A Data-Driven Approach to Improving Hospital Observation Unit Operations

Raheela Charania, Rajivini Tiruveedhula, Prabhu Shankar

Executive summary:
This project aims to improve the operational efficiency of Montanaro Hospital's 23-bed observation unit. The main issue of focus is the large number of patients who transferred from observation status to inpatient status, which led to the underutilization of the resources and extended length of stay. Our goal is to utilize machine learning techniques and predictive analytics to identify the key factors that led to patient transfer. The results of the analysis are expected to provide insights that can help increase operational efficiency and improve overall patient care and bed allocation. In order to predict the "flipped status" we created several models, including logistic regressions, decision tree and random forest and utilized different metrics were used to evaluate the performance criteria. The logistic regression serves as the best model for interpretation and performance. This is further discussed in the report.

Problem Description:

The main problem is to identify the primary factors that directly impact the transfer of patients, also called as "flipping" from observation to inpatient status frequently which ultimately leads to operation inefficiencies.

Methodology:

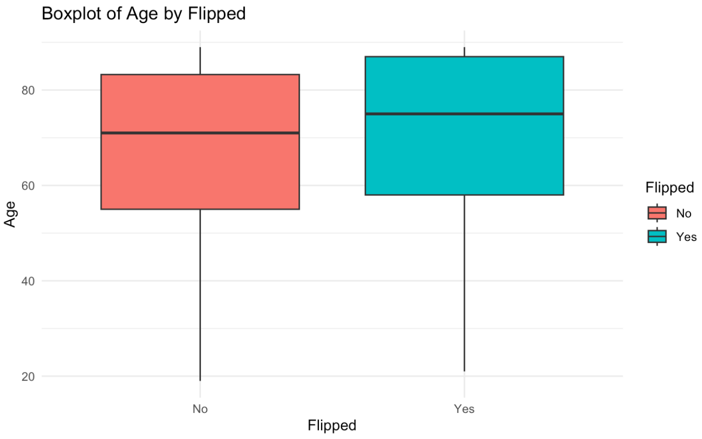
Data set: The dataset provides detailed patient information such as demographics, patient status, and patient vitals, which act as the variables in our model and will help predict the flipping status of new patients.

Data Cleaning and Transformation:

- The first step is to convert all categorical variables to numerical variables and discard any variables that are not important in our analysis. In this case, we decided to remove the 'observation record key' variable and then went ahead with imputing any missing values with the mean. Variables 'Gender' and 'insurance category' are converted to factors so that they can be used in the predictive models and creating dummy variables to facilitate a smoother analysis.
- Partitioned data into training and testing sets with a weightage of 80% and 20% respectively

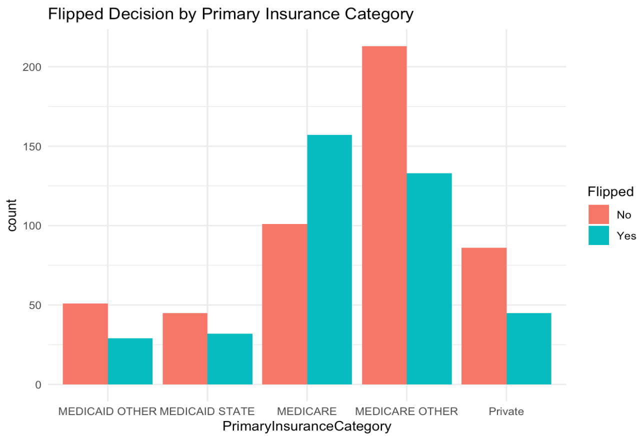
Exploratory Data Analysis:

1. Distribution of Age by Flipped Status: from the chart, it can be observed that patients 60 or older are most likely to transfer from observation to inpatient status in comparison to those that belong to the younger age groups. Patients under the age 50 do not show a high rate of transfer. From this we can conclude that patients belonging to older age groups have a high rate of "flipped" than those who are not, which shows that a relationship might exist between age and transfer status.

