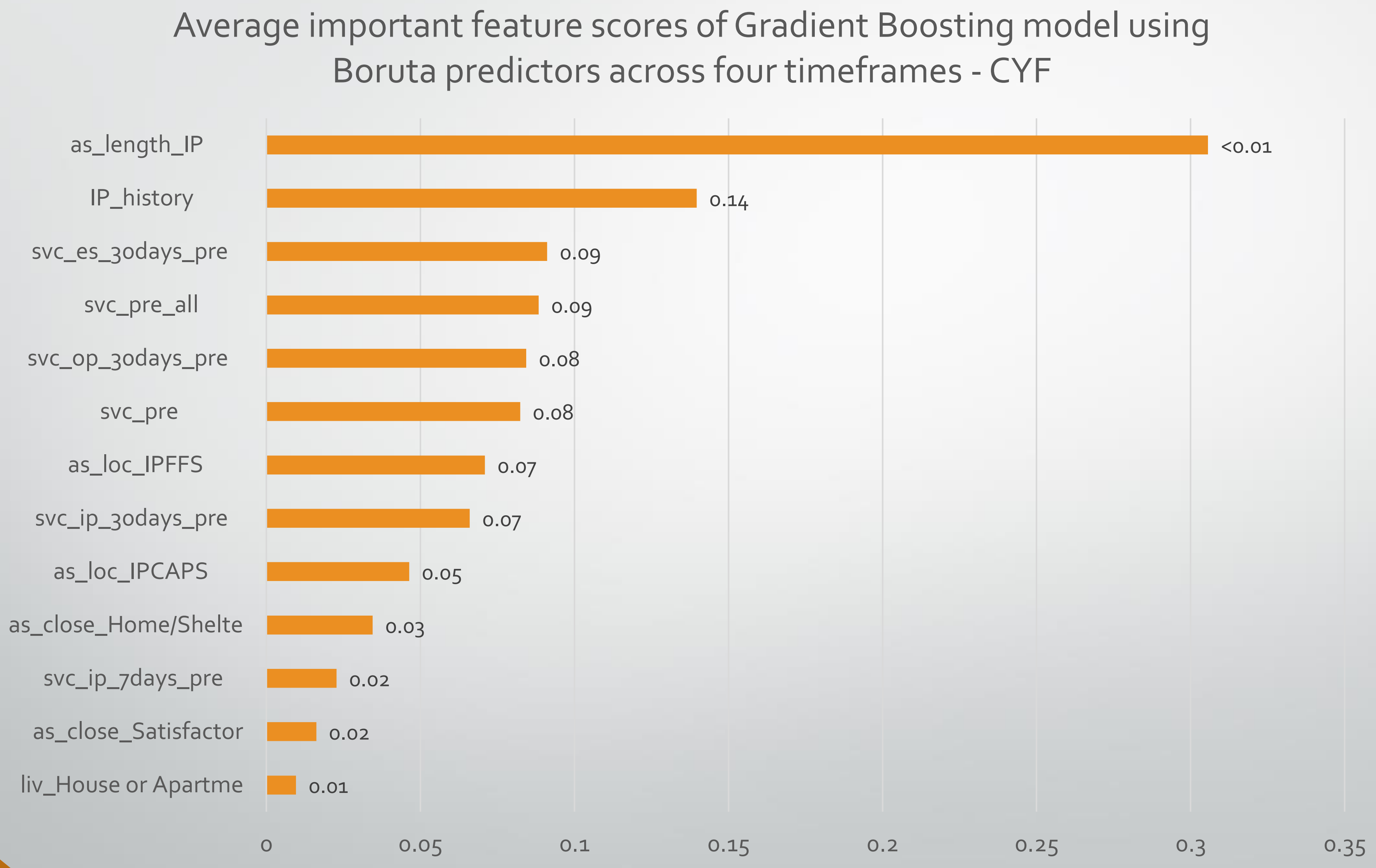


Final set of variables and model

The final model 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)

	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Learning rate	0.1	0.1	0.1	0.1
N_estimators	1000	500	500	1000
Max_depth	7	7	4	7
Accuracy	0.8680	0.7789	0.7653	0.7174
AUC	0.8090	0.7824	0.8077	0.7583



Average important feature scores of Gradient Boosting model using Boruta predictors across four timeframes - CYF
as_length_IP
IP_history
svc_es_30days_pre
svc_pre_all
svc_op_30days_pre
svc_pre
as_loc_IPFFS
svc_ip_30days_pre
as_loc_IPCAPS
as_close_Home/Shelte
svc_ip_7days_pre
as_close_Satisfactor
liv_House or Apartme
<0.01
0.14
0.09
0.09
0.08
0.08
0.07
0.07
0.05
0.03
0.02
0.02
0.01
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

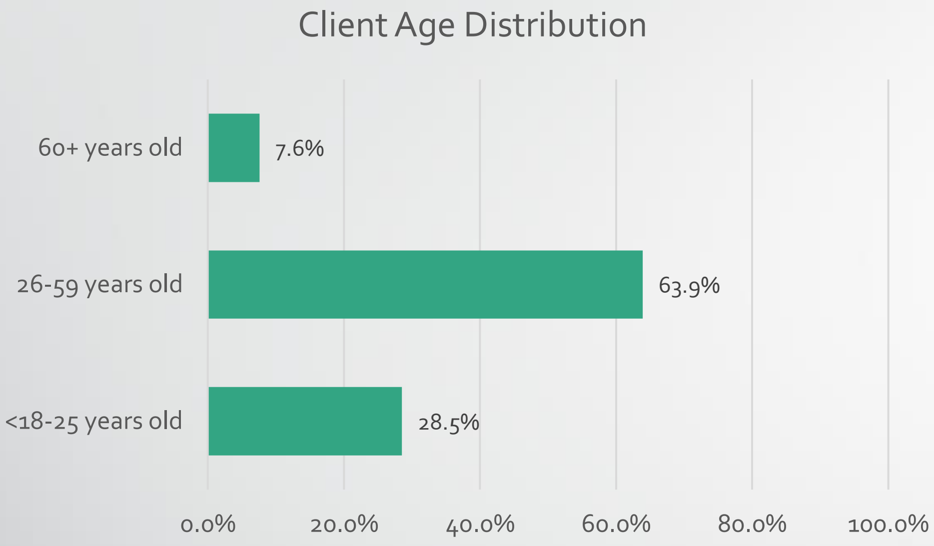
CYF coefficients (Coef) of Logistic Regression model using of Boruta predictors by timeframes

