**Capstone Project Final Report**

**Prediction of Inpatient Admission**

Yutong Wu, Ning Yang, Zihan Zheng

### **Introduction**

#### **1.1 Sponsor Background**

Community Health Plan of Washington (CHPW) is a local, Washington‐based Health Plan with long‐established ties to communities throughout the State and well‐equipped to facilitate and

coordinate with local resources on behalf of their members. CHPW is a not‐for‐profit company governed by community organizations (Community Health Centers) that are in turn governed by individuals that receive care within those organizations. They serve 255,000 Medicaid members and 10,000 Medicare members. Their programs are designed to proactively identify and address the behavioral, social, and medical needs of our members and to recognize the whole person’s needs.

#### **1.2 Problem Description and Previous Related Work**

Past studies have shown that most health care expenditure is usually concentrated on a small proportion of the population. For example, nearly 50% of health care spending can be attributed to the top 5% health care resource users in Ontario, Canada (Roselle et.al, 2018). Similar Patterns were observed in Australia, Japan, the European Union, as well as the U.S. (Roselle et.al, 2018). It is also well documented that the U.S. is an [outlier among peer countries](https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/) in terms of healthcare spending, and recent Peterson-KFF [analysis](https://www.healthsystemtracker.org/brief/what-drives-health-spending-in-the-u-s-compared-to-other-countries/) finds that the cost of inpatient care is one of the primary drivers behind the gap in health expenditures (McDermott et al., 2020). It is essential to develop a valid population-based risk prediction tool that helps identify population segments that will likely be admitted to the hospital in the near future so that the care management team in health insurance agencies can provide timely support for high-risk individuals and reduce the likelihood of adverse health outcomes.

When developing a prediction mode, if the training data is biased towards certain groups or classes of objects, the prediction model will also show discriminatory behavior towards that particular community. This will lead to biased outcomes when labeling unlabeled data objects (Kamiran & Calders, 2009). Impartial classification results are usually desired or even required for future data due to ethical and legal considerations (Kamiran & Calders, 2009). The issue of racial bias has been found in a prior predictive model developed by CHPW, which aimed to identify factors that can help predict the likelihood of being a high resource utilizer (HRU) of the health system. Members who experience barriers or discrimination to seeking health care, especially African Americans, were found to have a systematically lower likelihood of having the outcome of being a HRU. We anticipate that the same issue would likely come up again in a prediction model for inpatient admission, given that the study population behind the two predictions largely overlap. Hence, we will be adapting a Classification with No Discrimination (CND) technique to reduce the prejudicial behavior for outcome prediction with a minimal loss of prediction accuracy (Kamiran & Calders, 2009).

#### **1.3 Objectives**

The study aims to inform case management teams of Community Health Plan of Washington about enrollees who have a higher likelihood of inpatient admission in the following year so that the in-house care management team can reach out to the high-risk members, helping them achieve and maintain a healthier state and control the risk of adverse health consequences.

Primary Objective:

Build a prediction model for inpatient admission.

Secondary Objectives:

Apply Classification with No Discrimination (CND) Technique to correct for racial bias in the new prediction model.

### **2. Methods Overview**

#### **2.1 Study Population**

The target population includes all CHPW members enrolled from Jan 2018 to Dec 2019 who are legal Washington state residents and enrolled in Apple Health (Washington State’s Medicaid Program), Medicare, or other Washington state health care programs.

The purpose of developing the inpatient admission predictive model is to identify member segments that need attention and may benefit from care management services. Pregnant members and members with terminal illness are very likely to be admitted to inpatient care in a short period given their existing conditions, and case management is not expected to make a detectable difference for these individuals. Furthermore, newborns are impossible or highly unlikely to have experienced some of our predictors of interest, such as homelessness, smoking diagnosis and substance use disorder. Hence, participants that meet one or more of the listed criteria are excluded from the analysis: maternity members, members with severe cancer, with end-stage renal disease (ESRD), with transplant, with dialysis, with hospice, those who died during the period, and newborns with age < 1.

#### **2.2 Data Sources**

The project involves the use of in-house data on demographics information, healthcare cost, health resource utilization, health care program enrollment, disease diagnosis, social determinants of health (SDoH) and behavioral health information available within the health plan. The cost, utilization, diagnosis, SDoH and behavioral health data are collected through claims submitted by the health care providers to the health plan. The enrollment and demographic data are collected through enrollment files provided by Washington state.

The following is the a list of information available from the in-house data:

* Enrollment
* Medical claim (medical cost, health resource utilization, facility visits, etc.)
* Medical diagnosis (chronic conditions)
* Risk level (calculated using internal algorithms by CHPW)
* Demographic information
* Social determinants of health (SDoH)
* Behavioral health

#### 

#### **2.3 Dataset**

From the database, we have one aggregated dataset containing information on enrollment and claim information of each member that is maintained by CHPW and updated on a monthly basis; one dataset containing demographic information of each member; and one dataset with diagnosis of chronic conditions. We combine the information from all three datasets into one dataset to build the prediction model. Below is a description of the variables contained in the dataset:

Outcome variable

* Binary outcome of inpatient admission in the next 12 months (2019.1 - 2019.12).

Predictor variables (all information recorded in 2018)

* Demographic information:
  + Age (continuous)
  + Sex (binary)
  + Race (categorical)
  + Residence region (categorical)
* Enrollment information:
  + Enrolled plan (categorical)
  + Length of enrollment in months (count)
* Medical claim information:
  + Expenses for medical services (continuous)
  + Expenses for prescribed medicines (continuous)
  + Primary care visits (count)
  + Emergency room visits (count)
  + Inpatient admissions (count)
  + Specialty visits (count)
* SDoH and behavioral health information:
  + Homelessness (binary)
  + Behavioral health diagnosis (binary)
  + Substance abuse disorder (binary)
  + Severe mental illness (binary)
  + Smoking diagnosis (binary)
* Risk level (ordinal) calculated using the internal algorithm, based on healthcare cost, health resource utilization, as well as number and severity of chronic conditions from past 24 months
* Chronic conditions: there are a total of 60 chronic conditions, which are grouped into 20 categories (see Table 2 in the appendix for details):
  + Chronic infectious diseases
  + Cancer
  + Diabetes
  + Metabolic and endocrine disorders
  + Chronic liver diseases
  + Digestive system disorders
  + Chronic skeletal conditions
  + Hematological disorders
  + Mental health disorders
  + Substance used and abuse
  + Central nervous system disorders
  + Respiratory and lung issues
  + Cardiac diseases
  + Cerebrovascular diseases
  + Vascular diseases
  + Ophthalmological conditions
  + Chronic kidney diseases
  + Chronic skin conditions
  + Implant, transplant, and graft
  + Amputation

Characteristics of all predictors by outcome class is shown in Table 1 in the appendix.

#### **2.4 Prediction Models**

Given a large number of candidates for predictors of interest, the following modeling techniques are selected as primary candidates:

* Penalized Logistic Regression: Penalized logistic regression imposes a penalty for including too many predictors. The coefficients of less constructive predictors are forced to be 0, which encourages a simple, sparse model. The model is easy to implement, computationally inexpensive and straightforward to interpret. However, we expect its performance to be less strong than tree-based models, which can capture non-linear relationships between predictor and outcome.
* Gradient Boosting Trees (GBT): Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models. GBT is a tree-based model that incorporates interactions, GBT has shown great results in many classification benchmarks, and it is highly customizable and can be tuned for imbalanced data. However, it is less interpretable in nature, tends to overfit, and has lots of tuning parameters.
* Bayesian Additive Regression Trees (BART): Bayesian Additive Regression Tree is a sum of trees model adopting Bayesian framework. It uses Markov Chain Monte Carlo bayesian backfitting in training. The model is highly complex and it is regularized by imposing a prior probability. As a tree-based model, it naturally incorporates interaction between variables. It also provides confidence intervals in the prediction. It can be modified for classification, and it does not have a lot of parameters to tune in classification settings. According to the original BART paper (Chipman, 2010), the default model performs just as well as the tuned model using cross validation. The downside of BART is the high computational complexity and its non-interpretability.