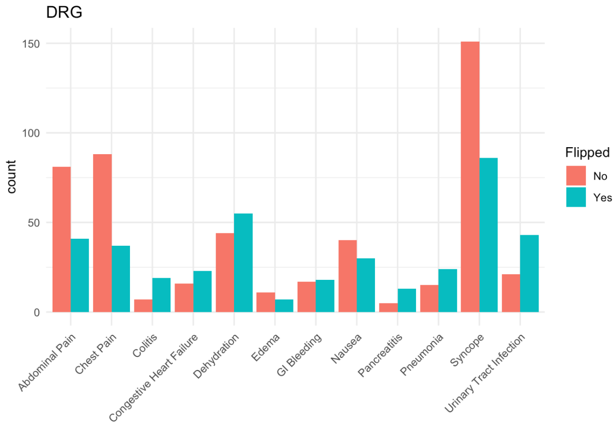


2. Flipped decision by gender: according to this chart, the number of female patients is higher than those of male patient cases, however, female patients were less likely to transfer and male patients showed an equal number of flipped and not-flipped cases.

3. Flipped decision by insurance category: Medicare Other shows the highest number of patients flipped, while other insurance categories like Medicaid state and Medicaid Other show a significantly lesser number of patients who transferred from observation to inpatient status.

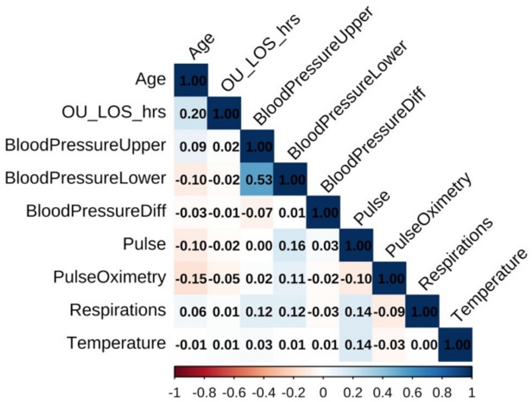


4. Flipped Decision by Disease: Patients that were admitted due to cognitive heart failure and GI bleeding have the highest number of patients who flipped, while patients admitted due to dehydration and urinary tract shows an equal number of patients who flipped and did not flip, while patients admitted due to abdominal and chest pain have more patients who did not flip.



Correlation Matrix

There is a strong correlation between upper blood pressure and lower blood pressure, which indicates a positive relation. Both, age and OU hours and pulse and lower blood pressure show a correlation in moderation.



Models:

To predict the "flipped" status, the following models were used

1. Logistic regression: This model predicts the outcome of "flipped" by estimating the probability of a patient transferring based on the variables age, blood pressure and pulse.
2. Decision tree: This model classifies patient transfer as "yes" or "no" based on decision rules that are derived using independent variables such as length of stay and insurance category that influence the decision.
3. Random forest: This method was employed to increase prediction accuracy by using results from 100 decision trees and using the average of the predictions. This method resulted in better accuracy in comparison to using individual decision trees.

Results:

Model 1: Logistic Regression Model
Model Performance:
AIC: 960.53
Model Performance on Test Dataset:
Accuracy: 64.80%
Sensitivity: 76.42%
Specificity: 47.95%
Interpretation: The first logistic model which included all variables except OU hours, primary insurance category and blood pressure upper , showed that variables DRG01Colitis, DRG01Congestive Heart Failure, DRG01Dehydration DRG01Pancreatitis, DRG01Pneumonia, DRG01Urinary Tract Infection are all statistically significant in predicting if a patient will flip or not. The overall accuracy of this model is 64.80%.

```
Call:
glm(formula = Flipped ~ . - OU_LOS_hrs - PrimaryInsuranceCategory -
    BloodPressureUpper, family = binomial, data = train_data)

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -14.406759  10.649904  -1.353  0.17613
Age                           0.006966   0.004868   1.431  0.15240
GenderMale                    0.256554   0.163703   1.567  0.11707
DRG01Chest Pain              -0.053486   0.313659  -0.171  0.86460
DRG01Colitis                  1.870622   0.553135   3.382  0.00072 ***
DRG01Congestive Heart Failure 1.094641   0.461269   2.373  0.01764 *
DRG01Dehydration              0.703824   0.321443   2.190  0.02855 *
DRG01Edema                    0.532328   0.607968   0.876  0.38126
DRG01GI Bleeding              0.663043   0.430598   1.540  0.12360
DRG01Nausea                   0.362599   0.343144   1.057  0.29065
DRG01Pancreatitis             1.629761   0.633143   2.574  0.01005 *
DRG01Pneumonia                1.043220   0.446515   2.336  0.01947 *
DRG01Syncope                 -0.098788   0.277910  -0.355  0.72224
DRG01Urinary Tract Infection  1.142180   0.374559   3.049  0.00229 **
BloodPressureLower           -0.008165   0.006166  -1.324  0.18544
BloodPressureDiff            -0.002173   0.004190  -0.519  0.60409
Pulse                         0.001472   0.005378   0.274  0.78431
PulseOximetry                 0.006850   0.034358   0.199  0.84196
Respirations                 -0.020955   0.033420  -0.627  0.53065
Temperature                   0.138721   0.099584   1.393  0.16362
```

