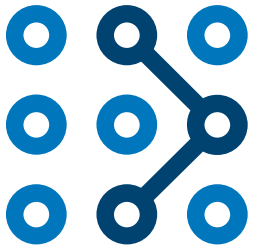


THE Data Science INSTITUTE

at Saint Peter's University

Interpretable Machine Learning for Gender Specific Prediction of Mortality

Student: Hulya Alpagu

Professor: Gulhan Bizel & Reshma Kar

Introduction

THE Data Science INSTITUTE at Saint Peter's University

- Numerous studies have shown that female tend to live longer than male, and this is believed to be due to physiological differences.
- This work aims to find the predictors of all-cause mortality by developing a machine learning model that is gender-specific and interpretable. To help personalized dissemination of urgent care.
- Factors like age, gender, surgical procedures, and disease progression are linked to patient mortality rates. Identifying personalized predictors of mortality can help manage it.

Problem Statement

THE Data Science INSTITUTE at Saint Peter's University

- **Q1** Can interpretable machine learning be used to
  - predict mortality accurately?
  - Explain the reason for mortality prediction?

- **Q2** Are there gender differences in the primary mortality predictors?

Data Source

THE Data Science INSTITUTE at Saint Peter's University

- The Texas Public Use Data File contains data on discharges from Texas hospitals.
- For this study, data from Texas hospital Inpatient Discharge General Use Data File containing 1,493,150 patient records of 710,385 male and 782,765 female were analyzed.
- As we review the patient records, it becomes evident that the data exhibits nearly equal distribution between male and female.

Input Variables

THE Data Science INSTITUTE at Saint Peter's University

Primary Variables

- **Admission type**
  - "Admission type" refers to the reason or method a person enters a hospital.
- **Race**
  - "Patient race" refers to a demographic characteristic related to an individual's racial or ethnic identity.
- **Admission Source**
  - "Admission source" is how a person enters at the hospital for treatment.
- **Length of Stay**
  - "Length of stay" in healthcare refers to how long a patient is in the hospital for treatment.
- **Age**
  - "Patient age" refers to the chronological age or number of years a person has lived.
- **Illness Severity**
  - "Illness severity" is how much a person is affected by a medical condition.

Derived Variables

- **Types of Complications Developed during Stay**
  - It refers to the length of the patient's hospital stay resulting from the development of complications.
- **Total Surgeries during The Stay**
  - It refers to the length of the patient's hospital stay resulting from the total surgeries.

Methodology

- We used three machine learning models : Logistic Regression with Ridge and Lasso, and Gradient Boosting with Ensemble learning to predict gender-specific mortality risk

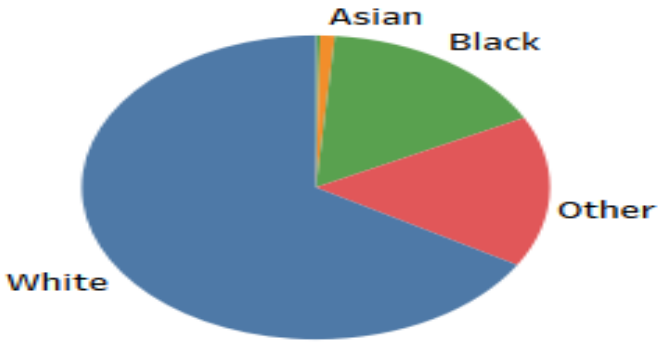
ML Algorithms	Description
Logistic Regression **+** **Lasso** **OR** **Ridge**	**Logistic regression - binary outcomes (yes/no)** **Logistic regression with Lasso and Ridge provide:** • **better model stability** • **prevents overfitting** • **performs feature selection** • **identify important predictors** • **flexibility to control model complexity** • **useful with high-dimensional data**
Gradient Boosting **+** **Ensemble Learning**	**Gradient Boosting - an ensemble machine learning technique.** • **builds predictive models sequentially to minimize errors** • **combines multiple weak models to create a strong predictor** • **used in various applications for its high accuracy.** **Ensemble learning enhances Gradient Boosting by:** • **improving predictive accuracy,** • **reducing overfitting and** • **increasing robustness.** **It combines the strengths of multiple models for better performance.**

