NO SECOND STAY: PREDICTING 30-DAY HOSPITAL READMISSIONS

DSC450 | Team Presentation

Team Members:

Tyler Heflin – Data Wrangler

Andres Melendez – Data Scientist

Carlos Escamilla – Data Visualizer

Delilah Slabaugh –Presenter

INTRODUCTION

Hospital readmissions within 30 days are costly and common

Goal: use data to predict patients at risk of being readmitted

Approach: machine learning on synthetic EHR dataset (30,000 patients)

BUSINESS PROBLEM / HYPOTHESIS

Can we predict 30-day readmission using clinical and demographic data?

Hypothesis: Discharge destination and chronic conditions increase risk

METHOD & ANALYSIS

1

Preprocessing: handled missing values, encoded variables, cleaned features

2

Exploratory Data Analysis (EDA) to check distributions and imbalance

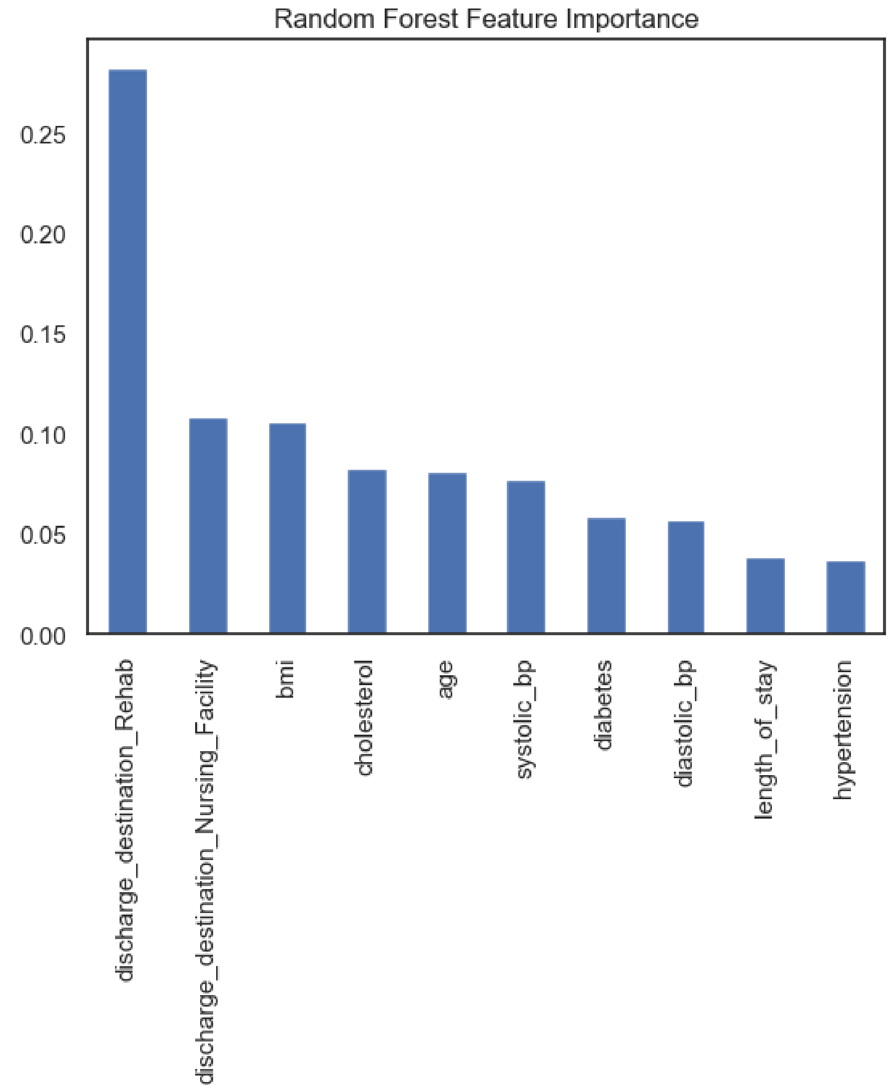
3

Modeling: Random Forest & XGBoost with GridSearchCV tuning

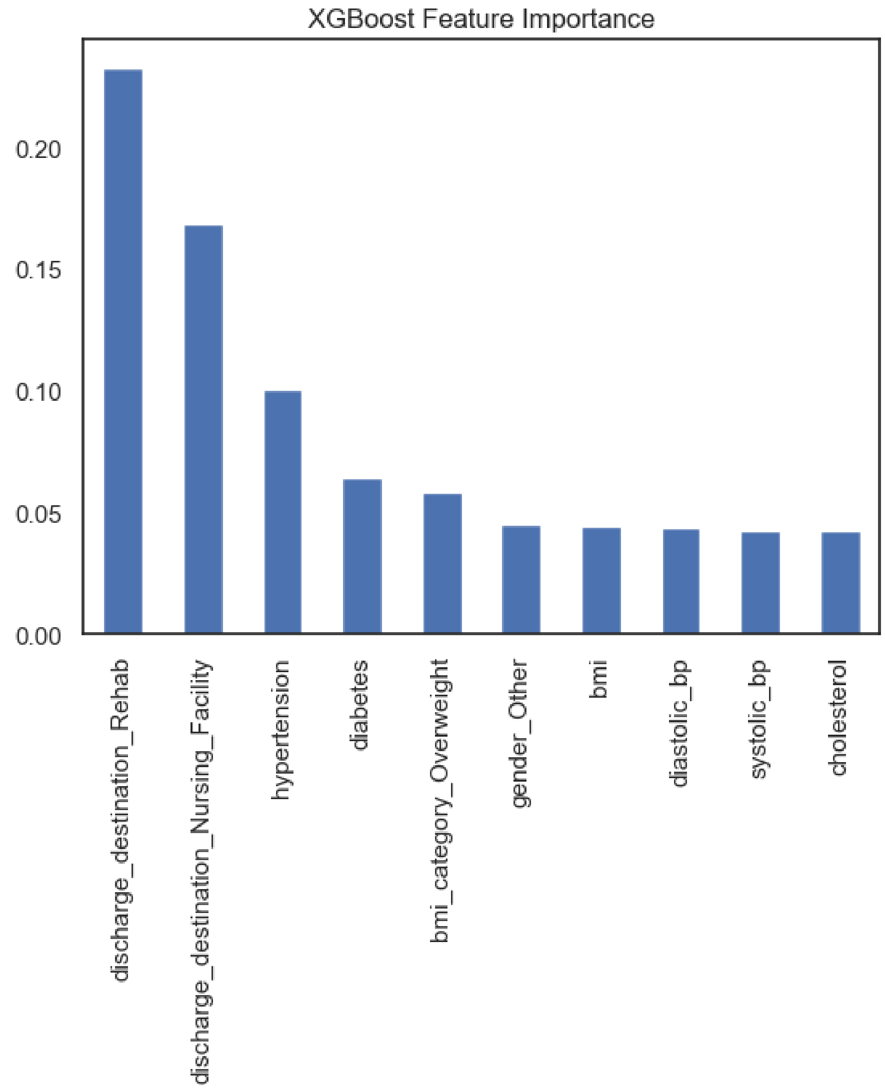
4

Evaluation: stratified 80/20 split, ROC-AUC, precision, recall, F1-score

FEATURE IMPORTANCE (RANDOM FOREST)



FEATURE IMPORTANCE (XGBOOST)

