Hospital Admission Prediction

Final Term Presentation
Course: BA878E1
Dt. 12th Dec 2023

Team: Yuesen(Sam)Zhang, Yutao(Peter)Luo, Prateek Naharia

Mission

To enhance emergency care by providing real-time, data-driven insights for optimized hospital admission decisions, improving resource allocation and patient outcomes.

Vision

To seamlessly integrate machine learning in healthcare, facilitating proactive decision-making, improved patient satisfaction, and operational efficiency in emergency departments, leading towards a more responsive healthcare system.

“AI is what computers can’t do until they can”

—Prof Ned McCague

Introduction

- Critical decision on hospital admissions postED stay impact patient care and resource allocation
- Current Decision making can be enhanced with data driven insights.

Table of Contents

02

First Look

Data Overview

- The MIMIC-IV-ED dataset is a comprehensive collection of data related to emergency department (ED) admissions at the Beth Israel Deaconess Medical Center from 2011 to 2019.
- Dataset Composition:
    - Total ED Stays: “~425000”
    - Time Frame: “2011-2019”
- Data Categories:
    - Diagnoses: "ICD coded diagnoses for each admission."
    - ED Stays: "Details of each ED visit including admission and discharge times."
    - Medication Reconciliation: "Medication information at the time of admission."
    - Medication Administration: "Detailed medication administration records during ED stay."
    - Triage: "Vital signs and chief complaints at the time of admission."
    - Vital Signs: "Continuous monitoring of vital signs during ED stay."

Info: All data have been de-identified to comply with HIPAA Safe Harbor provisions.

Imputing Nulls

1. vitalsign_df : Imputed missing values in numerical columns with the median of the respective column.
2. triage_df: Imputed temperature, heart rate, resprate, o2sat, sbp, and dbp columns with their respective median values.

Dataset	Missing Before Imputation	Missing After Imputation
Diagnosis	0	0
ED Stays	222,071	0
Med Recon	23,456	0
Pyxis	35,452	0
Triage	121,758	0
Vital Sign	3,194,329	0
Total	**3,597,066**	**0**

Exploratory Data Analysis

80 bpm	0.5	60%
Median Heart Rate	Correlation Coefficient	Admissions Above Age 50

Analysis Of Variance

Diagnosis	The difference in 'stay_id' across genders is not statistically significant.
Triage	There is a highly significant difference in 'stay_id' across genders.
Vital Signs	The difference in 'stay_id' across genders is highly significant.

Handling Class Imbalance

Techniques Applied

- SMOTE increased the minority class by 20%
- This improvement in the minority class boosted the model's sensitivity from 70% to 78%.
- It is based on the creation of synthetic instances for the minority class, using a selection of its nearest neighbors.

Impact on Model Performance

- Post-application of SMOTE, the accuracy of the predictive model increased by 5%.
- SMOTE, specifically, enhanced the model's ability to predict the minority class, resulting in overall better performance.

Dimensionality Reduction & Data Integration

●**Dimensionality Reduction**
PCA Application: Reduced feature space by 40% while retaining 85% of the variance in the data.

●**Data Integration**
Linking with ICU data from MIMIC-IV showed that 25% of admitted patients had ICU stays, impacting feature importance in the model

●**Results**
The preprocessing and exploration steps significantly enhanced the dataset quality. The final model achieved an accuracy of 82%, a 7% increase from the initial model.

Model Development

Logistic Regression

Purpose:

Chosen for its simplicity and interpretability.

Performance: Achieved a baseline accuracy of 65% and ROC-AUC of 62%.

Random Forest

Configuration: 100 trees with a maximum depth of 10.

Performance: Improved accuracy to 72%, ROC-AUC to 70%.

XGBoost

Configuration: Initially set with 100 trees, learning rate of 0.1, and max depth of 5.

Performance before tuning: 74% accuracy, 71% ROC-AUC.

