



**TEXAS**  
The University of Texas at Austin

THE UNIVERSITY OF TEXAS  
**MD Anderson**  
~~Cancer Center~~  
Making Cancer History®

**TEXAS ORIE**

# Leveraging Machine Learning to Predict Post Surgery Length - of - Stay

---

## ***Presented By:***

Prudhvinath G (gp23259)

Soorya Sriram (s9623)

## ***Supervised By:***

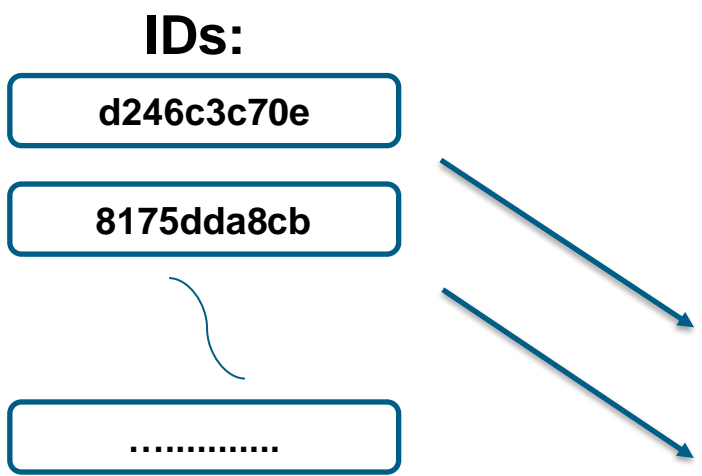
Dr Eric Bickel

Dr Erhan Kutanoglu

## Agenda:

- Background
- Solution
- Experiments
- Results
- Future Approaches

# Background:



Age	BMI	Surgery Diagnosis Name	LOS 4 groups - Target
35	24.29	Lymphede ma of right upper limb	60+ hours
57	29.54	Other acquired deformity of head	60+ hours

139,634 rows with 47634 unique IDs and 43 features of



Data

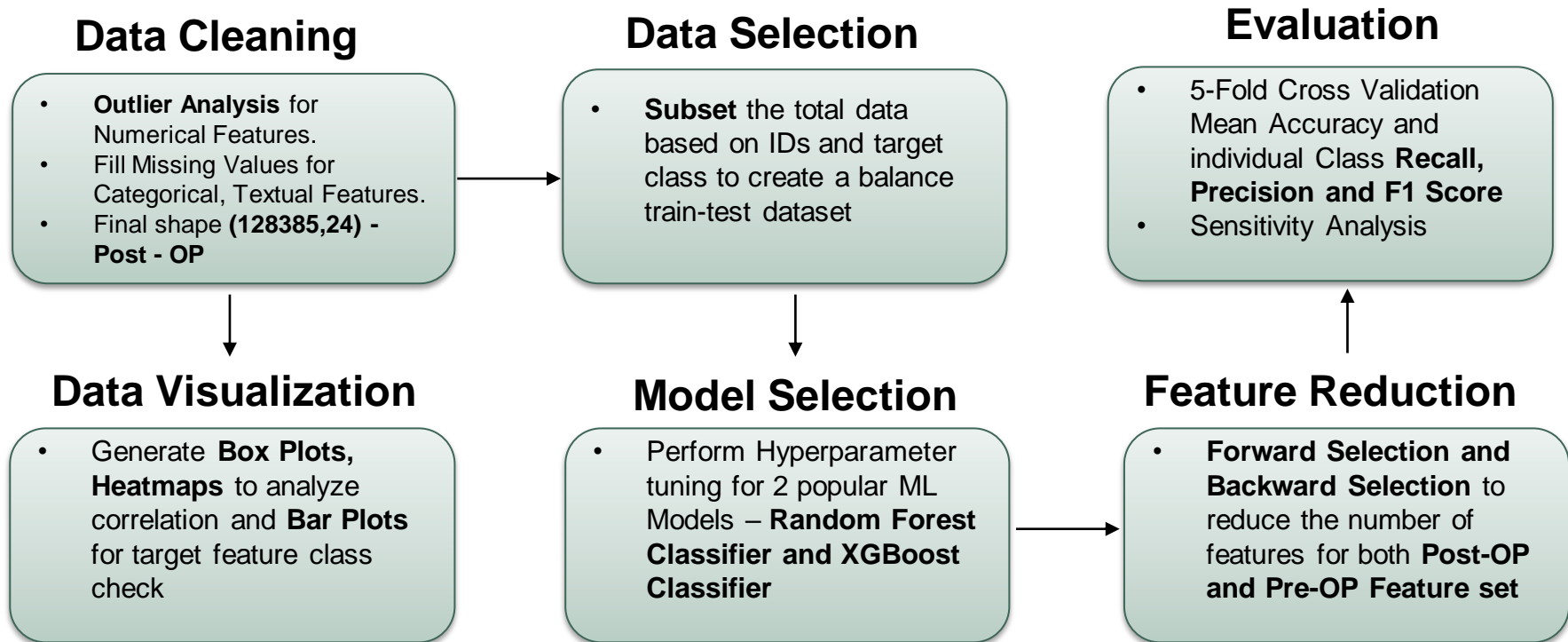
## Objective:

Develop a robust Machine Learning solution to predict the Post-Surgery Length of Stay into 4 categories to optimize the assignment of patients between inpatient and outpatient care based on a set of features known before, after and accommodating to the changes that might occur during the surgery.

## Problem Breakdown:

1. Preprocess data and perform EDA to make it suitable for training and testing.
2. Reduce the number of features using feature importance techniques based on which features are known before and after the surgery.
3. Evaluate using appropriate metrics
4. Explain the impact of features on the prediction

# Solution: End – To – End Flow



# Summary of PreOP Experiments:

	Model		Accuracy	Recall	F1 Score
Pre - OP	RFC	<i>Base</i>	89.39%	89%	89%
		<i>Tuned</i>	89.21%	89%	89%
	XGB	<i>Base</i>	78.49%	78%	78%
		<i>Tuned</i>	78.64%	79%	79%
	RFC with Selected Features (Forward Selection)	<b><i>Base</i></b>	<b>93.29%</b>	<b>93%</b>	<b>93%</b>
		<i>Tuned</i>	89.21%	89%	89%
	XGB with Selected Features (Forward Selection)	<i>Base</i>	77.72%	78%	78%
		<i>Tuned</i>	77.90%	78%	78%

- Initial features: **19**
- No. of features that got selected for RFC in Pre - Op are **11**.
- No. of features that got selected for XGB in Pre - Op are **15**.

**Recommendation:** RFC with 11 features after forward feature selection to predict the LOS target features has 93.29% accuracy as compared to all the other approaches.



# Summary of PostOP Experiments:

	Model		Accuracy	Recall	F1 Score
Post OP	RFC	<i>Base</i>	93.43%	93%	93%
		<i>Tuned</i>	87.63%	88%	88%
	XGB	<i>Base</i>	82.46%	83%	83%
		<i>Tuned</i>	83.60%	84%	84%
	RFC with Selected Features (Forward Selection)	<b><i>Base</i></b>	<b>96.18%</b>	<b>96%</b>	<b>96%</b>
		<i>Tuned</i>	95.24%	95%	95%
	XGB with Selected Features (Forward Selection)	<i>Base</i>	82.84%	84%	84%
		<i>Tuned</i>	83.09%	84%	84%

- Initial features: **23**
- No. of features that got selected for RFC in Post Op are **10**.
- No. of features that got selected for XGB in Post Op are **17**.

**Recommendation:** RFC with 10 features after forward feature selection to predict the LOS target features has 96.18% accuracy as compared to all the other approaches.

---

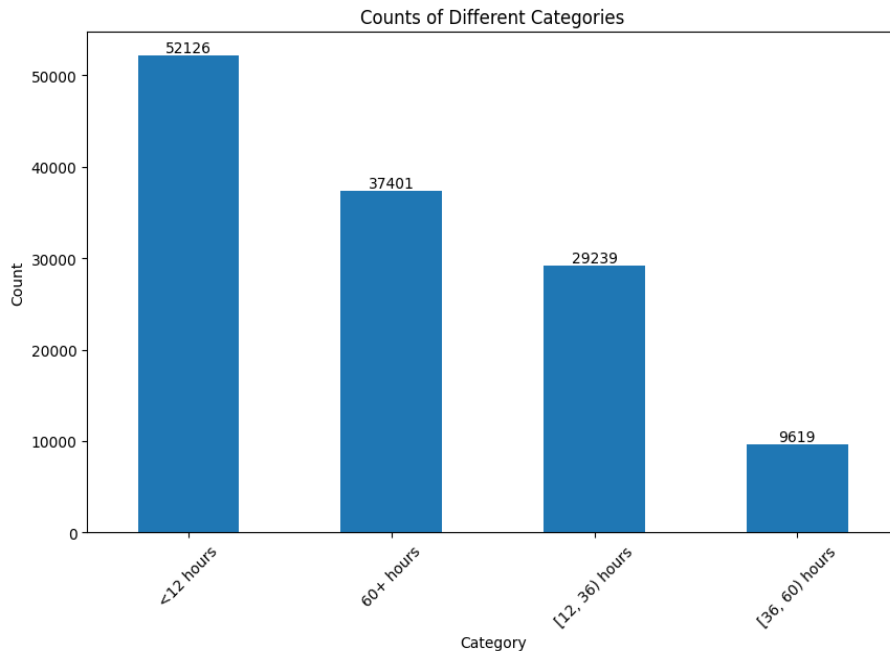
# Post – OP Features

---



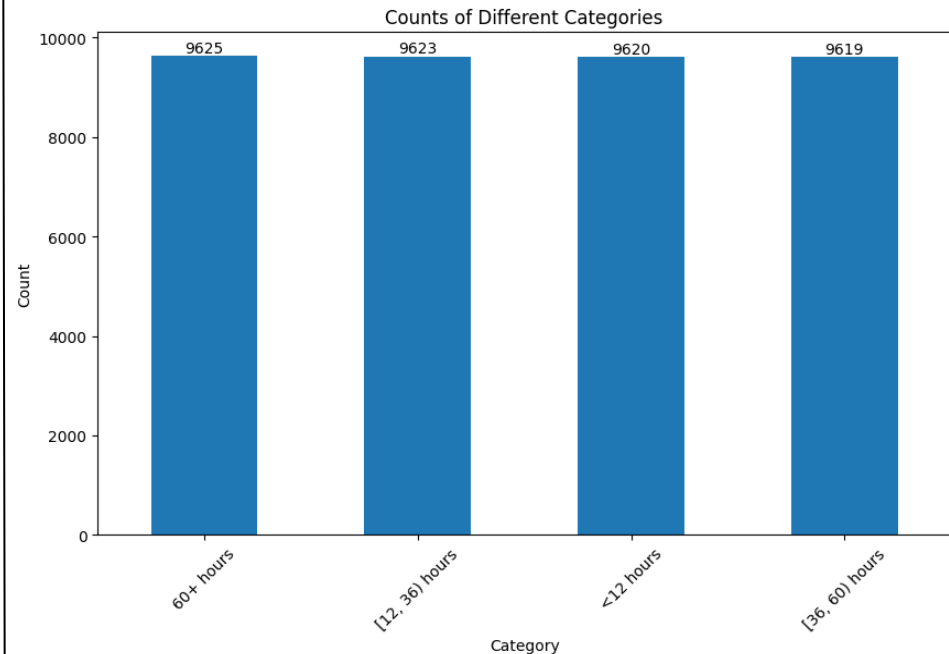
## Imbalanced Dataset

- The total number of rows after performing the cleaning process: **128385**
- Number of features selected: **24**
- Class [36,60) is **7.5%** of the total number of rows



