

**A Machine Learning Approach for Predicting Length of Stay at the Emergency**

**Department**

Maggie J. Xu, Qihan(Ella) Zhang

2024 Calgary Youth Science Fair

March 15th, 2024

## **Abstract**

This paper summarizes the research and analysis done with a machine learning algorithm, Random Forest, that was trained and tested on a de-identified emergency department (ED) dataset that contains 30% of ED visits from 16 Albertan hospitals from 2018 to 2020 to predict patient length of stay at the ED using eight different input variables. Random forest classification and regression models were constructed and evaluated using accuracy, out-of-bag error rate, confusion matrix, percent variability explained, and mean absolute residual. The classification models showed considerable accuracy in classifying stay times into above or below 4, 6, and 8 hours, while the regression model may have the potential to demonstrate more accurate results if more variables were available for training the model. Variable importance and partial dependence plots were generated from the models to show the importance and impact of each variable (ICD-10 category, CTAS score, and age group) in predicting length of stay, and how changes in each variable impact the model results. Future work includes acquiring more specific variables to further improve model performance and further analysis using other machine learning or statistics methods.

## Introduction

In urban Alberta hospitals, the average number of hours in which patients stay in the Emergency Department (ED) is around 6 hours (*Patient Emergency Department Total Length of Stay (LOS) | HQCA Focus*, 2024). The length of stay is dependent on a variety of variables, including the severity of the patient's condition and the crowdedness of the department. For some people, they may only stay in the ED for a few minutes, while some others may stay for up to 30 days. The current unpredictability of patient stay length creates the possibility of poor resource management, unstable expectations for patients, and disruptive patient flow in the department. The ability to forecast how long a patient will stay in the emergency department aids in resource allocation, capacity planning, and patient flow management, leading to both higher quality of care and more effective cost control. With the progressing advancement of Artificial Intelligence (AI), there is great potential to utilize this technology to predict the length in which a patient will stay in the ED. A common AI approach is a machine learning algorithm known as Random Forest, which combines multiple decision trees to make predictions. It aggregates the results of individual decision trees to improve accuracy and reduce overfitting, making it effective for tasks such as classification and regression (IBM, 2023). This project will utilize Random Forest to classify and predict patients' stay times using basic patient characteristics and hospital information.

## Method

### *Objectives*

The primary objectives of this study were to:

1. Predict whether a patient stays in the ED for more or less than:
  - a. 4 hours (government's target stay time for patients who do not need admission into hospital)
  - b. 8 hours (government's goal for patients who need admission into hospital)
  - c. 6 hours (average of the two goals)
2. Predict the stay time to the minute.

The secondary objectives of this study were to:

1. Determine the impact of using different time breaks for a two-category model.
2. Determine the impact of using a different number of categories on a model.
3. Determine the impact of using different numbers of variables.
4. Determine the impact changing the number of trees (ntree) and number of variables considered at each split (mtry) have on the results.
5. Determine the importance of variables and their impact on the prediction.
6. Predict whether the patient is admitted into hospital, discharged, or other (left before being seen, etc.).

### *Dataset*

The dataset was obtained from Alberta Health Services. It includes a random sample of 30% of emergency department visits from the 16 highest volume hospitals in Alberta in the years 2018 to 2020, amounting to 755,215 rows of data. To guarantee patient privacy, every patient

was given an encoded ID, and some variables were adjusted so that it is impossible to identify any particular patient. Variables that were recorded post-triage were excluded from the analysis, as the prediction is intended to be made at or before triage. Any data with missing or unreasonable values were also excluded.

### *Variables*

There are eight independent variables used in this study. The first was sex assigned at birth (female, male, or intersex), and the second was age, which was reported in 5 year intervals. Another variable was the ICD-10 (International Classification of Diseases, Tenth Revision) code, which was then further organized into chapters (World Health Organization, 2019). Categories XX and XXI were excluded due to their unpredictability and variability. Other variables also included the triage code (Canadian Triage and Acuity Scale score), previous hospitalization within three years, hospital type, patient district code (the first three characters of the patient's postal code), and hospital ID.

The main dependent variable is the patient length of stay in minutes, which was measured from seeing the triage nurse until leaving the emergency department.

As mentioned in the secondary objectives, this project also aimed to predict whether a patient is admitted to hospital, discharged home, or neither (labelled as "Other"; includes death, leaving against medical advice, left without being seen, or transferred). For this model, the input variables remained the same, but the goal for prediction was changed to the disposition group.

### *Random Forest*

Random Forest is a machine learning algorithm that aggregates multiple decision trees, made through training them on subsets of data, to create a final prediction with the combination of all trees (IBM, 2023). Random Forest is capable of producing two distinct types of models: classification, which classifies data points into categories, and regression models, which predicts the exact value of a data point.

The number of decision trees created (*ntree* value) and the number of variables tried at each tree's split (*mtry* value) are generally set to different values based on the type of model (*Random Forests · AFIT Data Science Lab R Programming Guide*, n.d.). This study utilized the package *RandomForest* in R version 4.3.2; for this package, the default for a classification model was that *ntree* = 500 and *mtry* =  $\sqrt{p}$ , with *p* being the number of variables, rounded down. On the other hand, the default for a regression model was *ntree* = 500, and *mtry* =  $p/3$  (*p* being the number of variables). Both the *ntree* and *mtry* values have an impact on the models' accuracy.

Random Forest is capable of creating variable importance plots on its models, which reveal the relative importance of each variable to the prediction of the model (*How Is Variable Importance Calculated for a Random Forest?*, 2018). To find the impact of each variable on the models, partial dependence plots (PDP) were created which display how much the model may *depend* on each variable (Molnar, n.d.). PDPs may be 1-variable (only showing the dependency of one variable on the target variable) or 2-variable (showing the interconnected dependencies of two variables and the target variable).

### *Classification Model*

The first type of Random Forest are classification models, where the algorithm assigned data points into a defined category. The performance of these models can be evaluated using the confusion matrix.

Figure 1. Confusion matrix of a two-category classification model.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

In the confusion matrix, the predicted and actual values are evaluated, with true positive (TP) being the number of correctly predicted values of the first category and true negative (TN) being the values correctly predicted of the second category (*Basic Evaluation Measures from the Confusion Matrix*, 2015). False positive (FP) then refers to the number of values incorrectly predicted to be in the first category while, in actuality, being in the second; conversely, false negative (FN) is the number of values predicted to be in the second category when it is in the first. Because TP, TN, FP, and FN value is calculated for each category in a classification model,

the number of columns and rows of the confusion matrix increases with the number of categories.

Using the confusion matrix, 4 different measurements can be calculated to evaluate the model's performance (*Basic Evaluation Measures from the Confusion Matrix*, 2015). Positive predictive value refers to the total number of true positives in all of the positive predictions, which is calculated by  $TP / (TP + FP)$ . Negative predictive value is the total number of true negatives in all of the negative predictions, calculated using the formula  $TN / (TN + FN)$ . Sensitivity is calculated using the formula  $TP / (TP + FN)$ . A high sensitivity is an indicator that there are few false negative results, thus fewer cases of long stay are missed. It also means that it has less false negatives. Specificity uses the formula  $TN / (TN + FP)$ . A high specificity indicates that the model is effective at correctly identifying instances of the negative class. It also means that the number of false positives is lower.

The accuracy of the model can then be calculated with the equation  $accuracy = (TP + TN) / (TP + FP + TN + FN)$ . Conversely, the out-of-bag (OOB) error rate is found by separating the dataset into two sets, where the "out of bag" set is used to test the accuracy of the model as trained on the "bagged" set, to find the overall error rate (*Out-of-Bag Error* | *Dremio*, n.d.). The error rate of a specific category is shown at the end of the confusion matrix under "class.error", where 0 = all correct predictions and 1 = all incorrect predictions.

### *Regression Model*

Random Forest is also capable of creating regression models, where data is predicted by the algorithm to the exact value; in this study, the regression model would be predicting for the exact minute (Sruthi, 2021).



The model can be evaluated via the % variability explained, which is the amount of variability in the data that can be “explained” by the model (Wickramasinghe, n.d.). The residual is the difference between the model’s predicted value and its actual value, where, for each datapoint, there is a different residual (Taylor, 2019). The mean absolute error is the mean of the absolute value of the residuals, where the closer the mean absolute error is to zero, the better the model performs (Chugh, 2020).

### *Procedure*

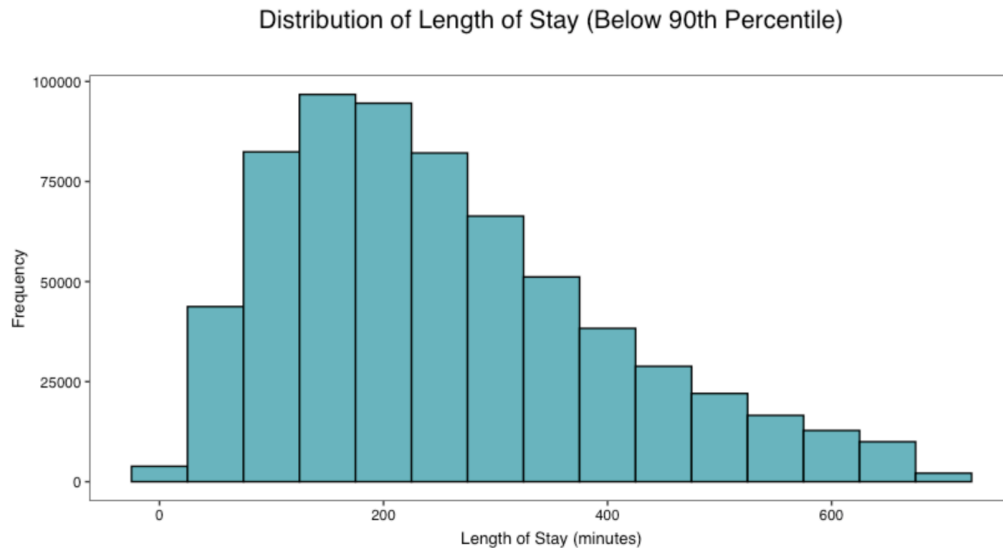
To resolve the first primary objective, 3 binary classification models with their time breaks at 4, 6, and 8 hours were made, each with a mtry value of 2 and a ntree value of 500. Although those were the emphasized models according to the primary objectives, more models with different time breaks (two categories, time break from 1 hour - 12 hours), number of categories (from 1 - 10), ntree values (from 100 to 1000, increasing in intervals of 100, with mtry = 3) and mtry values (from 1 to 7, with ntree = 500) were created to resolve the first four secondary objectives. Another classification model was created to predict which category of disposition group (admitted to hospital, discharged from hospital, or other) to determine the results of the sixth secondary objective.

One regression model was additionally made for the second primary objective, where, for the fifth secondary objective, variable importance plots were created, as well as both 1-variable and 2-variable PDPs.

## Results

### Dataset

Figure 2. Histogram of the distribution of length of stay, in minutes, for all data under the 90th percentile.



In this dataset, it was most common for patients to have a length of stay of roughly 150 minutes. On the other hand, it was least common for patients to have a length of stay of 700 minutes. However, it must be considered that, despite it being the least common, the y-axis of the graph is very extensive and there are still a substantial number of patients (25,000 or more) who have a 700 minute stay time.

### Baseline Characteristics

Table 1. Baseline characteristics table of patients, categorized by their length of stay into above 4 hours, below 4 hours, and in total.