THE Data Science INSTITUTE
at Saint Peter's University

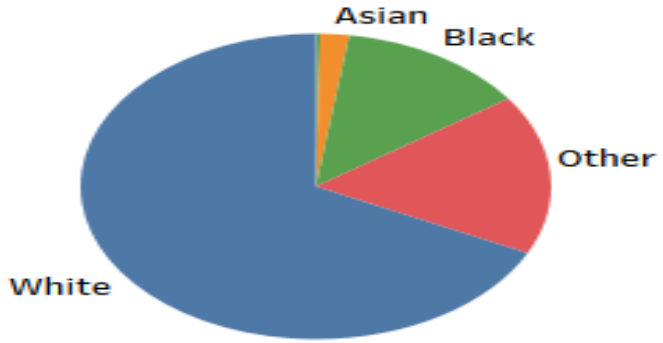
Demographic Information: Race and Gender

THE Data Science INSTITUTE at Saint Peter's University

FEMALE



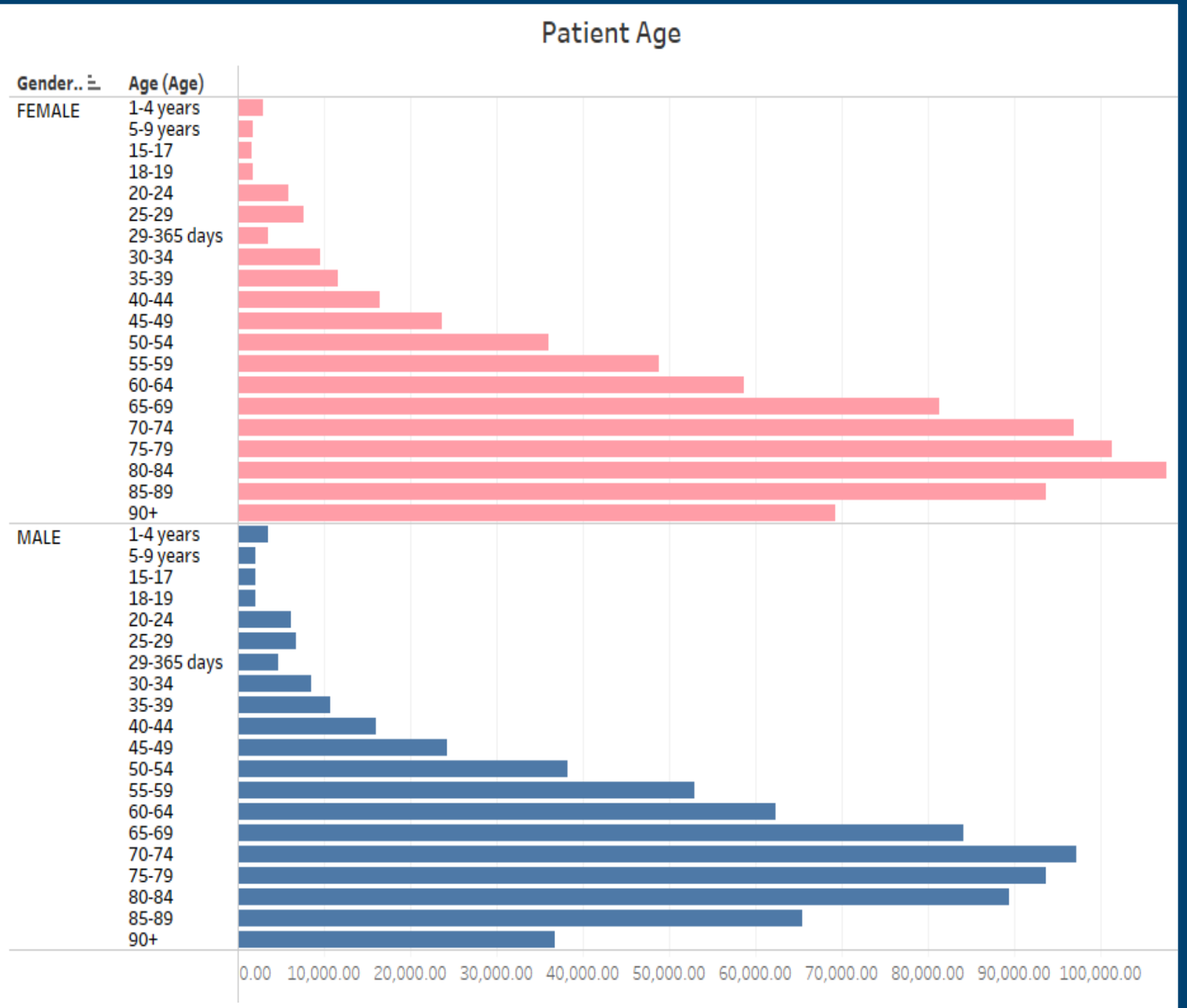
MALE