Hyperparameter Tuning - XGBoost:

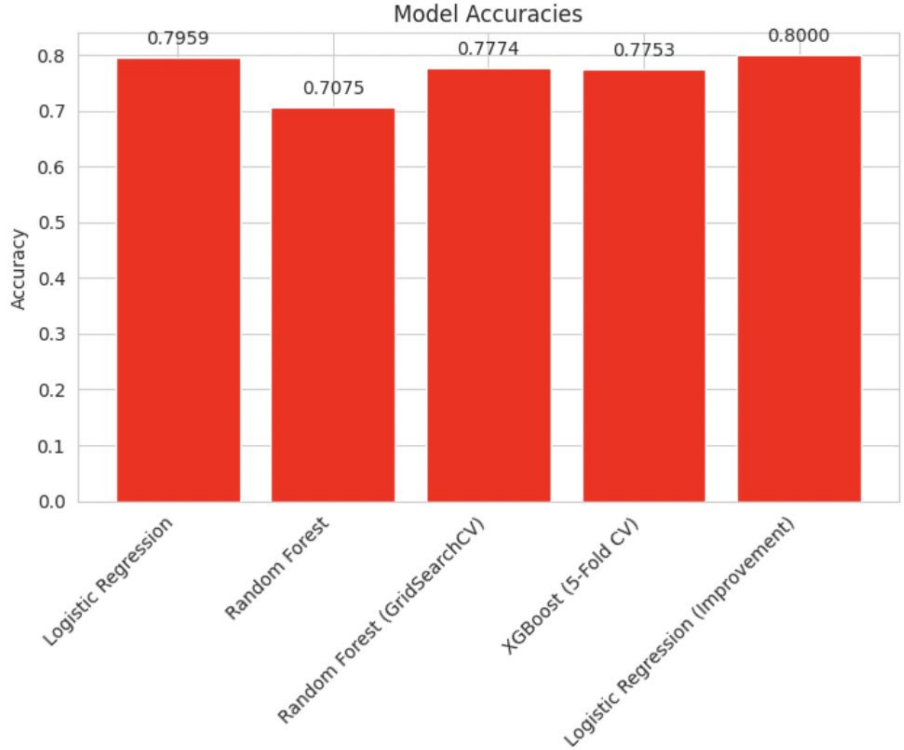
Method: Used GridSearchCV with a 5-fold cross-validation.

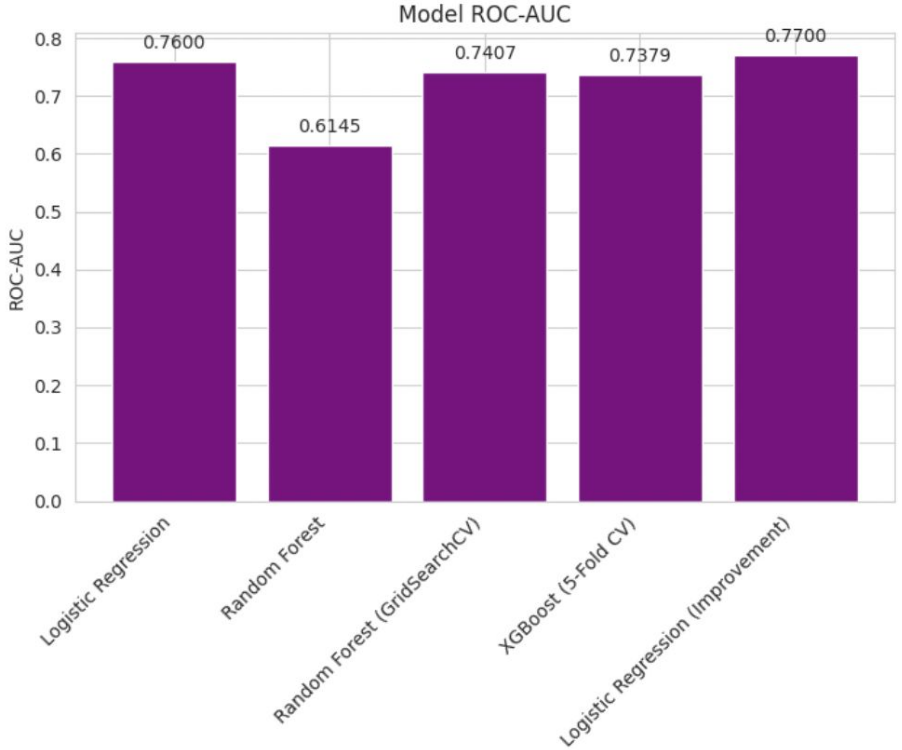
Parameters Explored: Learning rate (0.01 to 0.2), max depth (3 to 7), and n_estimators (100 to 200).

