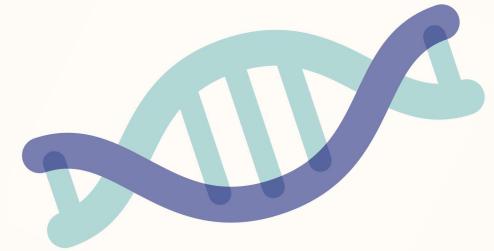


# PREDICTING HOSPITAL READMISSIONS

Mahi Sheth, Ashwin Ajit, Anupam Pradeep, Rebecca Han

STAT486 - December 19, 2025

# INDEX



**01**

Introduction

**02**

Dataset Description

**03**

EDA

**04**

Feature Engineering

**05**

Modeling

**06**

Results

01

# Introduction

## What is hospital readmission?

Hospital readmission is when a patient returns for an inpatient admission in a short time, usually within 30 days of being discharged.

## Why is this important to research?

Not only would it significantly decrease costs for hospitals, as well as avoid penalties for higher readmission rates, but it would also improve quality of life for patients, as they are given the proper care they need. Patients' mortality and morbidity are affected by hospital readmissions, so they must be prevented. Hospitals should optimize their resources to focus on vulnerable patients and prevent future problems.



02

## Dataset Description

The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. Each row contains hospital records of diabetes patients who underwent laboratory, medications, and stayed up to 14 days. There are 25,000 rows in total.

Relevant Columns:

- age
- time in hospital
- number of procedures during stay
- number of laboratory procedures
- number of medications administered
- primary diagnosis, secondary diagnosis, additional secondary diagnosis
- readmission



We cleaned the data by removing “other”, missing, and duplicate values. We also converted “age” to ordered factor and encoded our yes/no variables as binary.

## 04

# EDA - Correlation & Multicollinearity

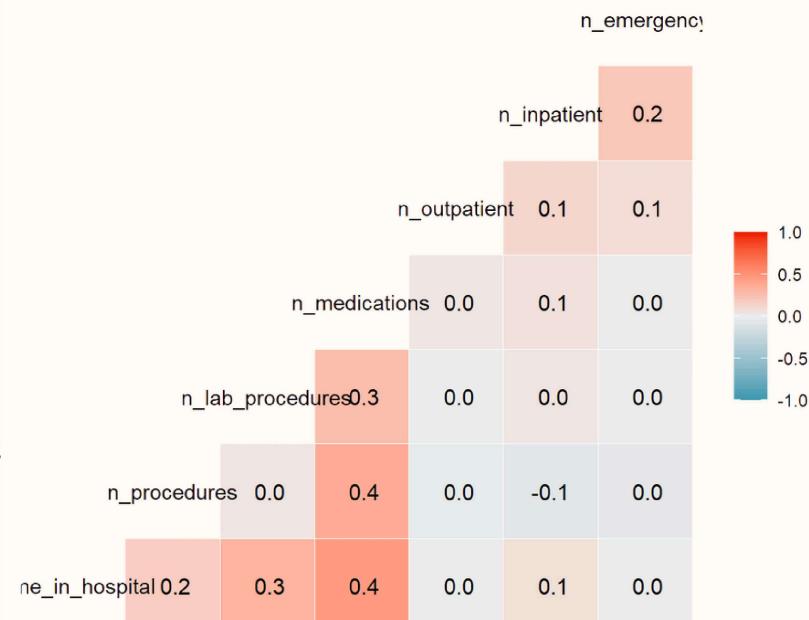
## What we observed

- Most correlations are low to moderate ( $\approx 0.1\text{--}0.4$ )
- Strongest relationships:
  - Time in hospital  $\leftrightarrow$  number of medications
  - Number of procedures  $\leftrightarrow$  number of medications
  - Time in hospital  $\leftrightarrow$  number of procedures
  - Time in hospital  $\leftrightarrow$  number of lab procedures

These relationships are expected, as longer hospital stays typically involve more tests, procedures, and medications. Correlated predictors raise concerns about multicollinearity in linear models.