## Balanced Dataset

- Subset the Imbalanced Dataset based on the **Target Class with the least occurrence.**
- The total number of rows after performing the cleaning process: **38487**
- Number of features selected: **24**
- Class [36,60) is **25%** of the total number of rows



# Final Selected Features PostOP:

## Numerical Features

- Age at Surgery
- BMI
- Scheduled Room Duration
- Intraop Minutes
- Total Scheduled Panel Length
- Total Panel Default Length

## Target Feature

- *LOS 4 Groups*

## ***One Hot Encoding:***

The process of converting categorical features into binary numerical columns.

One hot encoded categorical features and final dataset shape is *(38487, 6158)*.

## Categorical Features

- Sex
- Ethnicity
- Primary Race
- Surgery Diagnosis Name
- Location
- Primary Service
- Patient Class
- Anesthesia Type
- ASA Status
- Primary Surgeon ID
- Primary Procedure Name
- Robotic Case?
- Procedure Name
- Scheduled?
- Performed?
- Procedure Panel
- Panel Primary Surgeon ID

# First Iteration after Data Cleaning:

RFC	XGB																																								
<ul style="list-style-type: none"><li>Mean accuracy with Cross Validation of 5 sets was <b>93.43%</b>.</li><li>Strong precision and recall for classes 0 and 1.</li></ul>	<ul style="list-style-type: none"><li>Mean accuracy with Cross Validation of 5 sets was <b>82.46%</b>.</li><li>Lower Recall for Class 3</li></ul>																																								
<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th></tr><tr><td>0</td><td>0.96</td><td>0.94</td><td>0.95</td></tr><tr><td>1</td><td>0.95</td><td>0.98</td><td>0.96</td></tr><tr><td>2</td><td>0.92</td><td>0.91</td><td>0.92</td></tr><tr><td>3</td><td>0.90</td><td>0.91</td><td>0.91</td></tr></table>		precision	recall	f1-score	0	0.96	0.94	0.95	1	0.95	0.98	0.96	2	0.92	0.91	0.92	3	0.90	0.91	0.91	<table><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th></tr><tr><td>0</td><td>0.87</td><td>0.85</td><td>0.86</td></tr><tr><td>1</td><td>0.88</td><td>0.94</td><td>0.91</td></tr><tr><td>2</td><td>0.81</td><td>0.76</td><td>0.78</td></tr><tr><td>3</td><td>0.76</td><td>0.78</td><td>0.77</td></tr></table>		precision	recall	f1-score	0	0.87	0.85	0.86	1	0.88	0.94	0.91	2	0.81	0.76	0.78	3	0.76	0.78	0.77
	precision	recall	f1-score																																						
0	0.96	0.94	0.95																																						
1	0.95	0.98	0.96																																						
2	0.92	0.91	0.92																																						
3	0.90	0.91	0.91																																						
	precision	recall	f1-score																																						
0	0.87	0.85	0.86																																						
1	0.88	0.94	0.91																																						
2	0.81	0.76	0.78																																						
3	0.76	0.78	0.77																																						

Class	Label
0	60+ hours
1	<12 hours
2	[12, 36) hours
3	[36, 60) hours

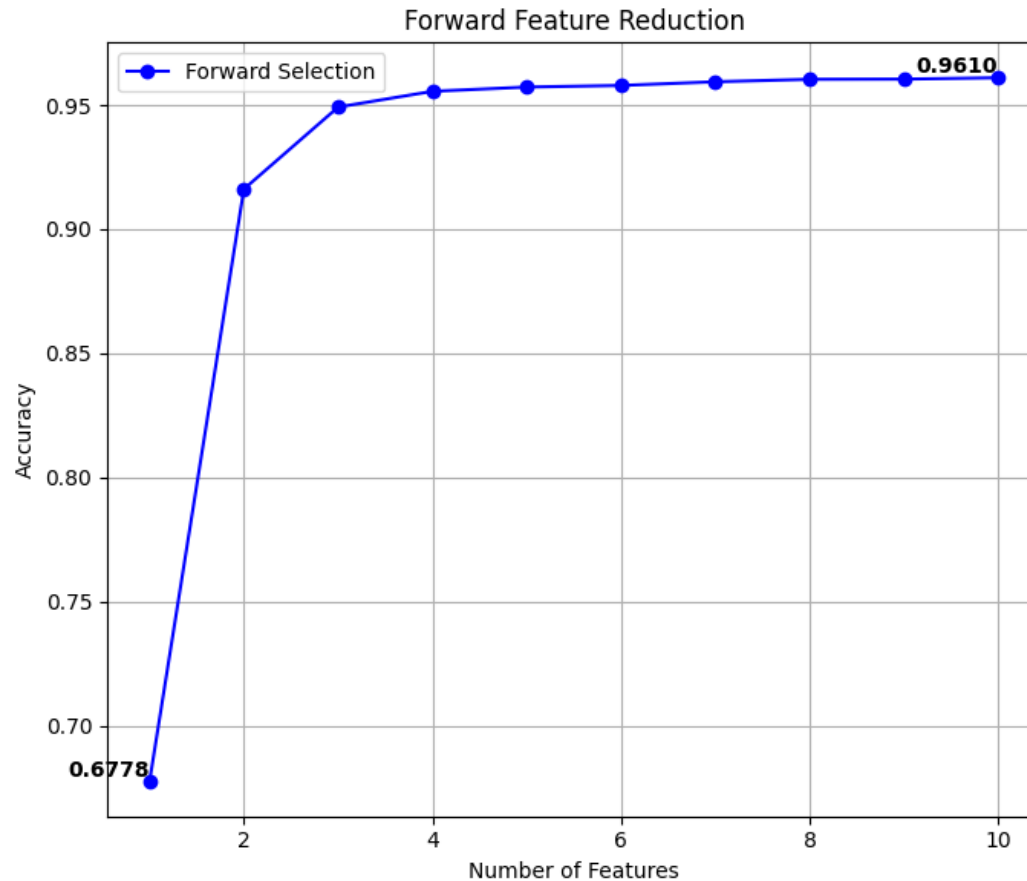
# Hyperparameter Tuning:

- Bayesian optimization for hyperparameter tuning.
- It builds a probabilistic model of the objective function, which estimates how changes in hyperparameters affect performance.
- It balances exploration (searching new regions) and exploitation (focusing on known good regions) to efficiently find optimal hyperparameter values.
- Better tuning compared to Grid Search and Random Search. Converges faster.

RFC Base	RFC Tuned	XGB Base	XGB Tuned
<b>Default Parameters:</b> <ul style="list-style-type: none"> <li>• N_estimators:100</li> <li>• min_samples_split: 2</li> <li>• min_samples_leaf: 1</li> </ul> <b>Range:</b> <ul style="list-style-type: none"> <li>• N_estimators: (50, 200),</li> <li>• min_samples_split: (2, 20),</li> <li>• min_samples_leaf: (1, 10)</li> </ul>	<b>Parameters Tuned:</b> <ul style="list-style-type: none"> <li>• <i>n_estimators,</i></li> <li>• <i>max_depth,</i></li> <li>• <i>min_samples_split,</i></li> <li>• <i>min_samples_leaf.</i></li> </ul> <b>Best Parameters found:</b> <ul style="list-style-type: none"> <li>• Max_depth: 50</li> <li>• N_estimators:200</li> <li>• min_samples_split: 2</li> <li>• min_samples_leaf: 1</li> </ul> <b>Runtime:</b> 3 hours	<b>Default Parameters:</b> <ul style="list-style-type: none"> <li>• Learning Rate: 0.3</li> <li>• Max_depth: 6</li> <li>• N_estimators:100</li> </ul> <b>Range:</b> <ul style="list-style-type: none"> <li>• N_estimators: (50, 125),</li> <li>• Max_depth: (4,7)</li> <li>• learning_rate: Real (0.05, 0.2, prior='log-uniform')</li> </ul>	<b>Parameters Tuned:</b> <ul style="list-style-type: none"> <li>• <i>Learning_rate,</i></li> <li>• <i>max_depth,</i></li> <li>• <i>n_estimators.</i></li> </ul> <b>Best Parameters found:</b> <ul style="list-style-type: none"> <li>• Learning Rate: 0.19</li> <li>• Max_depth: 7</li> <li>• N_estimators:102</li> </ul> <b>Runtime:</b> 4 hours

