Predicting 30-Day Readmission Risk Among Diabetes Patients

University of Pittsburgh HI2251

Soli Evans & Abhimaan Mayadam

Background & Motivation

- Diabetes accounts for a significant portion of hospital admissions.

- 30-day readmissions disrupt care continuity and increase costs.

- HRRP penalties create urgency to reduce preventable readmissions.

- **Our Objective:** Identify key predictors of 30-day readmission using a large, real-world dataset.

Problem Statement

- Patients with diabetes are at elevated risk of readmission due to complex management needs.

- Unplanned readmissions are costly and clinically impactful.

- **Guiding Question:**
  Which demographic, clinical, and utilization factors most strongly predict 30-day readmission?

Dataset Overview

UCI Diabetes 130-US Hospital Dataset

101,766 encounters, 51 variables

Includes demographics, diagnoses, utilization history, treatments

Readmission recoded into binary 0/1 (within 30 days)

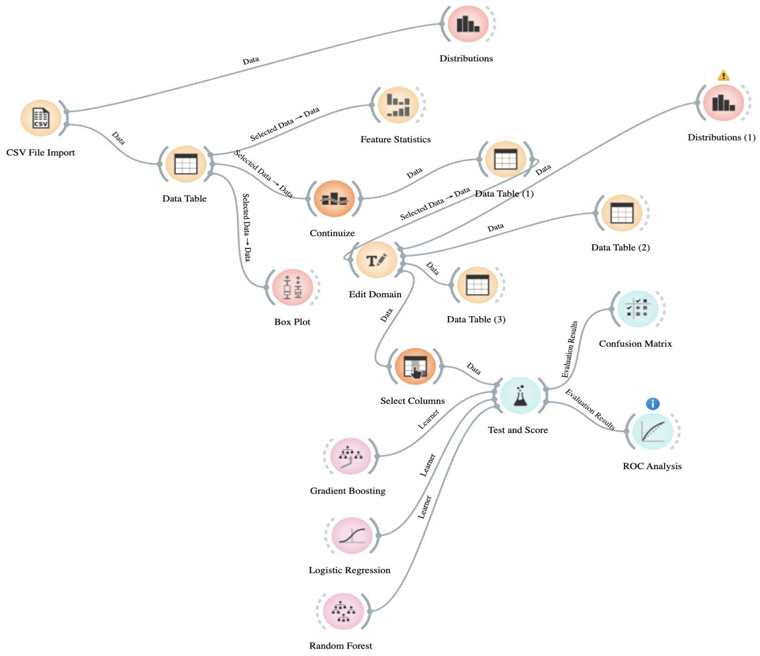
Tools Used

R: initial NA cleaning

Python: feature engineering, logistic regression, odds ratios, AUC

Orange: workflow diagram, machine learning models, ROC curves

Visualization tools: built-in Orange charts for EDA



Data Preparation

Removed missing/invalid values

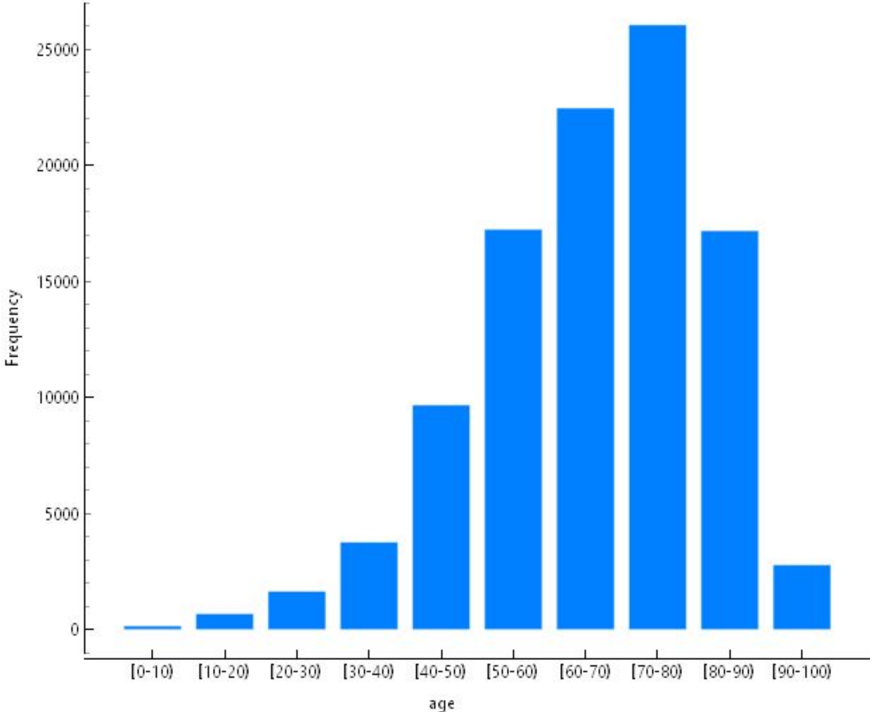
Converted age ranges to midpoints

Encoded categorical variables

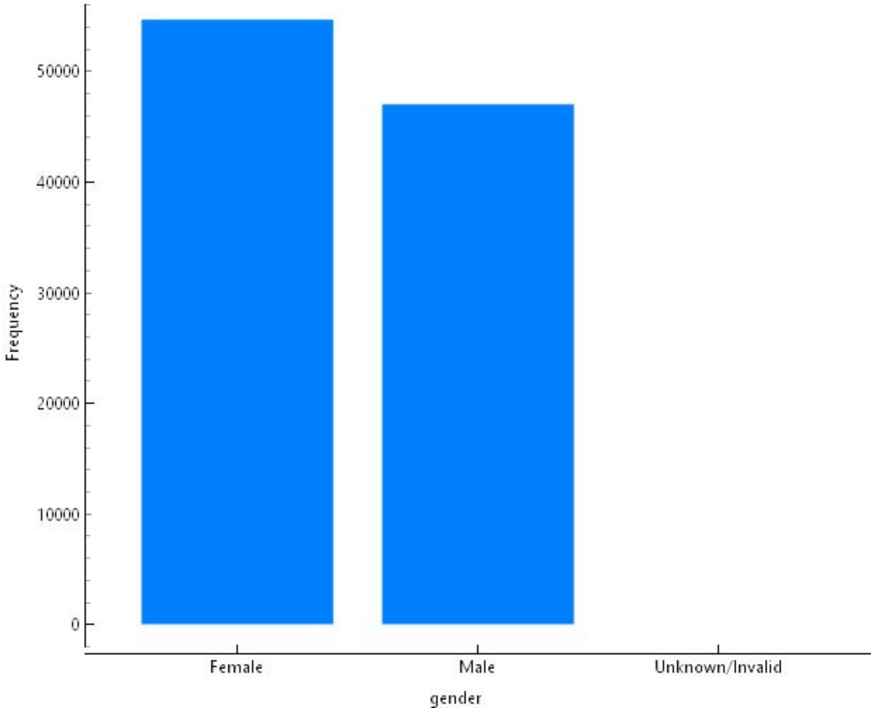
Selected clinically meaningful predictors