Predictive mortality race distribution by gender is almost the same for male and female.

Demographic Information: Age and Gender

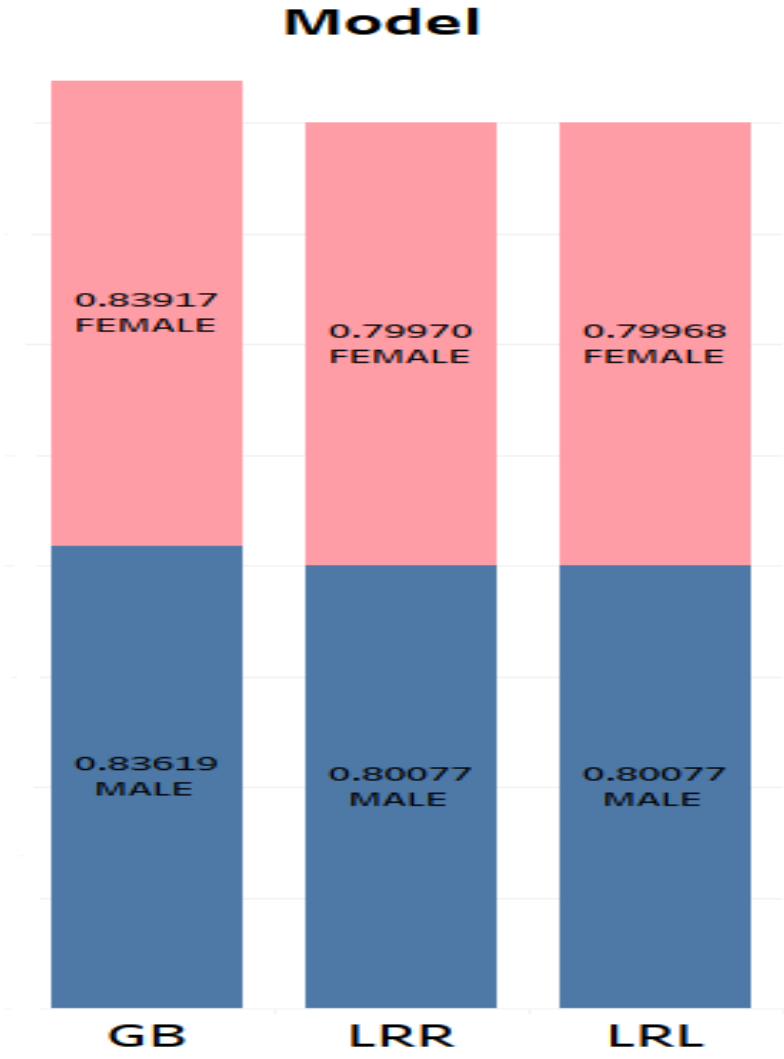
According to data visualization, evidence highlights that female tend to live longer than male.