# Feature Importance PostOP – RFC Classifier

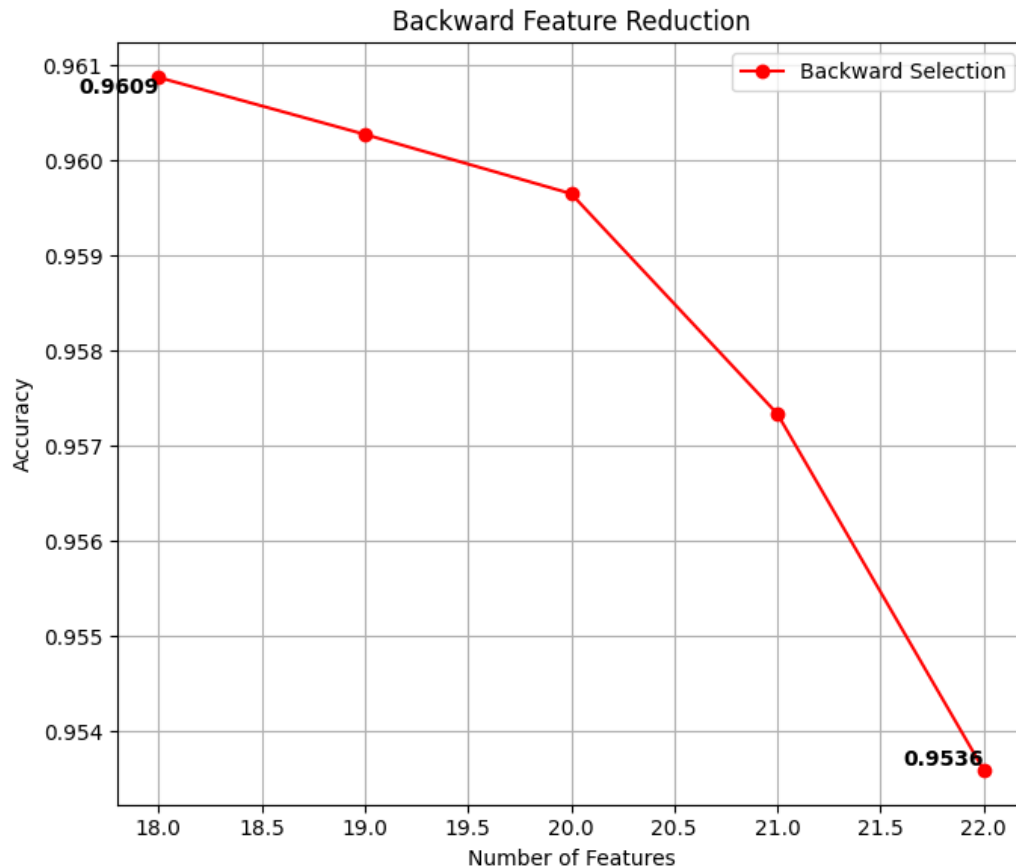
The *same set of Features* after data cleaning were used for Feature Reduction.



Features	Feature Name	Mean Accuracy
1	Primary Procedure Name	0.677
2	BMI	0.916
3	Intraop Minutes	0.949
4	Primary Surgeon ID	0.955
5	Total Panel Default Length	0.957
6	Patient Class	0.958
7	Sex	0.959
8	Primary Service	0.9603
9	Age at Surgery	0.9604
10	Primary Race	0.9609

- The total number features reduced from **23** to **10** in forward feature reduction.

# Feature Importance PostOP – RFC Classifier

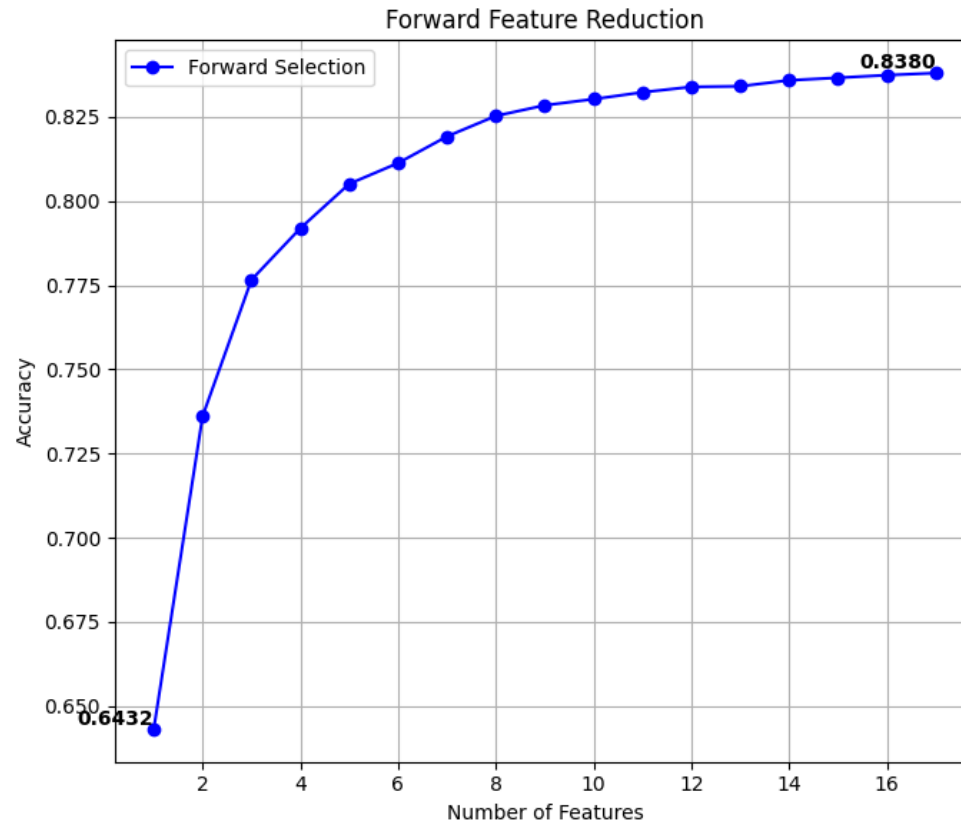


Removed Features	Feature Name	Mean Accuracy
1	Procedure Name	0.954
2	Scheduled?	0.957
3	Procedure Panel	0.959
4	Performed?	0.960
5	Ethnicity	0.961

## Important Features:

- 1. *Primary Procedure Name,*
  - 2. *BMI,*
  - 3. *Intraop Minutes,*
  - 4. *Primary Surgeon ID,*
  - 5. *Patient Class,*
  - 6. *Total Panel Default Length,*
  - 7. *Sex,*
  - 8. *Primary Service.*
- The total number features reduced from **23** to **18** in backward feature reduction.

# Feature Importance PostOP – XGB Classifier

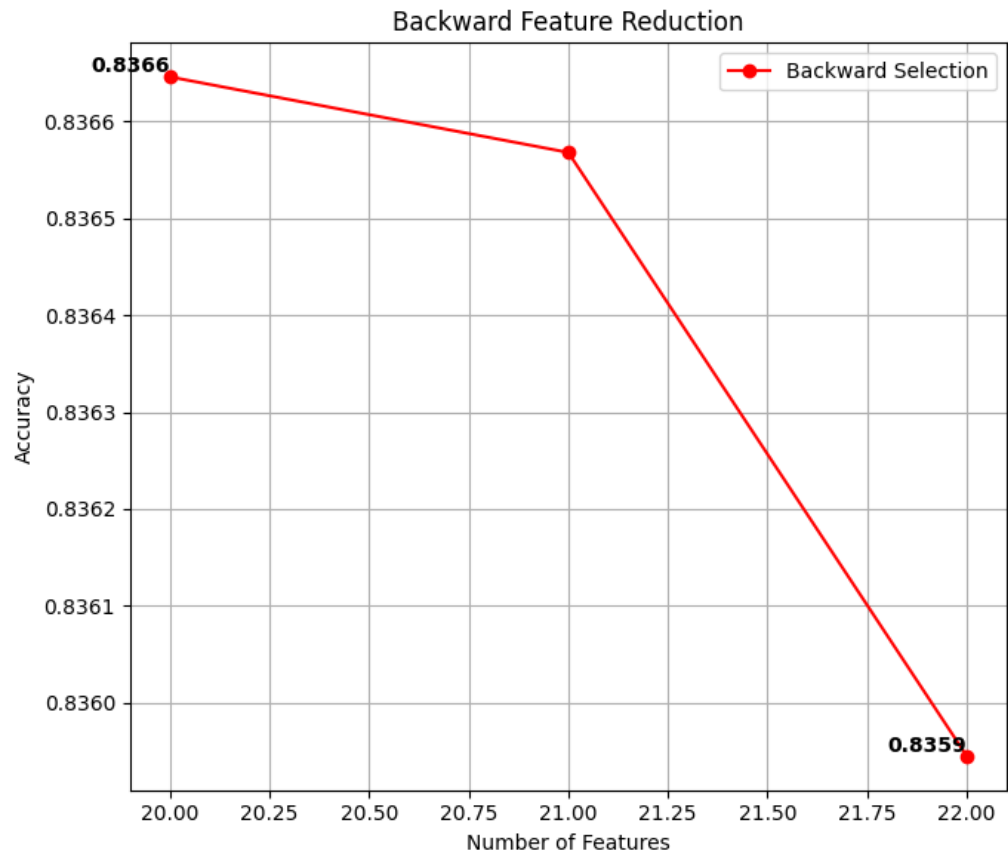


Features	Feature Name	Mean Accuracy
1	Primary Procedure Name	0.6431
2	Intraop Minutes	0.736
3	Primary Surgeon ID	0.776
4	Patient Class	0.791
5	Total Panel Default Length	0.805
6	Primary Service	0.811
7	Age at Surgery	0.819
8	BMI	0.825
9	Total Scheduled Panel Length	0.828
10	Robotic Case?	0.830
11	ASA Status	0.832
12	Primary Race	0.833
13	Surgery Diagnosis Name	0.834
14	Anesthesia Type	0.835
15	Sex	0.836
16	Scheduled Rood Duration	0.837
17	Procedure Panel	0.838

- The total number features reduced from **23** to **17** in forward feature reduction.



# Feature Importance PostOP – XGB Classifier



Removed Features	Feature Name	Mean Accuracy
1	Procedure Name	0.836
2	Procedure Panel	0.837
3	Anesthesia Type	0.837

## Important Features:

1. *Primary Procedure Name,*
2. *Intraop Minutes,*
3. *Primary Surgeon ID,*
4. *Patient Class,*
5. *Total Panel Default Length,*
6. *Primary Service,*
7. *Age,*
8. *BMI,*
9. *Total Scheduled Panel Length,*
10. *Robotic Case?*

- The total number features reduced from **23** to **20** in forward feature reduction.