	Below 4h (N=347444)	Above 4h (N=376649)	Total (N=724093)
<b>Biologically Assigned Sex</b>			
Male	171920 (49.5%)	200292 (53.2%)	372212 (51.4%)
Female	175520 (50.5%)	176348 (46.8%)	351868 (48.6%)

	<b>Below 4h (N=347444)</b>	<b>Above 4h (N=376649)</b>	<b>Total (N=724093)</b>
Intersex	4 (0.0%)	9 (0.0%)	13 (0.0%)
<b>Age Group</b>			
1 (1-5)	53454 (15.4%)	18581 (4.9%)	72035 (9.9%)
2 (6-10)	25677 (7.4%)	8990 (2.4%)	34667 (4.8%)
3 (11-15)	21937 (6.3%)	9591 (2.5%)	31528 (4.4%)
4 (16-20)	22195 (6.4%)	16143 (4.3%)	38338 (5.3%)
5 (21-25)	23143 (6.7%)	21674 (5.8%)	44817 (6.2%)
6 (26-30)	25753 (7.4%)	25886 (6.9%)	51639 (7.1%)
7 (31-35)	27041 (7.8%)	28561 (7.6%)	55602 (7.7%)
8 (36-40)	25004 (7.2%)	27893 (7.4%)	52897 (7.3%)
9 (41-45)	20202 (5.8%)	23930 (6.4%)	44132 (6.1%)
10 (46-50)	17921 (5.2%)	22918 (6.1%)	40839 (5.6%)
11 (51-55)	16655 (4.8%)	23438 (6.2%)	40093 (5.5%)
12 (56-60)	17094 (4.9%)	25319 (6.7%)	42413 (5.9%)
13 (61-65)	14361 (4.1%)	25023 (6.6%)	39384 (5.4%)
14 (66-70)	11396 (3.3%)	22379 (5.9%)	33775 (4.7%)
15 (71-75)	9191 (2.6%)	21535 (5.7%)	30726 (4.2%)
16 (76-80)	7029 (2.0%)	19263 (5.1%)	26292 (3.6%)
17 (81-85)	5152 (1.5%)	17602 (4.7%)	22754 (3.1%)
18 (86-90)	3762 (1.1%)	15568 (4.1%)	19330 (2.7%)
19 (90-95)	477 (0.1%)	2355 (0.6%)	2832 (0.4%)
<b>Previous Hospitalization Status</b>			
Has not been previously hospitalized	243900 (70.2%)	214451 (56.9%)	458351 (63.3%)
Has been previously hospitalized	103544 (29.8%)	162198 (43.1%)	265742 (36.7%)
<b>Triage Code</b>			
1	3064 (0.9%)	5653 (1.5%)	8717 (1.2%)
2	69391 (20.0%)	121737 (32.3%)	191128 (26.4%)
3	157351 (45.3%)	194476 (51.6%)	351827 (48.6%)
4	103274 (29.7%)	49619 (13.2%)	152893 (21.1%)
5	14364 (4.1%)	5164 (1.4%)	19528 (2.7%)
<b>ICD 10 Category</b>			
(I) Certain infectious and parasitic diseases	15899 (4.6%)	14089 (3.7%)	29988 (4.1%)
(II) Neoplasms	677 (0.2%)	3280 (0.9%)	3957 (0.5%)
(III) Diseases of the blood	601 (0.2%)	3160 (0.8%)	3761 (0.5%)
(IV) Endocrine, nutritional and metabolic diseases	1845 (0.5%)	8972 (2.4%)	10817 (1.5%)
(V) Mental and behavioral disorders	11963 (3.4%)	31793 (8.4%)	43756 (6.0%)
(VI) Diseases of the nervous system	5398 (1.6%)	9003 (2.4%)	14401 (2.0%)
(VII) Diseases of the eye and adnexa	5629 (1.6%)	2065 (0.5%)	7694 (1.1%)
(VIII) Diseases of the ear and mastoid process	6513 (1.9%)	2687 (0.7%)	9200 (1.3%)

	<b>Below 4h (N=347444)</b>	<b>Above 4h (N=376649)</b>	<b>Total (N=724093)</b>
(IX) Diseases of the circulatory system	8901 (2.6%)	20734 (5.5%)	29635 (4.1%)
(X) Diseases of the respiratory system	31398 (9.0%)	29045 (7.7%)	60443 (8.3%)
(XI) Diseases of the digestive system	15042 (4.3%)	31685 (8.4%)	46727 (6.5%)
(XII) Diseases of the skin and subcutaneous tissue	13484 (3.9%)	8894 (2.4%)	22378 (3.1%)
(XIII) Diseases of the musculoskeletal system and connective tissue	20948 (6.0%)	18455 (4.9%)	39403 (5.4%)
(XIV) Diseases of the genitourinary system	14390 (4.1%)	21546 (5.7%)	35936 (5.0%)
(XV) Pregnancy, childbirth and the puerperium	5259 (1.5%)	7153 (1.9%)	12412 (1.7%)
(XVI) Certain conditions originating in the perinatal period	1045 (0.3%)	485 (0.1%)	1530 (0.2%)
(XVII) Congenital malformations, deformations and chromosomal abnormalities	216 (0.1%)	239 (0.1%)	455 (0.1%)
(XVIII) Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	73788 (21.2%)	99868 (26.5%)	173656 (24.0%)
(XIX) Injury, poisoning and certain other consequences of external causes	114448 (32.9%)	63496 (16.9%)	177944 (24.6%)
<b>Urban / Rural</b>			
Urban	332244 (95.6%)	357413 (94.9%)	689657 (95.2%)
Rural	15200 (4.4%)	19236 (5.1%)	34436 (4.8%)
<b>Institutional ID</b>			
1	42624 (12.3%)	16537 (4.4%)	59161 (8.2%)
2	22797 (6.6%)	42663 (11.3%)	65460 (9.0%)
3	29556 (8.5%)	35856 (9.5%)	65412 (9.0%)
4	13698 (3.9%)	22369 (5.9%)	36067 (5.0%)
5	28468 (8.2%)	29261 (7.8%)	57729 (8.0%)
6	14713 (4.2%)	41455 (11.0%)	56168 (7.8%)
7	30460 (8.8%)	49994 (13.3%)	80454 (11.1%)
8	9230 (2.7%)	4859 (1.3%)	14089 (1.9%)
9	15057 (4.3%)	4724 (1.3%)	19781 (2.7%)
10	19844 (5.7%)	20998 (5.6%)	40842 (5.6%)
11	27557 (7.9%)	36968 (9.8%)	64525 (8.9%)
12	21696 (6.2%)	15525 (4.1%)	37221 (5.1%)
13	12757 (3.7%)	8704 (2.3%)	21461 (3.0%)
14	7334 (2.1%)	5495 (1.5%)	12829 (1.8%)
15	28219 (8.1%)	28530 (7.6%)	56749 (7.8%)
16	23434 (6.7%)	12711 (3.4%)	36145 (5.0%)
<b>Institutional Peer Group</b>			
Large Urban	142211 (40.9%)	194439 (51.6%)	336650 (46.5%)
Large Urban Ambulatory	21696 (6.2%)	15525 (4.1%)	37221 (5.1%)
Suburban / Rural	80322 (23.1%)	51996 (13.8%)	132318 (18.3%)

	<b>Below 4h (N=347444)</b>	<b>Above 4h (N=376649)</b>	<b>Total (N=724093)</b>
Teaching	103215 (29.7%)	114689 (30.4%)	217904 (30.1%)

A few rows are of note from the table above, such as that there is a considerable difference of ED stay length for patients with different ages. For individuals in age group 1 (ages 1 - 5), 74.2% of age group 1 patients stay for less than 4 hours. On the other hand, out of those in age group 19 (ages 85 - 90), 83.2% of age group 19 patients stay for longer than 4 hours. There is also a much larger number of younger patients who come to the ED than older patients. Additionally, for the stay time for patients who had or had not had a previous hospitalization within the past 3 years, patients who had not had a previous hospitalization make up 70.2% of those who stay for less than 4 hours, and 56.9% of patients who stay for longer than 4 hours, accumulating to 63.3% of the total dataset.

The majority of patients in the ED have triage code (CTAS score) of 2 and 3, making up 26.4% and 48.6% of the dataset respectively, and most individuals who stay in the ED for levels 1, 2, and 3 stay for longer than 4 hours, while more patients in levels 4 and 5 stay for less than 4 hours. Of all the ICD 10 categories, the highest number of patients were in categories 18 (unclassified symptoms and abnormal clinical/laboratory findings) and 19 (injury and poisoning). The lowest number of patients in the ICD 10 categories were of 16 (conditions from the perinatal period) and 17 (congenital malformations/deformations or chromosomal abnormalities).

For the baseline characteristics table, the 173 different types of patient district codes were collapsed into the categories “urban” and “rural”. Most patients of this dataset live in urban areas (95.2% of the dataset), and 46.5% come from large urban hospitals.

## Classification Model Results

### *Two-Category, 4 Hours Timebreak Model*

Table 2. Confusion matrix for the two-category 4-hour timebreak classification model.

	Actual Values			
		< 4 Hours	> 4 Hours	
Predicted Values	< 4 Hours	67212	36852	<b>PPV = 64.6%</b>
	> 4 Hours	27844	85528	<b>NPV = 75.4%</b>
		<b>Sensitivity = 70.7%</b>	<b>Specificity = 69.9%</b>	

\* The correctly predicted values for a stay time of less than 4 hours is highlighted in pink, and the correctly predicted values for a stay time of greater than 4 hours is highlighted in purple.

The accuracy of the model with 4 hours as the time break is 70.2% and, consequently, the OOB error rate is 29.8%. Comparatively, the model's negative predictive value is 10.8% higher than the positive predictive value, indicating that out of all the negative class predictions made by the model, more were correct compared to the positive class predictions. The sensitivity and specificity have similar values, so the model has consistent performance for identifying both categories.

### *Two-Category, 6 Hours Timebreak Model*

Table 3. Confusion matrix for the two-category 6-hour timebreak classification model.

	Actual Values			
		< 6 Hours	> 6 Hours	
Predicted Values	< 6 Hours	136910	15263	<b>PPV = 90.0%</b>
	> 6 Hours	37125	28138	<b>NPV = 43.1%</b>
		<b>Sensitivity =</b>	<b>Specificity =</b>	

		<b>78.7%</b>	<b>64.8%</b>	
--	--	--------------	--------------	--

\* The correctly predicted values for a stay time of less than 6 hours is highlighted in pink, and the correctly predicted values for a stay time of greater than 6 hours is highlighted in purple.

Accuracy for the model with 6 hours as the time break is 75.9%, with a 24.6% OOB error rate. The PPV and sensitivity both indicate that this model is performing better with predicting the people with stay times of less than 6 hours.

### *Two-Category, 8 Hours Timebreak Model*

Table 4. Confusion matrix for the two-category 8-hour timebreak classification model.

		Actual Values		
		< 8 Hours	> 8 Hours	
Predicted Values	< 8 Hours	171114	6245	<b>PPV = 96.5%</b>
	> 8 Hours	30680	9397	<b>NPV = 23.4%</b>
		<b>Sensitivity = 84.8%</b>	<b>Specificity = 64.8%</b>	

\* The correctly predicted values for a stay time of less than 8 hours is highlighted in pink, and the correctly predicted values for a stay time of greater than 8 hours is highlighted in purple.

The accuracy of this model is 83.0%, demonstrating higher overall performance compared to the previous models. There is a considerable difference between the model's ability to correctly predict the positive class and the negative class. As indicated by the PPV and sensitivity, it is significantly better at identifying patients with a stay time of less than 8 hours.

### *Classification Model Trends*

Figure 3. Accuracy across two-category classification models with different numbers of trees tested.