All three model classes have feature selection and regularization mechanisms, and they are proven to work well in binary classification.

#### **2.5 Overview of Modeling Procedure**

Workflow:

* As described in Figure 1, after data extraction, processing, and exploration, we split the data into training data and testing data in a ratio of 80:20 stratified on inpatient admission.
* For each of the three model types (penalized logistic regression, GBT BART):
  + We train the model on training data and optimize the performance by cross validation. This is model training stage 1.
  + Then, using the optimized model from stage 1, we make predictions on the training data, and get the predicted probabilities to guide the implementation of CND (see specific implementation in section 2.6).
  + Next, we use the best hyperparameters found in the stage 1 cross-validation, to train the model again on CND-massaged training data, and get the final model from each class. This is model training stage 2.
* Lastly, we evaluate the three final models on testing data, and elect the overall best candidate model for deployment.

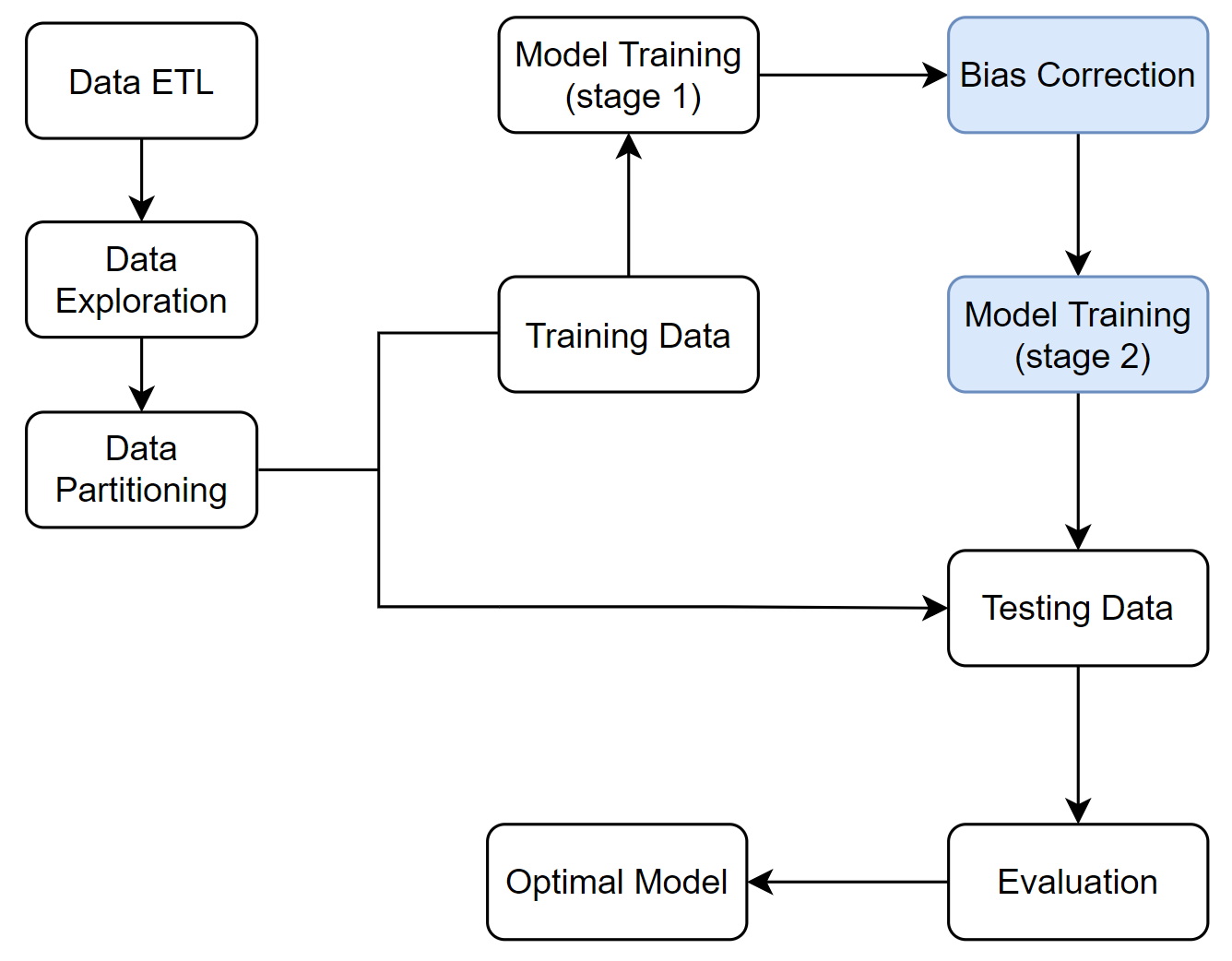


Figure 1. Flowchart of Modeling Procedure

Evaluation

* Model performance is evaluated using threshold-based classification accuracy metrics defined below. In particular, the F1 score is used because of the imbalance of the inpatient admission. Given that the binary outcome of inpatient admission has a very unbalanced distribution, it is more important to correctly identify true positives, as opposed to true negatives. Hence, we focused on the F1 score, which balances sensitivity and positive predictive value.
* CND results are evaluated by Chi-squared test on the equality of proportions of inpatient admissions across racial groups.

Performance metrics are defined as:

* True Positives (TP): Predicted label and actual label are both ‘positive’
* False Positives (FP): Predicted label is ‘positive’ but actual label is ‘negative’
* True Negatives (TN): Predicted label and actual label are both ‘negative’
* False Negatives (FN): Predicted label is ‘negative’ but actual label is ‘positive’
* Sensitivity: Proportion of actual positives captured by the model: TP / (TP + FN)
* Specificity: Proportion of actual negatives captured by the model: TN / (TN + FP)
* Positive Predictive Value (PPV): Proportion of predicted positives that are true positives: TP / (TP + FP)
* Negative Predictive Value (NPV): Proportion of predicted negatives that are true negatives: TN / (TN + FN)
* F1: The harmonic mean of Sensitivity and PPV: 2\*Sensitivity\*PPV / (Sensitivity + PPV)

#### **2.6 CND Technique**

The Classification with No Discrimination (CND) is a technique that correct the bias in training data by massing the dataset in the least intrusive way (Kamiran & Calders, 2009). We adopted the original technique that is designed for correcting bias over two classes to our model building process to correct the bias across eight racial groups. The process goes as follows:

1. Examine the bias in the training data, by calculating and comparing the event rate across racial groups. If bias exist, then we proceed to CND
2. Use the proportion of population in each race and total inpatient admissions to calculate the expected number of admissions
3. Take the difference between actual admissions and expected admission for each race. This difference will be the number of modifications needed in each race.
4. If a race has a positive difference, assign observations in that race with admission to domotion group; if that difference is negative, assign observations in that race with no admission to promotion group
5. Train and tune the prediction model on the original training data
6. Use the predicted probability of inpatient admission to rank the training data within each group
7. Switch the outcome label from 1 to 0 for observations in the demotion group with the lowest rank; switch the outcome label from 0 to 1 for observations in the promotion group with the highest rank. The number of switches for each group is determined in step 3
8. Finally, after all modifications are done, use the massaged data to train and tune the model again, as the final prediction model.

**3. Results**

#### **3.1 Descriptive Analysis**

The processed data contains 179899 observations and 85 variables.

Upon initial examination of the data, we found the outcome class is highly unbalanced, with only 4.3% of the whole population being admitted to inpatient facilities in 2019, shown in Figure 2.