# Summary of PostOp Experiments:

	Model		Accuracy	Recall	F1 Score
Post OP	RFC	<i>Base</i>	93.43%	93%	93%
		<i>Tuned</i>	87.63%	88%	88%
	XGB	<i>Base</i>	82.46%	83%	83%
		<i>Tuned</i>	83.60%	84%	84%
	RFC with Selected Features (Forward Selection)	<b>Base</b>	<b>96.18%</b>	<b>96%</b>	<b>96%</b>
		<i>Tuned</i>	95.24%	95%	95%
	XGB with Selected Features (Forward Selection)	<i>Base</i>	82.84%	84%	84%
		<i>Tuned</i>	83.09%	84%	84%

- Initial features: **23**
- No. of features that got selected for RFC in Post Op are **10**.
- No. of features that got selected for XGB in Post Op are **17**.

**Recommendation:** RFC with 10 features after forward feature selection to predict the LOS target features has 96.18% accuracy as compared to all the other approaches.

---

# Pre – OP Features

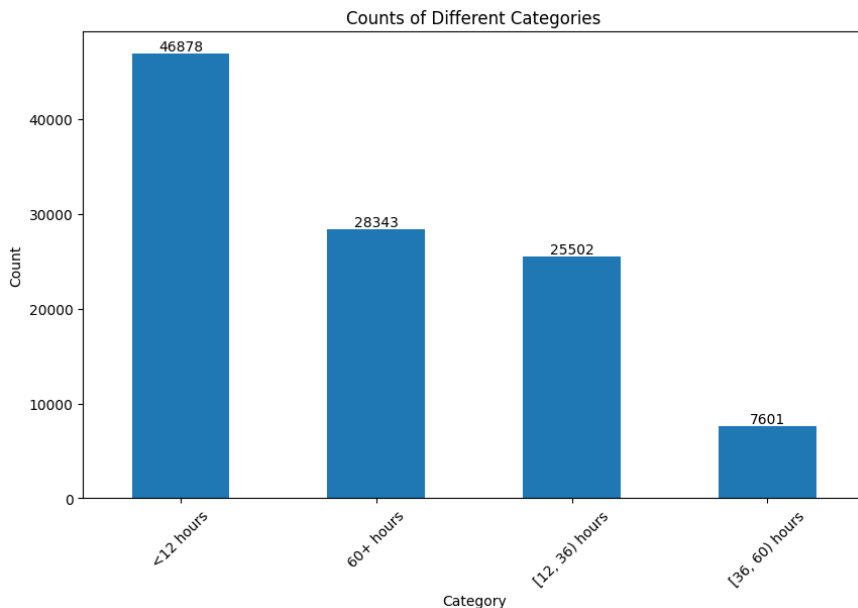
---

# Data Cleaning – Feature Selection

Features dropped for Pre – OP - 'Intraop Minutes', 'Scheduled?', 'Location', 'Patient Class'

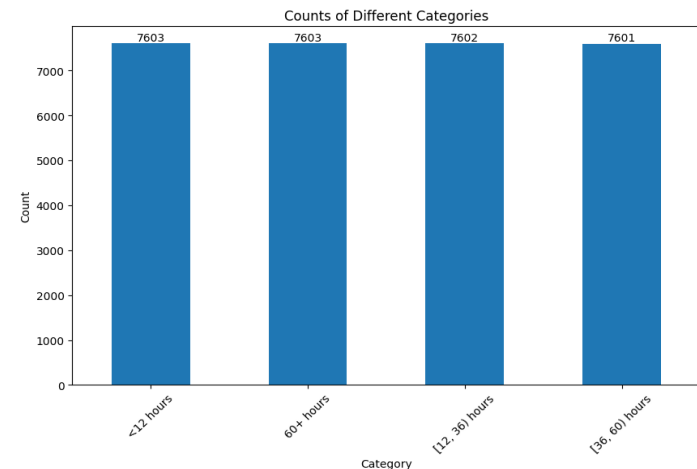
## Imbalanced Dataset

- The total number of rows after performing the cleaning process: **108324**
- Number of features selected: **19**
- Class [36,60) is **7%** of the total number of rows



## Balanced Dataset

- Subset the Imbalanced Dataset based on the **Target Class with the least occurrence.**
- The total number of rows after performing the cleaning process: **30409**
- Number of features selected: **19**
- Class [36,60) is **25%** of the total number of rows



# Final Selected Features PreOP:

## Numerical Features

- Age at Surgery
- BMI
- Scheduled Room Duration
- Total Scheduled Panel Length
- Total Panel Default Length

## Target Feature

- *LOS 4 Groups*

## Categorical Features

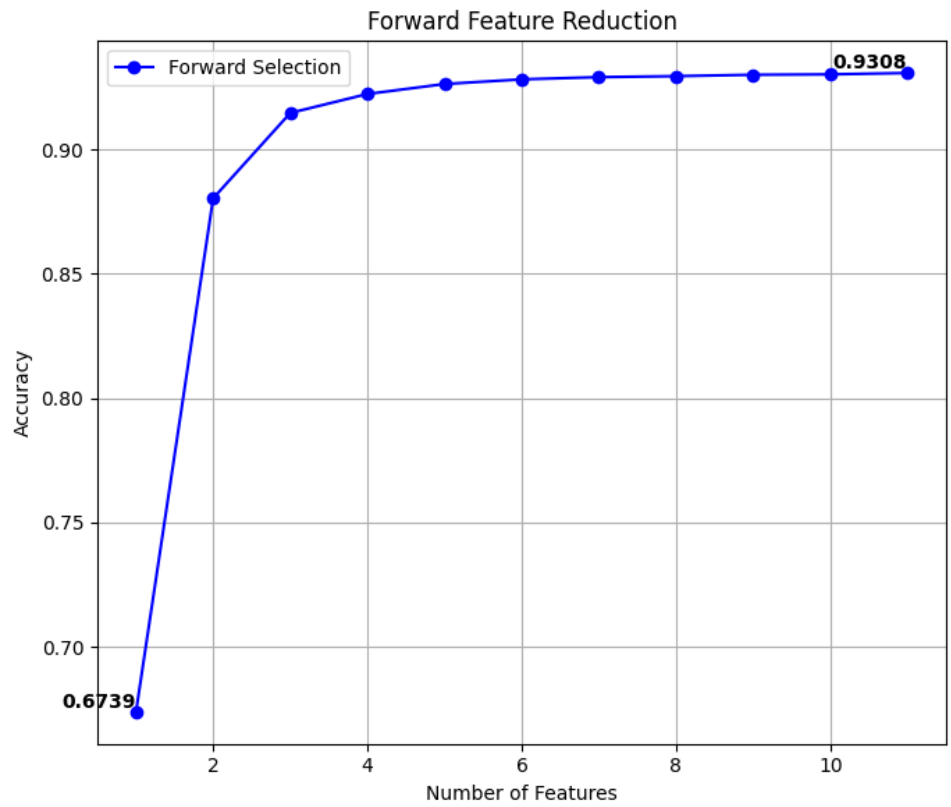
- Sex
- Ethnicity
- Primary Race
- Surgery Diagnosis Name
- Primary Service
- Anesthesia Type
- ASA Status
- Primary Surgeon ID
- Primary Procedure Name
- Robotic Case?
- Procedure Name
- Performed?
- Procedure Panel
- Panel Primary Surgeon ID

## ***One Hot Encoding:***

The process of converting categorical features into binary numerical columns.

One hot encoded categorical features and final dataset shape is (30409, 5831).

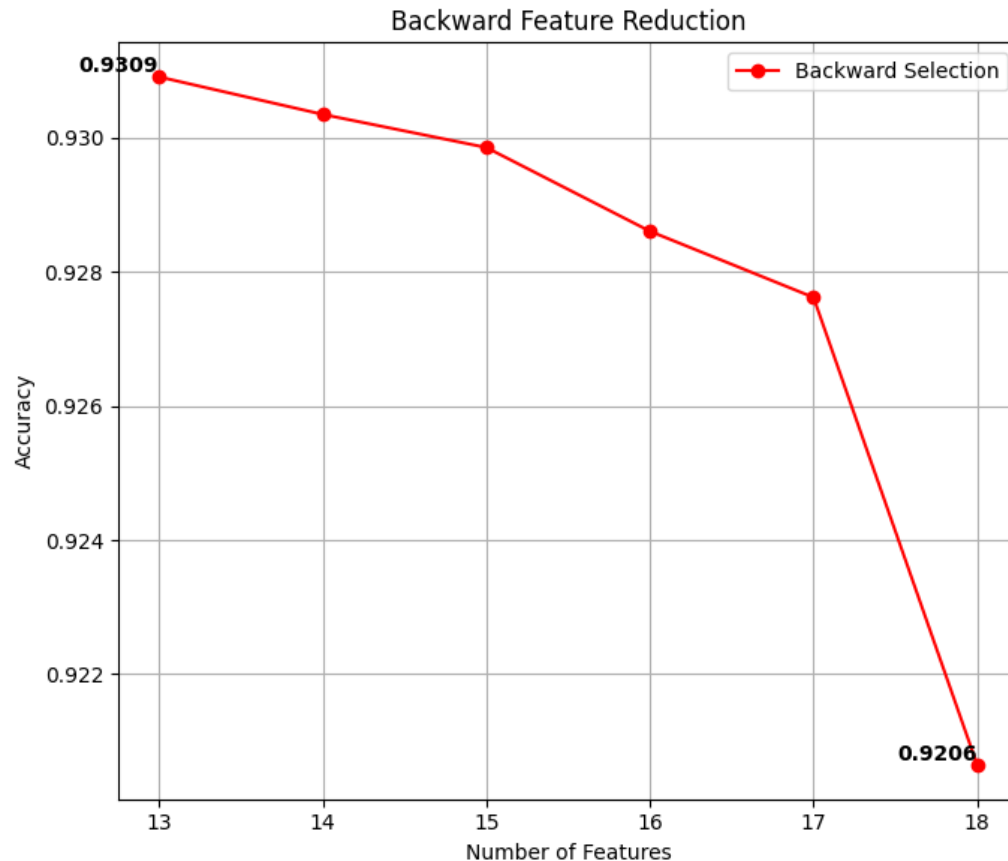
# Feature Importance PreOP – RFC Classifier



Features	Feature Name	Mean Accuracy
1	Primary Procedure Name	0.674
2	BMI	0.880
3	Scheduled Room Duration	0.915
4	Primary Surgeon ID	0.922
5	Surgery Diagnosis Name	0.926
6	Age at Surgery	0.928
7	Total Panel Default Length	0.9291
8	Primary Service	0.9295
9	ASA Status	0.9300
10	Total Scheduled Panel Length	0.9302
11	Ethnicity	0.9308

- The total number features reduced from **19** to **11** in forward feature reduction.