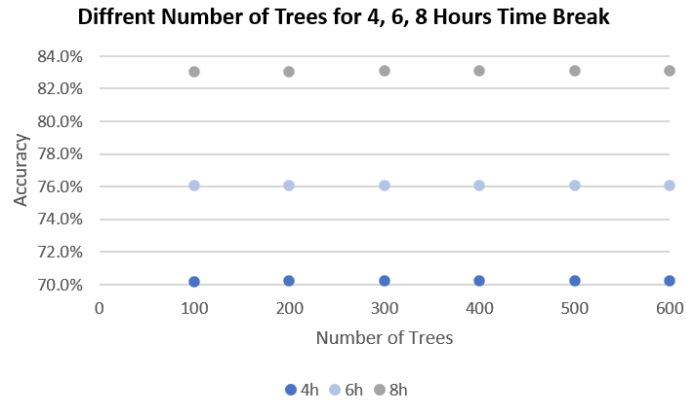
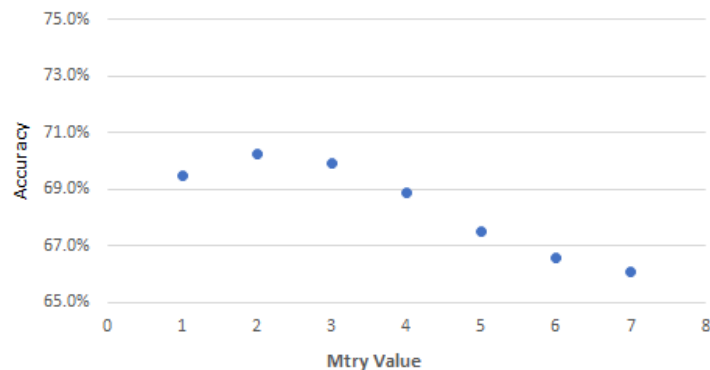


Figure 4. Accuracy across two-category classification models with different numbers of variables considered at each split.

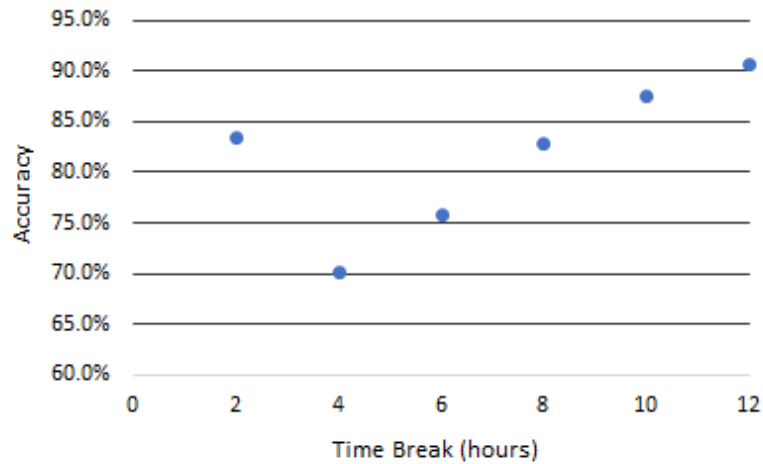


Although the results were mostly minimal, it can be seen that there is some impact that both mtry and ntree have on the accuracy of the model. Ntree had consistent accuracies across all values, while the mtry values were generally better at 2, although only by  $\sim 1\%$ . These results somewhat align with the default values for ntree and mtry. However, due to the model only utilizing 8 variables and the formula for mtry being  $\text{sqrt}(p)$  ( $p$  being the number of variables, rounded down), different mtry values may have much higher impact on the accuracy with a larger number of variables.

Another trend was found between the classifications models: as the timebreak reaches an extreme (either long or short), the accuracy increases.



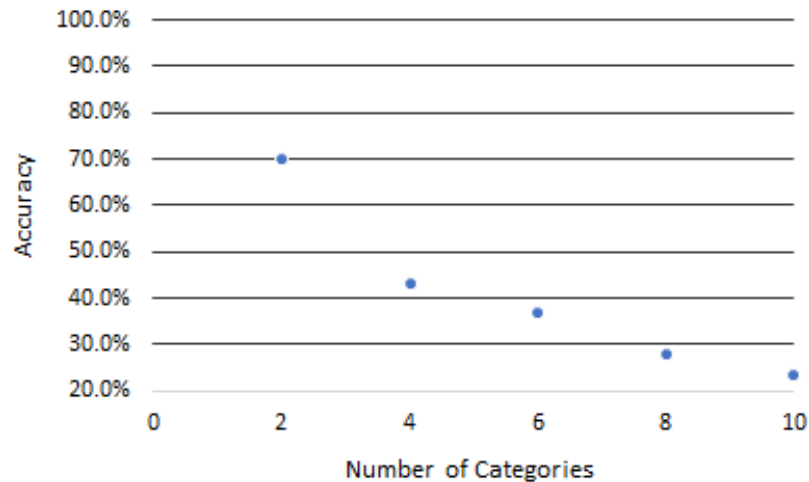
Figure 5. Accuracy across two-category classification models with 2, 4, 6, 8, 10, and 12 hour time breaks.



As the time break reached the median of the dataset (4 hours), it reached a depression: the accuracy dipped from the 2-hour timebreak model (nearly 85%) to 70%. After the 4 hour time break, the accuracy increases as the time break increases, with the highest accuracy being at the longest extreme end, being the models with a 12 hour timebreak.

Multiple models with a different number of categories were also considered, with the results summarized in the following graph:

Figure 6. Accuracy across classification models with 2, 4, 6, 8, and 10 categories.



Across this graph, the general trend is observable: as the number of categories increases, the accuracy decreases, going from about 70% for a 2-category model to nearly 20% for a 10-category model.

*Characteristics of Misclassified Patients for 2-Category Classification Models.*

Table 5. Baseline characteristics of correctly and incorrectly patients for the 4-hour timebreak classification model.

Characteristics of Correctly and Incorrectly Predicted Classification Model Patients with 4 and 6 Hours Timebreak					
	Correctly Predicted For 4 Hours Timebreak (N=515925)	Incorrectly Predicted For 4 Hours Timebreak (N=208168)	Correctly Predicted For 6 Hours Timebreak (N=556408)	Incorrectly Predicted For 6 Hours Timebreak (N=167685)	Overall (N=724093)
<b>Biologically Assigned Sex</b>					
Male	253285 (49.1%)	98583 (47.4%)	273711 (49.2%)	78157 (46.6%)	351868 (48.6%)
Female	262631 (50.9%)	109581 (52.6%)	282688 (50.8%)	89524 (53.4%)	372212 (51.4%)
Intersex	9 (0.0%)	4 (0.0%)	9 (0.0%)	4 (0.0%)	13 (0.0%)
<b>Age Groups (years)</b>					
1 (1 - 5)	54348 (10.5%)	17687 (8.5%)	65592 (11.8%)	6443 (3.8%)	72035 (9.9%)
2 (6 - 10)	26339 (5.1%)	8328 (4.0%)	31830 (5.7%)	2837 (1.7%)	34667 (4.8%)
3 (11 - 15)	23474 (4.5%)	8054 (3.9%)	27799 (5.0%)	3729 (2.2%)	31528 (4.4%)
4 (16 - 20)	26476 (5.1%)	11862 (5.7%)	31051 (5.6%)	7287 (4.3%)	38338 (5.3%)

Characteristics of Correctly and Incorrectly Predicted Classification Model Patients with 4 and 6 Hours Timebreak

	Correctly Predicted For 4 Hours Timebreak (N=515925)	Incorrectly Predicted For 4 Hours Timebreak (N=208168)	Correctly Predicted For 6 Hours Timebreak (N=556408)	Incorrectly Predicted For 6 Hours Timebreak (N=167685)	Overall (N=724093)
5 (21 - 25)	30081 (5.8%)	14736 (7.1%)	34794 (6.3%)	10023 (6.0%)	44817 (6.2%)
6 (26 - 30)	34305 (6.6%)	17334 (8.3%)	39526 (7.1%)	12113 (7.2%)	51639 (7.1%)
7 (31 - 35)	37148 (7.2%)	18454 (8.9%)	42012 (7.6%)	13590 (8.1%)	55602 (7.7%)
8 (36 - 40)	35609 (6.9%)	17288 (8.3%)	39778 (7.1%)	13119 (7.8%)	52897 (7.3%)
9 (41 - 45)	29941 (5.8%)	14191 (6.8%)	32823 (5.9%)	11309 (6.7%)	44132 (6.1%)
10 (46 - 50)	27968 (5.4%)	12871 (6.2%)	30131 (5.4%)	10708 (6.4%)	40839 (5.6%)
11 (51 - 55)	27831 (5.4%)	12262 (5.9%)	29040 (5.2%)	11053 (6.6%)	40093 (5.5%)
12 (56 - 60)	29758 (5.8%)	12655 (6.1%)	30636 (5.5%)	11777 (7.0%)	42413 (5.9%)
13 (61 - 65)	28239 (5.5%)	11145 (5.4%)	27810 (5.0%)	11574 (6.9%)	39384 (5.4%)
14 (66 - 70)	24733 (4.8%)	9042 (4.3%)	23454 (4.2%)	10321 (6.2%)	33775 (4.7%)
15 (71 - 75)	22957 (4.4%)	7769 (3.7%)	21200 (3.8%)	9526 (5.7%)	30726 (4.2%)
16 (76 - 80)	20238 (3.9%)	6054 (2.9%)	18013 (3.2%)	8279 (4.9%)	26292 (3.6%)
17 (81 - 85)	18175 (3.5%)	4579 (2.2%)	15580 (2.8%)	7174 (4.3%)	22754 (3.1%)
18 (86 - 90)	15890 (3.1%)	3440 (1.7%)	13328 (2.4%)	6002 (3.6%)	19330 (2.7%)
19 (90 - 95)	2415 (0.5%)	417 (0.2%)	2011 (0.4%)	821 (0.5%)	2832 (0.4%)
<b>Previous Hospital Admission</b>					
Yes	198095 (38.4%)	67647 (32.5%)	194807 (35.0%)	70935 (42.3%)	265742 (36.7%)
No	317830 (61.6%)	140521 (67.5%)	361601 (65.0%)	96750 (57.7%)	458351 (63.3%)
<b>CTAS Score</b>					
Level 1	6345 (1.2%)	2372 (1.1%)	6176 (1.1%)	2541 (1.5%)	8717 (1.2%)
Level 2	137063 (26.6%)	54065 (26.0%)	136043 (24.5%)	55085 (32.9%)	191128 (26.4%)
Level 3	243880 (47.3%)	107947 (51.9%)	263430 (47.3%)	88397 (52.7%)	351827 (48.6%)
Level 4	113267 (22.0%)	39626 (19.0%)	133105 (23.9%)	19788 (11.8%)	152893 (21.1%)
Level 5	15370 (3.0%)	4158 (2.0%)	17654 (3.2%)	1874 (1.1%)	19528 (2.7%)
<b>ICD 10 Category</b>					
(I) Certain infectious and parasitic diseases	21691 (4.2%)	8297 (4.0%)	24409 (4.4%)	5579 (3.3%)	29988 (4.1%)
(II) Neoplasms	3402 (0.7%)	555 (0.3%)	2955 (0.5%)	1002 (0.6%)	3957 (0.5%)
(III) Diseases of the blood	3225 (0.6%)	536 (0.3%)	2827 (0.5%)	934 (0.6%)	3761 (0.5%)
(IV) Endocrine, nutritional and metabolic diseases	9154 (1.8%)	1663 (0.8%)	7696 (1.4%)	3121 (1.9%)	10817 (1.5%)
(V) Mental and behavioral disorders	32776 (6.4%)	10980 (5.3%)	29191 (5.2%)	14565 (8.7%)	43756 (6.0%)
(VI) Diseases of the nervous system	9903 (1.9%)	4498 (2.2%)	10228 (1.8%)	4173 (2.5%)	14401 (2.0%)
(VII) Diseases of the eye and adnexa	5843 (1.1%)	1851 (0.9%)	6999 (1.3%)	695 (0.4%)	7694 (1.1%)
(VIII) Diseases of the ear and mastoid process	7000 (1.4%)	2200 (1.1%)	8247 (1.5%)	953 (0.6%)	9200 (1.3%)

Characteristics of Correctly and Incorrectly Predicted Classification Model Patients with 4 and 6 Hours Timebreak