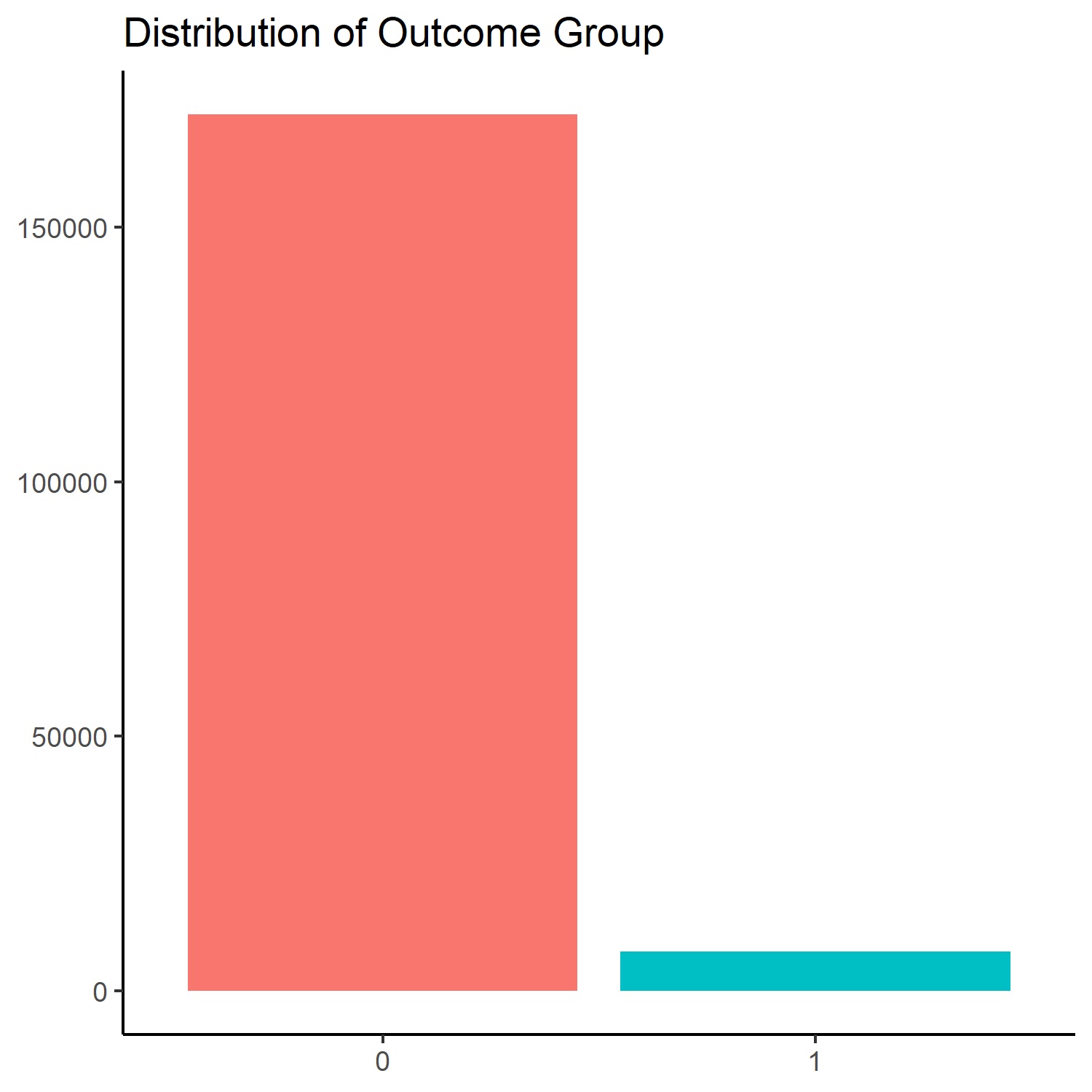


Figure 2. Distribution of outcome group

We also examined the proportion of people who are admitted to inpatient facilities in 2019 in different racial groups, shown in Figure 3. We observe that the incidence rate in different racial groups is quite different, ranging from 2.8% in Asians to 6.6% in American Indian or Alaska Natives. More formally, we performed a Chi-squared test to test the equality of proportions of positive groups across eight racial groups, and we found enough evidence that these proportions are not equal (p < 0.0001), and concluded that there is racial bias in inpatient admissions in the data. The number of observations in each group that was used to perform the Chi-squared test are shown in Table 3.

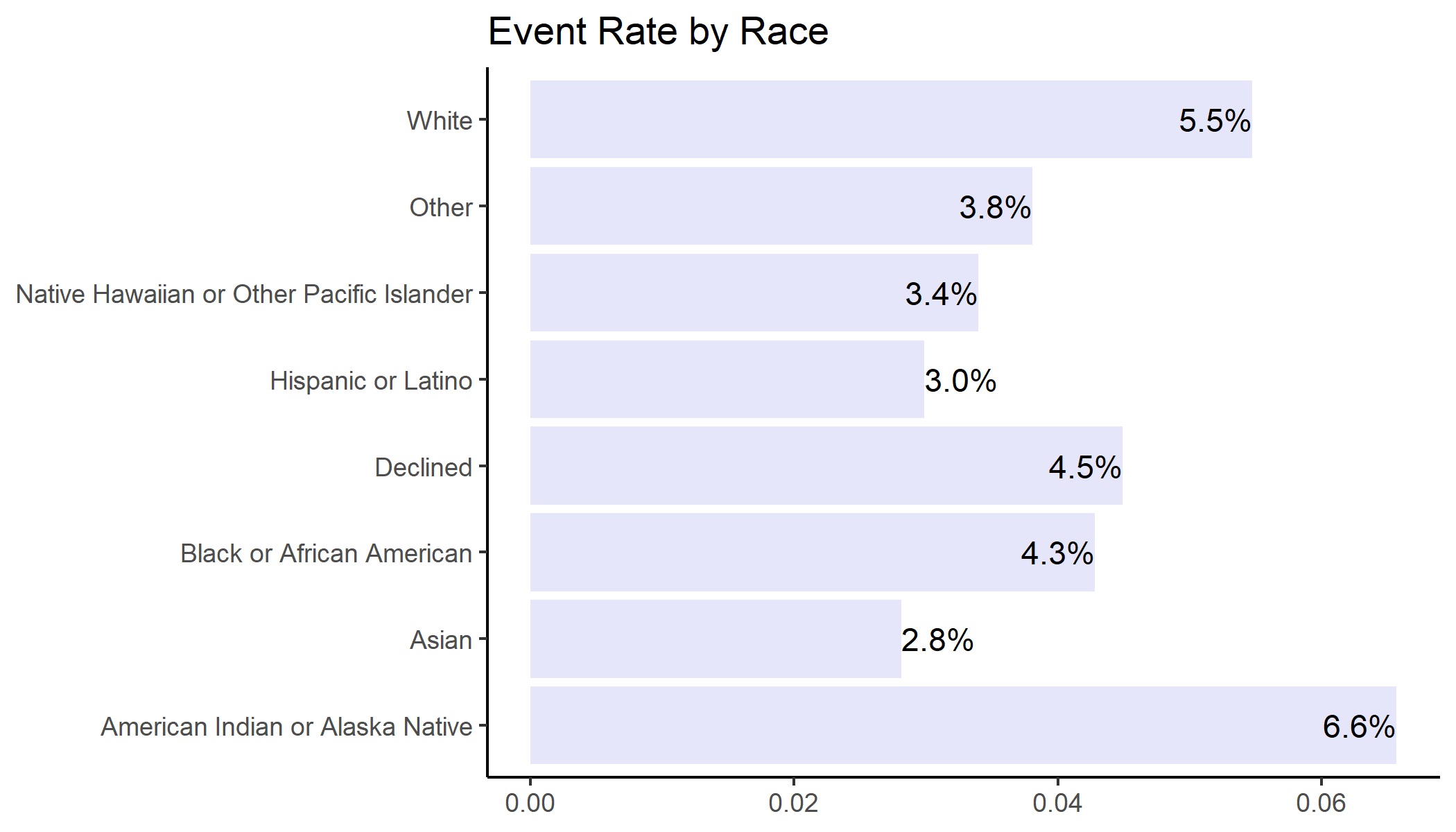


Figure 3. Proportion of Inpatient Admissions by Race

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | American Indian / Alaska Native | Asian | Black / African American | Declined | Hispanic or Latino | Native Hawaiian / Pacific Islander | Other | White |
| Overall count | 1872 | 12229 | 13739 | 4518 | 57306 | 4536 | 2811 | 82888 |
| Positive count | 123 | 344 | 588 | 203 | 1711 | 154 | 107 | 4537 |

Table 3. Racial bias evaluation.