Created binary readmission variable

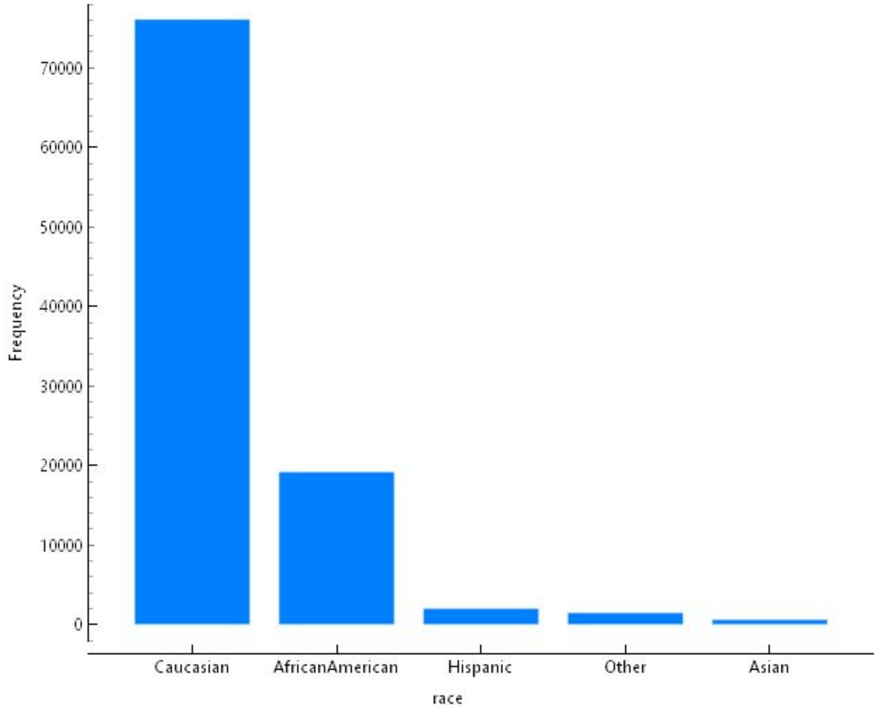
Age Distribution



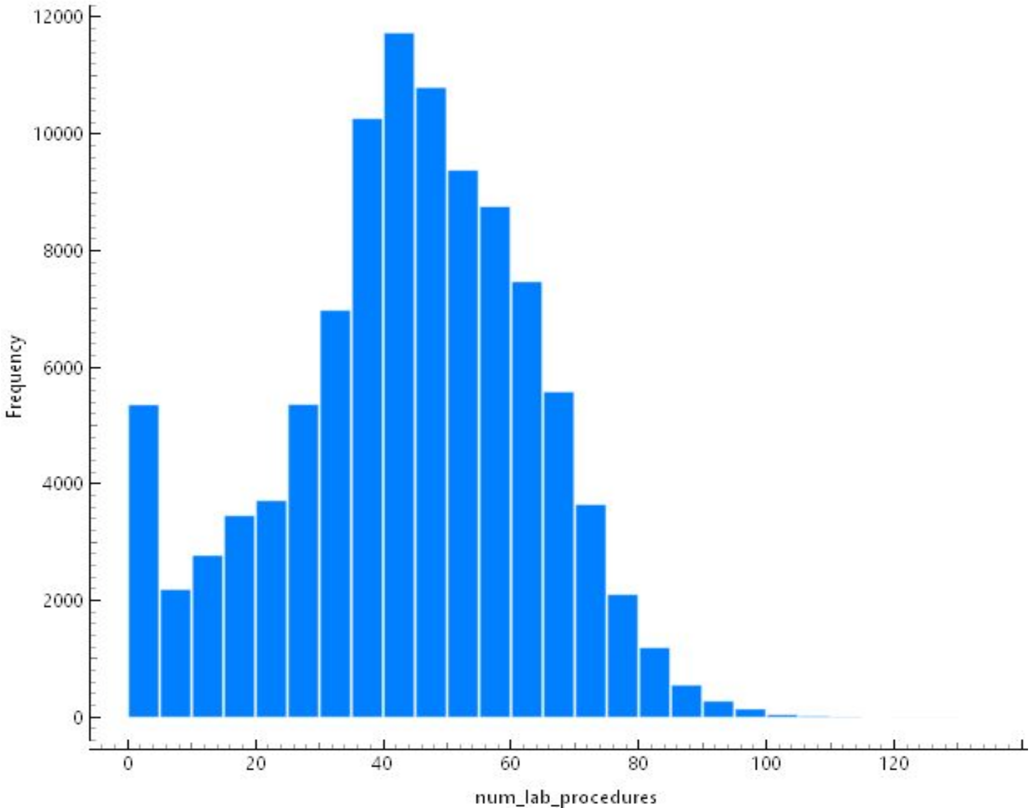
Gender Distribution



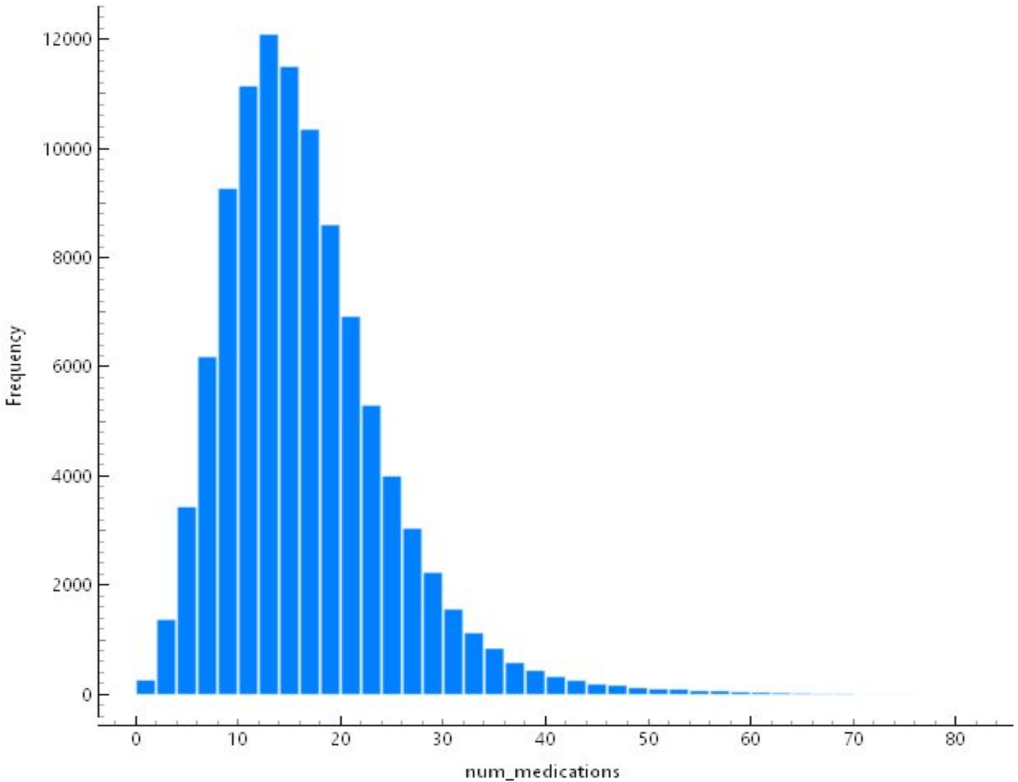