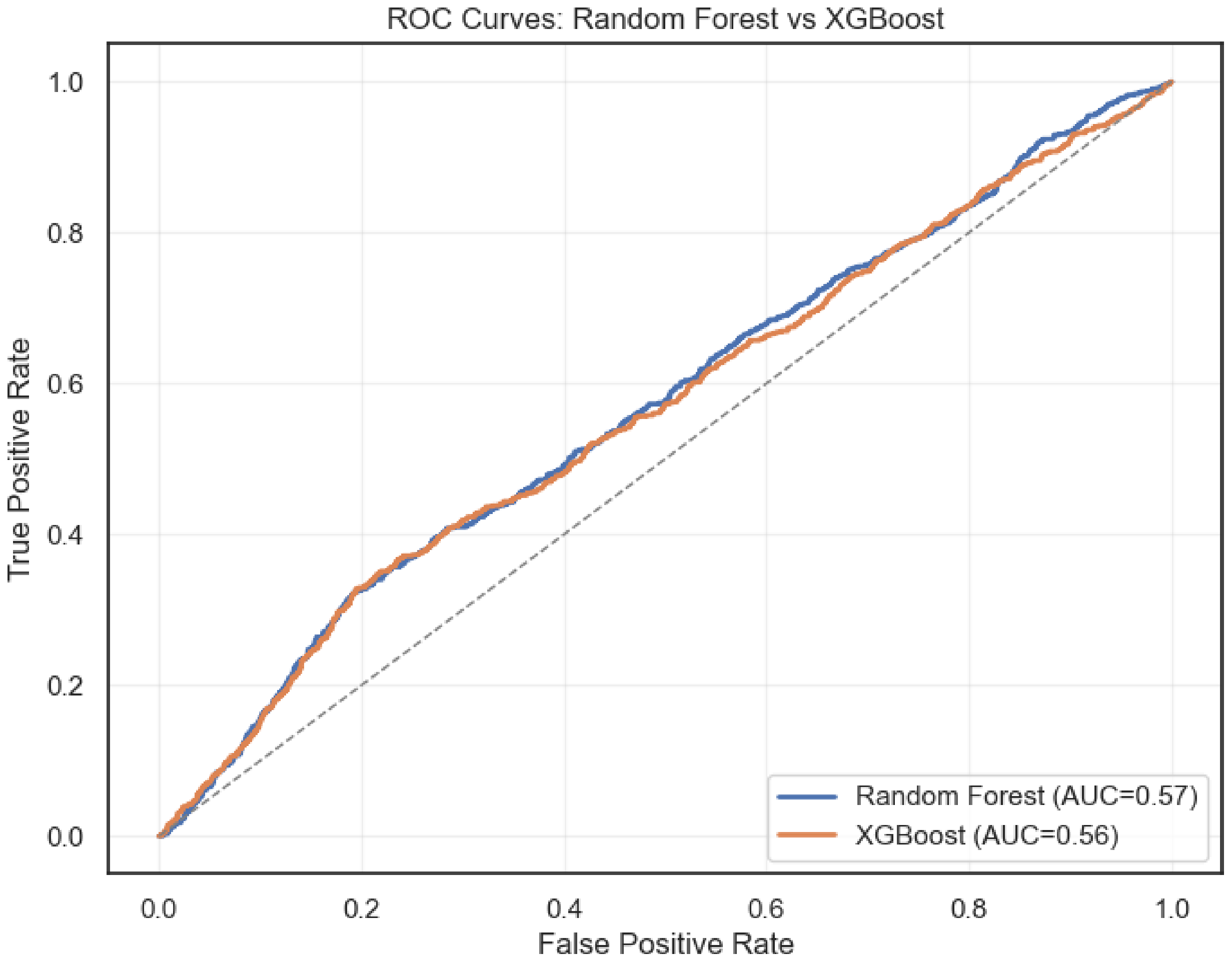


KEY FEATURE INSIGHTS

Top predictor: discharge to rehab or nursing facility