We noticed that many indicator variables have low event rate, especially in the chronic condition section, with many of the event rates being less than 0.05 even after grouping (see Table 1. in appendix).

#### **3.2 Model Performance**

For each prediction modeling method (Penalized Logistic Regression, XGBoost, BART), cross-validation was performed to tune the model hyperparameters. The performance of the best model in each method, measured in sensitivity, specificity, PPV, NPV, and F1 score are listed in Table 4(a).

|  |  |  |  |
| --- | --- | --- | --- |
| Performance  Metric | Penalized Logistic Regression | XGBoost | BART |
| Sensitivity  TP/(TP+FN) | 0.239 (0.218, 0.261) | 0.346 (0.321, 0.371) | 0.342 (0.318, 0.365) |
| Specificity  TN/(TN+FP) | 0.968 (0.966, 0.970) | 0.970 (0.968, 0.971) | 0.970 (0.962, 0.979) |
| PPV  TP/(TP+FP) | 0.249 (0.228, 0.272) | 0.323 (0.299, 0.347) | 0.342 (0.318, 0.365) |
| NPV  TN/(TN+FN) | 0.966 (0.964, 0.968) | 0.973 (0.971, 0.974) | 0.970 (0.962, 0.979) |
| F1 | 0.242 | 0.334 | 0.342 |

Table 4(a). Model Performance Before CND Implementation

(with estimated 95% CI for all metrics except F1 score)

The best model of each class was used to guide the data massaging in CND. The models with the optimal hyperparameters (Table 6 in Appendix) were trained again on CND-modified training data. The performance of the final models based on the testing data after implementation of CND are shown in Table 4(b). We noticed that CND implementation has only a small impact on the performance.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Penalized Logistic Regression | XGBoost | BART |
| Sensitivity  TP/(TP+FN) | 0.243 (0.221, 0.265) | 0.343 (0.319, 0.368) | 0.334 (0.311, 0.358) |
| Specificity  TN/(TN+FP) | 0.965 (0.963, 0.967) | 0.969 (0.967, 0.971) | 0.970 (0.961, 0.978) |
| PPV  TP/(TP+FP) | 0.238 (0.217, 0.259) | 0.317 (0.294, 0.341) | 0.334 (0.311, 0.358) |
| NPV  TN/(TN+FN) | 0.966 (0.964, 0.968) | 0.972 (0.971, 0.974) | 0.970 (0.961, 0.978) |
| F1 | 0.240 | 0.330 | 0.334 |

Table 4(b). Model Performance After CND Implementation

(with estimated 95% CI for all metrics except F1 score)

#### **3.3 CND Results**

We evaluated the effect of CND based on predictions on the testing data. Before the application of CND, we found significant racial bias in the predictions (Chi-squared test p-value < 0.0001). After the application of CND there is much less racial bias in the predictions, and we do not have statistical evidence that the proportions of inpatient admissions in each racial group are different (Chi-squared test p-value = 0.20). Specific predicted counts by race are shown in Table 5.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Overall  Counts | Overall  Proportion | Positive Counts | Expected counts | Predicted counts Before CND | Difference before CND | Predicted counts After CND | Difference after CND |
| American Indian Alaska Native | 398 | 1.1% | 33 | 17 | 31 | 14 | 21 | 4 |
| Asian | 2417 | 6.7% | 61 | 105 | 42 | -63 | 85 | -20 |
| African American | 2825 | 7.9% | 110 | 123 | 114 | -9 | 114 | -9 |
| Declined | 890 | 2.5% | 38 | 39 | 36 | -3 | 38 | -1 |
| Hispanic or Latino | 11361 | 31.6% | 369 | 493 | 293 | -200 | 492 | -1 |
| Native Hawaiian Pacific Islander | 889 | 2.5% | 25 | 39 | 30 | -9 | 36 | -3 |
| Other | 572 | 1.6% | 29 | 25 | 27 | 2 | 34 | 9 |
| White | 16627 | 46.2% | 897 | 722 | 989 | 277 | 742 | 20 |
| Total | 35979 | 100% | 1562 | 1563 | 1562 | -1 | 1562 | -1 |

Table 5. Effect of Racial Bias Correction

#### **3.4 Feature Importance**

For penalized logistic regression, importance was measured by the magnitude of the absolute value of the coefficient on the predictor variable. Top 10 most important predictors are:

* Insurance plan of BH Service Only
* Risk level of Catastrophic
* Chronic digestive illness
* American Indian and Alaska Native race
* Risk level of High
* Hispanic and Latino race
* Service region of Pierce
* Insurance plan of Medicare SNP
* Spanish language
* Sex

(shown in Figure 4(a)).

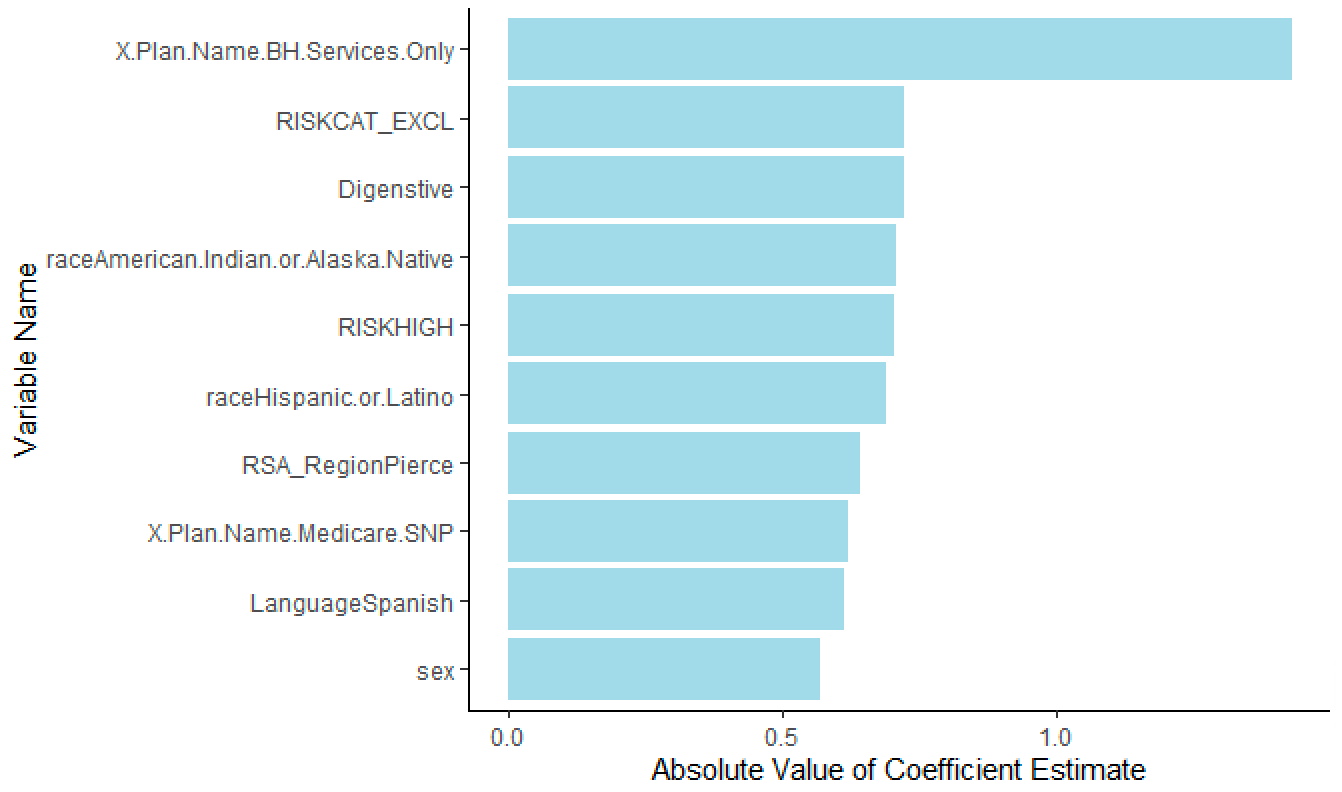


Figure 3(a). Penalized logistic regression importance plot

For XGBoost, importance was measured by information gains, which implied the relative contribution of the corresponding feature to the model calculated by taking each feature’s contribution for each tree in the model. A higher value of this metric when compared to another feature implies it is more important for generating a prediction. Top 10 important predictors are:

* Age
* Number of chronic conditions in the last 12-month period
* Number of membership months in the last 12-month period
* Insurance plan of Apple Health (Medicaid)
* Medical expense
* Medical expense ratio
* Revenue
* Total expense
* Sex
* Number of inpatient admissions in the last 12-month period

(shown in Figure 3(b).