Race Distribution



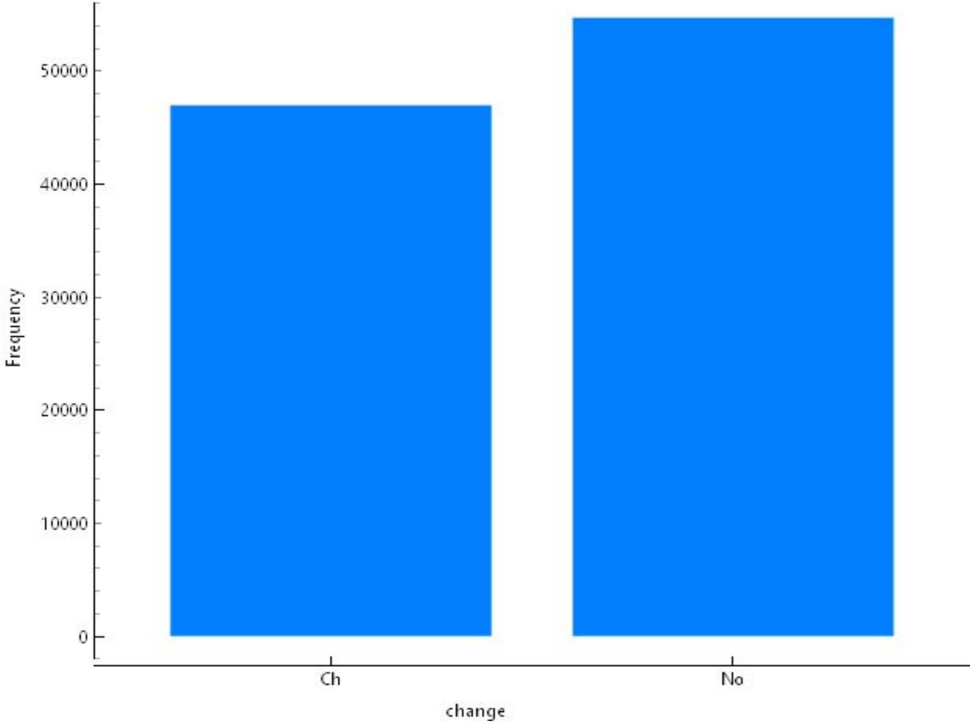
Clinical Variables: Lab Procedures



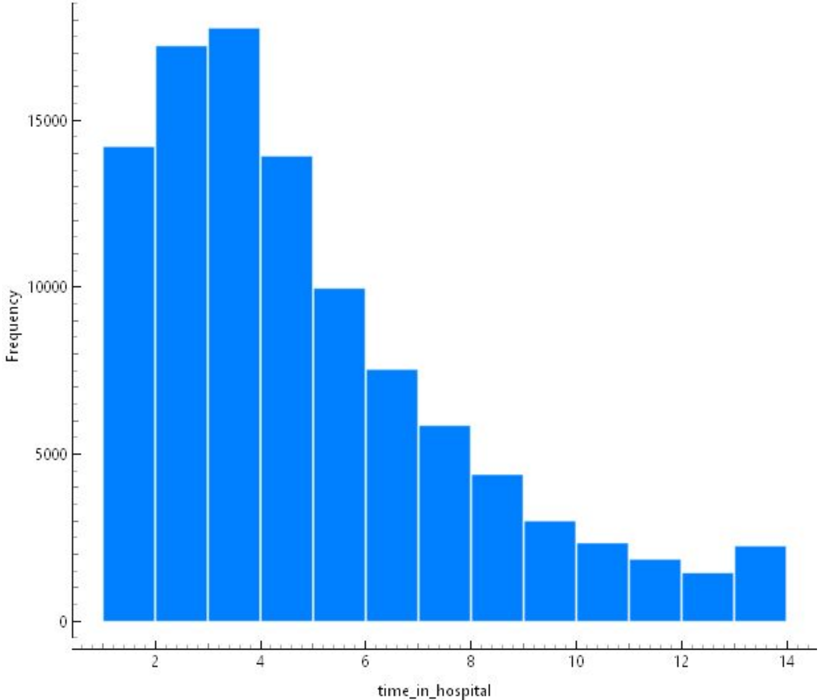
Clinical Variables: Medication

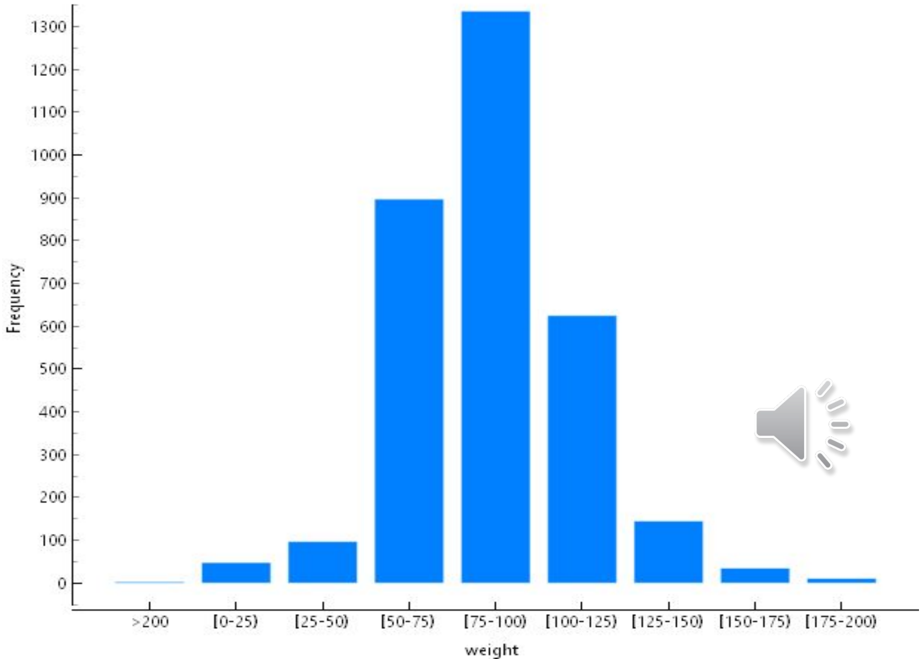


Treatment Indicators



Clinical Utilization





Modeling Approach

Baseline: Logistic Regression (interpretability)

Machine Learning models:

- Random Forest
- Gradient Boosting (best performance)

Evaluated with:

- ROC curves
- AUC
- Classification accuracy, precision, recall, MCC

Logistic Regression Results

Strongest predictors:

- **Inpatient visits — OR ~1.3189**
- **Emergency visits — small but meaningful effect**

Age, meds, lab procedures had small effects

Points to utilization as the best predictor of readmission risk

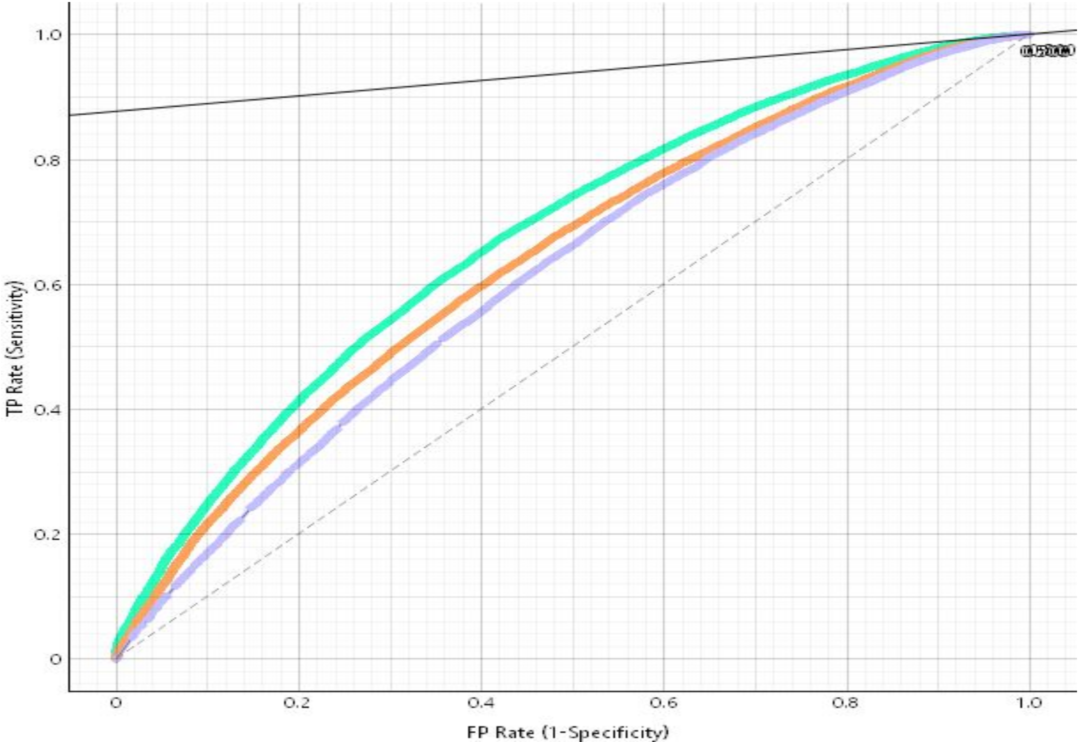
Model Performance - Comparative AUC

Gradient Boosting AUC
0.647 (best)