Other drivers: BMI, cholesterol, hypertension, diabetes

XGBoost also picked up gender and BMI category



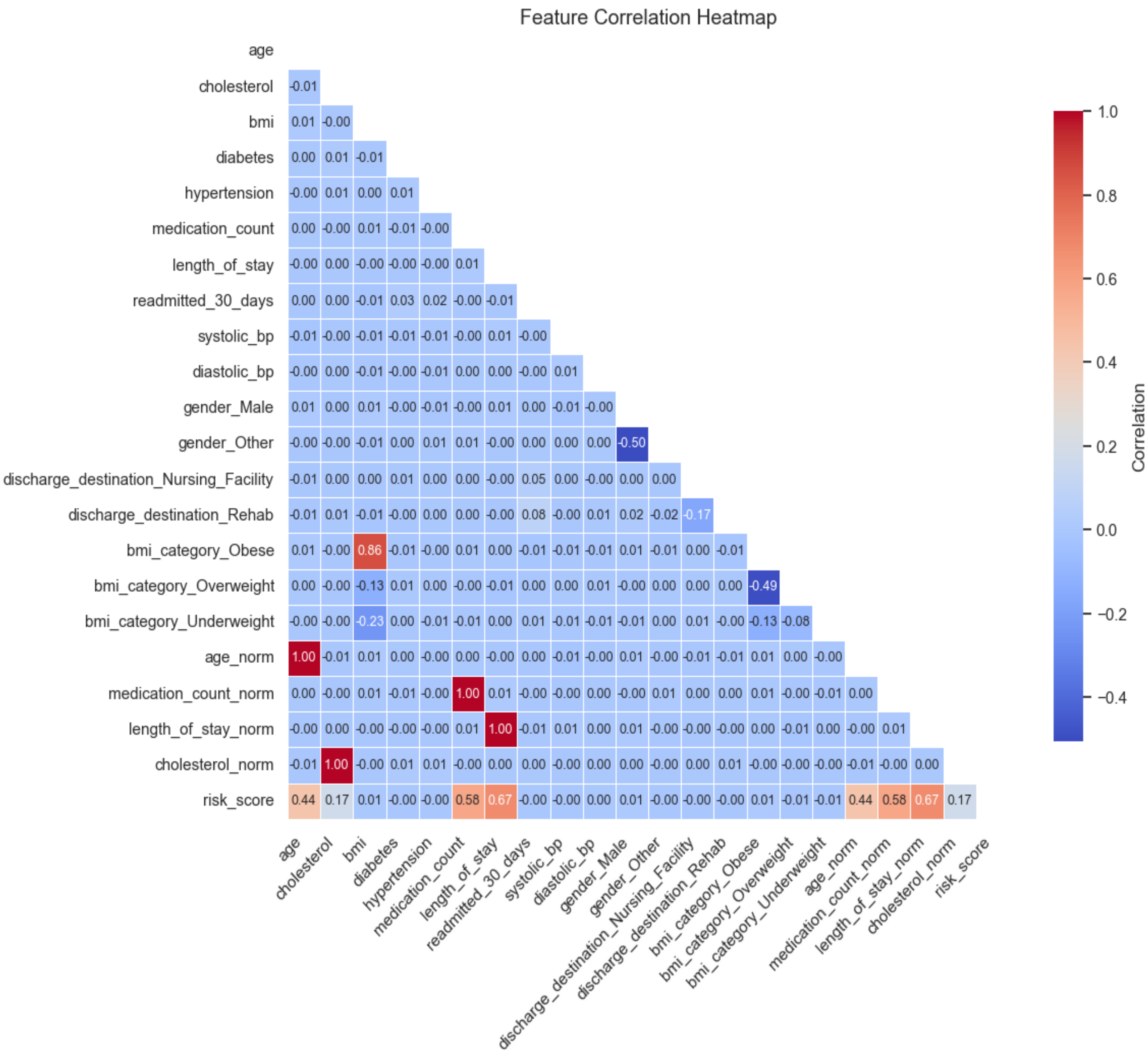
ROC CURVE: RF VS XGB