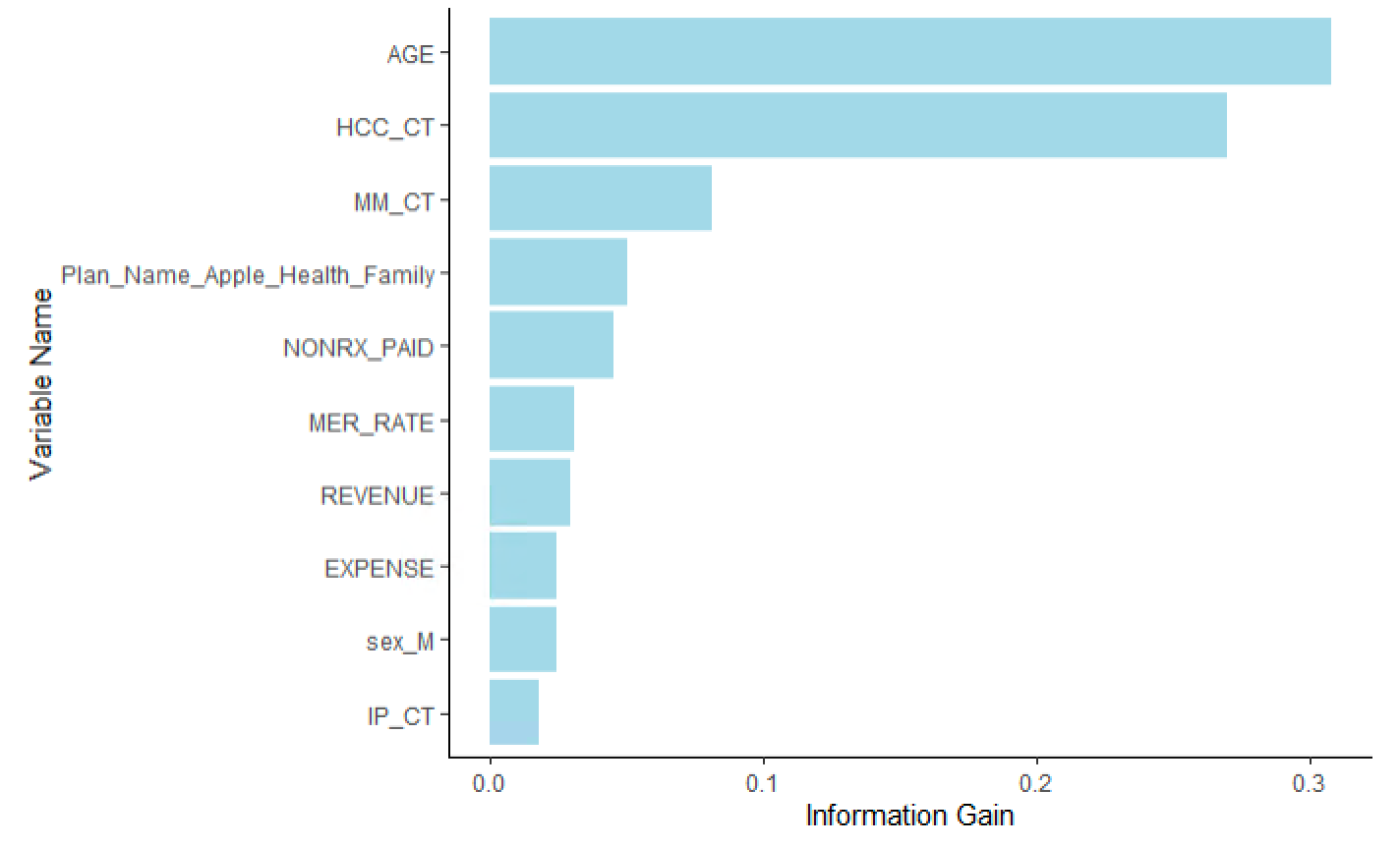


Figure 3(b). XGBoost importance plot

For BART, importance was measured by the proportion of inclusion in tree splits of the predictor. A feature is more important if it was included in more splitting in the sum-of-trees model. Top 10 important features are:

* Number of membership months in the last 12-month period
* Age
* Revenue
* Medical Expense
* Chronic cardiac conditions
* Hispanic and Latino race
* Number of chronic conditions in the last 12-month period
* Number of inpatient admission in the last 12-month period
* Insurance plan of Apple Health (Medicaid)
* Chronic infectious diseases (shown in Figure 3(c))