## Modeling implication

- Logistic regression remains a valid baseline
- However, coefficient stability may be affected
- We therefore perform VIF checks and use regularized models (LASSO / Ridge)



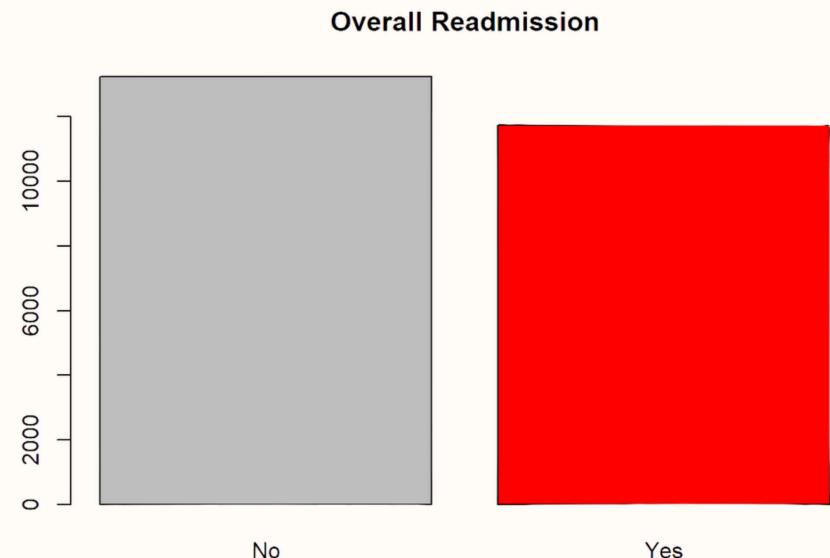
# 04

## EDA - Overall Readmission Rate

Readmissions are less frequent than non-readmissions, but there is still a substantial amount that raises concerns.

### What it means for modeling:

- There is moderate class imbalance in the outcome - models might favor majority class ("No") and under-predict the minority class ("Yes")
- Accuracy alone can be misleading



# 05

# Feature Engineering

## 01

### Log Transformations

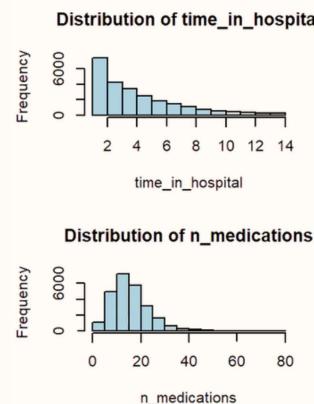
In our EDA, we noticed that variables such as “time in hospital” and “number of medications” were right skewed. To combat that, we performed logarithmic transformations, also limiting the influence of extreme-case patients.

## 02

### Rates instead of Counts

Rates reflect treatment intensity better than counts. One patient could receive 10 medications in 2 days, while another receives 10 medications in 10 days.

These are clinically very different!



## 03

### FDR Control

Since we are testing for multiple engineered features, we used the Benjamini-Hochberg procedure to control the false discovery rate. This allowed us to identify robust predictors of readmission while limiting false positives, ensuring that only statistically meaningful features were used in our modeling process.

# 04

# Feature Engineering

## VIF Test (Variance Inflation Factor)

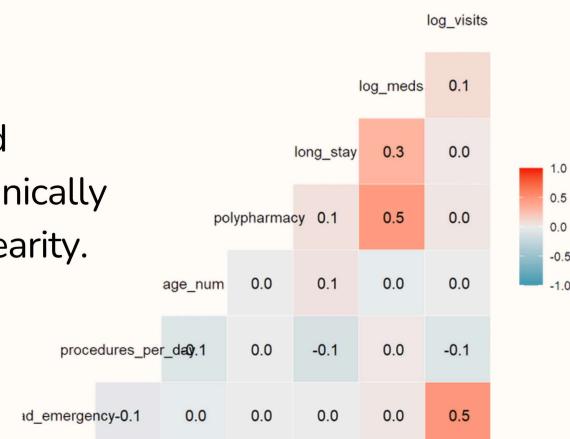
```
##      total_visits      had_emergency      meds_per_day  procedures_per_day
##      1.245451          1.241105          1.455027          1.265822
##      polypharmacy      elderly          long_stay
##      1.035896          1.003679          1.166417
```

All VIFs are close to 1, meaning there is no evidence of multicollinearity. Logistic regression coefficients should be stable and interpretable.