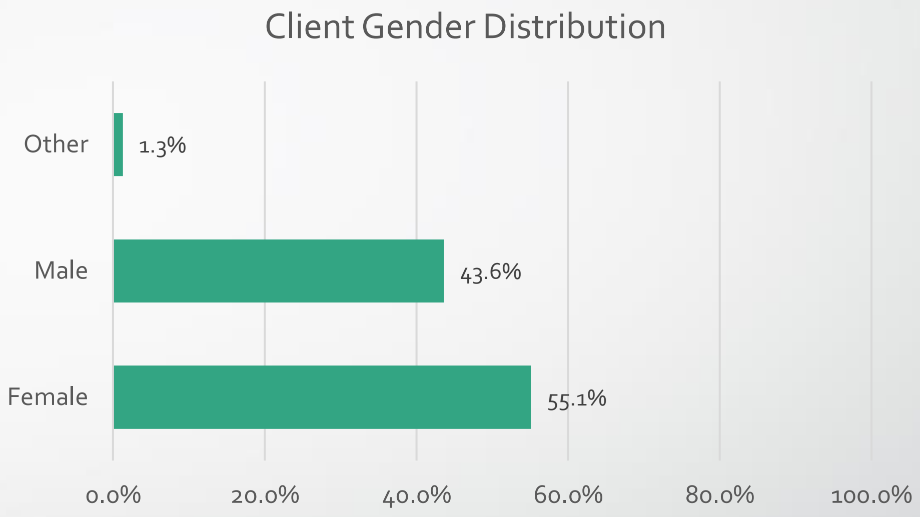
	7-day readmission	30-day readmission	60-day readmission	90-day readmission
svc_ip_30days_pre	3.799367	4.401424	4.107667	4.100226
IP_history	2.304024	2.346932	2.529734	2.668832
svc_pre_all	0.396693	0.642057	0.404826	0.384198
svc_pre	-0.015104	-1.053096	-1.314142	-0.631956
as_close_Home/Shelter	-0.491407	-0.628238	-0.226659	-0.243241
as_loc_IPCAPS	-2.148970	-1.665435	-1.369292	-1.236845
as_loc_IPFFS	-2.421381	-1.817746	-1.567102	-1.385462
as_length_IP	-2.954058	-1.632578	-1.596086	-1.110309
svc_ip_7days_pre	-3.492471	-3.465995	-3.258150	-3.341092
svc_op_30days_pre	-	1.787100	2.313307	1.905521
svc_es_30days_pre	-	1.096795	1.367971	1.157350
Discharge _Satisfactorily Achieved Goals	-	-0.438637	-	-
Live at House or Apartment	-	-	-0.304524	-
loc2_pre_None	-	-	-	-0.428993

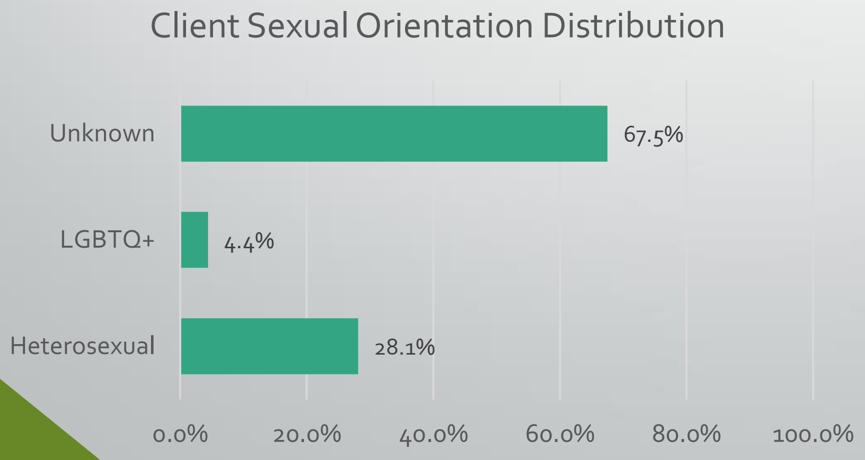
Results

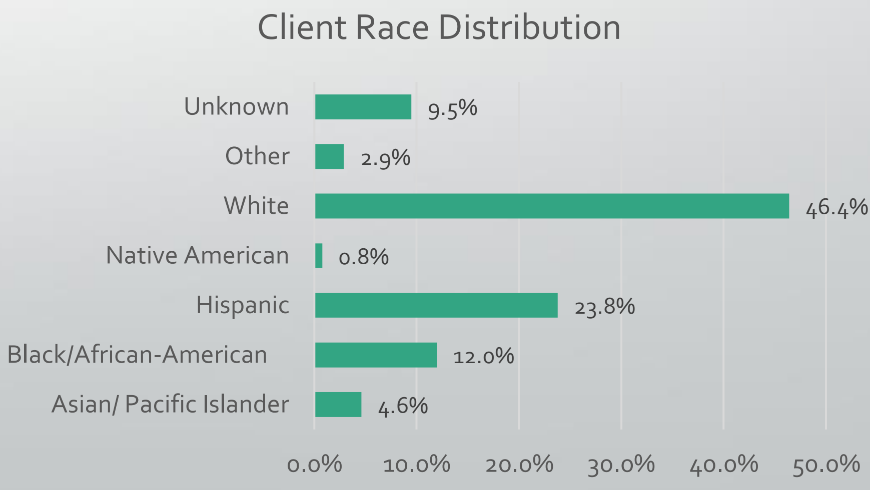
AOA Findings

AOA demographics
(Total unique clients: 38,272)