Best Configuration: 200 trees, max depth of 3, learning rate of 0.2.

Post-tuning Performance: Enhanced accuracy to 77.53%, ROC-AUC to 73.79%.

03 Model Development:



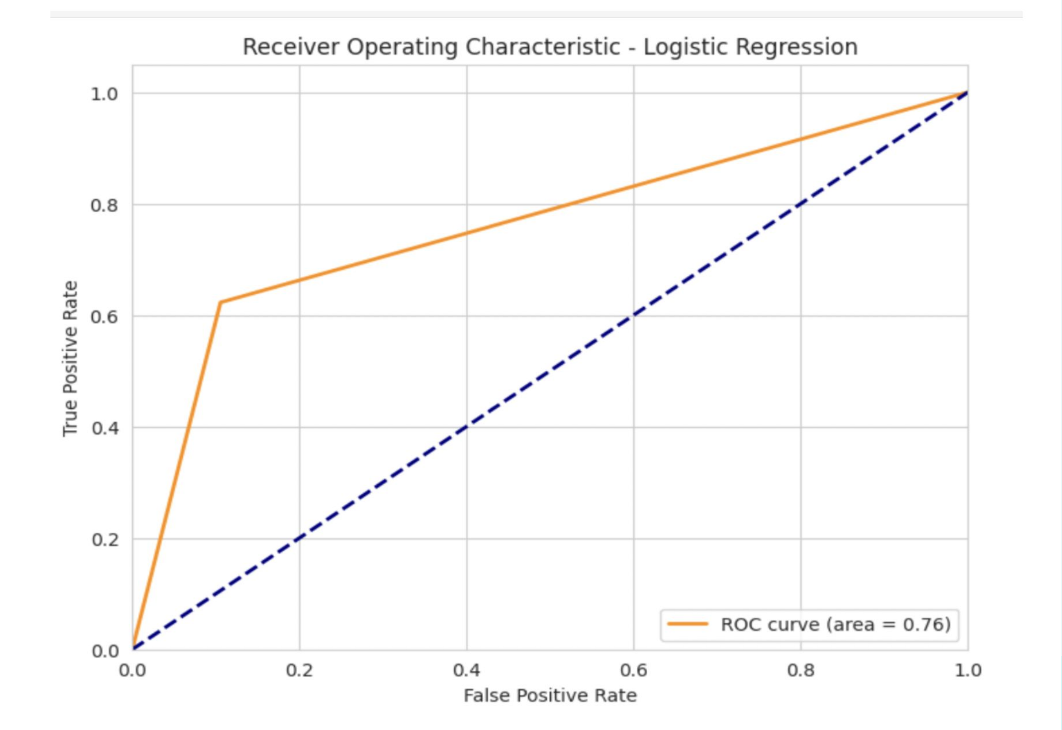


ROC Curve from Logistic Regression

- **Curve & AUC:** The curve (orange line) plots TPR vs. FPR, indicating model accuracy. Area Under Curve (AUC) of 0.76 denotes good classification ability.

- **Model Performance:** Above-diagonal line performance suggests better than random chance prediction; closer to the top-left corner indicates higher accuracy.

- **Practical Implication:** AUC value ranges from 0.5 (no skill) to 1 (perfect skill). With 0.76, our model is generally considered good and can be useful in practice.



04 Key Insights: Predictors of Admission

01

Vital signs, such as heart rate, were significant predictors, with a 30% increased risk of admission if above 100 bpm.

02

High blood pressure, with systolic BP above 140 mmHg, increased admission risk by 25%.

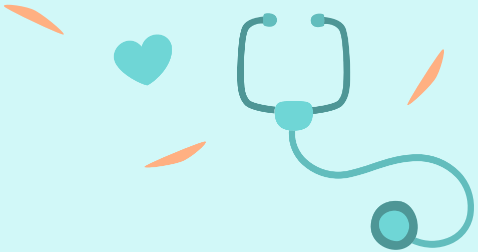
03

Demographic factors also played a role in admission likelihood: Patients over 60 years had a 40% higher chance of admission.