## Correlation Heat Map

Pairwise correlations among engineered features are relatively low, indicating limited linear dependence. A small number of moderate correlations are present between clinically related variables, but no extreme correlations that would indicate serious multicollinearity.

The rate and log transformations appear successful in reducing feature dependence.



# 05

# Logistic Regression



## Validation Table

##	Accuracy	Sensitivity	Specificity	AUC
##	0.6072000	0.3555841	0.8280135	0.6347430

## Performance Table

##	Accuracy	Sensitivity	Specificity	AUC
##	0.5994000	0.3395168	0.8240119	0.6241304

The model shows moderate discriminative ability (AUC = 0.62 - 0.63) , meaning it can separate high-risk patients from low-risk patients better than chance, but not with confidence. High specificity indicates strong performance at identifying non-readmitted patients. Lower sensitivity reflects difficulties predicting readmission risks.

## Key Takeaways

- Logistic regression model captures meaningful signal but misses many true readmissions
- Performance is stable across validation and test sets – no overfitting
- There is a need for more flexible models.

# 05

# Penalized Logistic: LASSO/Ridge



## LASSO vs. Ridge

	##	Accuracy	Sensitivity	Specificity	AUC
LASSO	##	0.6022000	0.3304573	0.8370619	0.6250457
Ridge	##	0.6028000	0.3429681	0.8273676	0.6241023

LASSO (L1) encourages sparsity and feature selection. Ridge (L2) shrinks coefficients to reduce variance. Both methods achieved nearly identical AUC values (0.62 - 0.63) to standard logistic regression, suggesting limited multicollinearity and a well-specified base model. Our penalized logistic models did not significantly improve discrimination, but it did confirm the robustness of our feature engineering.

## Key Takeaways

- Penalization does not meaningfully increase the model's ability to distinguish between readmitted and non-readmitted patients.
- Feature engineering successfully reduced noise, skewness, and redundancy.

## 05

# Random Forest (Nonlinear)



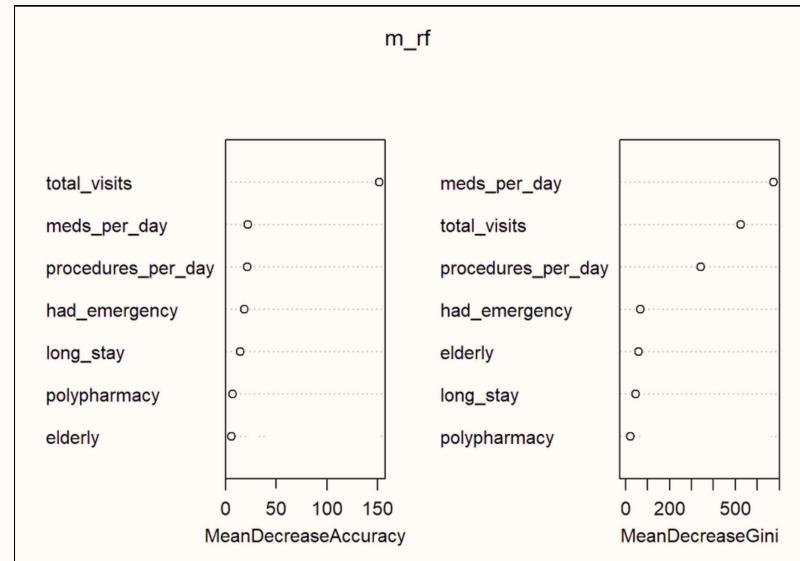
The Random Forest model achieves similar accuracy (0.60) and AUC (0.61) to logistic and penalized models, suggesting that allowing nonlinear interactions does not meaningfully improve predictive performance.

total\_visits has the highest Mean Decrease Accuracy, meaning the model becomes much worse at predicting readmissions when this variable is shuffled. meds\_per\_day and total\_visits have high Mean Decrease Gini, meaning they were used frequently and often in the decision trees.