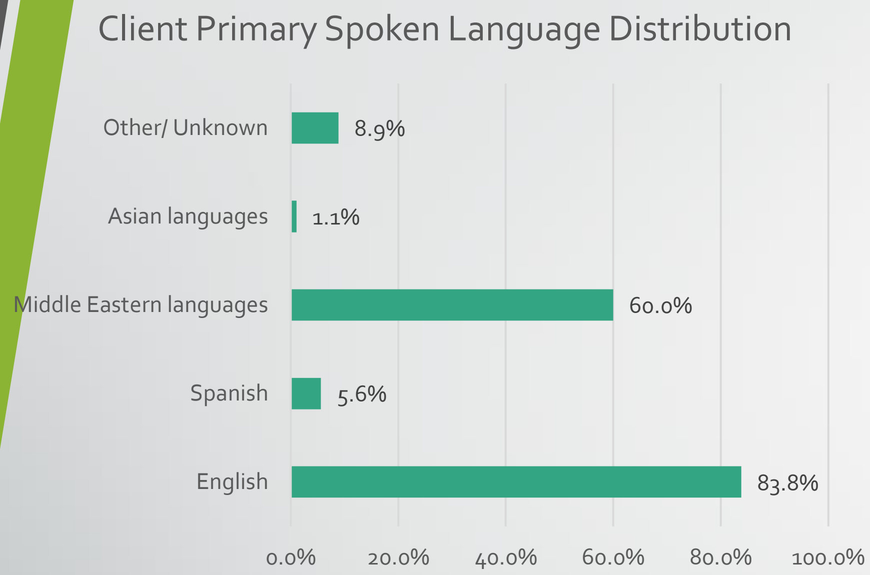


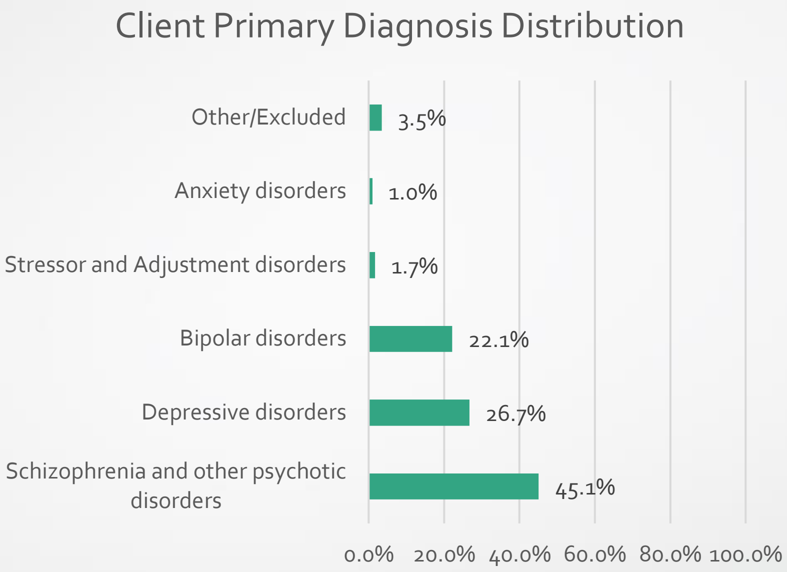


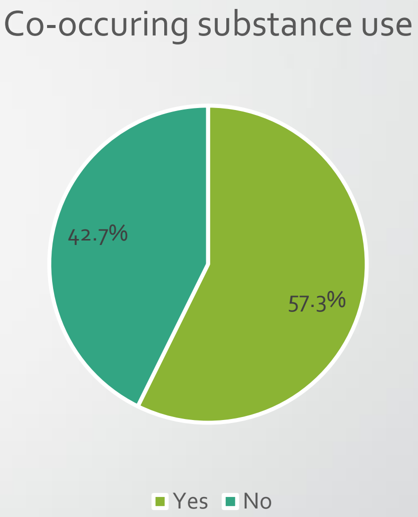


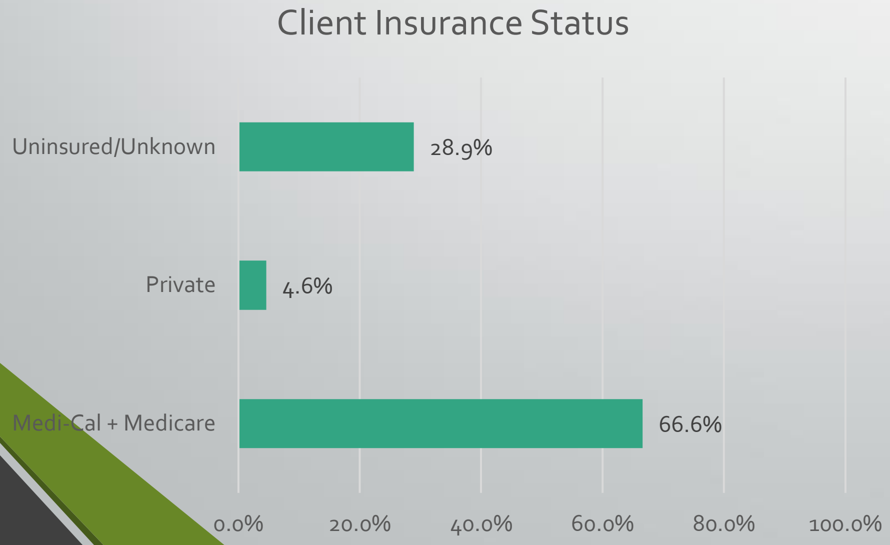


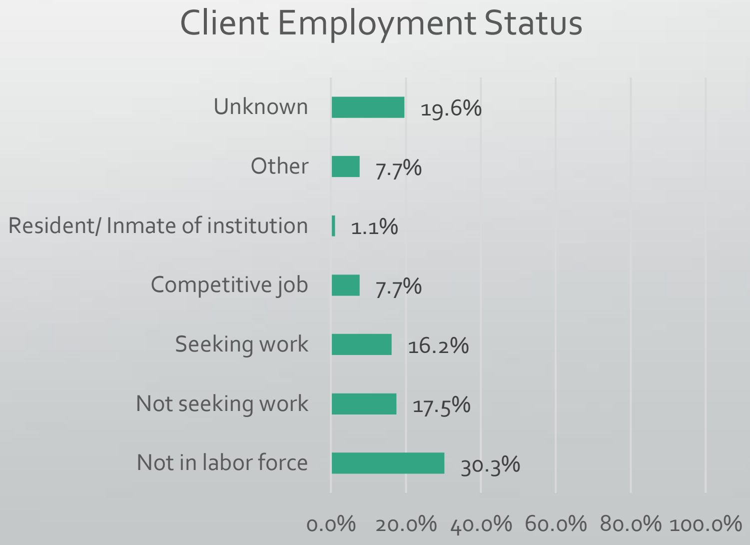
AOA demographics
(Total unique clients: 38,272)