# Feature Importance PreOP – RFC Classifier



Removed Features	Feature Name	Mean Accuracy
1	Procedure Name	0.9206
2	Panel Primary ID	0.9276
3	Performed?	0.9286
4	Anesthesia Type	0.9299
5	Primary Race	0.9303
6	Procedure Panel	0.9309

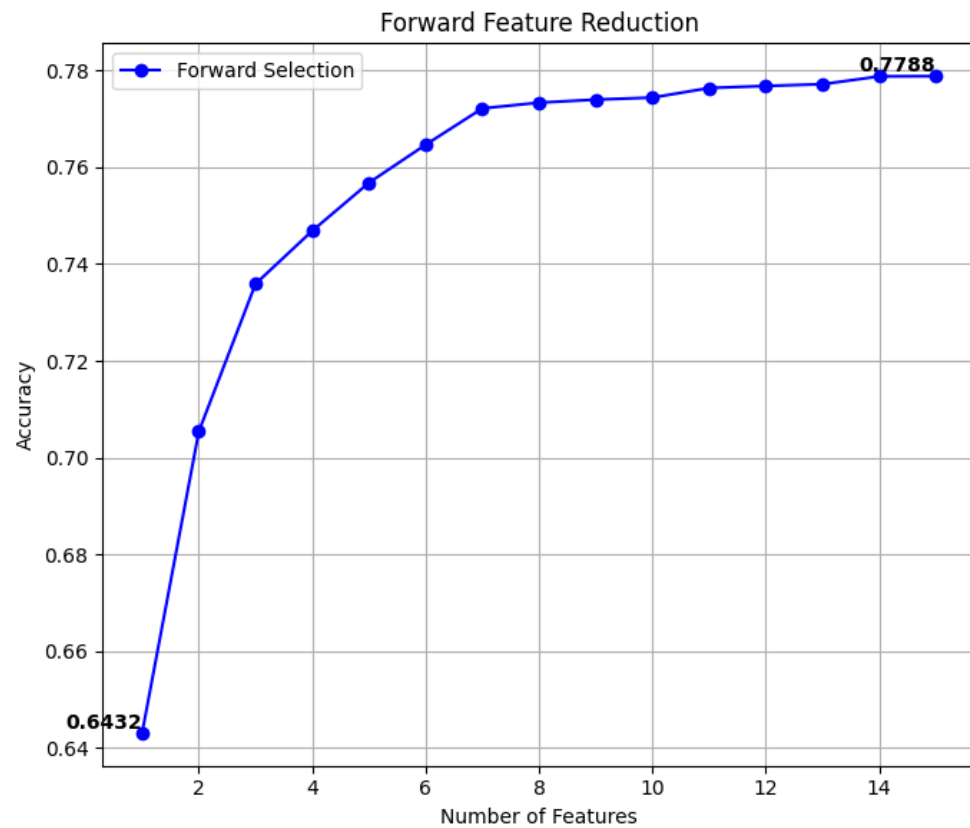
## Important Features:

1. *Primary Procedure Name,*
2. *BMI,*
3. *Scheduled Room Duration,*
4. *Primary Surgeon ID,*
5. *Surgery Diagnosis Name,*
6. *Age,*
7. *Total Panel Default Length.*

- The total number features reduced from **19** to **13** in backward feature reduction.



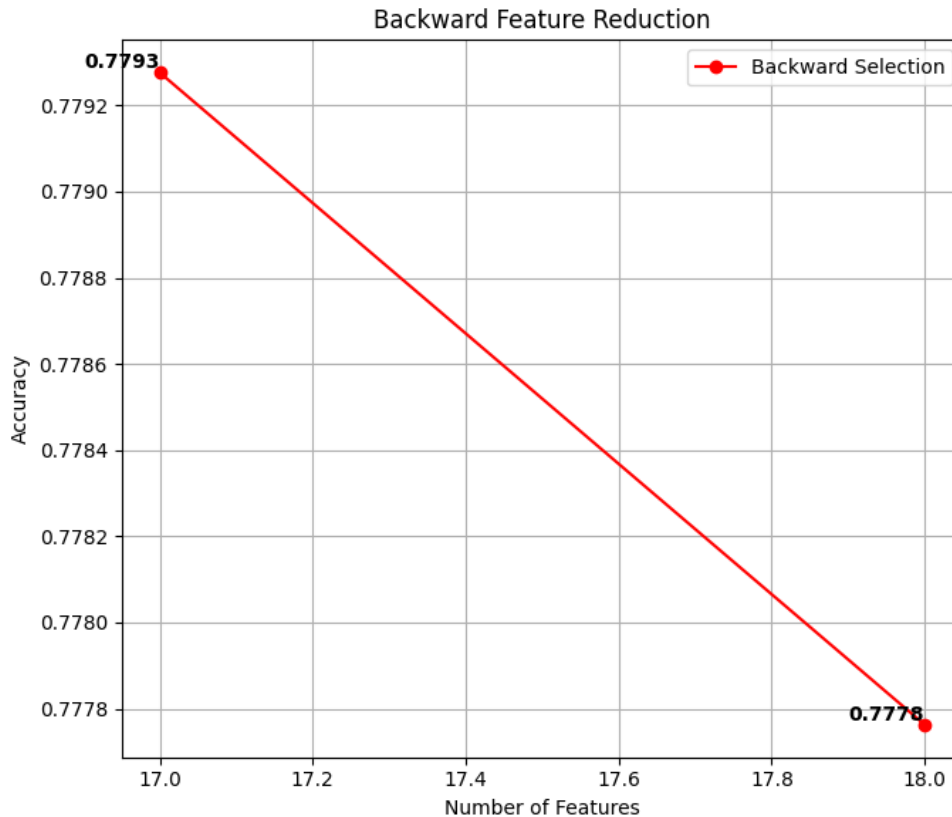
# Feature Importance PreOP – XGB Classifier



Features	Feature Name	Mean Accuracy
1	Primary Procedure Name	0.6431
2	Total Panel Default Length	0.705
3	Primary Surgeon ID	0.735
4	Scheduled Rood Duration	0.746
5	Primary Service	0.756
6	BMI	0.764
7	Age at Surgery	0.772
8	Primary Race	0.773
9	Sex	0.7739
10	Surgery Diagnosis Name	0.774
11	ASA Status	0.7763
12	Total Scheduled Panel Length	0.7767
13	Robotic Case?	0.777
14	Anesthesia Type	0.778
15	Procedure Panel	0.779

- The total number features reduced from **19** to **15** in forward feature reduction.

# Feature Importance PreOP – XGB Classifier



Removed Features	Feature Name	Mean Accuracy
1	Procedure Name	0.777
2	Panel Primary ID	0.779

## Important Features:

1. *Primary Procedure Name,*
2. *Total Panel Default Length,*
3. *Primary Surgeon ID,*
4. *Scheduled Room Duration,*
5. *Primary Service*
6. *BMI,*
7. *Age,*
8. *Primary Race,*
9. *Sex,*
10. *Surgery Diagnosis Name.*

- The total number features reduced from **19** to **17** in forward feature reduction.

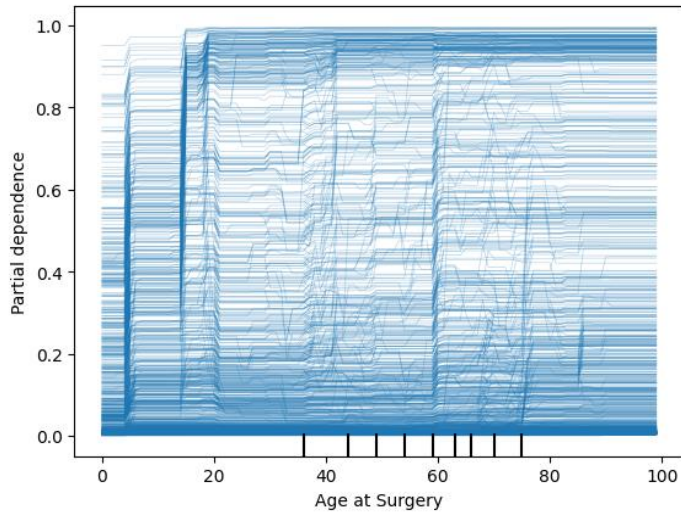
# Summary of PreOP Experiments:

	Model		Accuracy	Recall	F1 Score
Pre - OP	RFC	<i>Base</i>	89.39%	89%	89%
		<i>Tuned</i>	89.21%	89%	89%
	XGB	<i>Base</i>	78.49%	78%	78%
		<i>Tuned</i>	78.64%	79%	79%
	RFC with Selected Features (Forward Selection)	<b>Base</b>	<b>93.29%</b>	<b>93%</b>	<b>93%</b>
		<i>Tuned</i>	89.21%	89%	89%
	XGB with Selected Features (Forward Selection)	<i>Base</i>	77.72%	78%	78%
		<i>Tuned</i>	77.90%	78%	78%

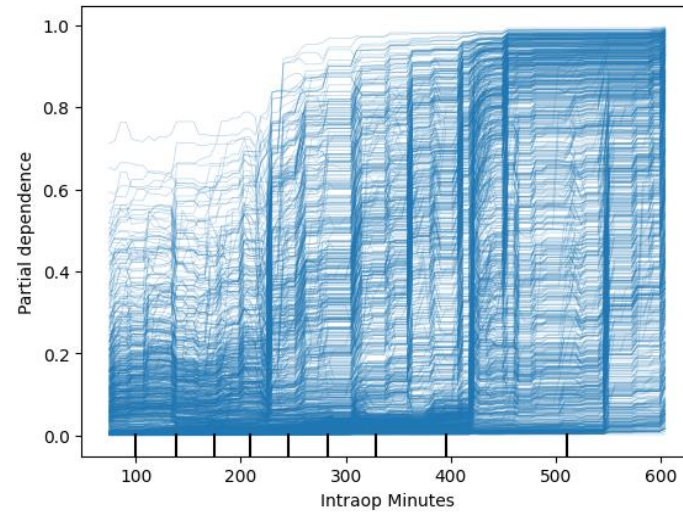
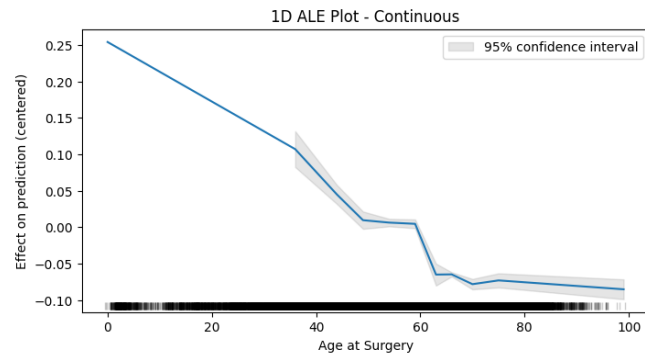
- Initial features: **19**
- No. of features that got selected for RFC in Pre - Op are **11**.
- No. of features that got selected for XGB in Pre - Op are **15**.