## Key Takeaways

- Currently, healthcare utilization intensity is the strongest predictor of readmission.
- Linear and nonlinear models identify the same dominant predictors

	Accuracy	Sensitivity	Specificity	AUC
##	0.6034000	0.4535730	0.7348855	0.6184759
##	Accuracy	Sensitivity	Specificity	AUC
##	0.6014000	0.4572908	0.7259508	0.6059777



# 05

# XGBoost

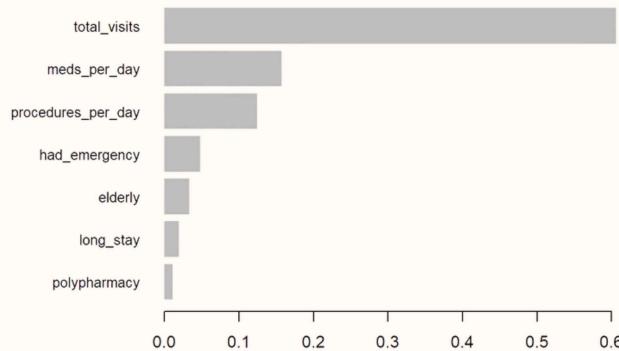


## Validation Table

```
##   Accuracy Sensitivity Specificity      AUC
## 0.6100000 0.5207531 0.6883214 0.6395974
```

## Performance Table

```
##   Accuracy Sensitivity Specificity      AUC
## 0.6002000 0.5198447 0.6696495 0.6297672
```



XGBoost achieves a modest improvement in AUC ( $\approx 0.63\text{--}0.64$ ) with stable validation and test performance. Although boosting captures nonlinear effects by correcting prior errors, gains in sensitivity remain limited. Feature importance is dominated by utilization measures, reinforcing that readmission risk is driven more by care intensity than complex interactions.

## Key Takeaways

- Boosting yields only modest gains over other models
- Utilization intensity remains the dominant signal in prediction
- Boosting confirms limited benefit from complex interactions

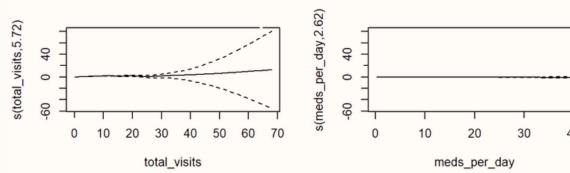
# 05

# GAM (Generalized Additive Model)



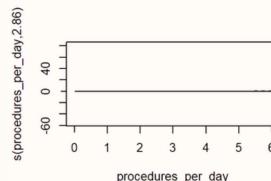
## Validation Table

```
##   Accuracy Sensitivity Specificity      AUC
## 0.6108000 0.4728284 0.7318813 0.6404809
```



## Performance Table

```
##   Accuracy Sensitivity Specificity      AUC
## 0.6084000 0.4745470 0.7240865 0.6259336
```



GAM allows smooth nonlinear effects while preserving interpretability. Performance is similar to linear models, suggesting nonlinearities are present but weak. Utilization intensity remains the dominant driver of readmission risk.

## Key Takeaways

- GAM captures some nonlinear patterns but does not greatly improve prediction
- Results are similar to logistic regression
- Readmission risk is mainly driven by how much care a patient uses

# 05

# SVM (RBF kernel)



## Validation Table

```
##    Accuracy Sensitivity Specificity          AUC
##  0.6052000  0.5220368   0.6781825  0.6241290
```

## Performance Table

```
##    Accuracy Sensitivity Specificity          AUC
##  0.6000000  0.5241588   0.6655481  0.6169184
```

The RBF SVM achieves similar performance to the other models, with AUC around 0.62 and accuracy near 0.60 on both validation and test sets. Results are stable across splits, indicating no overfitting, but improvements over simpler models are modest.

## Key Takeaways

- Nonlinear SVM provides little improvement over other approaches
- Scaling is important for SVM
- Model predicts non-readmissions better than readmissions

**05**

# XGBoost Tuned



## Validation Table

##	Accuracy	Sensitivity	Specificity	AUC
##	0.6052000	0.5220368	0.6781825	0.6241290

## Performance Table

##	Accuracy	Sensitivity	Specificity	AUC
##	0.6000000	0.5241588	0.6655481	0.6169184

The tuned XGBoost model shows a slight performance improvement, with validation AUC around 0.64 and test AUC near 0.63, while accuracy remains about 0.60. Similar validation and test results indicate no overfitting.

Compared to the untuned model, tuning adjusts tree depth, learning rate, and subsampling, leading to a small but consistent gain in discrimination, though overall improvements remain modest.