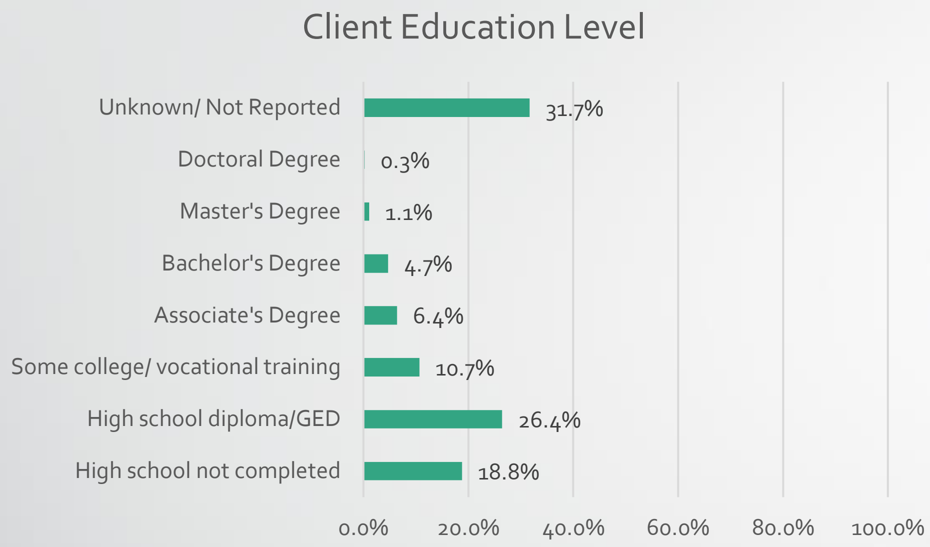


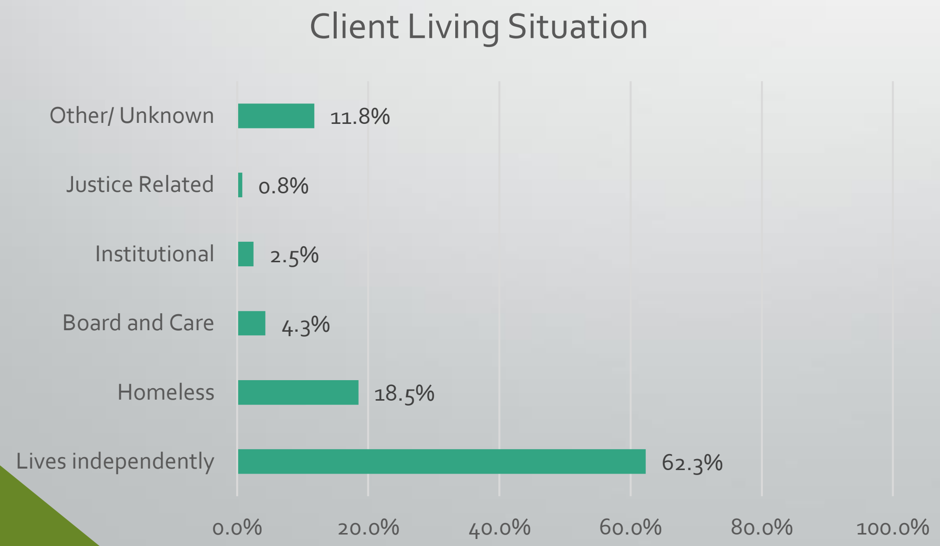


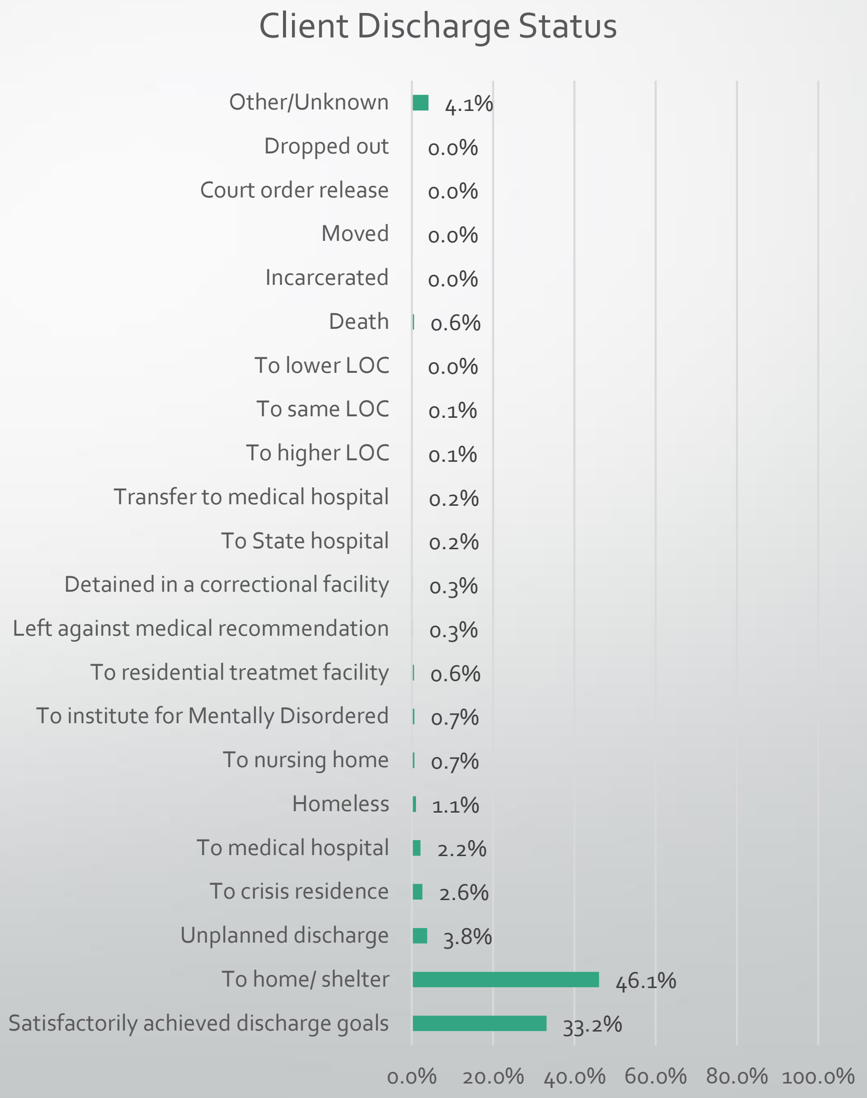




AOA demographics
(Total unique clients: 38,272)







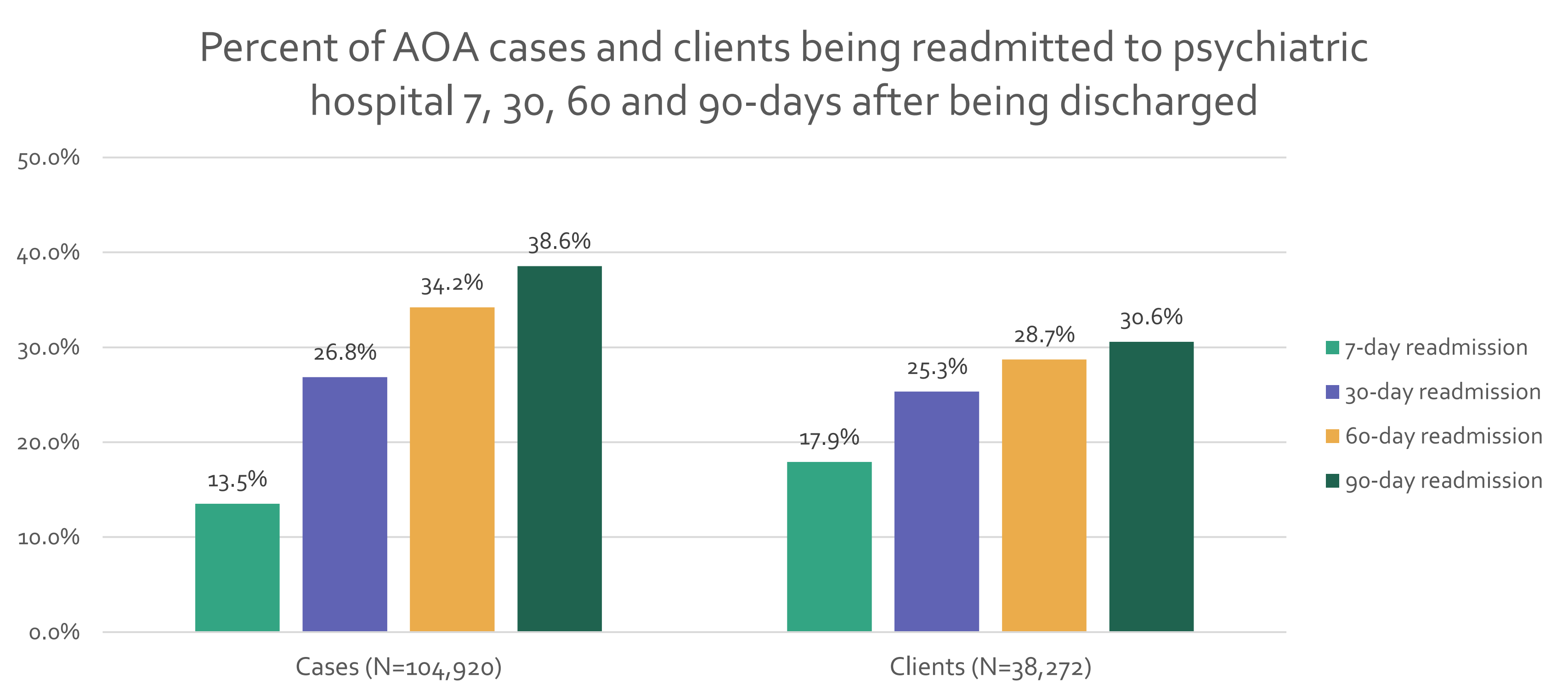
AOA services

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.

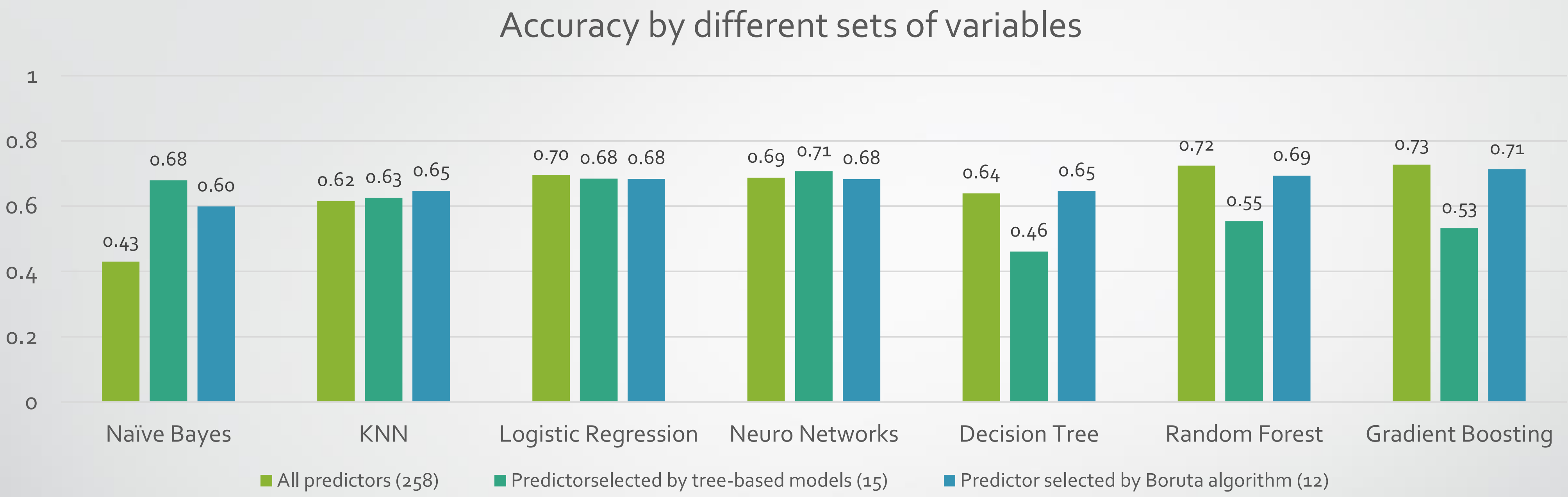
Group of Services	Metrics	Within 90 days	Within 60 days	Within 30 days	Within 7 days
Outpatient Services	Total visits:	491,466	346,263	192,847	62,783
	Total clients:	2,076	2,289	2,676	3,219
	Percent user:	4.8	5.2	6.1	7.4
Emergency Services	Total visits:	312,409	258,965	193,694	123,911
	Total clients:	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