	Correctly Predicted For 4 Hours Timebreak (N=515925)	Incorrectly Predicted For 4 Hours Timebreak (N=208168)	Correctly Predicted For 6 Hours Timebreak (N=556408)	Incorrectly Predicted For 6 Hours Timebreak (N=167685)	Overall (N=724093)
(IX) Diseases of the circulatory system	21686 (4.2%)	7949 (3.8%)	19721 (3.5%)	9914 (5.9%)	29635 (4.1%)
(X) Diseases of the respiratory system	44152 (8.6%)	16291 (7.8%)	49558 (8.9%)	10885 (6.5%)	60443 (8.3%)
(XI) Diseases of the digestive system	34939 (6.8%)	11788 (5.7%)	32346 (5.8%)	14381 (8.6%)	46727 (6.5%)
(XII) Diseases of the skin and subcutaneous tissue	16025 (3.1%)	6353 (3.1%)	18541 (3.3%)	3837 (2.3%)	22378 (3.1%)
(XIII) Diseases of the musculoskeletal system and connective tissue	26224 (5.1%)	13179 (6.3%)	31158 (5.6%)	8245 (4.9%)	39403 (5.4%)
(XIV) Diseases of the genitourinary system	25145 (4.9%)	10791 (5.2%)	26183 (4.7%)	9753 (5.8%)	35936 (5.0%)
(XV) Pregnancy, childbirth and the puerperium	7916 (1.5%)	4496 (2.2%)	8927 (1.6%)	3485 (2.1%)	12412 (1.7%)
(XVI) Certain conditions originating in the perinatal period	1063 (0.2%)	467 (0.2%)	1337 (0.2%)	193 (0.1%)	1530 (0.2%)
(XVII) Congenital malformations, deformations and chromosomal abnormalities	309 (0.1%)	146 (0.1%)	348 (0.1%)	107 (0.1%)	455 (0.1%)
(XVIII) Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	115355 (22.4%)	58301 (28.0%)	125727 (22.6%)	47929 (28.6%)	173656 (24.0%)
(XIX) Injury, poisoning and certain other consequences of external causes	130117 (25.2%)	47827 (23.0%)	150010 (27.0%)	27934 (16.7%)	177944 (24.6%)
<b>Patient Postal District</b>					
Rural	25118 (4.9%)	9330 (4.5%)	25527 (4.6%)	8921 (5.3%)	34448 (4.8%)
Urban	490807 (95.1%)	198838 (95.5%)	530881 (95.4%)	158764 (94.7%)	689645 (95.2%)
<b>Hospital ID</b>					
1	44161 (8.6%)	15000 (7.2%)	52989 (9.5%)	6172 (3.7%)	59161 (8.2%)
2	46649 (9.0%)	18811 (9.0%)	44901 (8.1%)	20559 (12.3%)	65460 (9.0%)
3	44593 (8.6%)	20819 (10.0%)	48217 (8.7%)	17195 (10.3%)	65412 (9.0%)
4	25238 (4.9%)	10829 (5.2%)	25699 (4.6%)	10368 (6.2%)	36067 (5.0%)
5	40006 (7.8%)	17723 (8.5%)	45095 (8.1%)	12634 (7.5%)	57729 (8.0%)
6	42576 (8.3%)	13592 (6.5%)	37018 (6.7%)	19150 (11.4%)	56168 (7.8%)
7	59868 (11.6%)	20586 (9.9%)	58053 (10.4%)	22401 (13.4%)	80454 (11.1%)
8	10363 (2.0%)	3726 (1.8%)	11982 (2.2%)	2107 (1.3%)	14089 (1.9%)
9	15783 (3.1%)	3998 (1.9%)	17978 (3.2%)	1803 (1.1%)	19781 (2.7%)
10	29577 (5.7%)	11265 (5.4%)	31546 (5.7%)	9296 (5.5%)	40842 (5.6%)

Characteristics of Correctly and Incorrectly Predicted Classification Model Patients with 4 and 6 Hours Timebreak

	Correctly Predicted For 4 Hours Timebreak (N=515925)	Incorrectly Predicted For 4 Hours Timebreak (N=208168)	Correctly Predicted For 6 Hours Timebreak (N=556408)	Incorrectly Predicted For 6 Hours Timebreak (N=167685)	Overall (N=724093)
11	42443 (8.2%)	22082 (10.6%)	47704 (8.6%)	16821 (10.0%)	64525 (8.9%)
12	24706 (4.8%)	12515 (6.0%)	31517 (5.7%)	5704 (3.4%)	37221 (5.1%)
13	14887 (2.9%)	6574 (3.2%)	17787 (3.2%)	3674 (2.2%)	21461 (3.0%)
14	8416 (1.6%)	4413 (2.1%)	10216 (1.8%)	2613 (1.6%)	12829 (1.8%)
15	40758 (7.9%)	15991 (7.7%)	44137 (7.9%)	12612 (7.5%)	56749 (7.8%)
16	25901 (5.0%)	10244 (4.9%)	31569 (5.7%)	4576 (2.7%)	36145 (5.0%)
<b>Hospital Type</b>					
Large Urban Ambulatory	24706 (4.8%)	12515 (6.0%)	31517 (5.7%)	5704 (3.4%)	37221 (5.1%)
Large Urban	235614 (45.7%)	101036 (48.5%)	247870 (44.5%)	88780 (52.9%)	336650 (46.5%)
Teaching	159094 (30.8%)	58810 (28.3%)	166159 (29.9%)	51745 (30.9%)	217904 (30.1%)
Suburban / Rural	96511 (18.7%)	35807 (17.2%)	110862 (19.9%)	21456 (12.8%)	132318 (18.3%)

Although most baseline characteristics showed a considerable increase in correct predictions from the 4-hour timebreak model to the 6-hour timebreak model, some variables showed a sizable decrease, such as a patient having a previous hospital admission having a correct prediction of 38.4 in the 4 hour timebreak model, which declined to 35.0% in the 6 hour timebreak model. The percentages for sex, ICD categories and patient postal district displayed a nominal change from category to category, indicating that they have minimal impact on the accuracy of the predictions.

#### *Disposition Group Prediction Model*

This model had a confusion matrix as follows:

Table 6. Confusion matrix for the classification model predicting the disposition of a patient in the ED.

		Actual Values			
		Admitted	Discharged	Other	Predictive Values
Admitted		8278	52208	22	13.7%

Predicted Values	<b>Discharged</b>	4108	169920	128	<b>97.6%</b>
	<b>Other</b>	430	14708	198	<b>1.3%</b>
	<b>Accuracy Out of Actual Values</b>	<b>64.5%</b>	<b>71.7%</b>	<b>56.9%</b>	

\* The correctly predicted value for being admitted into the hospital is highlighted in pink, the correctly predicted value for being discharged from the ED is highlighted in purple, and the correctly predicted value of others is highlighted in blue.

The accuracy of the model is 71.4%, with an OOB error rate of 28.6%. The model is the best at identifying people who were discharged home.

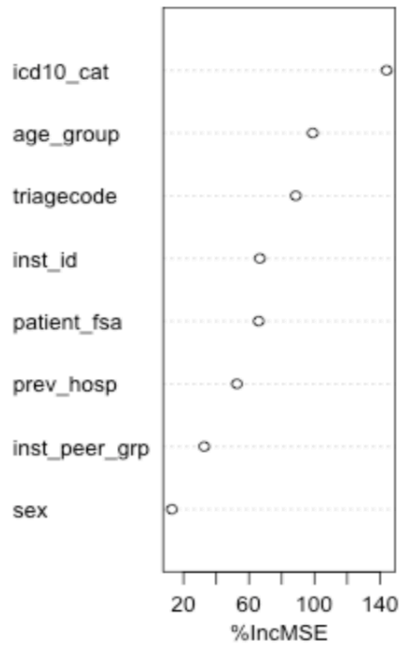
### Regression Model

The regression model that was made displayed a 20.1% variability explained, with the minimum absolute residual being 0.001 minutes, the maximum absolute residual being 14525.4 minutes, and the mean absolute residual being 205.5 minutes.

### *Variable Importance*

A variable importance plot was created on the regression model to demonstrate which variables contributed the most to the model:

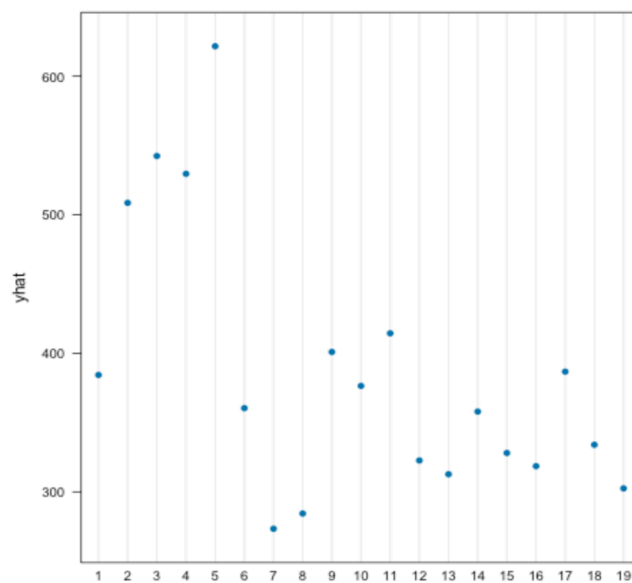
Figure 7. Variable importance plot for the regression model.



### *1-Variable Partial Dependence Plots*

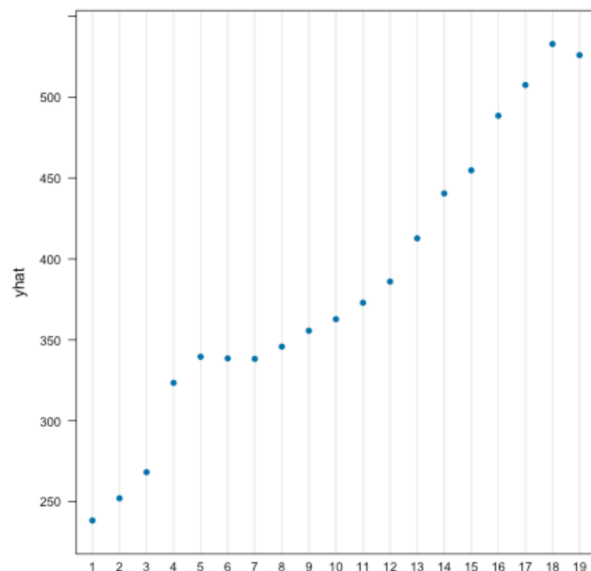
For the three most important variables, partial dependence plots (PDP) were created to show how the model predictions *depends* on each variable, such as for the ICD 10 category:

Figure 8. Partial dependence plot for the ICD 10 category.



The plot indicates that, of every ICD 10 category, the ones that have the longest length of stay, averaging more than 10 hours of stay time in the ED, are mental and behavioral disorders under category 5. However, categories 2 (neoplasms), 3 (diseases of the blood), and 4 (endocrine/nutritional/metabolic diseases) also constitute considerable ED stay lengths of ~520 minutes (8.7 hours). Conversely, the categories that lead to the shortest stay times are categories 7 (diseases of the eye) and 8 (diseases of the ear).

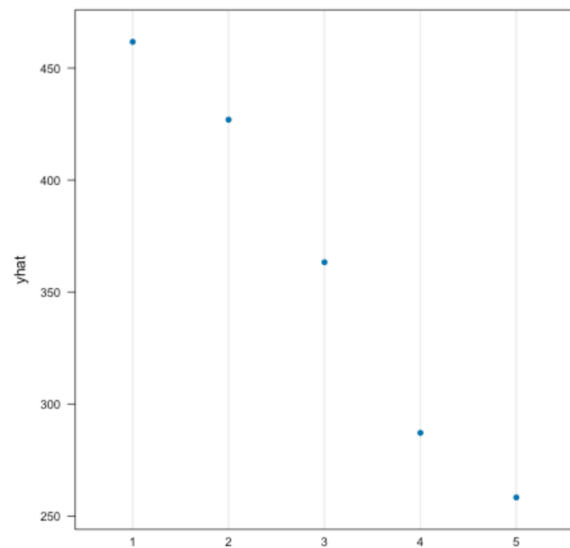
Figure 9. Partial dependance plot for the age group.



