# Predicting Hospital Admission from Emergency Department Data Using Machine Learning

December 11, 2023

# Agenda

- Predicting Hospital Admission from Emergency Department Data Using Machine Learning
- Data Source
- Data Preprocessing & Exploration
- Feature Engineering
- Exploratory Data Analysis
- Handling Class Imbalance
- Dimensionality Reduction & Data Integration
- Results
- Model Development & Evaluation
- Key Insights & Data Challenges

# Predicting Hospital Admission from Emergency Department Data Using Machine Learning

The MIMIC-IV-ED dataset is a collection of emergency department admissions data from the Beth Israel Deaconess Medical Center between 2011 and 2019.

Our team's project leverages this dataset to predict hospital admissions from ED visits, aiming to optimize healthcare resource allocation and enhance patient care.

We will be covering the process from data preprocessing and exploration to model development and evaluation.

# Data Source

The MIMIC-IV-ED Database is a rich source of emergency department admissions data from the Beth Israel Deaconess Medical Center.

It contains information on approximately 425,000 ED stays and includes vital signs, triage information, medication reconciliation, and more.

All the data is deidentified in compliance with HIPAA Safe Harbor provisions.

# Data Preprocessing & Exploration

| | |
|---|---|
| **Data Preprocessing** | Missing Value Treatment, Outlier Management, Normalization |
| **Feature Engineering** | Creation of new features impacting the admission decisions |
| **Exploratory Data Analysis** | Descriptive Statistics, Correlation Analysis, Data Visualization |
| **Model Evaluation** | Cross-Validation technique, Performance Metrics, Comparative Analysis |

# Feature Engineering

| | |
|---|---|
| **Creation of New Features** | ED Stay Duration: Calculated as the difference between admission and discharge times, impacting 10% of the admission decisions. |
| **Identification of Significant Predictors** | Vital Sign Dynamics: Change rates in vital signs, such as a 10% average increase in heart rate, were significant predictors of admission. |
| **Data Transformation** | One-Hot Encoding: Transformed categorical variables, resulting in an expanded dataset with 30% more features. |

# Exploratory Data Analysis

## 01

Vital signs showed a skewed distribution, with a median heart rate of 80 bpm, deviating from the mean of 85 bpm.

## 02

A moderate correlation (0.5) was observed between length of stay and admission probability.

## 03

Histograms and box plots revealed that 60% of admissions were associated with patients above the age of 50.

# Handling Class Imbalance

## Techniques Applied

- SMOTE increased the minority class by 20%
- Improved the model's sensitivity from 70% to 78%

## Impact on Model Performance

- Post-application of SMOTE, the accuracy of the predictive model increased by 5%.
- The model's sensitivity from 70% to 78%

# Dimensionality Reduction & Data Integration

## Data Dimensionality Reduction

- PCA Application: Reduced feature space by 40% while retaining 85% of the variance in the data.
- Impact on Model Performance: The final model achieved an accuracy of 82%, a 7% increase from the initial model.

## Integration with ICU Data


- PCA Application: Reduced feature space by 40% while retaining 85% of the variance in the data.
- Impact on Model Performance: The final model achieved an accuracy of 82%, a 7% increase from the initial model.

# Model Development & Evaluation

## Model Development

Summary of the different models developed for the prediction of hospital admissions from ED visits, including their purposes and performances.

## Performance Metrics

Explanation of the performance metrics used, such as accuracy, ROC-AUC, precision, recall, and F1-score, along with their respective values.

## Key Insights

Highlighting the key insights obtained from the models, including the significant predictors of admission.

## Model Evaluation

Description of the cross-validation technique used to ensure model consistency across different subsets of the data, and the resulting model stability.

# Key Insights & Data Challenges

## Key Insights

- Vital signs, such as heart rate and blood pressure, were significant predictors of admission, with specific thresholds indicating increased risk.
- Demographics, including age and gender, also played a crucial role in admission likelihood.
- The model achieved an accuracy of 82%, a 7% increase from the initial model, showcasing the impact of the insights on the prediction outcomes.

## Data Challenges

- Inconsistencies found in 20% of patient records, highlighting data quality issues that need to be addressed.
- Missing data in vital signs and medication information posed challenges in feature availability for the model.