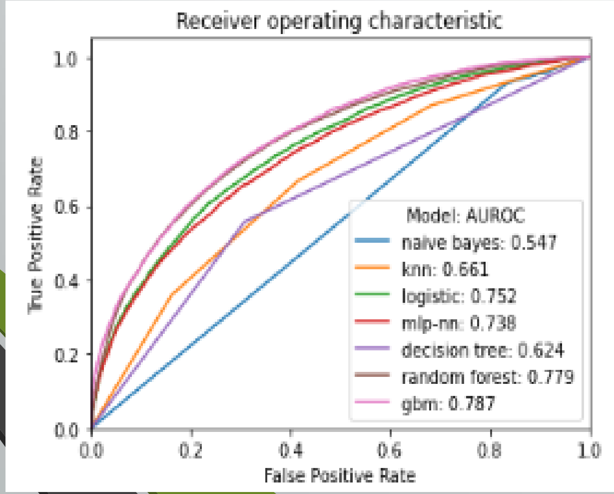
AOA hospitalization and readmission rates

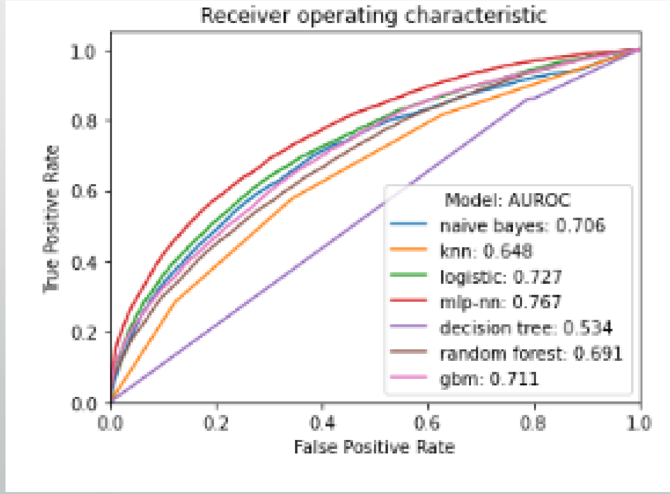
The current AOA client sample has 1.8 psychiatric hospitalization (4.79 standard deviation) in average. The average length of stay in the psychiatric hospital is 11.3 days (38.8 days standard deviation). The gap between hospitalizations is 619.2 days or about 1.7 years.