The PDP for age group suggests that, as the patient's age increases, their corresponding ED length of stay increases as well. It is of note that there is a large jump of ~90 minutes between age groups 3 (ages 10 - 15) and 4 (16 - 20).

Figure 10. Partial dependance plot for the triage code (CTAS score).

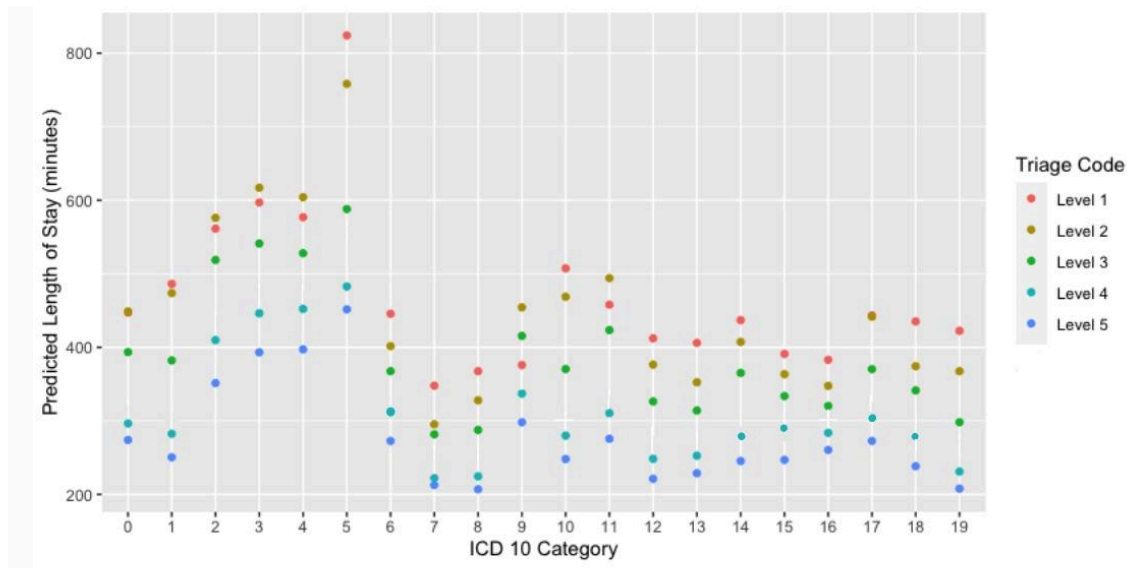




The correlation found with the PDP for triage code (CTAS score) displays that the more severe the condition, the longer a patient stays in the ED. Interestingly enough, this reveals that patients with the most severe conditions have nearly double the length of stay as patients with the least severe conditions.

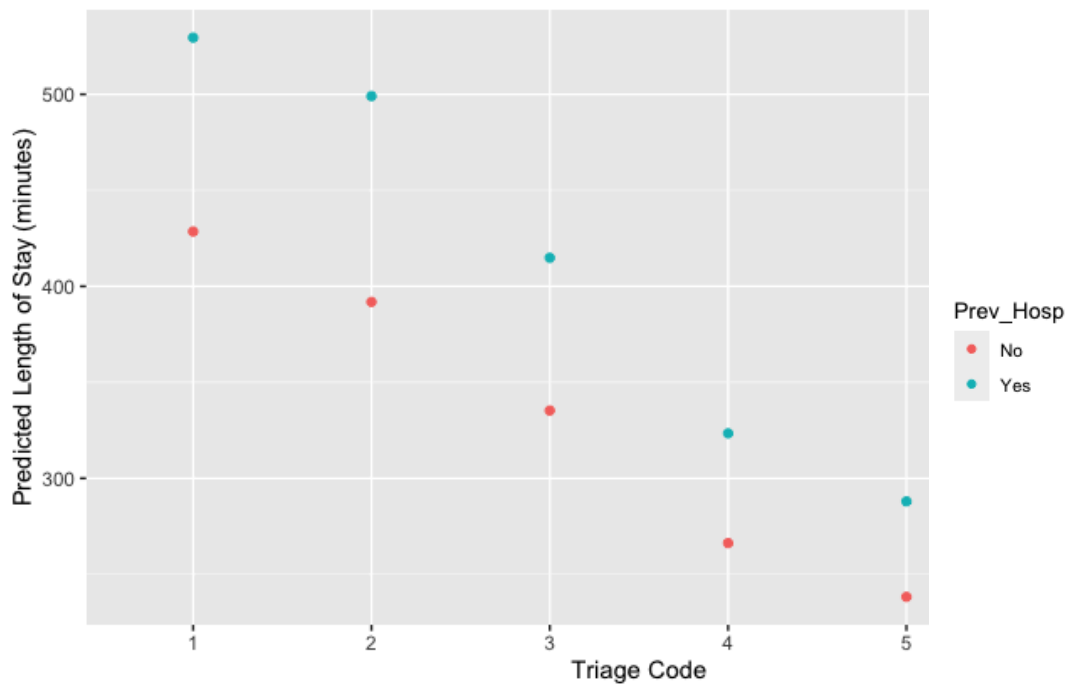
### *2-Variable Partial Dependence Plots*

Figure 11. 2-Variable partial dependence plot for ICD 10 category and triage code, correlated with the predicted length of stay in minutes.



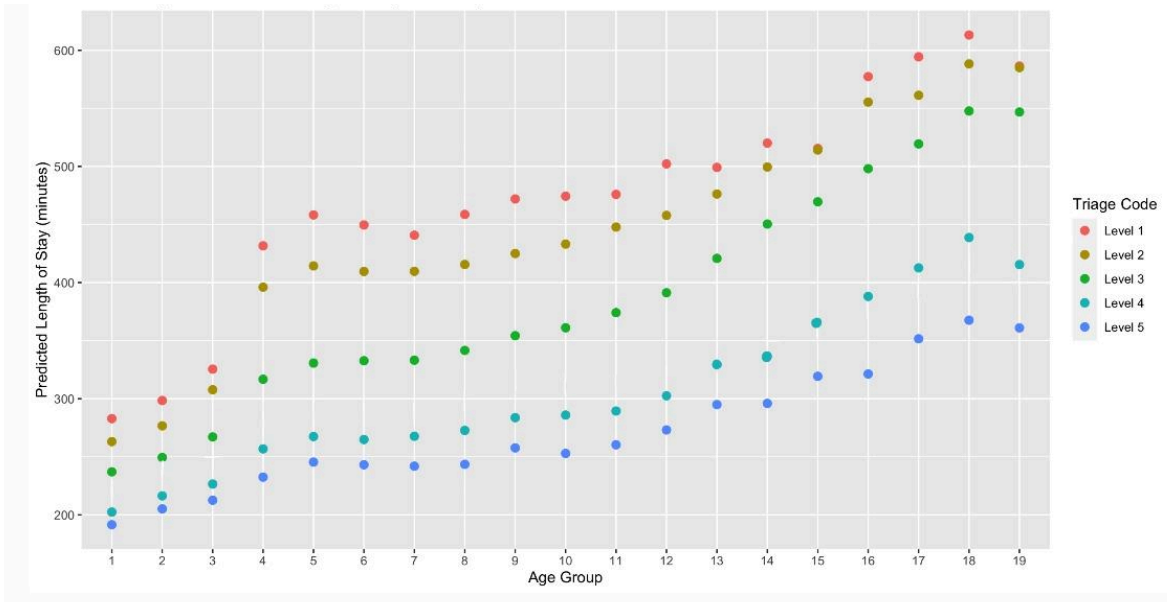
A PDP conducted on two of the three most important variables, as according to the variable importance plot, is with the ICD-10 category with the triage code. The observable trend is that a decrease in triage level (and thus an increase in severity) is correlated with an increase in the predicted length of stay across all the ICD-10 categories. It is notable, however, that there are some categories with a large difference of stay time between different triage code levels, such as between triage levels 1 and 3 for ICD-10 category 5 (mental and behavioral disorders), which has an estimated 200 minute gap. On the other hand, for category 6 (diseases of the nervous system), 7 (diseases of the eye), 8 (diseases of the ear), and 16 (conditions originating in the perinatal period), there is very little difference in stay time between levels 1, 2, and 3. Some ICD-10 categories, such as categories 2 (neoplasms), 3 (diseases of the blood), 4 (endocrine, nutritional, and metabolic diseases), 9 (diseases of the circulatory system), 11 (diseases of the digestive system), and 17 (diseases of the skin) actually indicate a longer stay time for CTAS score 2 rather than 1.

Figure 12. 2-Variable partial dependence plot for triage code and previous hospitalization status, correlated with the predicted length of stay in minutes.



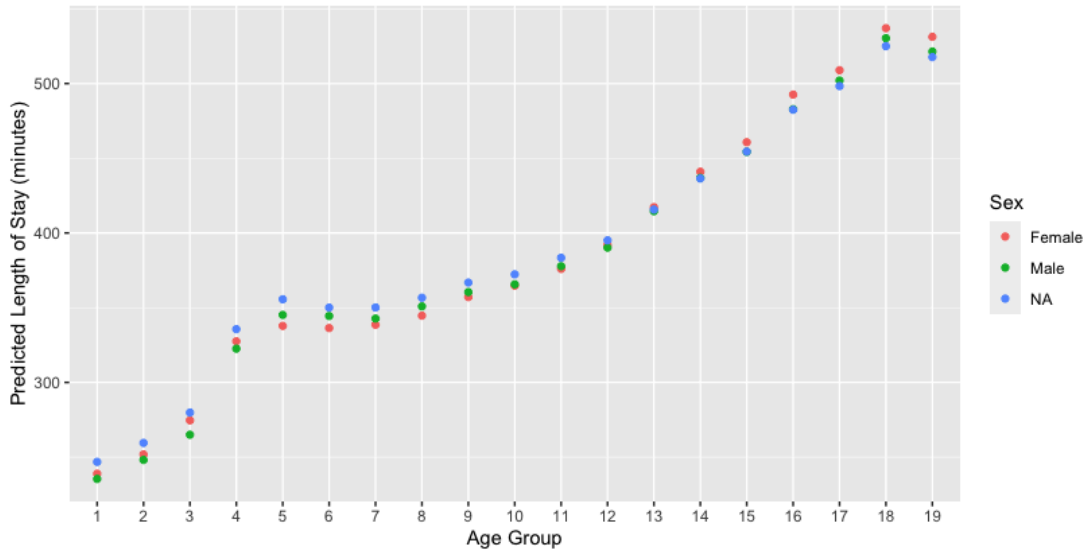
The graph shows that as the triage code level increases (and thus severity decreases), the length of stay decreases. Patients with a previous hospitalization status have longer stay times than people who have not been hospitalized before, regardless of the triage code level; the difference between the stay times of patients who have and have not been previously hospitalized, however, decreases per triage code level.

Figure 13. 2-Variable partial dependence plot for age group and triage code, correlated with the predicted length of stay in minutes.