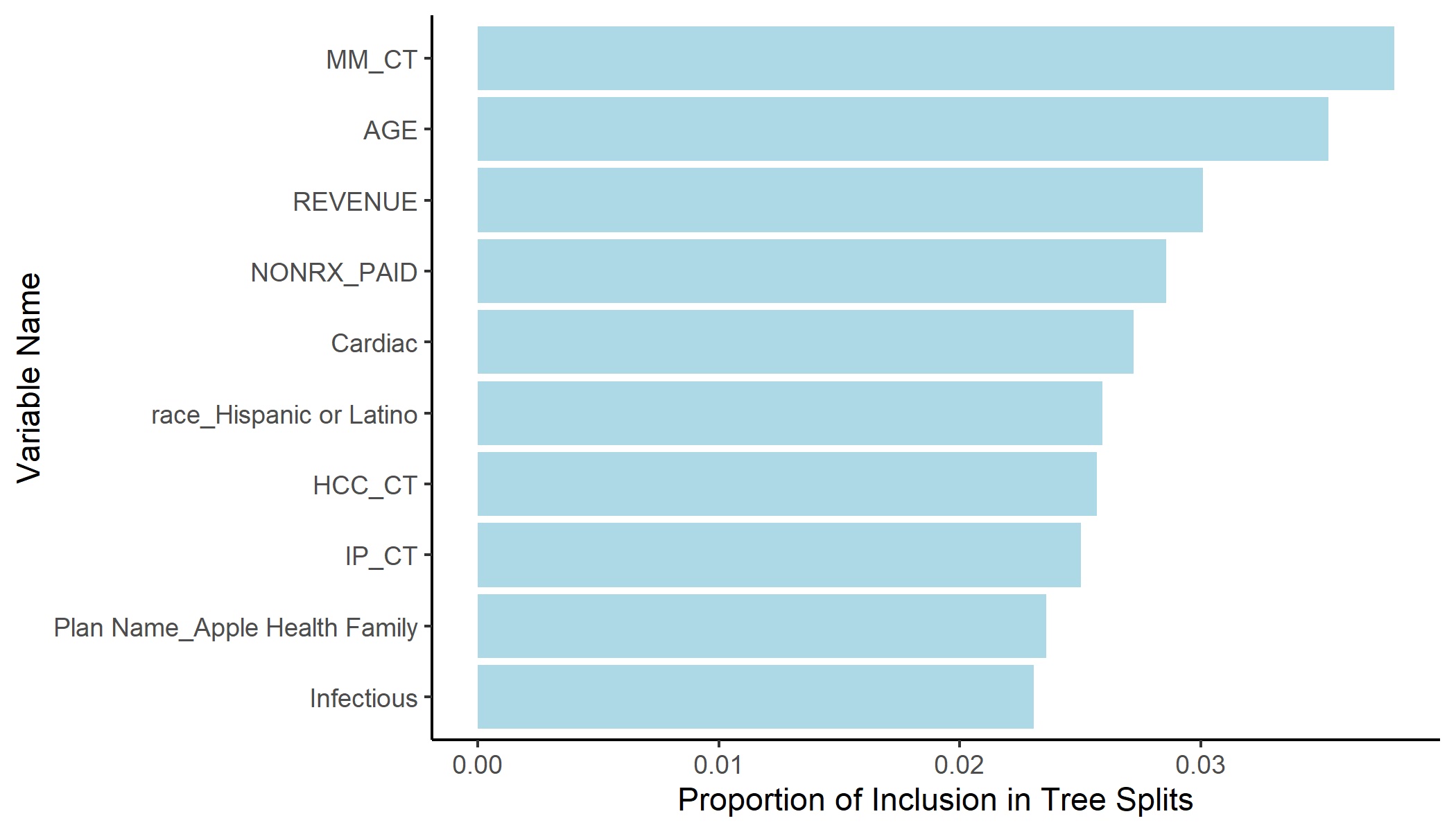


Figure 3(c). BART importance plot

**4. Discussion**

#### **4.1 Feature Importance**

According to Penalized logistic regression, the most important features for prediction are insurance plan (BH Service and Medicare SNP), risk level (catestrophic and high), digestive system disorder, race (native American and Latino), and sex.

This result is quite different from XGBoost and BART, while these two have very similar results. According to both XGBoost and BART: age, number of chronic conditions, membership month, medical expenses, revenue, insurance plan (Medicaid), and number of inpatient admissions in the previous year are among the most important features.

#### **4.2 Limitations**

Given that the binary outcome of inpatient admission has a highly unbalanced distribution, it is more important to correctly identify the true positives, as opposed to the true negatives. Hence, we choose F1 score as our main evaluation matrix, which combines sensitivity and positive predictive value. The F1 scores of the selected prediction models are not ideal in general, ranging from 0.24 to 0.34.

Regarding the non-ideal F1 score, we think the primary reason is still the imbalance of outcome distribution. We tried to deal with this issue through different approaches. For penalized logistic regression and BART, we tried over-sampling and under-sampling with class ratios of 1:1, 1:2, and 1:4, respectively. For XGBoost, we tried setting the *scale\_pos\_weight* parameter, which controls the balance of positive and negative outcome classes, to be the ideal value (21.76), given that 4% of participants have the outcome. However, model performance does not have any meaningful improvement after implementing these techniques.

Additionally, the high accuracy but non-ideal F1 score might indicate that the model is underfitting the unbalanced data. Thus, some of the future improvements can be:

1. Modify model: Use a better classification algorithm and better hyper-parameters, such as neural networks.

2. Solve the data unbalanced problem using other oversampling, undersampling, or weights approaches, like SMOTE.

3. Use a subset of features that contribute more to outcome prediction and exclude less important features.

Another limitation of the study is that the predictions from XGBoost and BART are not deterministic due to the stochastic nature of the algorithm themselves when using multi-thread parallelization. One way to avoid this issue is to set the number of CPU threads to 1 before running the algorithm. However, it would be extremely time consuming to do so given the size of the data and the limited computational resources.

#### **4.3 Other Discovery**

After massaging our training dataset using CND technique, we applied the best model from penalized logistic regression, XGBoost, and BART, respectively. According to Table 4, the test F1 score is 0.262 before CND and 0.25 after CND for penalized logistic regression, 0.328 before and 0.324 after for XGBoost, and 0.342 before and 0.334 after for BART. The train F1 score is 0.242 before CND and 0.308 after CND for penalized logistic regression, 0.381 before CND and 0.499 after CND for XGBoost, 0.338 before CND and 0.480 after CND for BART. The train F1 score increased a lot after racial bias correction while the test F1 score did not change much and even slightly decreased. The possible reason is that we only applied CND on the training dataset. This means, after racial bias correction on the dataset, the same model classifies more observations into the correct class. However, racial bias correction on training dataset does not necessarily improve the performance of the model.

#### **4.4 Project Impact and Future Direction**

Although our model performances are not ideal, they performed a lot better than the stepwise logistic regression model developed by our sponsors in their previous work (see Table 7 in the appendix for specifics). Our models shed light on factors that contribute the most to the prediction of inpatient admission, which should help case management teams of CHPW identify high-risk enrollees, improve their health status and reduce the risk of adverse health-related incidents. Given the large sample size and high diversity of our study population, future statisticians may consider splitting the study population into multiple subgroups based on certain characteristics and identify an optimal prediction model for each subgroup. Subgroup-specific model development can possibly improve the prediction accuracy for certain member segments or the entire population. Additionally, CHPW may consider collecting data on other variables that can contribute to the prediction of inpatient admission, such as participants’ education level, stress level and type of occupation. Besides, imbalanced outcome distribution is a major obstacle to higher model performance. If time and resources allow, it would be worthwhile to try different techniques to handle the imbalanced dataset, such as applying Random Over-Sampling Examples (ROSE) or Synthetic Minority Oversampling Technique (SMOTE). ROSE includes a combination of over and under-sampling. It creates a synthetic sample of data simulated based on a smoothed-bootstrap approach. SMOTE randomly samples data from the minority class and uses k-nearest neighbor algorithm to generate new synthetic data for the minority class. The procedure is repeated until the classification balance is reached (Wijaya, 2020). Finally, trying a wider range of statistical methods might also contribute to the boost of prediction accuracy. For example, neural networks might have the potential to outperform the current models, given that they have the ability to implicitly detect complex non-linear associations between predictors and outcome, ability to detect all possible interactions between predictors, and availability of multiple training algorithms.