MODEL PERFORMANCE

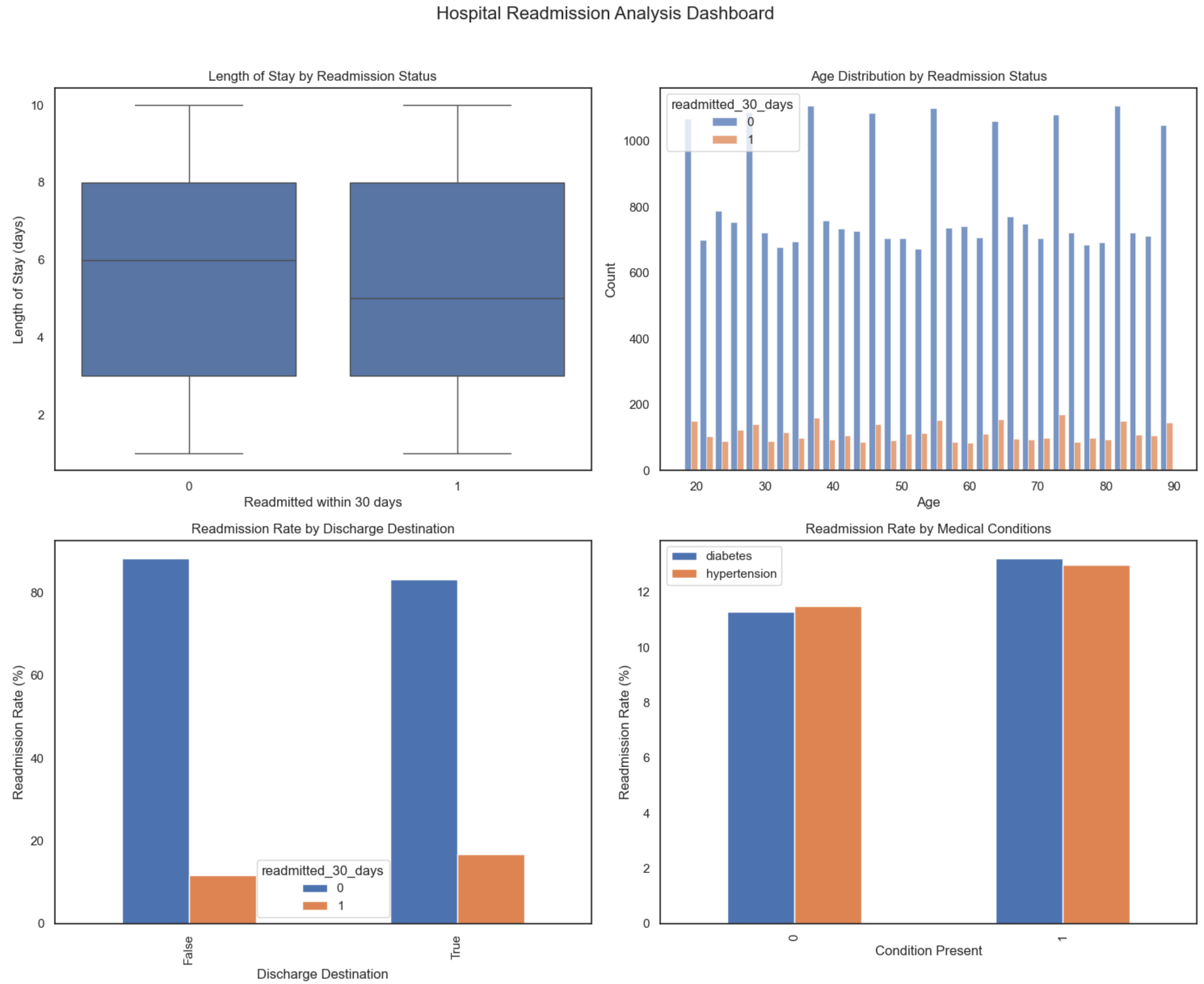
ROC curves show low improvement from random guessing