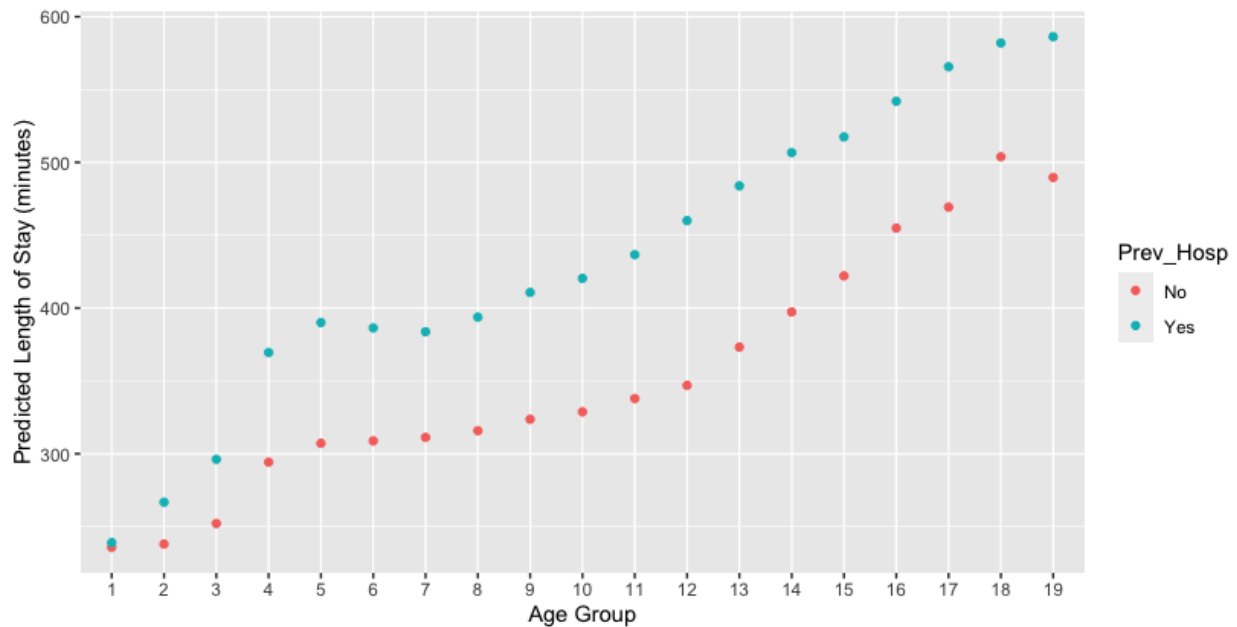
This PDP displays the general trend of the predicted length of stay increasing as the age group increases. Notably, however, there is a greater stay time difference between lower triage levels (1, 2, and 3) and higher triage levels (4 and 5) as the age group increases. There is also a notable gap in the predicted stay time of triage code levels 1 and 2 between age groups 3 (11-15) and 4 (16-20). Triage code level 1 is also the most observably sporadic of all the triage codes, while triage level 3, 4, and 5 had a more linear increase overall.

Figure 14. 2-Variable partial dependence plot for age group and sex, correlated with the predicted length of stay in minutes.



This PDP suggests that the general increase of stay time is related with the increase of the age groups. Interestingly, there is not a strict nor major difference between the length of stays of male and female patients across the different age groups. The 'N/A', or intersex category, shows having the longest predicted length of stay until age group 13 (ages 61 - 65), where it then has the shortest stay time. However, due to the limited number of intersex patients in the dataset, this observation cannot be fully dependable. On the other hand, while the first four age groups (ages 1 - 20) had patients assigned female at birth having a slightly longer stay time, the seven following age groups (ages 21 - 50) had male patients having a somewhat longer stay time. The last six age groups (ages 66 - 95), on the other hand, display a slightly larger difference between sexes, with female patients having an exceedingly longer stay time than males. However, as all observable differences are very slight (about 10 - 25 minutes), there is not any extremely significant impact of sex on the predicted stay time.

Figure 15. 2-Variable partial dependence plot for age group and previous hospitalization status, correlated with the predicted length of stay in minutes.



For this PDP, the general trend is that, again, the increase of age group is correlated with the increase of the predicted length of stay. However, there is a sizable difference of roughly 100 minutes between most of the stay times for individuals with and without a record of previous hospitalization, with the former having longer predicted times. There is a sudden noticeable difference between age groups 3 (11-15) and 4 (16-20) in the stay time for patients with a record of previous hospitalization, which is much smaller for patients without a record of previous hospitalization for that age group.

#### *Characteristics of mispredicted patients in the regression model*

Table 7. Baseline characteristics of patients in the regression model, categorized by the length of the mispredicted time in hours.

Characteristics of Mispredicted Patients in the Regression Model

	Mispredicted by <4 Hours (N=562953)	Mispredicted by 4-6 Hours (N=71748)	Mispredicted by 6-8 Hours (N=32534)	Mispredicted by 8-10 Hours (N=17122)	Mispredicted by >10 Hours (N=39736)	Overall (N=724093)
<b>Biologically Assigned Sex</b>						
Male	271840 (48.3%)	34673 (48.3%)	16409 (50.4%)	8679 (50.7%)	20267 (51.0%)	351868 (48.6%)
Female	291104 (51.7%)	37074 (51.7%)	16125 (49.6%)	8443 (49.3%)	19466 (49.0%)	372212 (51.4%)
Intersex	9 (0.0%)	1 (0.0%)	0 (0%)	0 (0%)	3 (0.0%)	13 (0.0%)
<b>Age (years)</b>						
1 - 5	68930 (12.2%)	1812 (2.5%)	558 (1.7%)	291 (1.7%)	444 (1.1%)	72035 (9.9%)
6 - 10	33292 (5.9%)	816 (1.1%)	240 (0.7%)	117 (0.7%)	202 (0.5%)	34667 (4.8%)
11 - 15	29126 (5.2%)	1239 (1.7%)	450 (1.4%)	212 (1.2%)	501 (1.3%)	31528 (4.4%)
16 - 20	32213 (5.7%)	2818 (3.9%)	1240 (3.8%)	687 (4.0%)	1380 (3.5%)	38338 (5.3%)
21 - 25	35973 (6.4%)	3945 (5.5%)	1658 (5.1%)	946 (5.5%)	2295 (5.8%)	44817 (6.2%)
26 - 30	41564 (7.4%)	4568 (6.4%)	1883 (5.8%)	1063 (6.2%)	2561 (6.4%)	51639 (7.1%)
31 - 35	44847 (8.0%)	5098 (7.1%)	2021 (6.2%)	1102 (6.4%)	2534 (6.4%)	55602 (7.7%)
36 - 40	42275 (7.5%)	4878 (6.8%)	1957 (6.0%)	1117 (6.5%)	2670 (6.7%)	52897 (7.3%)
41 - 45	34947 (6.2%)	4208 (5.9%)	1722 (5.3%)	933 (5.4%)	2322 (5.8%)	44132 (6.1%)
46 - 50	32033 (5.7%)	4181 (5.8%)	1637 (5.0%)	830 (4.8%)	2158 (5.4%)	40839 (5.6%)
51 - 55	30560 (5.4%)	4478 (6.2%)	1758 (5.4%)	940 (5.5%)	2357 (5.9%)	40093 (5.5%)
56 - 60	31358 (5.6%)	5035 (7.0%)	2211 (6.8%)	1100 (6.4%)	2709 (6.8%)	42413 (5.9%)
61 - 65	27511 (4.9%)	5452 (7.6%)	2383 (7.3%)	1204 (7.0%)	2834 (7.1%)	39384 (5.4%)
66 - 70	21695 (3.9%)	5249 (7.3%)	2603 (8.0%)	1343 (7.8%)	2885 (7.3%)	33775 (4.7%)
71 - 75	18971 (3.4%)	5124 (7.1%)	2577 (7.9%)	1219 (7.1%)	2835 (7.1%)	30726 (4.2%)
76 - 80	14568 (2.6%)	4649 (6.5%)	2736 (8.4%)	1356 (7.9%)	2983 (7.5%)	26292 (3.6%)
81 - 85	12046 (2.1%)	4117 (5.7%)	2410 (7.4%)	1283 (7.5%)	2898 (7.3%)	22754 (3.1%)
86 - 90	9598 (1.7%)	3547 (4.9%)	2200 (6.8%)	1218 (7.1%)	2767 (7.0%)	19330 (2.7%)
90 - 95	1446 (0.3%)	534 (0.7%)	290 (0.9%)	161 (0.9%)	401 (1.0%)	2832 (0.4%)
<b>Previous Hospital Admission</b>						
Yes	176040 (31.3%)	36269 (50.6%)	19335 (59.4%)	10430 (60.9%)	23668 (59.6%)	265742 (36.7%)
No	386913 (68.7%)	35479 (49.4%)	13199 (40.6%)	6692 (39.1%)	16068 (40.4%)	458351 (63.3%)
<b>CTAS Score</b>						
Level 1	4300 (0.8%)	1783 (2.5%)	948 (2.9%)	458 (2.7%)	1228 (3.1%)	8717 (1.2%)
Level 2	128474 (22.8%)	25193 (35.1%)	12435 (38.2%)	7022 (41.0%)	18004 (45.3%)	191128 (26.4%)
Level 3	272616 (48.4%)	36663 (51.1%)	16260 (50.0%)	8471 (49.5%)	17817 (44.8%)	351827 (48.6%)
Level 4	139348 (24.8%)	7346 (10.2%)	2629 (8.1%)	1084 (6.3%)	2486 (6.3%)	152893 (21.1%)
Level 5	18215 (3.2%)	763 (1.1%)	262 (0.8%)	87 (0.5%)	201 (0.5%)	19528 (2.7%)
<b>ICD 10 Category</b>						
(I) Certain infectious and parasitic diseases	24957 (4.4%)	2224 (3.1%)	991 (3.0%)	497 (2.9%)	1319 (3.3%)	29988 (4.1%)
(II) Neoplasms	1906 (0.3%)	783 (1.1%)	526 (1.6%)	267 (1.6%)	475 (1.2%)	3957 (0.5%)

Characteristics of Mispredicted Patients in the Regression Model

	Mispredicted by <4 Hours (N=562953)	Mispredicted by 4-6 Hours (N=71748)	Mispredicted by 6-8 Hours (N=32534)	Mispredicted by 8-10 Hours (N=17122)	Mispredicted by >10 Hours (N=39736)	Overall (N=724093)
(III) Diseases of the blood	1989 (0.4%)	749 (1.0%)	414 (1.3%)	217 (1.3%)	392 (1.0%)	3761 (0.5%)
(IV) Endocrine, nutritional and metabolic diseases	5170 (0.9%)	2076 (2.9%)	1365 (4.2%)	745 (4.4%)	1461 (3.7%)	10817 (1.5%)
(V) Mental and behavioral disorders	13697 (2.4%)	7006 (9.8%)	6525 (20.1%)	4984 (29.1%)	11544 (29.1%)	43756 (6.0%)
(VI) Diseases of the nervous system	11035 (2.0%)	1738 (2.4%)	596 (1.8%)	278 (1.6%)	754 (1.9%)	14401 (2.0%)
(VII) Diseases of the eye and adnexa	7204 (1.3%)	351 (0.5%)	62 (0.2%)	18 (0.1%)	59 (0.1%)	7694 (1.1%)
(VIII) Diseases of the ear and mastoid process	8677 (1.5%)	354 (0.5%)	75 (0.2%)	25 (0.1%)	69 (0.2%)	9200 (1.3%)
(IX) Diseases of the circulatory system	16301 (2.9%)	5803 (8.1%)	3043 (9.4%)	1365 (8.0%)	3123 (7.9%)	29635 (4.1%)
(X) Diseases of the respiratory system	47905 (8.5%)	4770 (6.6%)	2584 (7.9%)	1541 (9.0%)	3643 (9.2%)	60443 (8.3%)
(XI) Diseases of the digestive system	31187 (5.5%)	7112 (9.9%)	3397 (10.4%)	1654 (9.7%)	3377 (8.5%)	46727 (6.5%)
(XII) Diseases of the skin and subcutaneous tissue	19483 (3.5%)	1496 (2.1%)	526 (1.6%)	224 (1.3%)	649 (1.6%)	22378 (3.1%)
(XIII) Diseases of the musculoskeletal system and connective tissue	34034 (6.0%)	3146 (4.4%)	866 (2.7%)	346 (2.0%)	1011 (2.5%)	39403 (5.4%)
(XIV) Diseases of the genitourinary system	28728 (5.1%)	3495 (4.9%)	1421 (4.4%)	740 (4.3%)	1552 (3.9%)	35936 (5.0%)
(XV) Pregnancy, childbirth and the puerperium	10808 (1.9%)	1054 (1.5%)	250 (0.8%)	118 (0.7%)	182 (0.5%)	12412 (1.7%)
(XVI) Certain conditions originating in the perinatal period	1450 (0.3%)	44 (0.1%)	13 (0.0%)	5 (0.0%)	18 (0.0%)	1530 (0.2%)
(XVII) Congenital malformations, deformations and chromosomal abnormalities	386 (0.1%)	46 (0.1%)	3 (0.0%)	6 (0.0%)	14 (0.0%)	455 (0.1%)
(XVIII) Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	141383 (25.1%)	17622 (24.6%)	5998 (18.4%)	2501 (14.6%)	6152 (15.5%)	173656 (24.0%)
(XIX) Injury, poisoning and certain other consequences of external causes	156653 (27.8%)	11879 (16.6%)	3879 (11.9%)	1591 (9.3%)	3942 (9.9%)	177944 (24.6%)
<b>Patient Postal District</b>						
Rural	25079 (4.5%)	3995 (5.6%)	1958 (6.0%)	1042 (6.1%)	2374 (6.0%)	34448 (4.8%)
Urban	537874 (95.5%)	67753 (94.4%)	30576 (94.0%)	16080 (93.9%)	37362 (94.0%)	689645 (95.2%)



