

Analysis of ICU Mortality Based on Diagnoses and Lab Measurements

Abstract

This project analyzes demographic, clinical, and laboratory data from the MIMIC-III Clinical Database to identify factors influencing ICU mortality and length of stay (LOS). By examining patient age, gender, diagnoses, and lab results, we aim to provide insights that can optimize resource allocation and patient care protocols.

1. Introduction

The utilization of Intensive Care Units (ICUs) is critical for patient outcomes but resource-intensive. Understanding the key drivers of prolonged ICU stays and mortality is essential for hospital management. This study leverages the MIMIC-III dataset, a large, freely-available database comprising de-identified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.

2. Methodology

2.1 Data Source

The study utilizes the MIMIC-III Clinical Database Demo (v1.4), specifically the `PATIENTS`, `ADMISSIONS`, `ICUSTAYS`, `DIAGNOSES_ICD`, and `LABEVENTS` tables.

2.2 Data Preprocessing

- **Age Calculation:** Calculated from `D0B` and `ADMITTIME`. Patients aged >89 (shifted to 300) were capped at 90 for analysis.
- **LOS Calculation:** Derived from `INTIME` and `OUTTIME` in `ICUSTAYS`, aggregated per admission.
- **Categorization:** Patients were categorized into age groups (0-18, 19-40, 41-60, 61-80, 80+) and LOS categories (Short <2 days, Medium 2-7 days, Long >7 days).
- **Abnormalities:** Lab values were flagged based on the `FLAG` column.

2.3 Tools Used

- **Python (Pandas, NumPy)** for data cleaning and manipulation.
- **Matplotlib/Seaborn** for exploratory data analysis (EDA).
- **SQL** for data extraction and joining.
- **Tableau** (Designed) for interactive dashboard visualization.

3. Data Analysis

3.1 Exploratory Data Analysis (EDA)

- **Age Distribution:** The patient population is skewed towards the elderly (mean age approx. 70 years).
- **LOS Distribution:** The majority of stays are short (median < 3 days), with a long tail of complex cases.
- **Diagnoses:** The most frequent admitting diagnosis in the demo dataset is Sepsis, followed by Pneumonia.

3.2 Correlation Analysis

A negative correlation (-0.28) was observed between Age and LOS, suggesting that duration of stay might decrease slightly as age increases in this specific sample, possibly due to mortality or discharge practices.

4. Results

- **Key Finding 1:** Sepsis and Pneumonia account for the largest proportion of ICU admissions in the sample.
- **Key Finding 2:** Older age is associated with shorter ICU stays in this cohort.
- **Key Finding 3:** Lab values such as elevated Glucose and Creatinine are prevalent in Sepsis cases, indicating metabolic stress and potential kidney injury.

5. Conclusion

The analysis confirms the significant burden of infectious diseases (Sepsis, Pneumonia) on ICU resources. The demographic profile skew suggests the need for specialized geriatric care protocols. Further research with the full dataset is recommended to validate the age-LOS relationship and explore mortality predictors in greater detail.

6. Future Work

- Expand the analysis to the full MIMIC-III dataset for statistical significance.
- Implement machine learning models (e.g., Logistic Regression, Random Forest) to predict mortality risk based on admission variables.
- Conduct survival analysis (Kaplan-Meier curves) for different diagnosis groups.