Model 2: Reduced Logistic Regression Model
Model Performance:
AIC: 953.21
Model Performance on Test Dataset:
Accuracy: 63.69%
Sensitivity: 74.53%
Specificity: 47.95%
Interpretation: The reduced logistic regression model with only gender, disease type and temperature suggests that DRG01Colitis, DRG01Congestive Heart Failure, DRG01Dehydration, DRG01GI Bleeding, DRG01Pancreatitis, DRG01Pneumonia, DRG01Urinary Tract Infection are all significant predictors with an accuracy of 63.69% compared to the previous full model.

```
Call:
glm(formula = Flipped ~ Gender + DRG01 + Temperature, family = binomial,
    data = train_data)

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                  -13.99991    9.60856  -1.457 0.145109
GenderMale                     0.22775    0.16216   1.405 0.160169
DRG01Chest Pain               -0.04428    0.30711  -0.144 0.885360
DRG01Colitis                   1.93214    0.54916   3.518 0.000434 ***
DRG01Congestive Heart Failure  1.23537    0.44829   2.756 0.005856 **
DRG01Dehydration               0.79149    0.31119   2.543 0.010976 *
DRG01Edema                     0.57851    0.59708   0.969 0.332593
DRG01GI Bleeding               0.75844    0.42341   1.791 0.073253 .
DRG01Nausea                    0.39073    0.34013   1.149 0.250652
DRG01Pancreatitis              1.54971    0.63007   2.460 0.013910 *
DRG01Pneumonia                 1.17741    0.43444   2.710 0.006725 **
DRG01Syncope                   0.04758    0.25748   0.185 0.853386
DRG01Urinary Tract Infection   1.31367    0.35793   3.670 0.000242 ***
Temperature                    0.13546    0.09812   1.381 0.167422
```

Model 3: Decision Tree
Model Performance on Test Dataset:
Accuracy : 65.36%
Sensitivity (Ability to predict “No” correctly) : 74.53%
Specificity (Ability to predict “Yes” correctly) : 52.05%
Interpretation: The decision tree performs is almost similar to the logistic regression but the accuracy changes to 65.36%.

Model 4: Random Forest
A Random Forest model was created with 100 trees to predict whether the patient flipped based on Gender, Primary Insurance Category and DRG01.
Out of Bag Error Rate: 38.85%

Model Performance:
Accuracy: 61.45%
Sensitivity: 70.75%
Specificity: 47.95%
The random forest model shows an accuracy of 61.45%. Though the model has comparatively lower accuracy, it is more robust in prediction.

Model Evaluation and selection:

Model	Accuracy	Sensitivity	Specificity
Full logistic model	64.80 %	76.42%	47.95%
Reduced logistic model	63.69%	74.53%	47.95%
Random Forest	61.45%	70.75%	47.95%
Decision tree	65.36%	74.53%	52.05%

The final model selected is the full logistic model, which includes all variables except OU hours, primary insurance category and blood pressure upper as it shows the highest accuracy compared to all other models.
After selecting the final model that will be used for predicting "flipped" status, the results are then exported into an excel.
A cutoff point was selected which was, in this case, 0.3 and 0.45.

Decision	Before the cut off	After 0.4 cut off	After 0.35 cut off
No flip rate	0.556	0.714	0.755
Flip rate	0.443	0.285	0.244

The impact on OU utilization is show in the below table:

Type of care	Revenue saved
Post Surgery	10000
Inpatient	3000
Ward	2283

With a utilization of 73-74% from our simulation, we tend to save approximately $340,000 yearly with a prediction of using 0.35 as the cut-off rate.

Recommendations:

- By lowering the number of patients that flip from operating to inpatient status, we can increase the annual revenue to approximately 340,000$.
- By reducing the flipped rate, we can increase the utilization to 73-74%, which would mean we can serve more number of patients with a greater efficiency
- According to the simulation, we can accommodate approximately 55 patients which will increase the efficiency without impacting the quality of service.