Characteristics of Mispredicted Patients in the Regression Model

	Mispredicted by <4 Hours (N=562953)	Mispredicted by 4-6 Hours (N=71748)	Mispredicted by 6-8 Hours (N=32534)	Mispredicted by 8-10 Hours (N=17122)	Mispredicted by >10 Hours (N=39736)	Overall (N=724093)
<b>Hospital ID</b>						
1	55607 (9.9%)	1816 (2.5%)	757 (2.3%)	313 (1.8%)	668 (1.7%)	59161 (8.2%)
2	46341 (8.2%)	9131 (12.7%)	3645 (11.2%)	1686 (9.8%)	4657 (11.7%)	65460 (9.0%)
3	50610 (9.0%)	7055 (9.8%)	2865 (8.8%)	1482 (8.7%)	3400 (8.6%)	65412 (9.0%)
4	24018 (4.3%)	4676 (6.5%)	2385 (7.3%)	1448 (8.5%)	3540 (8.9%)	36067 (5.0%)
5	41811 (7.4%)	5849 (8.2%)	3123 (9.6%)	1940 (11.3%)	5006 (12.6%)	57729 (8.0%)
6	36712 (6.5%)	8960 (12.5%)	4273 (13.1%)	2193 (12.8%)	4030 (10.1%)	56168 (7.8%)
7	57355 (10.2%)	10790 (15.0%)	5111 (15.7%)	2415 (14.1%)	4783 (12.0%)	80454 (11.1%)
8	12318 (2.2%)	916 (1.3%)	302 (0.9%)	164 (1.0%)	389 (1.0%)	14089 (1.9%)
9	18412 (3.3%)	732 (1.0%)	201 (0.6%)	105 (0.6%)	331 (0.8%)	19781 (2.7%)
10	31748 (5.6%)	4098 (5.7%)	1787 (5.5%)	908 (5.3%)	2301 (5.8%)	40842 (5.6%)
11	47752 (8.5%)	7151 (10.0%)	3387 (10.4%)	1758 (10.3%)	4477 (11.3%)	64525 (8.9%)
12	35262 (6.3%)	1283 (1.8%)	356 (1.1%)	150 (0.9%)	170 (0.4%)	37221 (5.1%)
13	17246 (3.1%)	1806 (2.5%)	813 (2.5%)	454 (2.7%)	1142 (2.9%)	21461 (3.0%)
14	11355 (2.0%)	790 (1.1%)	253 (0.8%)	144 (0.8%)	287 (0.7%)	12829 (1.8%)
15	42049 (7.5%)	5687 (7.9%)	2985 (9.2%)	1776 (10.4%)	4252 (10.7%)	56749 (7.8%)
16	34357 (6.1%)	1008 (1.4%)	291 (0.9%)	186 (1.1%)	303 (0.8%)	36145 (5.0%)
<b>Hospital Type</b>						
Large Urban Ambulatory	35262 (6.3%)	1283 (1.8%)	356 (1.1%)	150 (0.9%)	170 (0.4%)	37221 (5.1%)
Large Urban	242952 (43.2%)	39378 (54.9%)	19018 (58.5%)	10597 (61.9%)	24705 (62.2%)	336650 (46.5%)
Teaching	170658 (30.3%)	22527 (31.4%)	9766 (30.0%)	4558 (26.6%)	10395 (26.2%)	217904 (30.1%)
Suburban / Rural	114081 (20.3%)	8560 (11.9%)	3394 (10.4%)	1817 (10.6%)	4466 (11.2%)	132318 (18.3%)

For the baseline characteristic to have little impact on the misprediction in the regression model, it should have a relatively consistent percentage across each of the mispredicted categories, such as biological assigned sex. However, for most variables, the misprediction time is diverse, ranging anywhere from generally decreasing percentage as the misprediction becomes greater (such as hospital type), generally increasing percentage as the misprediction becomes greater (such as rural patient postal districts), to sporadic (such as CTAS score). The most observable changes occurs in age group 1 (ages 1 - 5), where there is 12.2% mispredicted by less than 4 hours, but only 1.1% mispredicted by over 10 hours, category 5 for ICD-10 categories,

where it increases from only 2.9% mispredicted by less than 4 hours to 29.1% mispredicted by over 10 hours, and large urban hospitals that increase from 43.2% mispredicted by less than 4 hours to 62.2% mispredicted by more than 10 hours.

## Discussion

### Baseline Characteristics

There is not a notable difference in the sex of the patients who come to the ED: both sexes make up roughly 50% of the data, both for the over and under 4 hours stay time categories. This indicates that there is not a significant difference in the possibility of either sexes going into the ED for any length of stay. However, figure 2 does show a greater difference for different ages: a greater number of young individuals have a length of stay of less than 4 hours, while more aged people stay in the ED for longer than 4 hours. This may suggest that, as people age, they develop more complicated medical histories or conditions that prolong their stay.

A greater number of patients who come to the ED have not been previously hospitalized within 3 years; more of which stay for less than 4 hours. For people without hospitalization status, their condition is likely to be less severe, thus the stay time is generally shorter.

The baseline characteristics indicate that most individuals who stay in the ED for levels 1, 2, and 3 stay for longer than 4 hours, while more patients in levels 4 and 5 stay for less than 4 hours. This suggests that the more severe the condition, the longer a patient stays, likely because the treatment is more complex and time-consuming. Moreover, many hospitals have a “fast track” to rapidly run patients with lower CTAS levels through the hospital, which likely contributes to most patients with lower triage levels having shorter stay times.

The institutional ID shows that, of all the 16 encoded hospitals, hospital 1 has the most patients having a stay time of less than 4 hours while being the hospital with the tenth most patients having a stay time over 4 hours. This suggests that hospital 1 either has efficient patient control or most patients going to hospital 1 have conditions that require less time to treat. Contrastingly, hospital 7, which had the most patients of any hospital, had the most patients over

4 hours and the second most number of patients of all patients with a stay time of less than 4 hours. In consideration of the number of patients coming from this hospital, the institution is likely urban, and may have moderate management and resource control, or most of the patients in the urban area going to the hospital generally have conditions that take longer to treat, or some combination of both.

With institutional peer groups, it is of note that more patients in large urban hospitals (the number of patients make up for 50% of the dataset) stay for more than 4 hours than for less. This suggests that urban patients may go to the ED for diseases that are more lengthy to treat, or that urban hospitals may have worse resource allocation and management suitable for its large number of patients. However, for large urban ambulatory and rural/suburban hospitals, there is a greater number of patients who have a stay time of less than 4 hours than for more, which indicates that, with smaller hospitals, the resource management may be more efficient or patients have conditions that take less time to treat.

## **Classification Models**

### *Trends*

A trend was observed on the classification models: as the timebreak reaches an extreme high or low, the accuracy increases. This is likely because, as the timebreak becomes more extreme, there is a high difference in the number of data points in the two categories, so the model classifies most of the data points in the category with more people. Since one category may have significantly more people than the other one, classifying most people into the larger category is more likely to be correct, thus increasing the accuracy. As seen in the confusion matrices for the models with 6 and 8 hours as the time break, the model becomes better at

identifying the patients who stay less than the time break. For these two models, there are more people in the category of a short stay, meaning that most people are classified into that category, the predictions are more likely to be correct, thus increasing the overall accuracy.

When observing models with increasing numbers of categories, accuracy seems to decrease as the number of categories increases. This is likely due to the large range of factors that contribute to the length of stay's unpredictability, making it difficult for the model to classify accurately as the number of categories increase and the increment of each category decreases.

### *Misclassified Patients*

Although most of the variables showed an increase in accuracy in the 6-hour timebreak model, the age group in table 5. show an interesting trend, where, for age groups 13 - 19 (ages 61 - 95), the accuracy decreases from the 4-hour timebreak model to the 6-hour timebreak model. According to the 1-variable partial dependence plot with age group, the length of stay for groups 13 - 19 are all, on average, longer than 6 hours. Generally, mid-younger age groups hold a greater number of people, and the youngest patients have generally shorter stay times than patients who are in older age groups. Mid-older patients, however, are concentrated around the 6 hour mark, with only the highest of age groups having a more obviously long stay time. Due to this, the 4-hour timebreak classification model would likely categorize most of the older patients into the above 4 hours category as most of them do stay for longer than 4 hours. However, there is less certainty with the 6 hour-timebreak model as the mid-older patients are highly concentrated around the 6 hour stay time mark. Thus, the model is more likely to classify them into the incorrect category. This is potentially why there is a slight decrease in the correct prediction rate when the age group is increased.

Similarly, for patients who have been previously hospitalized in the past three years, the incorrect predictions also increase when the time break increases. This may be as patients who have been previously hospitalized are more likely to have more severe and sporadic conditions. The 4-hour timebreak model is likely to classify most of them as staying longer than 4 hours, but the 6 hour-timebreak model has to predict the data with previously hospitalized patients who stay for just around 6 hours, thus making it more difficult to predict.

For triage code (CTAS score), levels 1 and 2 show the number of incorrect predictions as somewhat consistent but with a slight increase. This may be due to the unpredictability of the patients who fall under this category, as the stay time is more likely to vary for severe conditions. For levels 3 to 5, the conditions are not as severe and the time needed to treat them decreases, which is why patients are more likely to fall under the shorter than 6 hours category for the 6 hour-timebreak model, thus increasing the accuracy for these 3 triage code levels.

### *Disposition Group*

The model has shown higher performance at identifying the people who were discharged home. One reason may be that people who were not admitted into the hospital generally have less severe conditions, which may potentially be more predictable than people who were admitted into the hospital. Another possible explanation may be that there are more people who were discharged home compared to the other two categories. Since there are more people in the discharged category, the predictions made by the model are more likely to be correct, thus increasing the overall accuracy.

## **Regression Model**

### *Variable Importance Plots*

As shown by the plot, the most important variable was the ICD 10 category; the type of disease or injury thus has some impact on the length of stay of a patient. The age group suggests that a patient's age is correlated with their ED stay time, as well as the severity of their condition as reflected in the triagecode. The institutional ID having some importance suggests that some hospitals may have conditions that cause longer or shorter stays than others (such as number of employees, location, etc.). The patient district code, in which their housing area could be correlated with their financial situation, shows to have moderate relation with the length of stay, suggesting that the economic status of a patient may impact the stay time. The previous hospitalization status has less of an impact on the model, suggesting that it does not contribute too much to a patient's length of stay. The type of hospital has minimal importance to the model, this indicates that hospitals may generally provide similar services without regard to the type of hospital. Lastly, sex assigned at birth is the least important to the model, which suggests that there is likely no bias for patients based on their sex.

### *1-Variable Partial Dependence Plots*

The PDP of the ICD-10 categories showed that mental and behavioral disorders, neoplasms, diseases of the blood, endocrine/nutritional/metabolic diseases had the highest length of stays. This may be because these are diseases or disorders that are difficult and lengthy to treat, or that it may require more observation and consideration for the treatment of. Conversely, diseases of the eye and the ear were the two categories that lead to the shortest stay times. This may be due to these diseases needing short treatment, resulting in a shorter ED stay length.