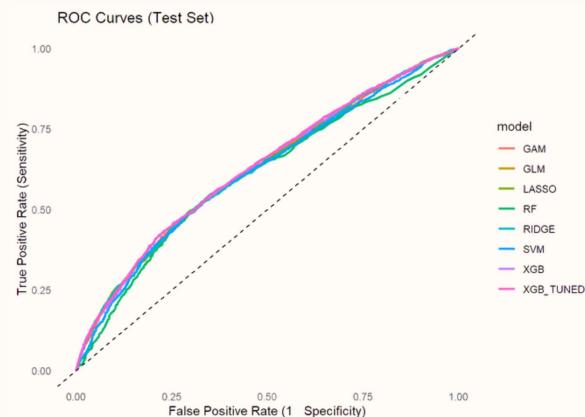
## Key Takeaways

- Hyperparameter tuning slightly improves XGBoost performance
- Gains over untuned XGBoost are small but consistent
- Still misses many true readmissions

06

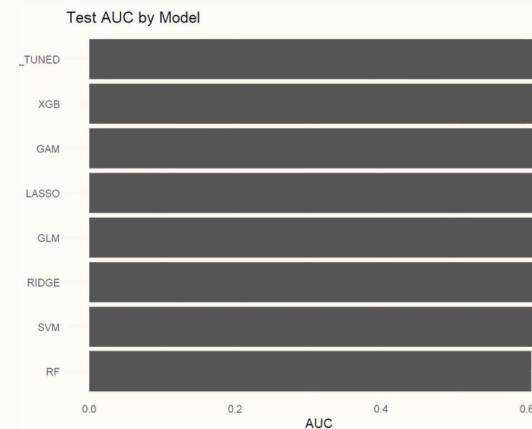
# Results

## ROC Curves



All models perform better than random guessing, as their ROC curves lie above the diagonal. The curves are closely grouped, indicating similar performance across models. Tuned XGBoost and GAM are slightly higher, consistent with their marginally higher AUC values.

## AUC bar chart

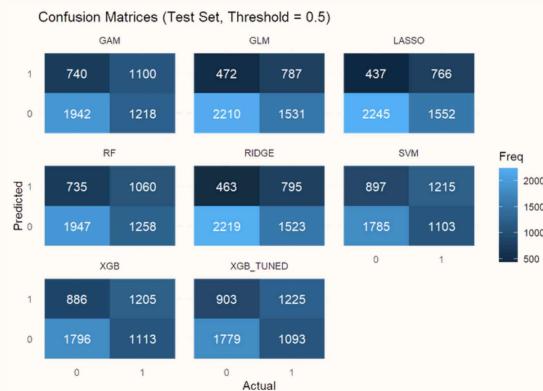


Test AUC values range narrowly from about 0.61 to 0.63. Tuned XGBoost achieves the highest AUC, followed closely by XGBoost, GAM, and the penalized linear models. Random Forest and SVM perform slightly worse but remain comparable.

# 06

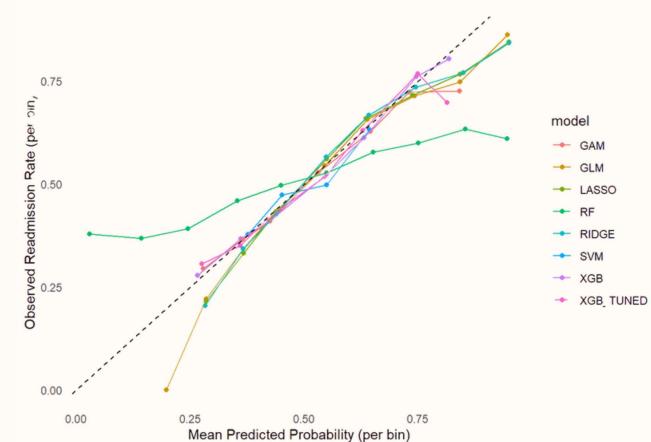
# Results (cont)

## Confusion Matrices (Threshold = 0.5)



At a standard threshold of 0.5, all models predict non-readmissions more accurately than readmissions. This is reflected in consistently higher specificity than sensitivity across models. Even the best-performing models still miss a substantial number of true readmissions.

## Calibration Plot



Most models are reasonably calibrated in the middle probability range, with some deviation at higher predicted risks. Random Forest shows weaker calibration than the other models.

# THANK YOU!