THE Data Science INSTITUTE at Saint Peter's University

Machine Learning Model Results

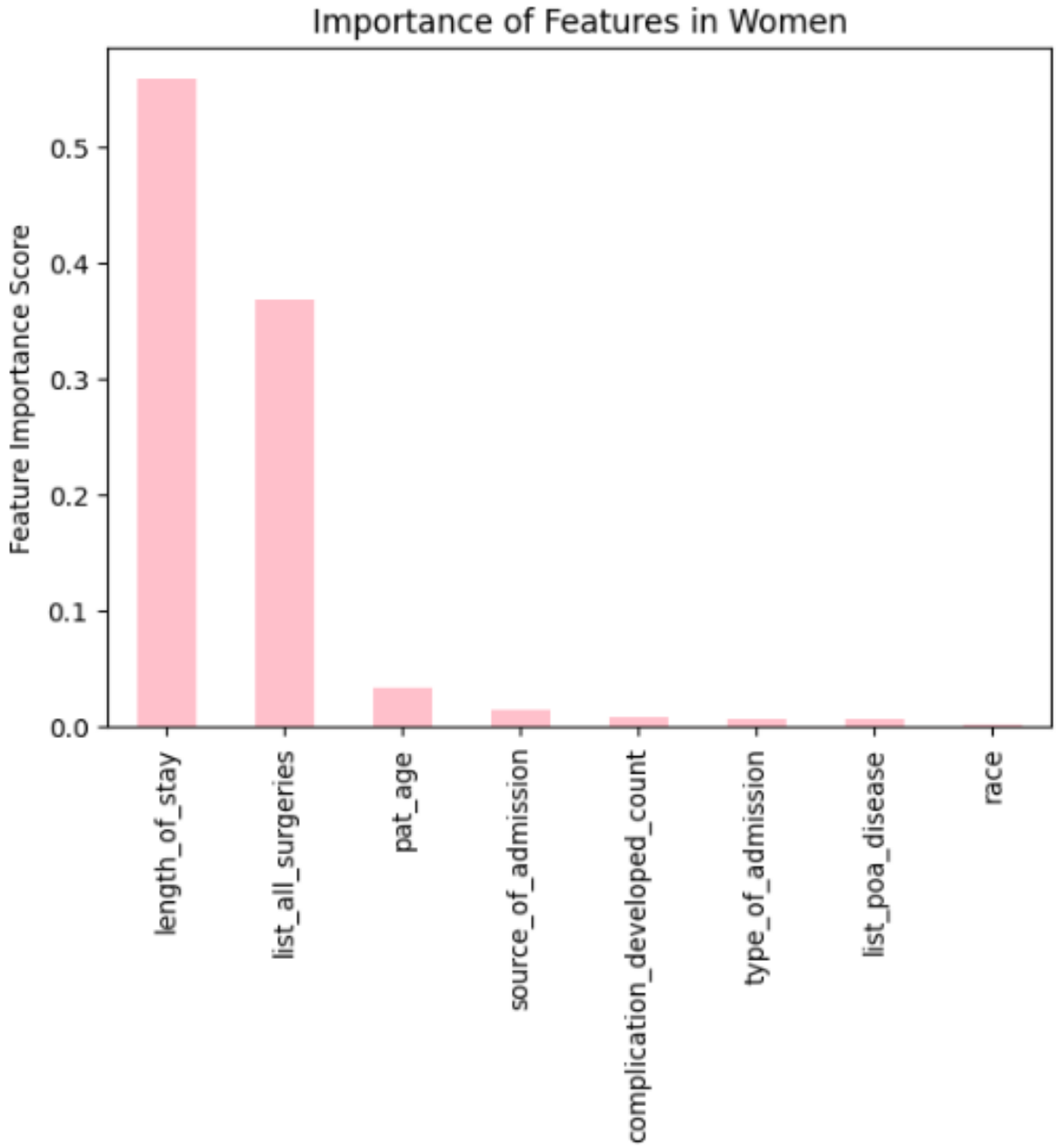
- In our study, we discovered that the most effective model is the **Gradient Boosting model** for both men and women.
- The best AUC-ROC with 5-fold cross validation was found to have a mean of **0.836** for male and **0.839** for female respectively using Gradient Boosting.