Random Forest AUC = 0.57, XGBoost AUC = 0.56

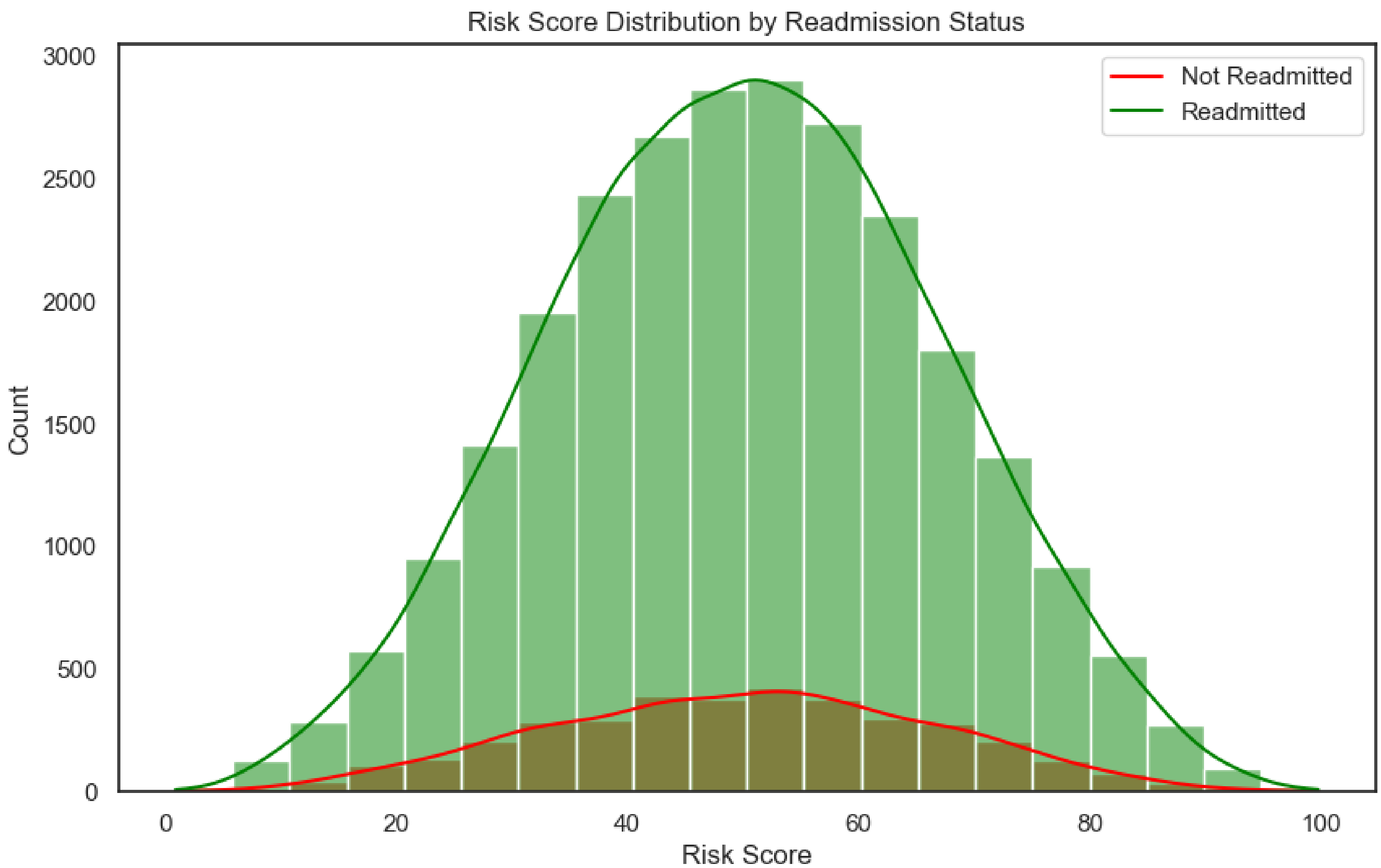
Low AUC suggests room for improvement in model design or data quality