**Recommendation:** RFC with 11 features after forward feature selection to predict the LOS target features has 93.29% accuracy as compared to all the other approaches.

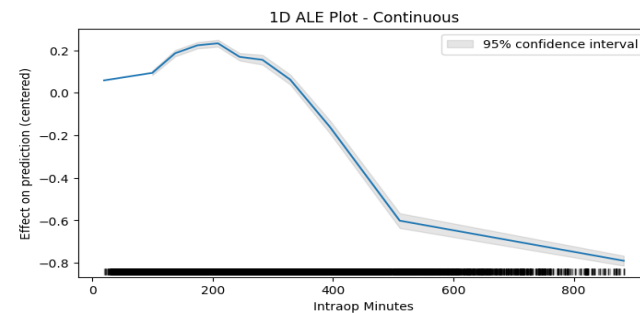
# Sensitivity Analysis – Numerical Features



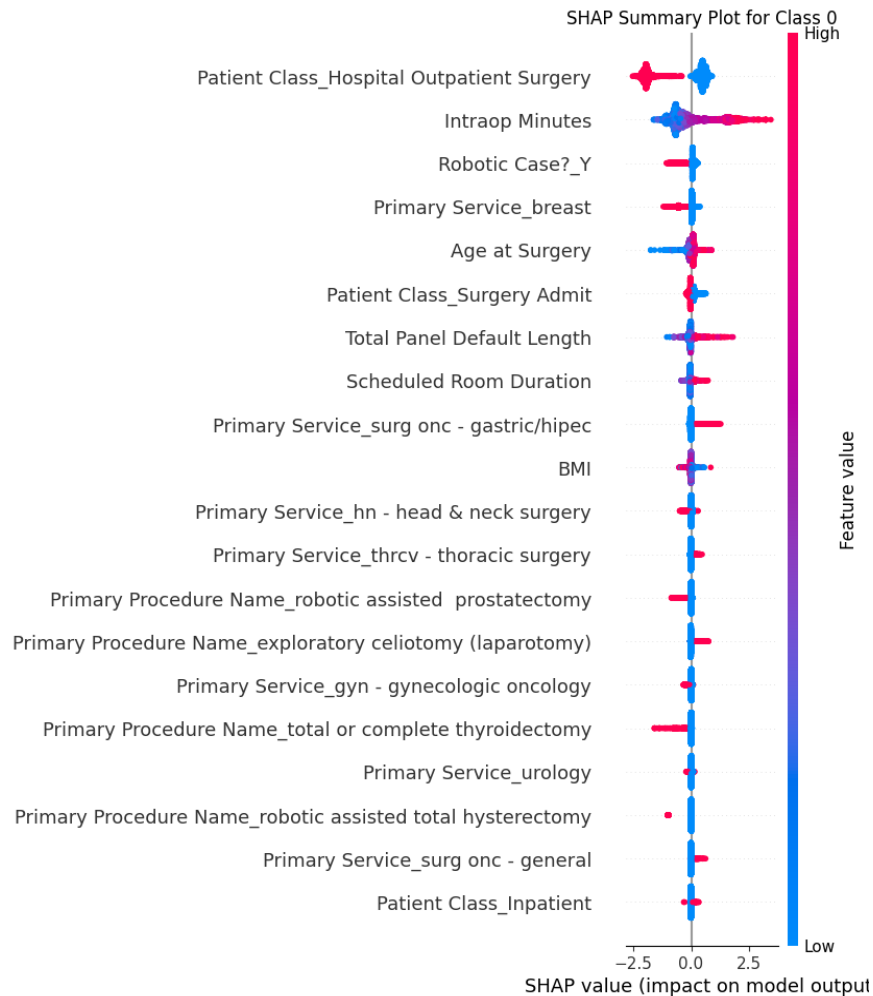
**Age at Surgery** affect the prediction for LOS group who have an age above 20 very minimally



**Intraop Minutes** affect the prediction for LOS group when the Intraop Minutes is less than 200 minutes.



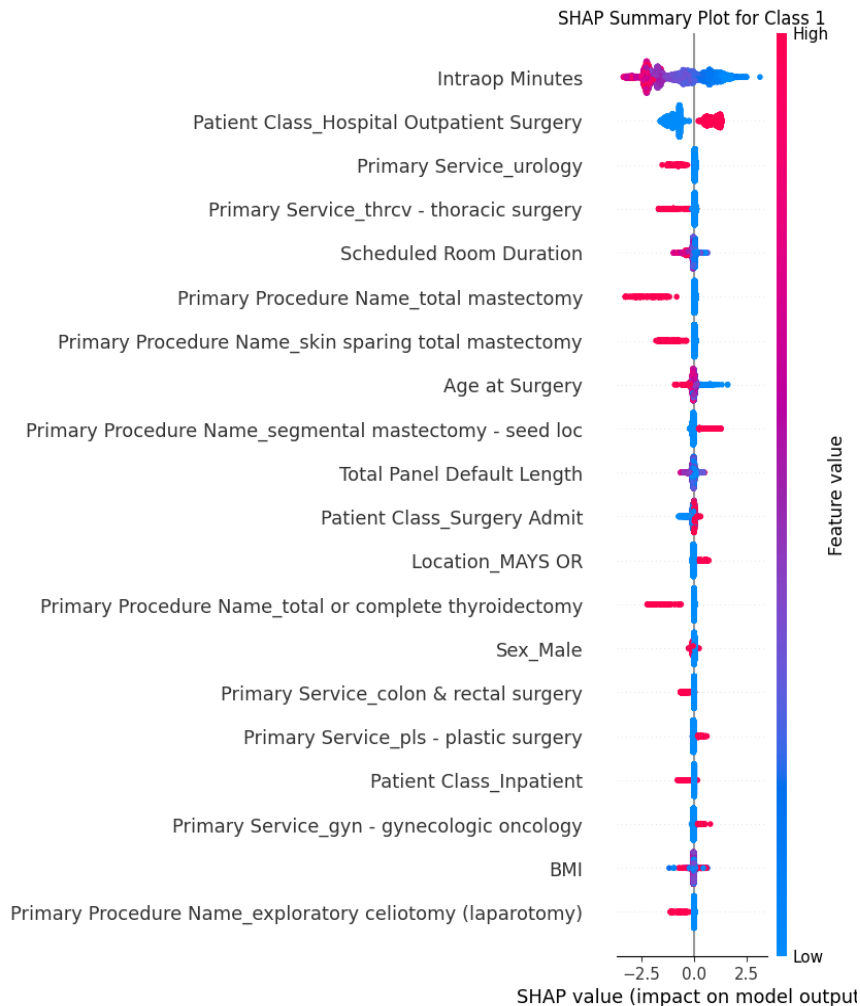
# Sensitivity Analysis – SHAP Analysis after Encoding for XGB Classifier on Post- OP Features



## Important Features after Prediction for LOS 60+ hours:

1. Patient Class
2. IntraOP Minutes
3. Robotic Case
4. Primary Service
5. Age at Surgery
6. Patient Class
7. Total Panel Default Length
8. Scheduled Room Duration
9. BMI
10. Primary Procedure Name

# Sensitivity Analysis – SHAP Analysis after Encoding for XGB Classifier on Post- OP Features

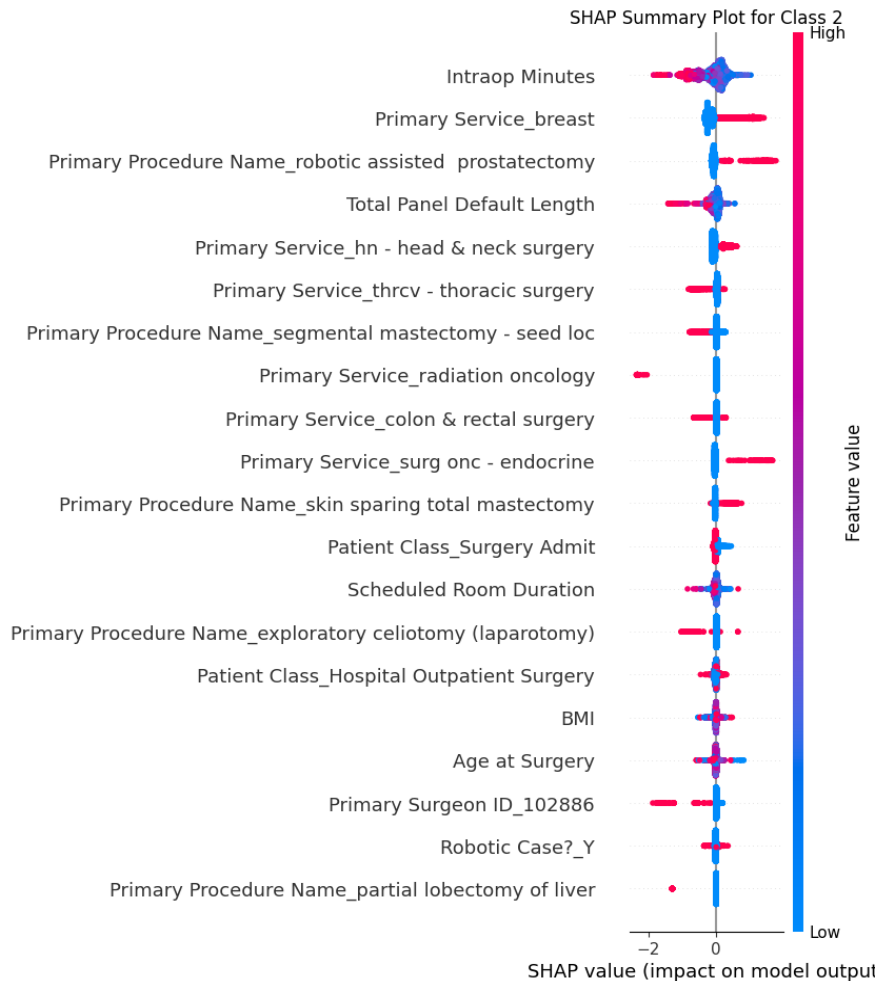


## Important Features after Prediction for LOS <12 hours:

1. Patient Class
2. IntraOP Minutes
3. Primary Service
4. Age at Surgery
5. Patient Class
6. Total Panel Default Length
7. Scheduled Room Duration
8. BMI
9. Primary Procedure Name
10. Location
11. Sex