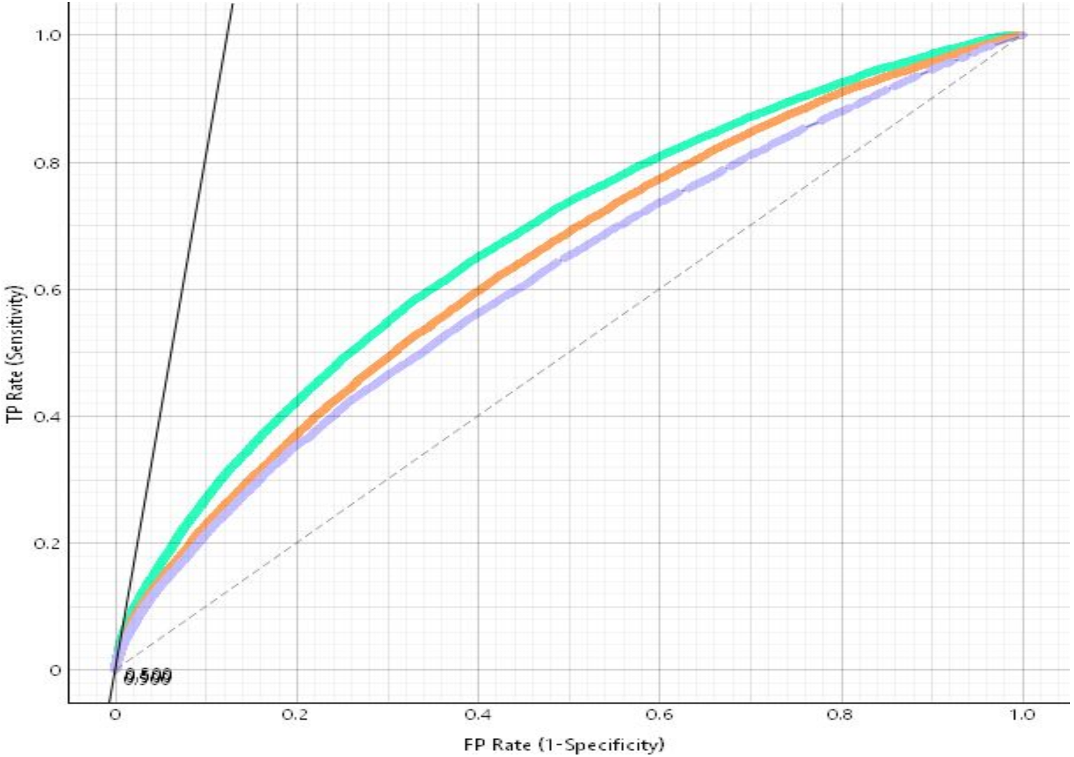
Logistic Regression AUC
0.640

Random Forest AUC
0.614

ROC Curve: Target 0 (Not Readmitted)



ROC Curve: Target 1(Readmitted)



Discussion

Utilization variables (inpatient & ER visits) strong predictors

Demographics alone not strong predictors

Machine learning better captures nonlinear complexity

Dataset imbalance likely reduced logistic regression performance

Conclusion

- Predictive modeling can meaningfully support early intervention
- Machine learning models offer moderate discriminatory ability
- Identified variables can guide targeted transitional care strategies

Limitations

Race imbalance

Missing data

Administrative (not clinical) dataset

Limited medication intensity details

Model generalizability concerns

Future Directions

- Improve dataset balance (SMOTE, stratified sampling)
- Add comorbidity indices (Charlson, Elixhauser)
- Integrate SDoH variables
- Test model in a real-world operational workflow
- Explore deep learning or ensemble stacks

Works Cited

American Diabetes Association. (2022). *The impact of diabetes on 30-day readmission*. *Diabetes, 71*(Supplement 1).

Agency for Healthcare Research and Quality. (2021). *Healthcare Cost and Utilization Project (HCUP): Readmissions and high utilizers*. https://www.ahrq.gov/

Centers for Medicare & Medicaid Services. (2023). *Hospital Readmissions Reduction Program (HRRP)*. 2https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/read missions-reduction-program

Fingar, K., Washington, R. (2015). Statistical Brief #196 Trends in Hospital Readmissions for Four High-Volume Conditions, 2009–2013. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs [Internet]*. https://www.ncbi.nlm.nih.gov/books/NBK338299/

Karunakaran, A., Zhao, H., & Rubin, D. J. (2018). Pre- and post-discharge risk factors for hospital readmission among patients with diabetes. *Medical Care, 56*(7), 634–642. https://doi.org/10.1097/MLR.0000000000000931

Rubin, D. J., Maliakkal, N., Zhao, H., & Miller, E. E. (2023). Hospital readmission risk and risk factors for people with a primary or secondary discharge diagnosis of diabetes. *Journal of Clinical Medicine, 12*(1274). https://doi.org/10.3390/jcm12041274

Talavera, M. J. (2017). *Hospital readmission prevention: A literature critique*. University of Pittsburgh.

University of California, Irvine. (n.d.). *Diabetes 130-US hospitals for years 1999–2008 data set*. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008

Van Walraven, C., Bennett, C., Jennings, A., Austin, P. C., & Forster, A. J. (2011). Proportion of hospital readmissions deemed avoidable: A systematic review. *CMAJ, 183*(7), E391–E402. https://doi.org/10.1503/cmaj.101860

Thank You!

Questions?