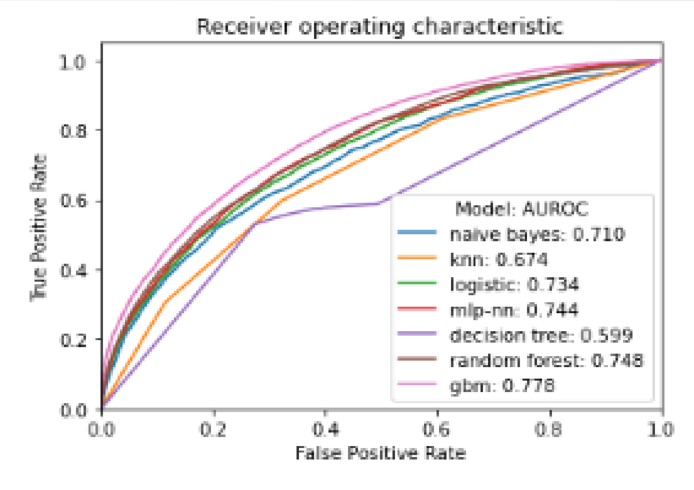


7-day readmission prediction

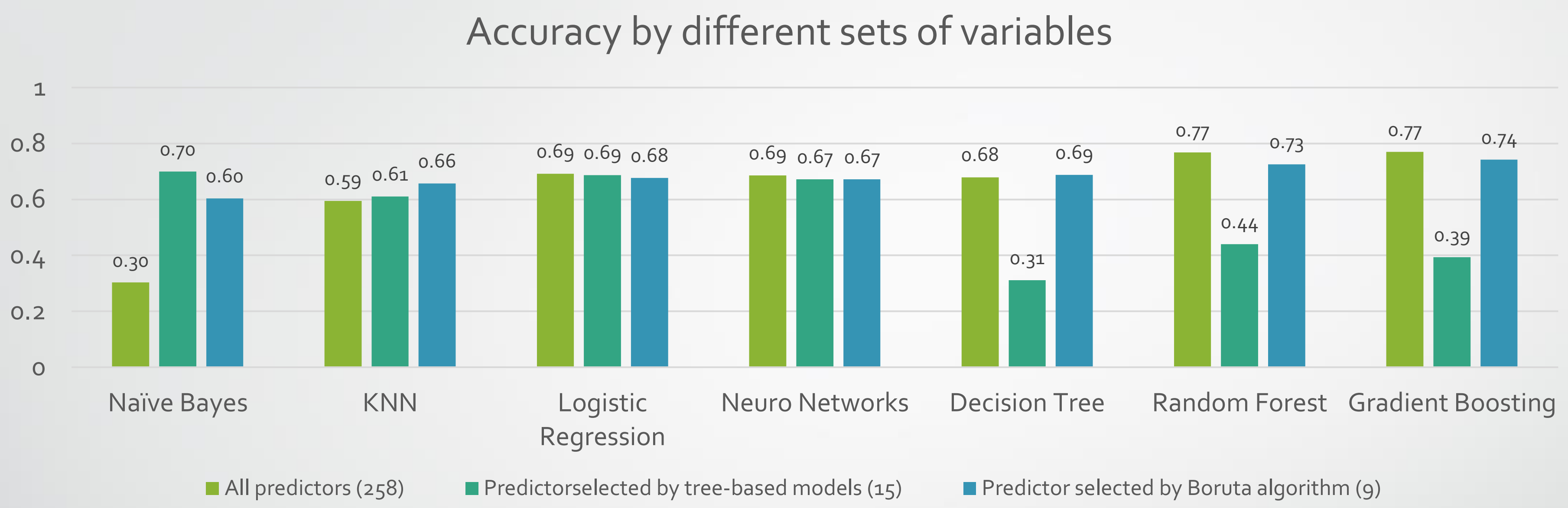


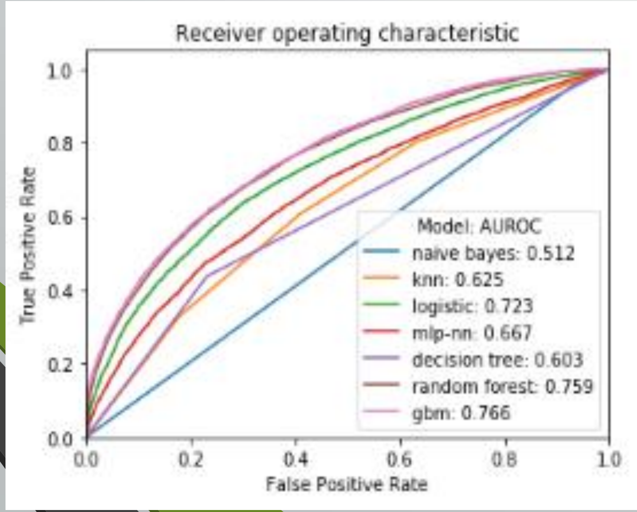


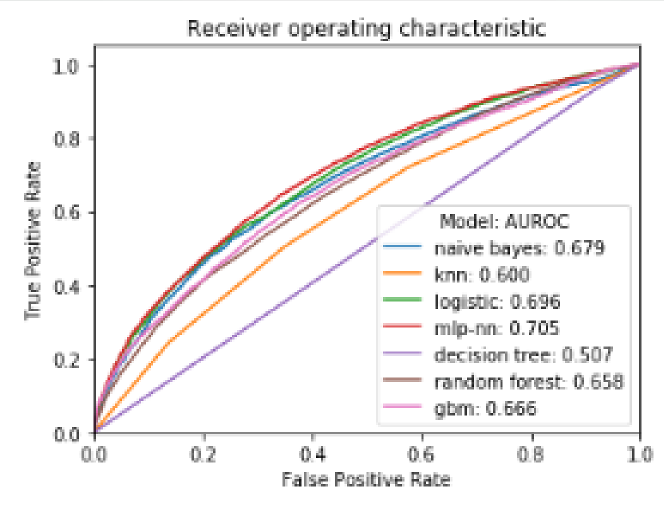


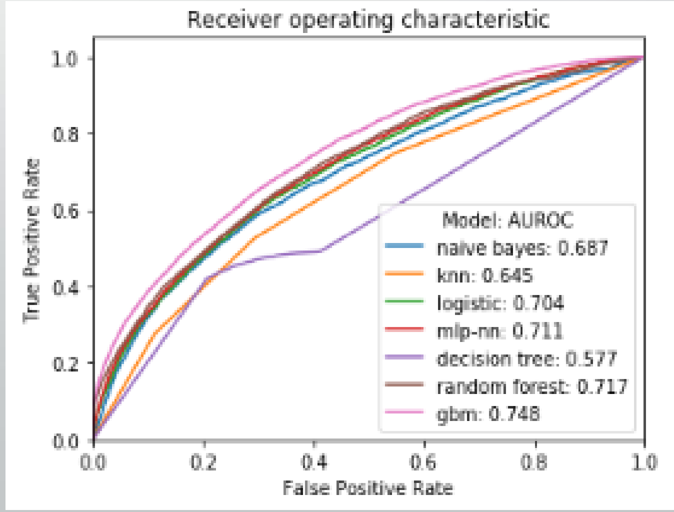


30-day readmission prediction

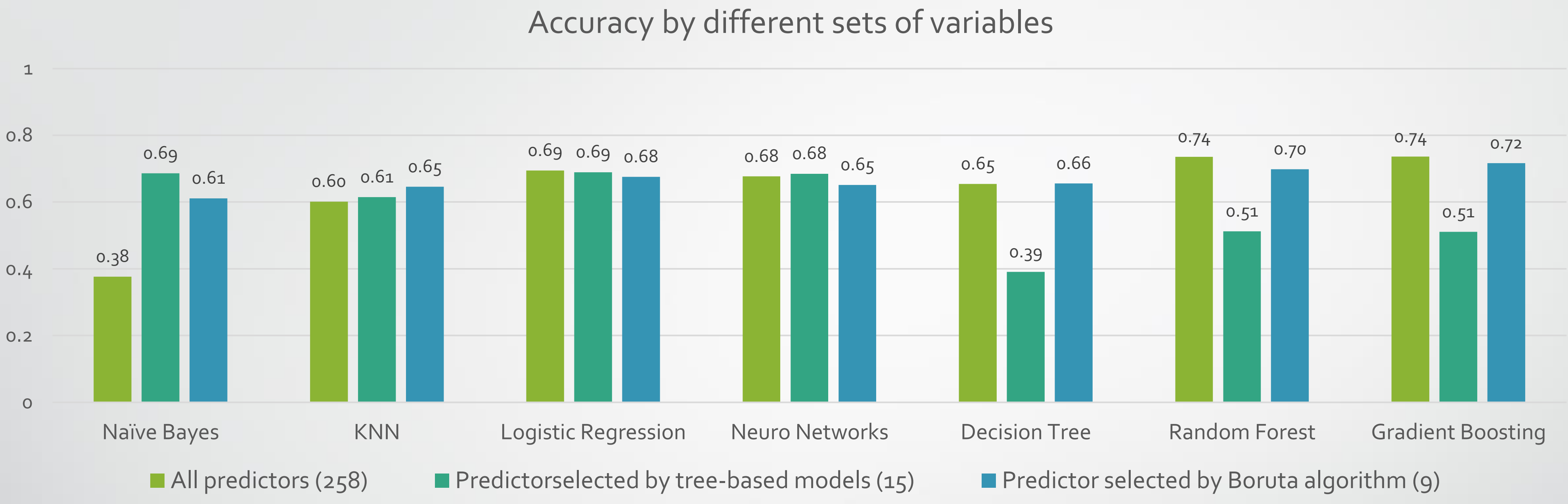


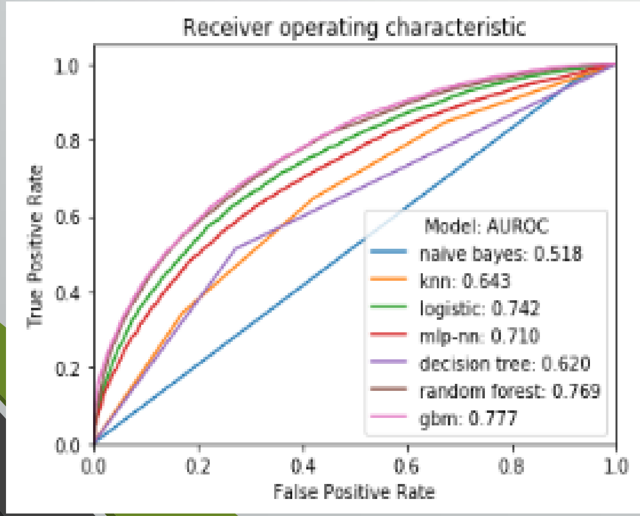


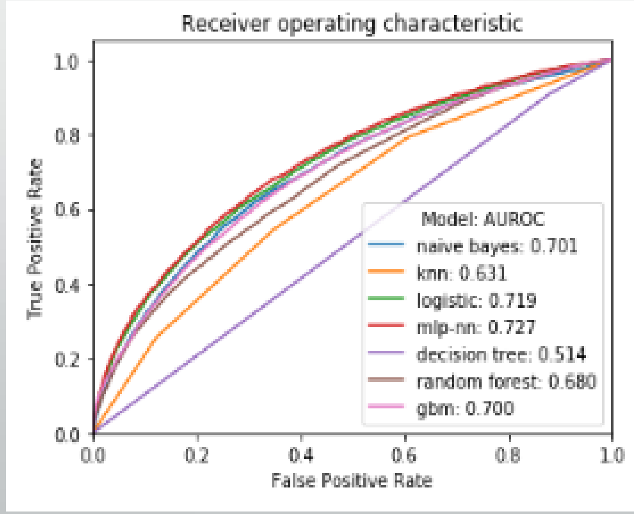


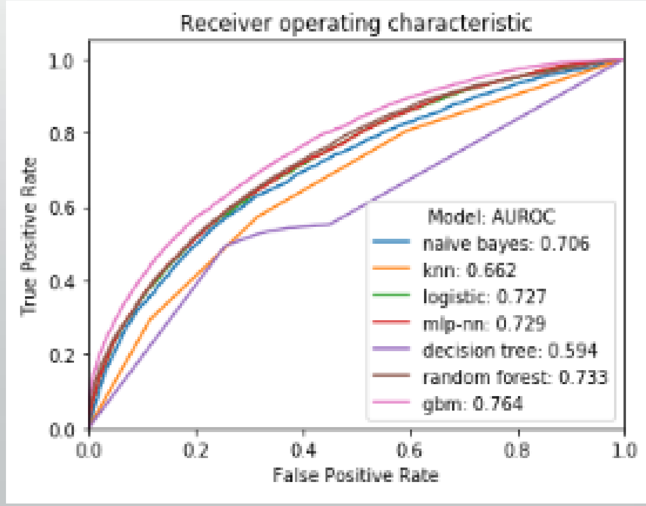


60-day readmission prediction

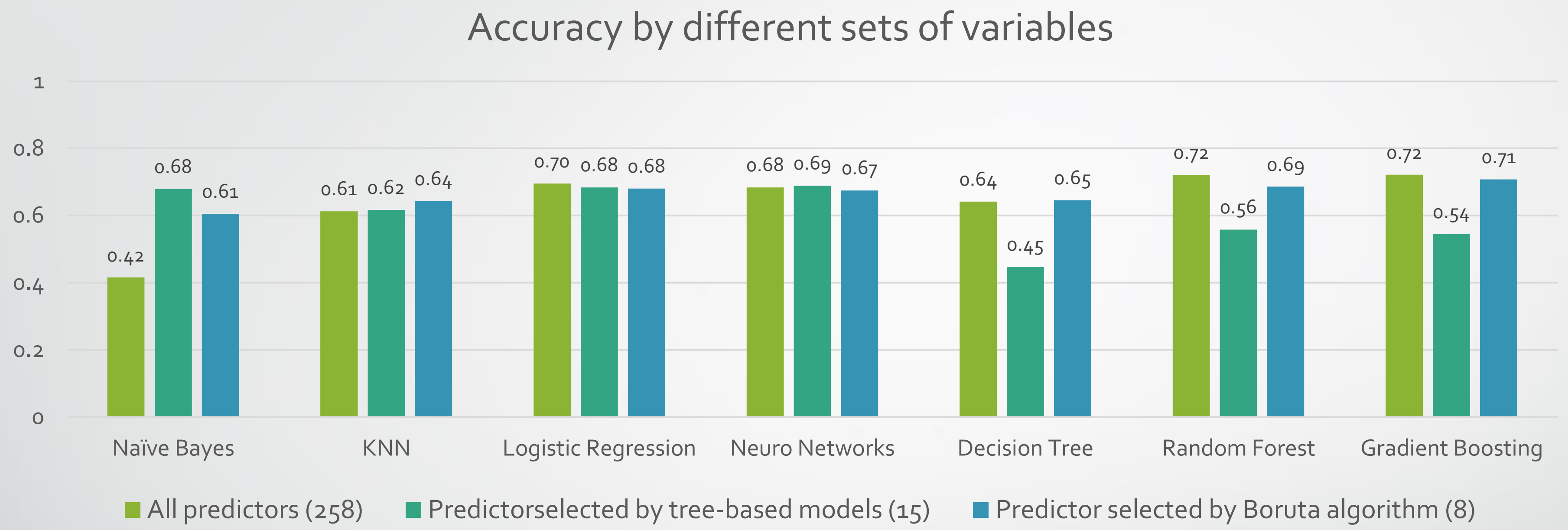


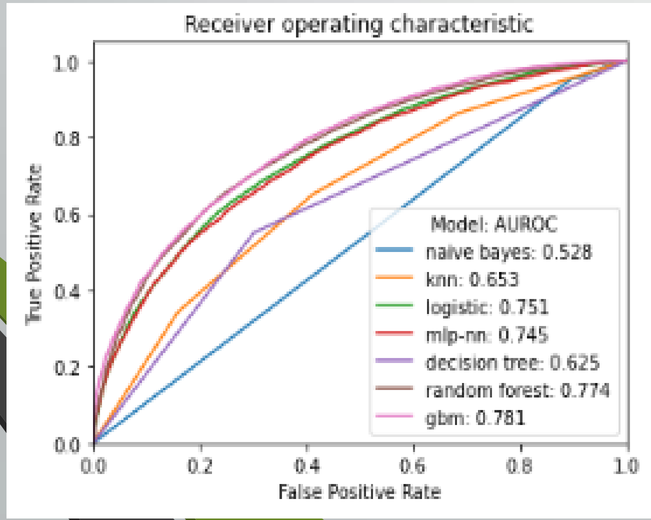


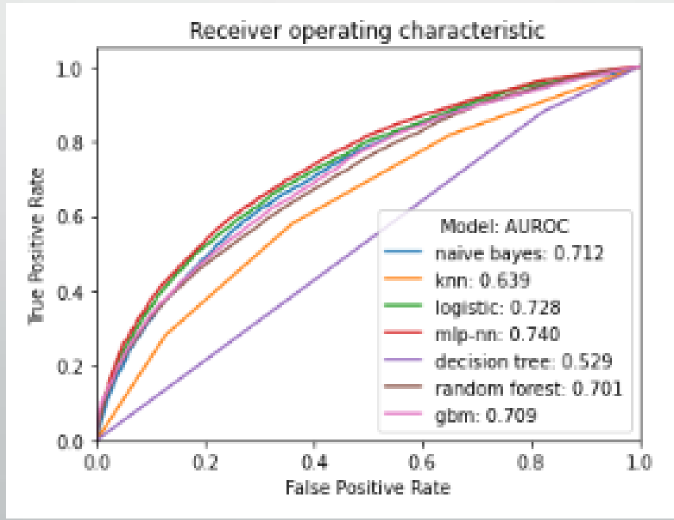


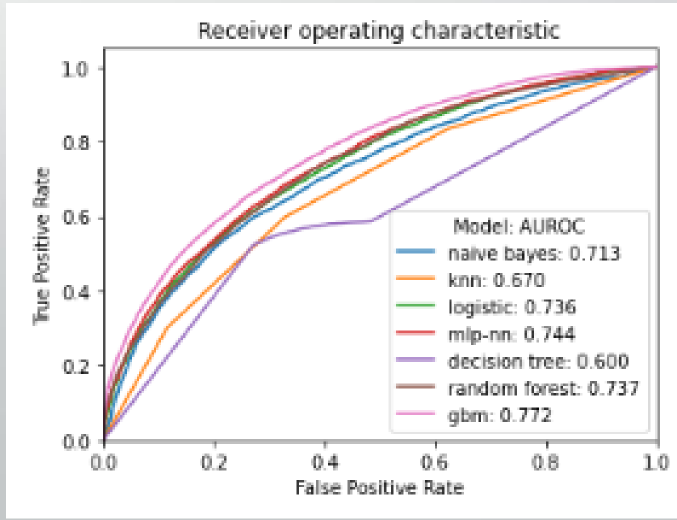


90-day readmission prediction









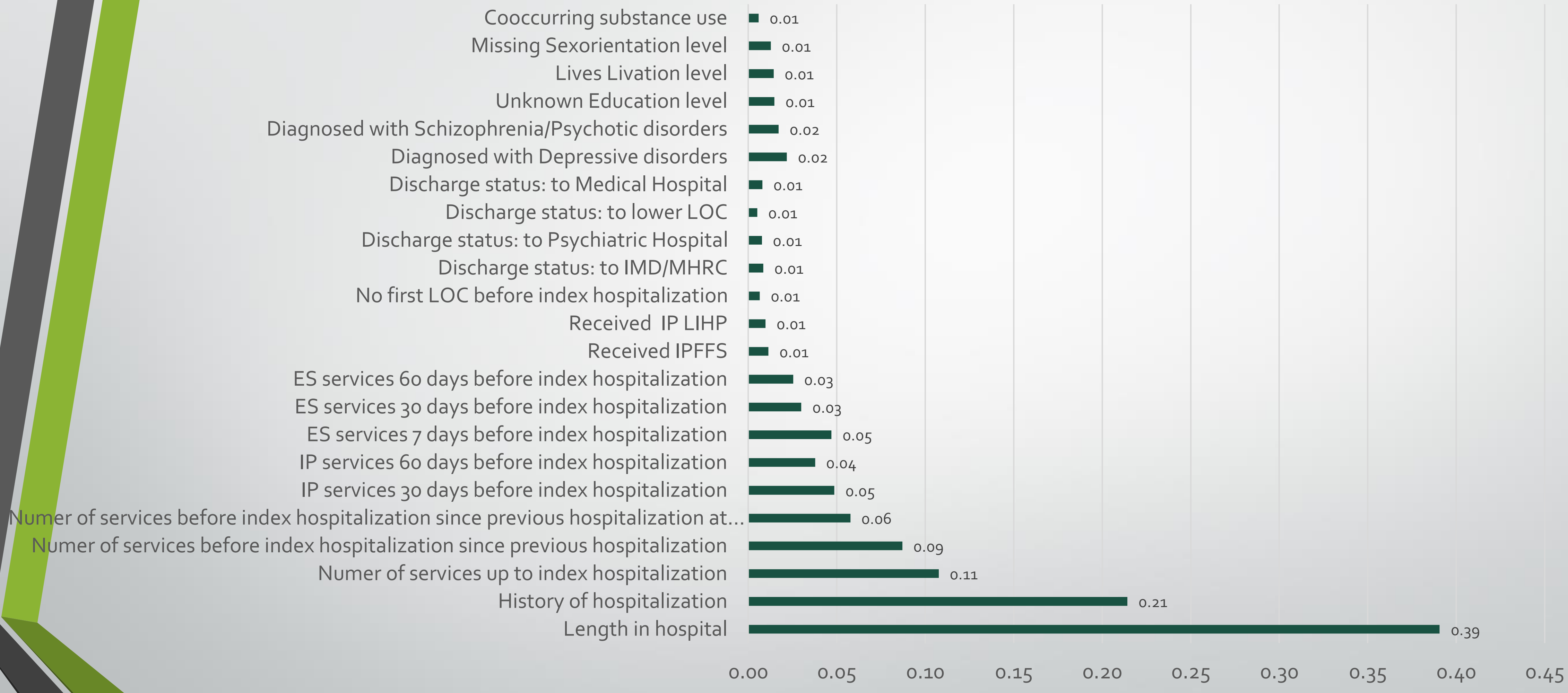
Final set of variables and model

The final model 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)

	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

Average important feature scores of Gradient Boosting model using Boruta predictors across four timeframes - AOA



AOA coefficients (Coef) of Logistic Regression model using of Boruta predictors by timeframes

	7-day readmission	30-day readmission	60-day readmission	90-day readmission
Length in hospital	-9.34	-3.20	-4.21	-3.93
History of hospitalization	3.52	4.75	6.10	7.13
Numer of services up to index hospitalization	-2.71	-1.95	-1.46	-1.24
Numer of services before index hospitalization since previous hospitalization	-	-3.57	-3.27	-2.95
Numer of services before index hospitalization since previous hospitalization at the same subunits	-0.40	-	-	-
IP services 30 days before index hospitalization	2.84	1.64	1.29	-