**Missing: Robotic Case**

# Sensitivity Analysis – SHAP Analysis after Encoding for XGB Classifier on Post- OP Features



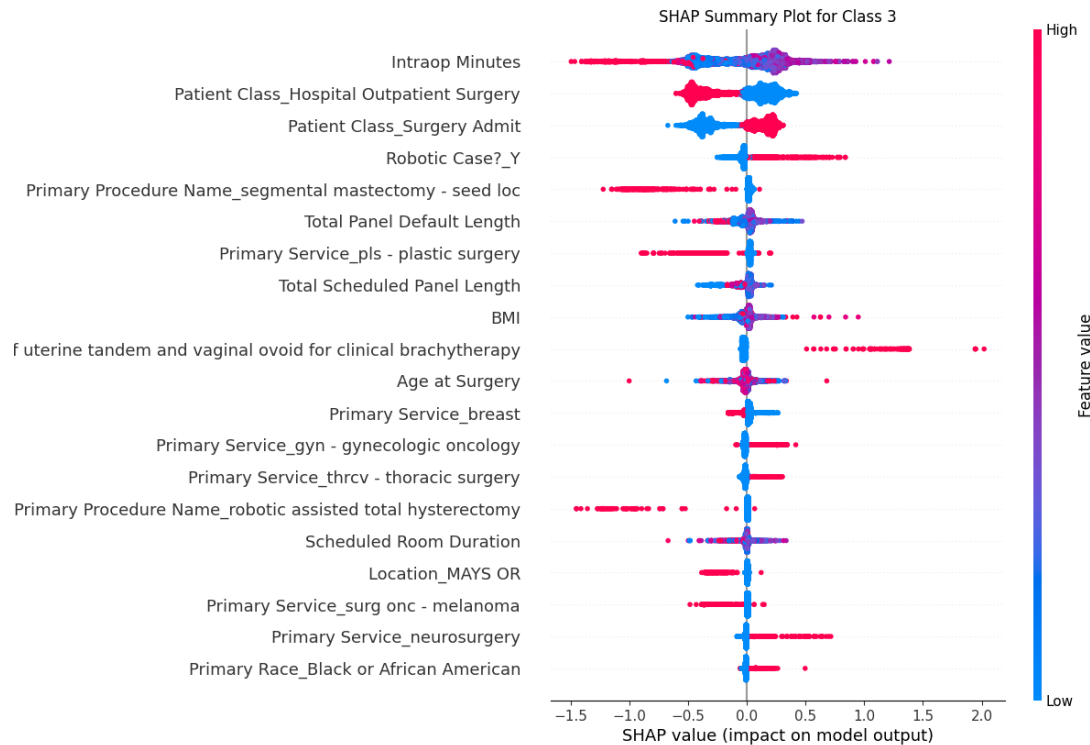
## Important Features after Prediction for LOS [12,36) hours:

1. Patient Class
2. IntraOP Minutes
3. Primary Service
4. Age at Surgery
5. Patient Class
6. Total Panel Default Length
7. Scheduled Room Duration
8. BMI
9. Primary Procedure Name
10. Robotic Case
11. Primary Surgeon ID

**Missing: Location, Sex**



# Sensitivity Analysis – SHAP Analysis after Encoding for XGB Classifier on Post- OP Features



## Important Features after Prediction for LOS [36,60) hours:

1. Patient Class
2. IntraOP Minutes
3. Primary Service
4. Age at Surgery
5. Patient Class
6. Total Panel Default Length
7. Scheduled Room Duration
8. BMI
9. Primary Procedure Name
10. Location
11. Primary Race

**Missing: Age at Surgery, Sex**

## Conclusion

1. Random Forest Classifier with 7-11 essential features with Post - OP as well as Pre – OP provides the best prediction for LOS groups.
2. However, if experiments needs to be performed only on Pre – OP features, Random Forest Classifier has the most accurate prediction for LOS groups.
3. Prediction is affected by the Numerical Features – Age at Surgery and Intra OP minutes significantly.
4. Tests such as Random State did not change the accuracy of prediction indicating no overfitting and bias within the model.

## Future Steps

1. Textual Columns instead of using one-hot encoding could be vectorized using transformers such as Clinical Bert preventing the explosion of features.
2. Categorical Columns such as Procedure Panel and Primary Service which impacts heavily as seen from SHAP needs to be explored further.
3. Accurately predict the actual number of minutes (regression) instead of classification.

**Thank You!**



THE UNIVERSITY OF TEXAS AT AUSTIN

# **Operations Research & Industrial Engineering**

---

# Appendix

---

# Data Cleaning of Numerical Features:

## Outlier Analysis:

- Objective is to identify a threshold for the IQR multiplier to balance outlier handling and data retention.
- Instead of using the standard 1.5 times IQR, we tested multipliers at 0.25 intervals from 1 to 3.
- 2.5 times IQR effectively identified outliers while minimizing data loss.

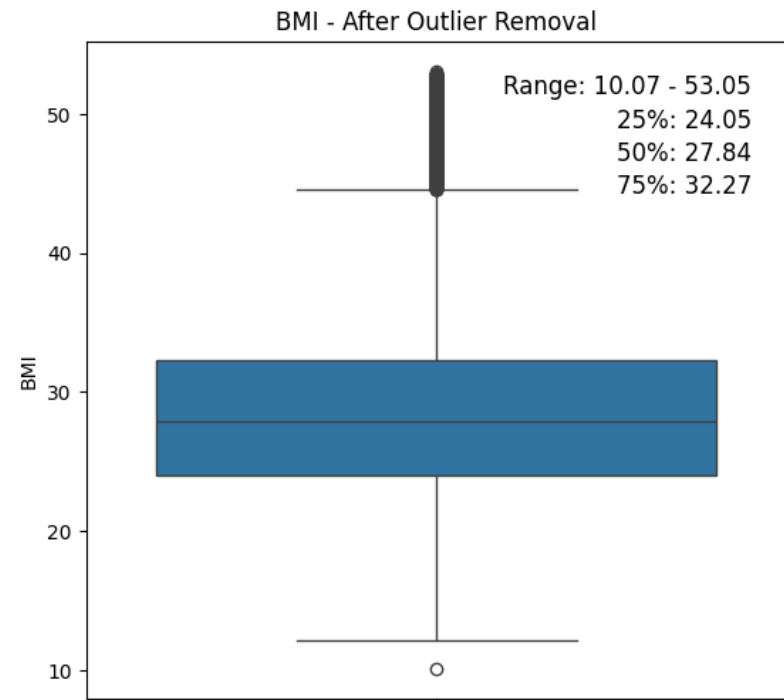
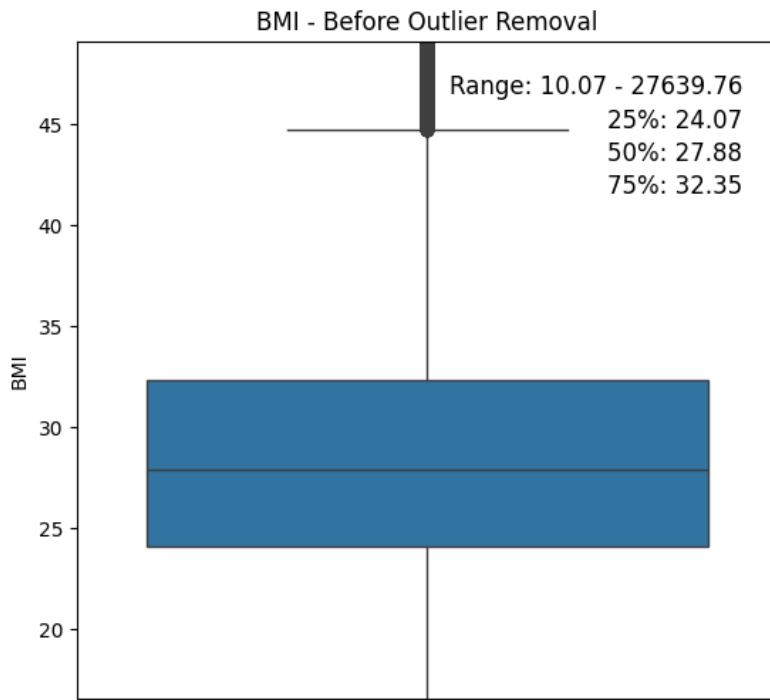


Fig 1. Box Plot of BMI

# Data Cleaning of Numerical Features:

Features	No. of Outliers	Missing Values after Outlier Analysis	Range of Values after Outlier Analysis
<i>BMI</i>	611	25024	10.07 to 52.77
<i>Scheduled Panel Length</i>	420	0	1 to 945 min
<i>Panel Default Length</i>	4300	0	0 to 965 min
<i>In Recovery to Discharge Time Minutes</i>	5026	5330	0 to 16614 min
<i>Intraop Minutes</i>	812	1863	13 to 886 min
<i>Scheduled Room Duration</i>	1352	1610	1 to 940 min

- A total of **12,521** rows were removed after outlier analysis and the range of these features were suitable to fill in missing values.