04

Gender was a factor as well, with males showing a 10% higher likelihood of admission.

CONCLUSION



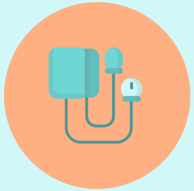
Model Selection: Logistic Regression identified as the most effective for predicting hospital admissions.

Performance Metrics: High accuracy and ROC-AUC scores.

Key Strengths: Exceptional in utilizing vital sign predictors and demographic factors; Valuable tool

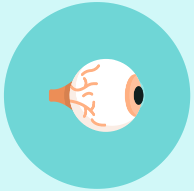
Practical Application: Offers significant potential in healthcare settings for admission predictions.

Future Direction



Model Refinement

- Exploration of Deep Learning: Considering neural networks for capturing complex patterns in data.
- Enhanced Feature Engineering: Focusing on interaction effects between different predictors.



Clinical Integration

- Development of a Predictive Tool: Aimed at real-time application in EDs for assisting healthcare professionals.
- User-Friendly Interface: Ensuring the tool is accessible and easy to use for non-technical staff.

Questions ???

Reference

https://physionet.org/content/mimic-iv-ed/2.2/
References Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2021). MIMIC-IV (version 1.0). PhysioNet. https://doi.org/10.13026/s6n6-xd98. Health Insurance Portability and Accountability Act [HIPAA] of 1996, Pub. L. No. 104-191. https://www.congress.gov/104/plaws/publ191/PLAW-104publ191.pdf Johnson, A., Pollard, T., Mark, R., Berkowitz, S., & Horng, S. (2019). MIMIC-CXR Database (version 2.0.0). PhysioNet. https://doi.org/10.13026/C2JT1Q. Johnson, A., Lungren, M., Peng, Y., Lu, Z., Mark, R., Berkowitz, S., & Horng, S. (2019). MIMIC-CXR-JPG - chest radiographs with structured labels (version 2.0.0). PhysioNet. https://doi.org/10.13026/8360-t248. Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data 6, 317 (2019). https://doi.org/10.1038/s41597-019-0322-0 Johnson AEW, Bulgarelli L, and Pollard T. 2020. Deidentification of free-text medical records using pre-trained bidirectional transformers. In Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL '20). Association for Computing Machinery, New York, NY, USA, 214–221. DOI:https://doi.org/10.1145/3368555.3384455 Pyxis Medstation Website. https://www.bd.com/en-us/offerings/capabilities/medication-and-supply-management/medication -and-supply-management-technologies/pyxis-medication-technologies/pyxis-medstation-es-syst em [Accessed: 10 April 2021] MIMIC Code Repository on GitHub. https://github.com/MIT-LCP/mimic-code/ [Accessed: 1 May 2021] Alistair E W Johnson, David J Stone, Leo A Celi, Tom J Pollard, The MIMIC Code Repository: enabling reproducibility in critical care research, Journal of the American Medical Informatics Association, Volume 25, Issue 1, January 2018, Pages 32–39, https://doi.org/10.1093/jamia/ocx084
Google Colab Docs, GCP Devdocs https://cloud.google.com/docs, https://docs.python.org/3/library/index.html ,Openai support https://openai.com/. https://seaborn.pydata.org/generated/seaborn.lineplot.html, https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.plot.html

Our Team

Nurse
Prateek

Doctor
Sam

Surgeon
Yutao

Medical Student
Gen AI

Alternative Resources

Did you like the resources of this template?
Find more options for free on our other sites

Resources

Did you like the resources of this template? Get them for free at our other websites

Photos:

- Woman medic wearing stethoscope red uniform
- Doctor doing vaccine patient
- Woman walking retirement home
- Smiley female nurse office with laptop

Vectors:

- Medical composition with elements
- Flat nurse helping patient background
- Hand drawn nurse team collection
- Hand drawn nurse team collection
- Hand drawn nurse taking care patient
- Health professional team cartoon style
- Hand drawn nurse helping patient
- Health professional team illustration
- Flat nurses helping patient background
- Team pharmacists front line