THE Data Science INSTITUTE at Saint Peter's University

Results: The top three variables for female

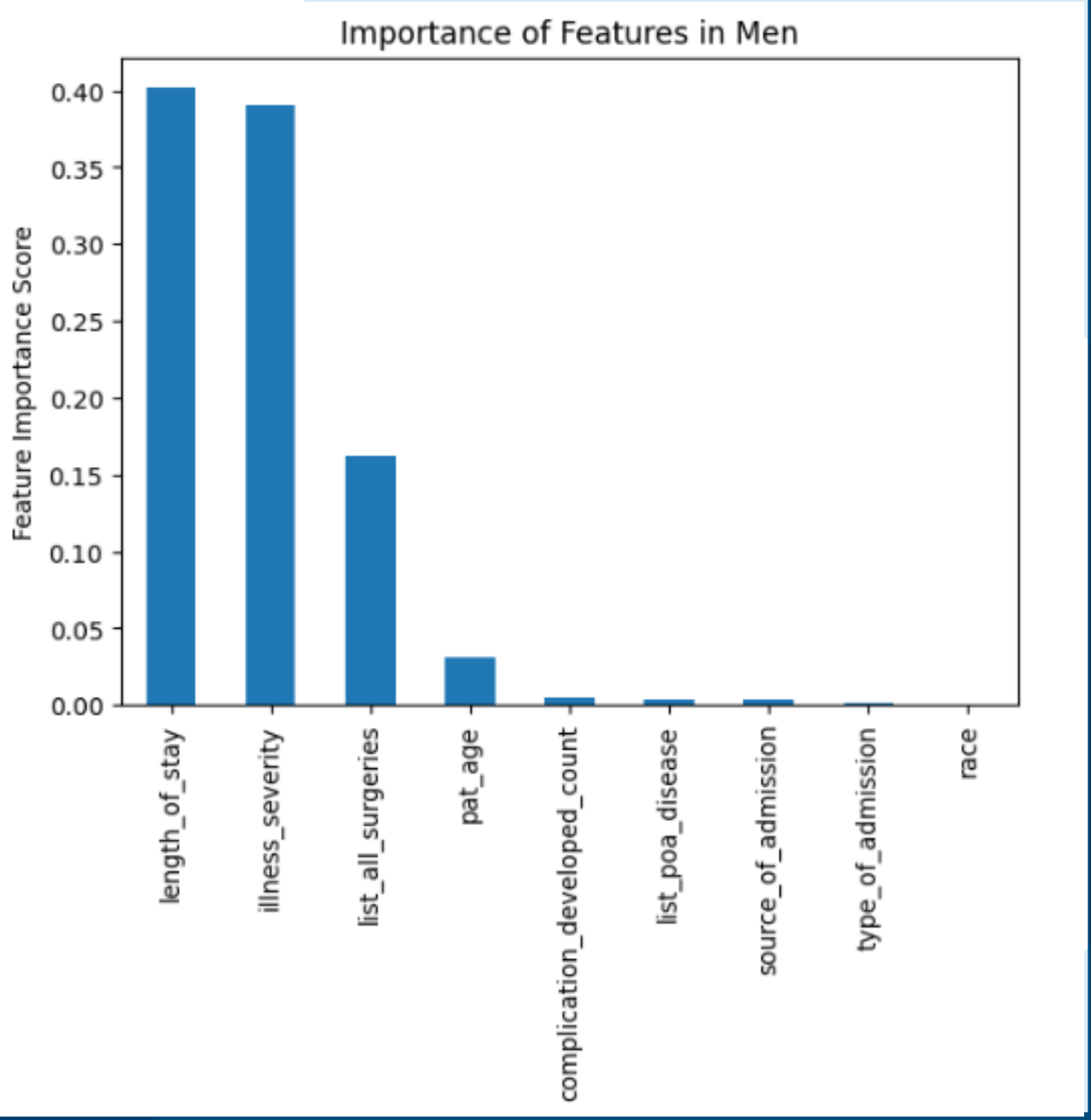
- Top three variables that impact mortality in female.
- The top 3 variables impacting mortality in women were **length of stay(day), total surgeries during the stay(day), and patient age** respectively.