**Appendix**

**Table 1. Characteristics of both outcome groups (descriptive statistics of selected predictors)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Overall | No Admission | Admission |
| Sample Size | 179,899 | 172,132 | 7,767 |
| Age, mean (sd) | 24.92 (20.50) | 24.38 (20.34) | 36.87 (20.50) |
| Plan Name, n (%) |  |  |  |
| Apple Blind and Disabled | 12,956 ( 7.20) | 11,792 ( 6.85) | 1,164 (14.99) |
| Apple Health Family | 103,636 (57.61) | 100,867 (58.60) | 2,769 (35.65) |
| BH Services Only | 3,237 ( 1.80) | 3,213 ( 1.87) | 24 ( 0.31) |
| Expansion | 52,720 (29.31) | 49,705 (28.88) | 3,015 (38.82) |
| Medicare | 3,924 ( 2.18) | 3,548 ( 2.06) | 376 ( 4.84) |
| Medicare SNP | 3,426 ( 1.90) | 3,007 ( 1.75) | 419 ( 5.39) |
| Medical Expense, mean (sd) | 1,950.00 (11611.42) | 1,617.37 (9319.56) | 9,321.83 (33783.09) |
| Rx Expense, mean (sd) | 827.46 (5114.72) | 754.28 (4847.11) | 2,449.26 (9083.07) |
| Total Expense, mean (sd) | 2,777.46 (13317.57) | 2,371.65 (11056.60) | 11,771.09 (36253.49) |
| Revenue, mean (sd) | 2,916.21 (3911.43) | 2,807.90 (3611.15) | 5,316.50 (7703.65) |
| Medical Expense Ratio, mean (sd) | 0.96 (8.82) | 0.86 (7.83) | 3.14 (20.97) |
| MER = 1, n (%) | 7,099 ( 3.95) | 6,304 ( 3.66) | 795 (10.24) |
| Primary Care Visit, mean (sd) | 2.50 (4.13) | 2.39 (3.93) | 4.92 (6.80) |
| ER Visit, mean (sd) | 0.52 (1.39) | 0.48 (1.26) | 1.38 (2.91) |
| Inpatient Visit, mean (sd) | 0.07 (0.36) | 0.05 (0.28) | 0.36 (1.02) |
| Number of Chronic Conditions, mean (sd) | 0.60 (1.25) | 0.54 (1.11) | 2.07 (2.56) |
| Behavioral Health Diagnosis, incidence rate (sd) | 0.33 (0.47) | 0.32 (0.46) | 0.62 (0.49) |
| Indicator of Rx > 8, incidence rate (sd) | 0.14 (0.35) | 0.13 (0.34) | 0.39 (0.49) |
| Indicator Obesity, incidence rate (sd) | 0.16 (0.36) | 0.15 (0.36) | 0.26 (0.44) |
| Hypertension, incidence rate (sd) | 0.11 (0.32) | 0.11 (0.31) | 0.31 (0.46) |
| Smoking, incidence rate (sd) | 0.15 (0.36) | 0.14 (0.35) | 0.42 (0.49) |
| Health Home Eligible, incidence rate (sd) | 0.07 (0.26) | 0.06 (0.24) | 0.24 (0.43) |
| Health Home Status, incidence rate (sd) | 0.01 (0.10) | 0.01 (0.09) | 0.04 (0.19) |
| MHIP, incidence rate (sd) | 0.02 (0.14) | 0.02 (0.13) | 0.03 (0.18) |
| Billing Provider Visit, mean (sd) | 2.62 (4.30) | 2.50 (4.09) | 5.22 (7.16) |
| Specialty Visit, mean (sd) | 0.68 (1.18) | 0.65 (1.14) | 1.25 (1.85) |
| RISK, n (%) |  |  |  |
| Catastrophe | 2,768 ( 1.54) | 2,118 ( 1.23) | 650 ( 8.37) |
| High | 6,158 ( 3.42) | 4,758 ( 2.76) | 1,400 (18.02) |
| Medium High | 12,692 ( 7.06) | 11,611 ( 6.75) | 1,081 (13.92) |
| Medium | 54,641 (30.37) | 52,189 (30.32) | 2,452 (31.57) |
| Medium Low | 3,1409 (17.46) | 30,617 (17.79) | 792 (10.20) |
| Low | 72,231 (40.15) | 70,839 (41.15) | 1,392 (17.92) |
| Avoidable ER Visit, mean (sd) | 0.07 (0.34) | 0.07 (0.33) | 0.12 (0.42) |
| Homeless, incidence rate (sd) | 0.09 (0.28) | 0.08 (0.28) | 0.17 (0.38) |
| Substance Use Disorder, incidence rate (sd) | 0.15 (0.35) | 0.14 (0.34) | 0.41 (0.49) |
| Severe Mental Illness, incidence rate (sd) | 0.15 (0.36) | 0.14 (0.35) | 0.34 (0.47) |
| Sex Male, n (%) | 86,483 (48.07) | 83,563 (48.55) | 2,920 (37.59) |
| Language, n (%) |  |  |  |
| English | 136,076 (75.64) | 129,349 (75.15) | 6,727 (86.61) |
| Other | 13,300 ( 7.39) | 12,875 ( 7.48) | 425 ( 5.47) |
| Spanish | 30,523 (16.97) | 29,908 (17.38) | 615 ( 7.92) |
| Membership Months, mean (sd) | 10.46 (3.38) | 10.52 (3.32) | 9.33 (4.39) |
| RSA\_Region, n (%) |  |  |  |
| Greater Columbia | 42,831 (23.81) | 41,005 (23.82) | 1,826 (23.51) |
| King | 43,667 (24.27) | 41,913 (24.35) | 1,754 (22.58) |
| North Central | 46 ( 0.03) | 38 ( 0.02) | 8 ( 0.10) |
| North Sound | 35,577 (19.78) | 34,227 (19.88) | 1,350 (17.38) |
| Peninsula | 4,562 ( 2.54) | 4,374 ( 2.54) | 188 ( 2.42) |
| Pierce | 91 ( 0.05) | 87 ( 0.05) | 4 ( 0.05) |
| Spokane | 18,102 (10.06) | 17,048 ( 9.90) | 1,054 (13.57) |
| SW WA | 14,097 ( 7.84) | 13,488 ( 7.84) | 609 ( 7.84) |
| Thurston-Mason | 8,315 ( 4.62) | 7,981 ( 4.64) | 334 ( 4.30) |
| Timberlands | 12,611 ( 7.01) | 11,971 ( 6.95) | 640 ( 8.24) |
| Race, n (%) |  |  |  |
| American Indian / Alaska Native | 1,872 ( 1.04) | 1,749 ( 1.02) | 123 ( 1.58) |
| Asian | 12,229 ( 6.80) | 11,885 ( 6.90) | 344 ( 4.43) |
| Black or African American | 13,739 ( 7.64) | 13,151 ( 7.64) | 588 ( 7.57) |
| Declined | 4,518 ( 2.51) | 4,315 ( 2.51) | 203 ( 2.61) |
| Hispanic or Latino | 57,306 (31.85) | 55,595 (32.30) | 1,711 (22.03) |
| Native Hawaiian /  Other Pacific Islander | 4,536 ( 2.52) | 4,382 ( 2.55) | 154 ( 1.98) |
| Other | 2,811 ( 1.56) | 2,704 ( 1.57) | 107 ( 1.38) |
| White | 82,888 (46.07) | 78,351 (45.52) | 4,537 (58.41) |
| Chronic conditions, incidence rate (sd) |  |  |  |
| Infectious | 0.01 (0.12) | 0.01 (0.10) | 0.09 (0.29) |
| Cancer | 0.01 (0.09) | 0.01 (0.08) | 0.04 (0.18) |
| Diabetes | 0.06 (0.23) | 0.05 (0.22) | 0.17 (0.38) |
| MetaEndo | 0.05 (0.21) | 0.04 (0.20) | 0.14 (0.35) |
| Liver | 0.02 (0.13) | 0.01 (0.12) | 0.07 (0.25) |
| Digestive | 0.00 (0.07) | 0.00 (0.06) | 0.02 (0.14) |
| Skeletal | 0.03 (0.16) | 0.02 (0.15) | 0.07 (0.26) |
| Blood | 0.02 (0.14) | 0.02 (0.13) | 0.09 (0.28) |
| Substance | 0.06 (0.25) | 0.06 (0.23) | 0.22 (0.41) |
| Mental | 0.11 (0.31) | 0.10 (0.30) | 0.25 (0.43) |
| Nerve | 0.03 (0.17) | 0.03 (0.17) | 0.10 (0.30) |
| Respiratory | 0.11 (0.31) | 0.11 (0.31) | 0.24 (0.43) |
| Cardiac | 0.02 (0.16) | 0.02 (0.14) | 0.13 (0.33) |
| Cerebro | 0.01 (0.08) | 0.01 (0.07) | 0.03 (0.17) |
| Vascular | 0.02 (0.14) | 0.02 (0.13) | 0.10 (0.30) |
| Eye | 0.00 (0.05) | 0.00 (0.04) | 0.01 (0.11) |
| Kidney | 0.00 (0.05) | 0.00 (0.04) | 0.02 (0.14) |
| Skin | 0.01 (0.08) | 0.01 (0.07) | 0.05 (0.22) |
| Implant | 0.01 (0.08) | 0.00 (0.07) | 0.03 (0.18) |
| Digestive | 0.00 (0.06) | 0.00 (0.05) | 0.03 (0.16) |
| Amputation | 0.00 (0.02) | 0.00 (0.02) | 0.00 (0.04) |