The PDP for age group showed a large jump of ~90 minutes between the age groups of 10 - 15 and 16 - 20). With 18 years old as the legal age, this difference may indicate a difference in wait time between regular hospitals and childrens' hospitals. Another possibility is that, in general, children may just be more prioritized than adults in the emergency department.

For the PDP for triage code (CTAS score), the indication was that patients with the most severe conditions have nearly double the length of stay as patients with the least severe conditions. This may be due to patients having more severe conditions requiring more complex and lengthy treatments as most high-urgency situations require immediate and lengthy action before it is possible for a patient to be transferred to another department. Contrarily, patients with the lowest scores and least severe conditions have the shortest ED stay time. This may be due in part to those patients requiring only small and fast treatments, and thus they may leave the emergency department faster. Another consideration is that this may be due to many hospitals having a "fast track" for patients with lower CTAS scores to manage them through the hospital's system faster.

### *2-Variable Partial Dependence Plots*

The PDP of ICD-10 categories and triage code levels show mostly sporadic points due to the ICD-10 categories being strictly classified and not progressive. The general trend shows that higher CTAS levels result in longer stay times, regardless of the ICD-10 category. However, the PDP shows an interestingly large spike in length of stay at category 5 (mental and behavioral disorders), of which the stay time increases as the severity increases. In consideration of the 1-variable PDP of ICD-10 category, where category 5 was the most lengthy condition to treat, this PDP indicates that the more severe the mental or behavioral disorder, the longer it will take



to treat, especially for the most severe of cases. For other conditions, such as the ones for categories 6 (diseases of the nervous system), 7 (diseases of the eye), 8 (diseases of the ear), and 16 (conditions originating in the perinatal period), the difference in time between difference triage levels is small, suggesting that, regardless of severity, these are all conditions that take similar amounts of time to treat. Moreover, for categories 2 (neoplasms), 3 (diseases of the blood), 4 (endocrine, nutritional, and metabolic diseases), 9 (diseases of the circulatory system), 11 (diseases of the digestive system), and 17 (diseases of the skin), the predicted stay time is longer for triage level 2 more than 1. One possibility is that there are external factors, such as more patients having those conditions triaged as CTAS score 2 go to a busier hospital, or having other complications, which prolong their stay time longer than patients with those conditions and triage level 1.

For the PDP of triage code and previous hospitalization status, there is an observable trend of the increase of triage levels (thus decrease of severity) being correlated with the decrease of stay time. However, it is noticeable that, for every triage code level, patients with a history of previous hospitalization stay, on average, just under 100 minutes longer than patients without. This suggests that, for all severities, there is a correlation between a patient's previous hospitalization status and their wait time. The decrease of difference in predicted wait time between previously and not previously hospitalized patients with the increase of triage levels suggests that, in more severe cases, the previous hospitalization yields more impact on the stay length, while in less severe cases the previous hospitalization has less impact. However, it is of consideration that the majority of patients in the data are in CTAS score 2; the large distribution of data may be contributing to the difference in stay time between those who have and have not been previously hospitalized in this level.

While the PDP of age group and triage codes shows a general trend of stay time increasing as the age group increases across all triage levels, there is a noticeable gap between the times of CTAS level 1 for age groups 3 (ages 11 - 15) and 4 (ages 16 - 20), with the latter having a roughly 100 minutes longer stay time. As these different groups outline the break between youth and adults, these results may be highlighting the different treatment between children and adults. It could be that children's hospitals have more effective management, resource allocation, and are overall more efficient than regular hospitals and are more prepared and prompt to remedy severe conditions; it could also be that regular hospitals may give children some priority ahead of adults, which is why there is such a great gap between the two stay times. The sporadicness of triage level 1 suggests that there are a variety of different factors and conditions that are correlated with the most severe conditions, causing the length of stay to be the most spread out. On the other hand, triage level 3, 4, and 5 are more linear and stable, indicating that for lower severities, the length of stay is more predictably longer as a patient grows older.

The PDP of sex and age group shows that, again, age group increase is directly correlated with the increase in stay time. With the limited number of intersex patients in the dataset, the increased trend proposed by the 'N/A', or intersex category, cannot be fully supported. However, male and female sex do show a change from age group to age group, where the youngest ages (groups 1 - 4; ages 1 - 20) have females having a slightly higher length of stay, the adult age groups of 5 - 13 (ages 21 - 65) have males having a higher stay time, and senior age groups including age groups 14 - 19 (ages 66 - 95) have females, again, having a higher stay time. This may suggest that younger males have a stronger immunity that/or requires shorter treatment, which changes as they come of age. This may be due to females obtaining a stronger immunity as they become adults, or males having other external factors that cause them to have more

severe and/or longer treatments at this time. The results then shift again as senior males seem to have shorter stay times than females, which becomes greater as the age groups reach 19, suggesting that males, as seniors, require treatments that are shorter than that of females, which may be due to external factors that cannot be accurately determined with the dataset. One reason may be, however, that females generally have lower income, especially in retirement and elder years; this may cause female seniors to stay away from the ED until absolutely necessary, of which their diseases may be severe and require a much longer treatment. With the differences between sexes being as slight as they are, however, there is not any extremely significant impact of sex on the predicted stay time.

The PDP of the age group with the previous hospitalization status shows that patients with a history of previous hospitalization do have longer length of stays than those without, across all ages (the stay time increases with the age as well). Interestingly, in referral to the baseline characteristics table, the majority of people who come to the ED are those without previous hospitalization, but patients who do have a history of previous hospitalization tend to stay for longer. There is a noticeably large increase for those with previous hospitalization between the age groups of 3 (11 - 15) and 4 (16 - 20), suggesting that coming-of-age patients who have been previously hospitalized have a longer stay time; however, this gap is evidently smaller for patients without a previous hospitalization status. One rationale may be because both children and adults, although there is some great gap between the two, who have not been previously hospitalized are more likely to have less severe conditions that require less time to treat. There is also a large difference in the points of more senior patients: this indicates that there is some significant factor, most likely external, that causes previously hospitalized senior patients to have a longer stay time. One reason may be that having been previously hospitalized

within three years indicates that the senior patient has already had a severe disease, which is more likely to relapse in older patients in intense severity and causing longer stay time; another conclusion could be that seniors who have not been previously hospitalized are likely to be healthier with less severe conditions. For either case, if the patient was aware of a condition that would require them to go to the ED, the stay time pre-transfer may be shorter.

### *Mispredicted Patients*

For all the variables, the number of patients mispredicted by less than 4 hours is the highest across all the different columns. However, for some sections of the table, such as biologically assigned sex and previous hospitalization status, more people are mispredicted for longer than 10 hours compared to the 4 - 10 hours categories. It may indicate that these characteristics have complex external factors that are difficult to completely predict with limited variables, likely causing a higher absolute residual value in the model.

## **Conclusion**

### **Conclusion**

Classification models have demonstrated considerable accuracy, especially the binary classification models. Regression models have the potential to make more accurate predictions when there are both a sufficient number of variables and data points. Due to the high complexity of the factors that contribute to a patient's stay time in the ED, it is difficult to make specific and accurate predictions with the limited number and precision of variables. The three most impactful variables were ICD 10 category, triage code, and age group. Categories 2 (neoplasms), 3 (diseases of the blood), 4 (endocrine, nutritional and metabolic diseases), and 5 (mental and behavioral disorders) generally had longer stay times; while categories 7 (diseases of the eye) and 8 (diseases of the ear) had the shortest length of stays comparatively. There may also be some difference in the treatment of underage and of age patients, where underage patients tend to have shorter length of stays regardless of the severity or previous hospitalization status; children's hospitals may have more effective management and resource allocation, or children may be just given priority in some hospitals. Generally, the increase of age and an increase in severity of condition is also directly correlated with the inclination of predicted length of stay; patients who have been previously hospitalized within 3 years also have a higher chance of longer stay times than patients without.

### **Limitations**

In the dataset, there was a rather limited number of variables compared to the large number of data. The input variables are categorical, making it difficult for the model to predict numerical values as the output. Potential variables that could be considered include the exact age

instead of age group, laboratory and screening test requested, physical activity rate of the patient, BMI, smoking and alcohol consumption, medical history, economic status, and level of education. Moreover, the dataset consists of only 30% of ED visits in the years 2018 - 2020, which may limit the generalizability of the findings.

### **Future Work**

Areas for improvement of the study would be to have a larger number of variables and more specific ones. For future analysis, the same dataset can be analyzed using other machine learning methods, where the results may be compared to Random Forest results to evaluate the efficacy of each machine learning model. Statistical analysis may also be employed to find correlations and the significance of independent variables.

## Acknowledgements

Our deepest gratitude and appreciation to:

**Dr. Jessalyn Holodinsky** for mentoring us through this project,

**Dr. Catherine Eastwood** for helping us find a mentor at the department,

**Dr. Bing Li** for helping us find and use the dataset,

**Ms. Jillian Vandenbrand** for helping us schedule meetings with Dr. Jessalyn,

**Our school teachers** for the project registration,

and, lastly, **our parents and friends** for all of their support throughout this study.

## References

- Basic evaluation measures from the confusion matrix.* (2015, June 3). Classifier Evaluation with Imbalanced Datasets; Classifier evaluation with imbalanced datasets.  
<https://classeval.wordpress.com/introduction/basic-evaluation-measures/>
- Chugh, A. (2020, December 8). *MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?* Medium.  
<https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>
- How is Variable Importance Calculated for a Random Forest?* (2018, July 30). Displayr.  
<https://www.displayr.com/how-is-variable-importance-calculated-for-a-random-forest/>
- IBM. (2023). *What is Random Forest? | IBM.* Wwww.ibm.com.  
<https://www.ibm.com/topics/random-forest>
- Molnar, C. (n.d.). 5.1 Partial Dependence Plot (PDP) | Interpretable Machine Learning. In *christophm.github.io*. <https://christophm.github.io/interpretable-ml-book/pdp.html>
- Out-of-Bag Error | Dremio.* (n.d.). Wwww.dremio.com. Retrieved March 14, 2024, from  
<https://www.dremio.com/wiki/out-of-bag-error/>
- Patient emergency department total length of stay (LOS) | HQCA Focus.* (2024). Focus.hqca.ca.  
<https://focus.hqca.ca/charts/total-length-of-patients-emergency-department-stay-los/>
- Random Forests · AFIT Data Science Lab R Programming Guide.* (n.d.). Afit-R.github.io.  
[https://afit-r.github.io/random\\_forests](https://afit-r.github.io/random_forests)
- Sruthi, E. R. (2021, June 17). *Random Forest | Introduction to Random Forest Algorithm.* Analytics Vidhya.  
<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>



Taylor, C. (2019, January 27). *What Are Residuals?* ThoughtCo.

<https://www.thoughtco.com/what-are-residuals-3126253>

*THE CANADIAN TRIAGE AND ACUITY SCALE Combined Adult/Paediatric Educational Program PARTICIPANT'S MANUAL Triage Training Resources.* (2007).

[http://ctas-phctas.ca/wp-content/uploads/2018/05/participant\\_manual\\_v2.5b\\_november\\_2013\\_0.pdf](http://ctas-phctas.ca/wp-content/uploads/2018/05/participant_manual_v2.5b_november_2013_0.pdf)

What Is a Confusion Matrix in Machine Learning? (2022, August 18). Plat.AI.

<https://plat.ai/blog/confusion-matrix-in-machine-learning/>

Wickramasinghe, S. (n.d.). *Bias & Variance in Machine Learning: Concepts & Tutorials.* BMC

Blogs. <https://www.bmc.com/blogs/bias-variance-machine-learning/>

World Health Organization. (2019). *ICD-10 Version:2019.* Icd.who.int.

<https://icd.who.int/browse10/2019/en>