IP services 60 days before index hospitalization	1.76	3.09	3.33	4.30
ES services 7 days before index hospitalization	-0.02	-	-	-
ES services 30 days before index hospitalization	1.43	-	-	-
ES services 60 days before index hospitalization	0.41	-	-	-
Received IP-FFS	0.15	0.15	0.22	0.27
Received - IP LIHP	-0.23	-0.38	-0.34	-0.40
No first LOC before index hospitalization	-0.14	-0.13	-0.22	-0.24
Discharge status: to IMD/MHRC	-	-2.97	-3.24	-3.27
Discharge status: to Psychiatric Hospital	2.73	1.94	1.66	1.56
Discharge status: to lower LOC	1.13	0.72	-	-
Discharge status: to Medical Hospital	1.96	-	-	-
Diagnosed with Depressive disorders	-	-0.26	-0.27	-0.30
Diagnosed with Schizophrenia/Psychotic disorders	-	0.16	0.16	0.17
Unknown Education level	-	-0.30	-0.38	-0.38
Lives Livation level	-	-	-0.32	-0.32
Missing Sexorientation level	-	-0.35	-0.37	-0.40
Co-occurring substance use	-	0.27	0.25	0.25

Discussion

Summary of findings

- 168 models: 2 populations (CYF and AOA) x 4 timeframes (7, 30, 60 and 90-day readmissions) x 3 set of variables (all predictors, predictors from tree-based models, and predictors from Boruta algorithm) x 7 ML algorithms
- There are not many differences in the overall model performance between the CYF and AOA samples.
- Performances of MLs:
  - Naïve Bayes, KNN and Decision Tree models had poor performances (accuracy and AUC < 0.65)
  - Logistic Regression and Neural Networks had a moderate performance (0.65 – 0.75 accuracy and 0.68 - 0.72 AUC)
  - Random Forest and Gradient Boosting had good performance (accuracy 0.70-0.86 and AUC 0.70 – 0.80)
- Performances of set of variables:
  - all predictors (185 for CYF and 258 for AOA) yielded the best performance but requiring high computing power
  - predictors from tree-based model yielded low performance
  - Predictors from Boruta algorithm yielded satisfactory to good performance and have much fewer variables

  Final model: Gradient Boosting with predictors selected from Boruta algorithm
- The model performed very well for the 7-day readmission but its performance decreased in 30, 60 and 90 days.

Strengths and limitations

- Strengths:
  - analyzed two relatively big datasets for both the CYF and AOA systems at four readmission timeframes (7, 30, 60 and 90 days)
  - application of multiple ML models with different sets of number of variables
- Limitations:
  - could not use all variables in the original dataset, especially demographics variables, due to a large portion of missing data
  - default set-up for ML models
  - class imbalance; there are much fewer rehospitalized cases than non-rehospitalized ones
  - only used administrative data

Next Steps

- Incorporate outcome measure data
- Analyze the data by individual FY or select the FYs that have more available demographics data
- Sensitivity analysis – remove some variables with high importance scores
- Deep learning model

References

- 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.
- Li, Q., Yao, X., & Échevin, D. (2020). How Good Is Machine Learning in Predicting All-Cause 30-Day Hospital Readmission? Evidence From Administrative Data. *Value in Health*, *23*(10), 1307-1315.
- 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.
- 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.
- 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.