THE Data Science INSTITUTE at Saint Peter's University

Results: The top thre variables for male

- We identified the top three variables that impact mortality in male.
- The top 3 variables impacting mortality in men were **severity of illness, length of stay(day), and total surgeries during the stay(day)**.

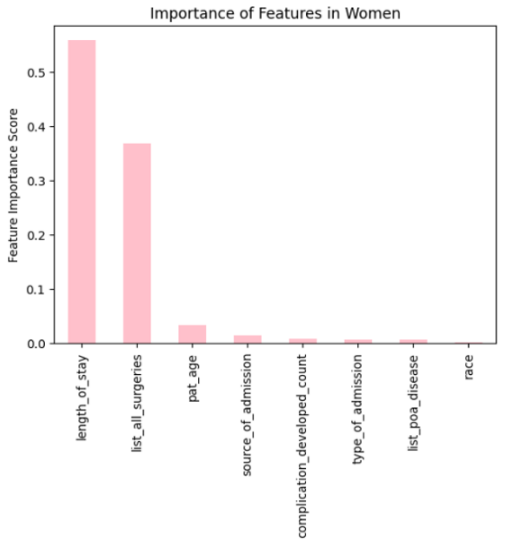


THE Data Science INSTITUTE at Saint Peter's University

- Comparison of top three variables affecting the mortality rate in male and female.
  Outcome: female's **patient_age** and male's **severity_of_illness** were different.

Result

- We identified the top three variables that impact mortality in women.
- The top 3 variables impacting mortality in women were length of stay, total surgeries conducted during the stay, and patient age respectively.

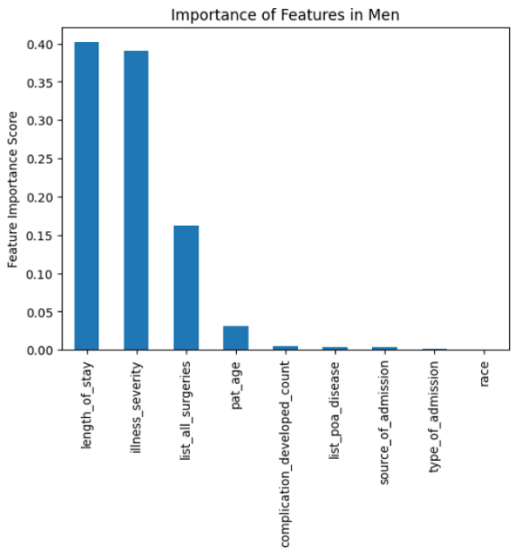


THE Data Science INSTITUTE at Saint Peter's University

10

Result

- We identified the top three variables that impact mortality in women.
- The top 3 variables impacting mortality in men were severity of illness, length of stay, and total surgeries during the stay.



THE Data Science INSTITUTE at Saint Peter's University

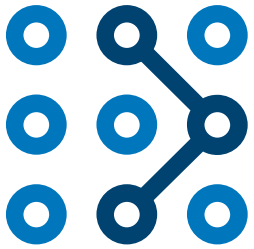
11

THE Data Science INSTITUTE at Saint Peter's University

THE Data Science INSTITUTE at Saint Peter's University

Conclusion

Our research suggests that considering gender-specific input variables might give insights to manage mortality.



THE Data Science INSTITUTE

at Saint Peter's University

Thank you for your attention!

Any Questions / Comments