# Data Cleaning of Numerical Features:

## *Handling Missing Values:*

- *BMI*: Missing Values were replaced by calculating the median within 10 age buckets, ranging from 1 to 100, and then further grouped by sex.
- *Scheduled Room Duration*: Subset the data by Primary Service and filled missing values with the median value for each subset.
- *Intraop Minutes*: Subset the data by Panel Service and used the median value to fill in missing entries.
- *In Recovery to Discharge Time Minutes*: Used absolute values and filled in missing values based on Panel Service.

# Data Cleaning of Categorical Features:

## *Handling Missing Values:*

- *Ethnicity*: Relabeled 'Unknown Ethnicity' and 'Decline to Answer' as 'Unknown'.
- *Primary Race*: Replaced missing values, 'Declined to Answer,' and 'Unknown' categories with 'Unknown'.
- *Anesthesia Type*: Filled with 'General' (74.5% Prevalence)
- *ASA Status*: Filled with 'Severe Systemic Disease' (68.3% Prevalence).
- *Patient Class*: Missing values belonged to a single category and were filled with Surgery Admit.



# Data Cleaning of Categorical Features:

## *Handling Missing Values:*

- ***Surgery Diagnosis Name:*** The rows with missing values were grouped by the Primary Surgeon ID and were filled with commonly occurring Surgery Diagnosis Name for that Primary Surgeon ID.
- ***Primary Procedure Name:*** Missing Primary Procedure Names were grouped by the Surgery Diagnosis Name. The Most commonly occurring primary procedure name for that surgery Diagnosis Name was then used to fill the missing values.

Feature	Unique Items
Surgery Diagnosis Name	5928
Surgery Diagnosis Code	2498
Primary Procedure Name	1309
Primary Procedure CPT Code	430
Procedure Name	1804
Procedure CPT Code	694

## *Dropped Features:*

Surgery Diagnosis Code, Procedure CPT Code, Primary Procedure CPT Code, Rolled up columns, Estimated Case Duration, 'Height', 'Weight'.

# Data Inconsistency:

## Categorical Feature - Procedure Panel:

### Issue:

- Inconsistent numbering of Procedure Panels.

### Solution:

- Step 1: If Primary Service matches Panel Service, assign Procedure Panel as 1.
- Step 2: For every unique Panel Service, find the most commonly occurring Procedure Panel Number.
- Step 3: If the Procedure Panel Number is unique for that Panel Service, assign it else assign the next largest Panel Number.

Before Cleaning				After Cleaning			
ID	Primary Service	Procedure Panel	Panel Service	ID	Primary Service	Procedure Panel	Panel Service
359e43a1a3	UROLOGY	1	urology	359e43a1a3	UROLOGY	1	urology
359e43a1a3	UROLOGY	1	urology	359e43a1a3	UROLOGY	1	urology
359e43a1a3	UROLOGY	2	urology	359e43a1a3	UROLOGY	1	urology
359e43a1a3	UROLOGY	2	urology	359e43a1a3	UROLOGY	1	urology
359e43a1a3	UROLOGY	1	surg onc - gastric/hipec	359e43a1a3	UROLOGY	2	surg onc - gastric/hipec
359e43a1a3	UROLOGY	3	pls - plastic surgery	359e43a1a3	UROLOGY	3	pls - plastic surgery
359e43a1a3	UROLOGY	3	pls - plastic surgery	359e43a1a3	UROLOGY	3	pls - plastic surgery
359e43a1a3	UROLOGY	2	pls - plastic surgery	359e43a1a3	UROLOGY	3	pls - plastic surgery

Fig 2. Inconsistent Numbering of Procedure Panel for ID: 359e43a1a3

# Post Operative:

Classification Report:					
	precision	recall	f1-score	support	
0	0.94	0.97	0.95	7480	
1	0.94	0.98	0.96	10425	
2	0.93	0.87	0.90	5848	
3	0.93	0.73	0.82	1924	
accuracy			0.94	25677	
macro avg	0.93	0.89	0.91	25677	
weighted avg	0.94	0.94	0.93	25677	
Accuracy Score: 0.9361685555166102					

Fig 1. RFC with unbalanced dataset

Classification Report:					
	precision	recall	f1-score	support	
0	0.97	0.97	0.97	1925	
1	0.95	0.98	0.96	1924	
2	0.96	0.91	0.93	1925	
3	0.94	0.95	0.95	1924	
accuracy			0.95	7698	
macro avg	0.95	0.95	0.95	7698	
weighted avg	0.95	0.95	0.95	7698	
Accuracy Score: 0.9524551831644583					

Fig 3. Hyperparameter Tuned RFC (balanced dataset)

Classification Report:					
	precision	recall	f1-score	support	
0	0.97	0.97	0.97	1925	
1	0.95	0.98	0.96	1924	
2	0.96	0.91	0.93	1925	
3	0.94	0.95	0.95	1924	
accuracy			0.95	7698	
macro avg	0.95	0.95	0.95	7698	
weighted avg	0.95	0.95	0.95	7698	
Accuracy Score: 0.9524551831644583					

Fig 2. RFC with balanced dataset

Classification Report:					
	precision	recall	f1-score	support	
0	0.98	0.97	0.97	1925	
1	0.97	0.98	0.98	1924	
2	0.96	0.95	0.96	1925	
3	0.95	0.96	0.95	1924	
accuracy			0.96	7698	
macro avg	0.96	0.96	0.96	7698	
weighted avg	0.96	0.96	0.96	7698	
Accuracy Score: 0.9644063393089114					

Fig 4. Best RFC model for PostOP (Forward Selection)

```
Classification Report:
              precision    recall  f1-score   support

     0       0.84         0.94         0.89       7528
     1       0.90         0.97         0.93      10456
     2       0.84         0.75         0.79       5782
     3       0.71         0.37         0.49       1911

 accuracy          0.86         0.86         0.86      25677
 macro avg         0.83         0.76         0.78      25677
 weighted avg      0.86         0.86         0.86      25677

Accuracy Score: 0.8648596019784243
```

Fig 5. XGB with unbalanced dataset

```
Classification Report:
              precision    recall  f1-score   support

     0       0.87         0.86         0.86       1925
     1       0.88         0.94         0.91       1924
     2       0.83         0.77         0.80       1925
     3       0.77         0.79         0.78       1924

 accuracy          0.84         0.84         0.84       7698
 macro avg         0.84         0.84         0.84       7698
 weighted avg      0.84         0.84         0.84       7698

Accuracy Score: 0.8393089114055599
```

Fig 6. Hyperparameter Tuned XGB (balanced dataset)

```
Classification Report:
              precision    recall  f1-score   support

     0       0.88         0.86         0.87       1925
     1       0.88         0.94         0.91       1924
     2       0.82         0.76         0.79       1925
     3       0.76         0.79         0.78       1924

 accuracy          0.84         0.84         0.84       7698
 macro avg         0.84         0.84         0.84       7698
 weighted avg      0.84         0.84         0.84       7698

Accuracy Score: 0.8360613146271759
```

Fig 7. XGB with balanced dataset

# Pre Operative:

Classification Report:					
	precision	recall	f1-score	support	
0	0.89	0.93	0.91	5669	
1	0.92	0.97	0.94	9376	
2	0.91	0.84	0.87	5100	
3	0.88	0.61	0.72	1520	
accuracy			0.91	21665	
macro avg	0.90	0.84	0.86	21665	
weighted avg	0.91	0.91	0.90	21665	
Accuracy Score: 0.9065312716362797					

Fig 8. RFC with unbalanced dataset

Classification Report:					
	precision	recall	f1-score	support	
0	0.92	0.90	0.91	1521	
1	0.83	0.96	0.89	1521	
2	0.95	0.83	0.89	1520	
3	0.88	0.88	0.88	1520	
accuracy			0.89	6082	
macro avg	0.90	0.89	0.89	6082	
weighted avg	0.90	0.89	0.89	6082	
Accuracy Score: 0.8921407431765866					

Fig 10. Hyperparameter Tuned RFC (balanced dataset)

Classification Report:					
	precision	recall	f1-score	support	
0	0.96	0.93	0.94	1521	
1	0.92	0.97	0.94	1521	
2	0.93	0.91	0.92	1520	
3	0.92	0.92	0.92	1520	
accuracy			0.93	6082	
macro avg	0.93	0.93	0.93	6082	
weighted avg	0.93	0.93	0.93	6082	
Accuracy Score: 0.932916803682999					

Fig 9. RFC with balanced dataset

Classification Report:					
	precision	recall	f1-score	support	
0	0.96	0.93	0.94	1521	
1	0.92	0.97	0.94	1521	
2	0.93	0.91	0.92	1520	
3	0.92	0.92	0.92	1520	
accuracy			0.93	6082	
macro avg	0.93	0.93	0.93	6082	
weighted avg	0.93	0.93	0.93	6082	
Accuracy Score: 0.932916803682999					

Fig 11. Best RFC model for PreOP (Forward Selection)

Classification Report:				
	precision	recall	f1-score	support
0	0.79	0.89	0.84	5669
1	0.87	0.95	0.91	9376
2	0.84	0.72	0.78	5100
3	0.68	0.29	0.41	1520
accuracy			0.84	21665
macro avg	0.80	0.71	0.73	21665
weighted avg	0.83	0.84	0.82	21665
Accuracy Score: 0.8354488806831295				

Fig 12. XGB with unbalanced dataset

Classification Report:				
	precision	recall	f1-score	support
0	0.83	0.78	0.80	1558
1	0.79	0.91	0.85	1499
2	0.78	0.72	0.75	1488
3	0.74	0.74	0.74	1537
accuracy			0.79	6082
macro avg	0.79	0.79	0.79	6082
weighted avg	0.79	0.79	0.79	6082
Accuracy Score: 0.7864189411377837				

Fig 13. XGB with balanced dataset

Classification Report:				
	precision	recall	f1-score	support
0	0.82	0.78	0.80	1558
1	0.78	0.90	0.84	1499
2	0.79	0.71	0.74	1488
3	0.73	0.73	0.73	1537
accuracy			0.78	6082
macro avg	0.78	0.78	0.78	6082
weighted avg	0.78	0.78	0.78	6082
Accuracy Score: 0.7790200591910555				

Fig 14. Hyperparameter Tuned XGB (balanced dataset)