CORRELATION BETWEEN FEATURES

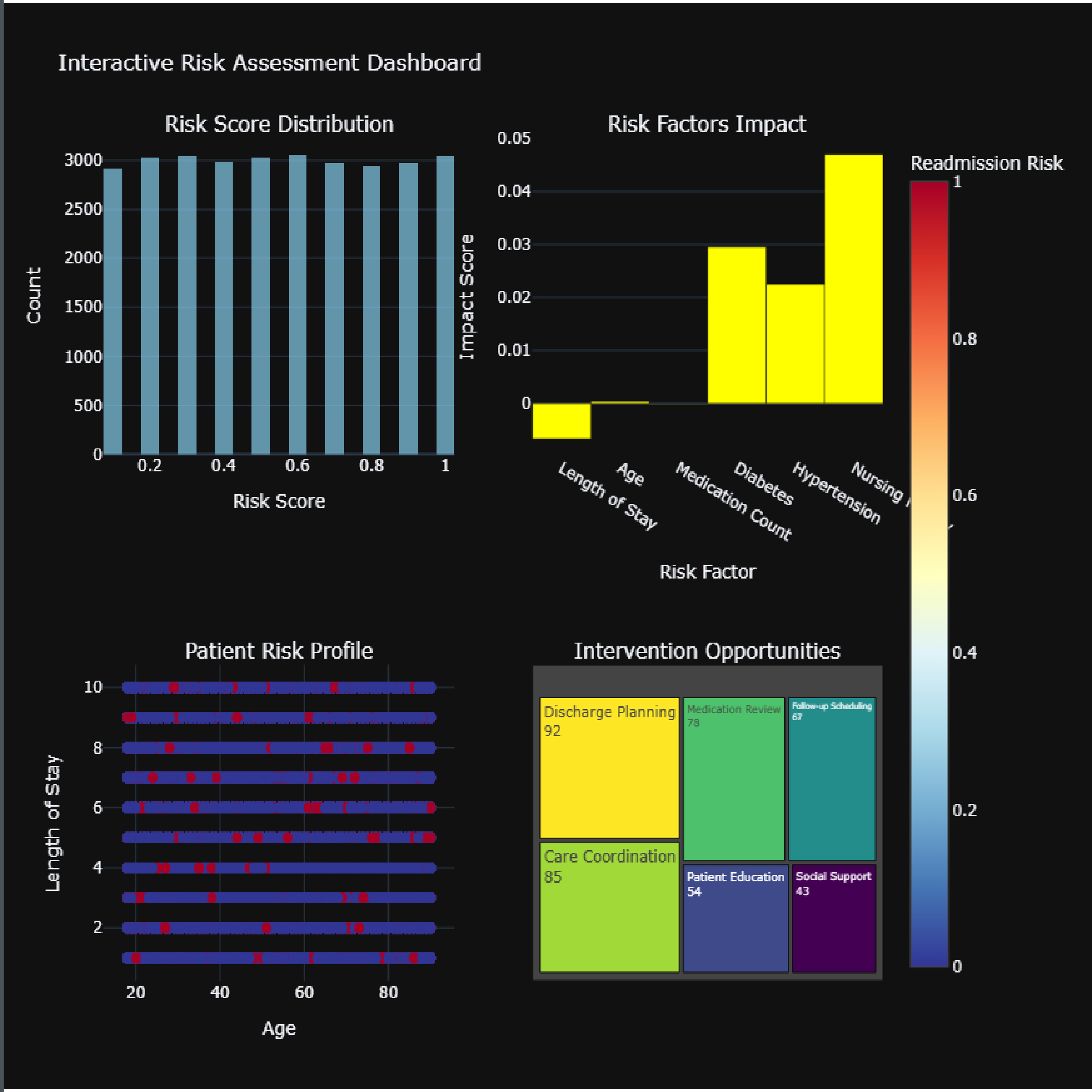


READMISSION ANALYSIS DASHBOARD



RISK SCORE DISTRIBUTION BY READMISSION STATUS

RISK & INTERVENTION OPPORTUNITIES



RECOMMENDATIONS & ETHICS

Use	Avoid	Validate
Use predictions to trigger follow-ups, not automate care decisions	Avoid bias and protect patient privacy	Validate models with real-world data before deployment

CONCLUSION

- OUR HYPOTHESIS WAS SUPPORTED: DISCHARGE DESTINATION AND CHRONIC CONDITIONS ARE KEY RISK FACTORS
- BOTH MODELS IDENTIFIED STRONG PREDICTORS LIKE REHAB DISCHARGE, HYPERTENSION, AND BMI
- MODEL PERFORMANCE WAS MODEST (AUC ~0.56–0.57), BUT SHOWED MEANINGFUL PATTERNS
- RISK SCORING AND DASHBOARDS OFFER PRACTICAL TOOLS FOR CLINICAL DECISION-MAKING
- FUTURE IMPROVEMENTS INCLUDE ADDING NEW FEATURES, VALIDATING WITH REAL DATA, AND REFINING MODEL DESIGN
- PREDICTIVE MODELING CAN HELP REDUCE UNNECESSARY READMISSIONS—AND SUPPORT BETTER OUTCOMES

REFERENCES

1. Bauder, R. A. & Khoshgoftaar, T. M. (2022). A survey of machine learning techniques for patient readmission prediction. Health Information Science and Systems, 10(1), 1-17. – Survey of ML algorithms used in readmission prediction including preprocessing, feature selection, and model evaluation.

2. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital readmission. Proceedings of the 21st ACM SIGKDD International Conference of Knowledge Discovery and Data Mining, 1721-1730 – Demonstrates trade-offs between black-box and interpretable models in critical care predictions.

3. Centers for Medicare & Medicaid Services. (2022). Hospital Readmissions Reduction Program (HRRP). Retrieved from http://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program - Background for policy and financial implications of hospital readmission rates.

4. Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2016). Doctor AI: Predicting clinical events via recurrent neural networks. Machine Learning for Healthcare Conference (MLHC). – Deep learning-based model that simulates clinical decision-making with sequential EHR data.

5. Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. Journal of Biomedical Informatics, 56, 229-238. – Applicable for benchmarking, comparing logistic regression, decision trees, and ensemble methods.

6. Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., … & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. – Real-world EHR dataset used for benchmarking of AI healthcare models.

7. Kaggle. (2022). Synthetic hospital readmission prediction dataset. Retrieved from https://www.kaggle.com/datasets/siddharth0935/hospital-readmission-predictionsynthetic-dataset - Dataset used for this project that, while synthetic, simulates real-world hospital data while preserving patient privacy.

8. Lundberg, S. M., & Lee, S. -I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30 (NeurIPS).

9. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future – big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216-1219. – Explores ethical concerns, clinical applications, and limitations of machine learning in medicine.

10. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine 380(14), 1347-1358. – A broad overview of ML applications across clinical domains with a focus on integration and usability.