Your Paper

You

May 18, 2025

Abstract

Your abstract.

1 Introduction

- source: 10.3389/fdgth.2024.1455446

One of the factors that affects the quality and cost of care is patient flow, which refers to the efficient use of resources and time during the patient's stay in the hospital (9, 10). A common practice for planning the discharge of patients is to rely on the clinicians' estimates of the discharge date. However, this practice has some drawbacks, such as consuming the clinicians' time that could be spent on other tasks or direct patient care, and having low accuracy (11, 12). A possible alternative is to apply machine learning models to predict the discharge date of patients based on their length of stay (LOS) in the hospital. Several studies have proposed and developed different LOS models for this purpose (11, 13–18).

1.1 Context

background information to set the stage for the research

Iezzoni et al., Med Care, 1988 – severity adjustment for LOS and discharge status in Medicare patients.

Severity matters—but it doesn't tell the whole story. Even with detailed physiologic data, over 80% of LOS variation remained unexplained. Hospitals can likely reduce pneumonia LOS through care-process changes (e.g., standardized clinical pathways, early mobilization, streamlined discharge). For policymakers, simple claims-based risk adjustment is inadequate when using LOS as a quality or efficiency metric.

Across 11 different severity measures—ranging from claims-based comorbidity counts to physiology-based scores—only 10-15% of the variation in individual patients' length of stay (LOS) was explained (trimmed data, R^2 0.10–0.15).

1.2 State of the art

review of the current research, findings, and technologies in the field

---

1.3 Research gap

Identifies gaps in existing research that the project aims to address

---

- source: 10.3389/fdgth.2024.1455446

The majority of published articles in healthcare machine learning have failed to showcase successful implementations of proposed models, resulting in a substantial gap between theoretical concepts and practical applications. This underscores the pressing need for practical research that addresses the entire life cycle of predictive models, spanning from their initial conceptualization to their effective integration into operational systems (19, 20).

1.4 Study objective, scope and deliverables

Clearly define what the research intends to accomplish, its limitations, and what you will deliver (e.g., findings, models, recommendations).

2 Methodology

research design, methods, and approaches used in the study

Methodology This study followed a structured data science workflow, consisting of data cleaning, merging, exploratory analysis, and preparation for modeling. The methodology is summarized as follows:

1. Laboratory Data Cleaning and Exploration The laboratory dataset, spanning approximately 16 years, was first loaded and inspected $(1_{d}ata_{c}leaning_{e}da_{l}ab.ipynb).Datacleaningstepsincluded:$

Renaming columns for clarity and consistency. Handling missing values, especially in the numeric$_{r}esultandtext_{r}esultcol$

Standardizing column names and formats. Handling missing or inconsistent entries. Removing duplicates and ensuring data integrity. Extracting relevant features for downstream analysis. 3. Data Merging The cleaned laboratory and clinical datasets were merged on patient and case identifiers $(3_{m}erge_{d}ata.ipynb):$

Only cases with sufficient laboratory and clinical information were retained. A reference table of laboratory tests was created to facilitate feature selection. The merged dataset was saved for further analysis. 4. Outlier Analysis Outlier detection and handling were performed to improve data quality $(4_{o}utlier_{a}nalysis.ipynb):$

Statistical methods and visualizations were used to identify extreme values in key variables. Outliers were either removed or capped, depending on their nature and impact. 5. Exploratory Data Analysis (EDA) of Merged Data Comprehensive EDA was conducted on the merged dataset $(5_{e}da_{m}erged_{d}ata.ipynb):$

Descriptive statistics and visualizations (e.g., boxplots, barplots) were used to summarize patient demographics, laboratory results, and clinical outcomes. The distribution of length of stay (LOS) and discharge types was analyzed. Relationships between variables were explored to inform feature engineering and modeling. This methodology ensured a robust and reproducible pipeline for preparing the data for predictive modeling of hospital length of stay and discharge type. Each step was documented in the corresponding Jupyter notebook for transparency and reproducibility.

3 Evaluation

how I will assess or measure the effectiveness or success of the research

3.1 Evaluation Methodology

3.2 Results

4 Conclusions

References