**Table 2. List of chronic conditions and corresponding categories used as predictor candidates**

|  |  |
| --- | --- |
| Condition Description | Category |
| HIV/AIDS | Infectious Diseases |
| Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock | Infectious Diseases |
| Opportunistic Infections | Infectious Diseases |
| Metastatic Cancer and Acute Leukemia | Cancer |
| Lung and Other Severe Cancers | Cancer |
| Lymphoma and Other Cancers | Cancer |
| Colorectal, Bladder, and Other Cancers | Cancer |
| Breast, Prostate, and Other Cancers and Tumors | Cancer |
| Diabetes with Acute Complications | Diabetes |
| Diabetes with Chronic Complications | Diabetes |
| Diabetes without Complication | Diabetes |
| Protein-Calorie Malnutrition | Metabolic & Endocrine Disorders |
| Morbid Obesity | Metabolic & Endocrine Disorders |
| Other Significant Endocrine and Metabolic Disorders | Metabolic & Endocrine Disorders |
| End-Stage Liver Disease | Liver Disease |
| Cirrhosis of Liver | Liver Disease |
| Chronic Hepatitis | Liver Disease |
| Intestinal Obstruction/Perforation | Digestive System Disorders |
| Chronic Pancreatitis | Digestive System Disorders |
| Inflammatory Bowel Disease | Digestive System Disorders |
| Bone/Joint/Muscle Infections/Necrosis | Skeletal |
| Rheumatoid Arthritis and Inflammatory Connective Tissue Disease | Skeletal |
| Severe Hematological Disorders | Hematological Disorders |
| Disorders of Immunity | Hematological Disorders |
| Coagulation Defects and Other Specified Hematological Disorders | Hematological Disorders |
| Dementia With Complications | Mental Health |
| Dementia Without Complication | Mental Health |
| Substance Use with Psychotic Complications | Substance Use & Abuse |
| Substance Use Disorder, Moderate/Severe, or Substance Use with Complications | Substance Use & Abuse |
| Substance Use Disorder, Mild, Except Alcohol and Cannabis | Substance Use & Abuse |
| Schizophrenia | Mental Health |
| Reactive and Unspecified Psychosis | Mental Health |
| Major Depressive, Bipolar, and Paranoid Disorders | Mental Health |
| Personality Disorders | Mental Health |
| Quadriplegia | Central Nervous System |
| Paraplegia | Central Nervous System |
| Spinal Cord Disorders/Injuries | Central Nervous System |
| Amyotrophic Lateral Sclerosis and Other Motor Neuron Disease | Central Nervous System |
| Cerebral Palsy | Central Nervous System |
| Myasthenia Gravis/Myoneural Disorders and Guillain-Barre Syndrome/Inflammatory and Toxic Neuropathy | Central Nervous System |
| Muscular Dystrophy | Central Nervous System |
| Multiple Sclerosis | Central Nervous System |
| Parkinson's and Huntington's Diseases | Central Nervous System |
| Seizure Disorders and Convulsions | Central Nervous System |
| Coma, Brain Compression/Anoxic Damage | Central Nervous System |
| Respirator Dependence/Tracheostomy Status | Respiratory & Lung Issues |
| Respiratory Arrest | Respiratory & Lung Issues |
| Cardio-Respiratory Failure and Shock | Respiratory & Lung Issues |
| Congestive Heart Failure | Cardiac Disease |
| Acute Myocardial Infarction | Cardiac Disease |
| Unstable Angina and Other Acute Ischemic Heart Disease | Cardiac Disease |
| Angina Pectoris | Cardiac Disease |
| Specified Heart Arrhythmias | Cardiac Disease |
| Intracranial Hemorrhage | Cerebrovascular Disease |
| Ischemic or Unspecified Stroke | Cerebrovascular Disease |
| Hemiplegia/Hemiparesis | Cerebrovascular Disease |
| Monoplegia, Other Paralytic Syndromes | Cerebrovascular Disease |
| Atherosclerosis of the Extremities with Ulceration or Gangrene | Vascular Disease |
| Vascular Disease with Complications | Vascular Disease |
| Vascular Disease | Vascular Disease |
| Cystic Fibrosis | Respiratory & Lung Issues |
| Chronic Obstructive Pulmonary Disease | Respiratory & Lung Issues |
| Fibrosis of Lung and Other Chronic Lung Disorders | Respiratory & Lung Issues |
| Aspiration and Specified Bacterial Pneumonias | Respiratory & Lung Issues |
| Pneumococcal Pneumonia, Empyema, Lung Abscess | Respiratory & Lung Issues |
| Proliferative Diabetic Retinopathy and Vitreous Hemorrhage | Ophthalmology |
| Exudative Macular Degeneration | Ophthalmology |
| Dialysis Status | Kidney Disease |
| Acute Renal Failure | Kidney Disease |
| Chronic Kidney Disease, Stage 5 | Kidney Disease |
| Chronic Kidney Disease, Severe (Stage 4) | Kidney Disease |
| Chronic Kidney Disease, Moderate (Stage 3) | Kidney Disease |
| Pressure Ulcer of Skin with Necrosis Through to Muscle, Tendon, or Bone | Skin |
| Pressure Ulcer of Skin with Full Thickness Skin Loss | Skin |
| Pressure Ulcer of Skin with Partial Thickness Skin Loss | Skin |
| Chronic Ulcer of Skin, Except Pressure | Skin |
| Severe Skin Burn or Condition | Skin |
| Severe Head Injury | Skeletal |
| Major Head Injury | Skeletal |
| Vertebral Fractures without Spinal Cord Injury | Skeletal |
| Hip Fracture/Dislocation | Skeletal |
| Traumatic Amputations and Complications | Skeletal |
| Complications of Specified Implanted Device or Graft | Implant & Transplant & Graft |
| Major Organ Transplant or Replacement Status | Implant & Transplant & Graft |
| Artificial Openings for Feeding or Elimination | Digestive System Disorders |
| Amputation Status, Lower Limb/Amputation Complications | Amputation |

**Table 6. Final Model Optimal Hyperparameters**

|  |  |
| --- | --- |
| Penalized Logistic Regression | LASSO penalty  Best lambda: 0.00012  Best probability threshold: 0.075 |
| XGBoost | max.depth = 3  eval\_metric = "aucpr"  eta = 0.3  min\_child\_weight = 0 |
| BART | alpha = 0.95  beta = 2  k = 2  number of trees (m) = 200 |

**Table 7: Previous similar model**

|  |  |
| --- | --- |
| Outcome | Inpatient Admission in Three Months |
| Sample Size | 55999 |
| Number of actual events | 889 |
| Number of predicted events | 107 |
| Number of events predicted correctly | 52 |
| Prediction accuracy | 0.984 |
| Sensitivity | 0.058 |
| Specificity | 0.999 |
| PPV | 0.486 |
| F-1 score | 0.104 |

**\***Study Population: CHPW members excluding Apple Health members

### 

### 

### 

### 

### 

### 

### 

### 

### **Team member contributions**

Data cleaning and extraction: Ning

Exploratory data analysis: Ning

Predictive model development:

* Penalized logistic regression: Zihan
* XGBoost: Yutong
* BART: Ning

Final report: all

Final slides: all

Capstone presentation: all

Communication with sponsors: all

References

Chipman HA, George EI, McCulloch RE (2010). “BART: Bayesian Additive Regression Trees.” The Annals of Applied Statistics, 4(1), 266–298. doi:10.1214/09-aoas285.

F. Kamiran and T. Calders, "Classifying without discriminating," 2009 2nd International Conference on Computer, Control and Communication, 2009, pp. 1-6, doi: 10.1109/IC4.2009.4909197.

McDermott, D., Hudman, J., Cotliar, D., Claxton, G., Cox, C., & Rae, M. (2020, November 4). *How costly are common health services in the United States?* Peterson-KFF Health System Tracker. Retrieved November 7, 2021, from https://www.healthsystemtracker.org/chart-collection/how-costly-are-common-health-services-in-the-united-states/#item-start.

Rosella, L. C., Kornas, K., Yao, Z., Manuel, D. G., Bornbaum, C., Fransoo, R., & Stukel, T. (2018). Predicting High Health Care Resource Utilization in a Single-payer Public Health Care System: Development and Validation of the High Resource User Population Risk Tool. *Medical care*, *56*(10), e61–e69. <https://doi.org/10.1097/MLR.0000000000000837>

Wijaya, C. Y. (2021, October 12). *5 smote techniques for oversampling your imbalance data*. Medium. Retrieved March 1, 2022, from https://towardsdatascience.com/5-smote-techniques-for-oversampling-your-imbalance-data-b8155bdbe2b5