Comprehensive Hospital Outcome Prediction System (CHOP)

PITCH

This project aims to predict facets of a patient's admission such as length of stay and cost. It is our vision that armed with this information, a hospital can better plan and allocate its resources. When a patient is admitted, the system will be able to predict the stay length and green or red light the admission based on resources needed.

Business Understanding

Overview

The CHOP Model aims to predict patient outcomes and optimize resource allocation in healthcare settings. By providing accurate predictions regarding patient stays, CHOP will serve as a decision-support tool for healthcare administrators.

Problem Statement

Hospitals struggle to accurately forecast patient recovery times, potential complications, and necessary resources such as staff and beds. This uncertainty can result in either ***over-preparation*** (wasted resources), or **under-preparation**(poor service).

Proposed Solution

Our solution is to develop a model that given patient information, can predict aspects of their stay including duration and cost. By having a better picture of patient needs, the hospital can more efficiently plan for allocation of resources such as beds and staff.

The system will be able to monitor inpatient numbers to flag if available resources will meet demand in the near future. As an example, with projected number of patients in the coming days, if the system finds that staff on duty on a specific date falls short of demand, it will make this known to administrators.

Objectives

- **Dummy Hospital:** A fake hospital complete with room, patient and staff records will be created.

- **Data preparation:** The data is split into dozens of CSV files. Information in the relevant files will be extracted and joined in a way that is usable for model training.

- **Develop a Predictive Model:** Create a model capable of accurately predicting patient outcomes, including recovery time and likelihood of complications after the doctor has logged their diagnosis in the system.

- ***Model Accuracy:*** 80% has been set as the baseline for a successful model.

- **Hospital Management System:** A hospital management suite will be created featuring patient and staff management.

- **Deploy Model:** The model will be deployed to the management software.

- **Optimize Resource Allocation:** Use the model's predictions to optimize the allocation of hospital resources, such as staff, equipment, and beds, ensuring they are used efficiently and effectively.

Metrics of Success

1. **Model Accuracy**: Achieving high accuracy in predicting patient outcomes is imperative.
2. **Compliance and Security**: Adherence to ethical standards and data privacy regulations.
3. **User Acceptance**: Positive feedback from hospital administration regarding the model's usability and insights.

Challenges

- **Data Availability and Quality:** Obtaining comprehensive, high-quality patient data that covers a wide range of variables is essential.

- **Complexity of Healthcare Data:** Medical data is complex. It includes medical records, lab results, and physician notes among others. Cleaning the data into a usable format will be a big undertaking.

- **Privacy and Security Concerns:** Handling sensitive patient data requires strict adherence to privacy regulations.

Data Understanding

Data Sources

The **MIMIC-IV** Clinical Database provides comprehensive clinical information on patients admitted to Beth Israel Deaconess Medical Center. The database contains de-identified patient data in compliance with HIPAA standards. Several subsets of the data have been pinpointed for use.

This dataset is a relational database where each table ha been stored as a CSV.

Relevance of Data

The selected datasets collectively provide a holistic view of patient information, covering demographics, clinical diagnoses, lab results, and treatment services.

Tables Utilized

For the Comprehensive Patient Outcome and Resource Prediction (CPORP) Model, we have selected specific datasets from the MIMIC-IV database. These datasets provide a rich set of features relevant to predicting patient outcomes and optimizing healthcare resource allocation. Below is a description of each dataset and its relevance to the project:

Admissions.csv

Description: This dataset contains information about patient admissions, including admission and discharge times, admission type (e.g., emergency, elective), and hospital stay details.

Patients.csv

Description: Includes demographic information such as patient identifiers, birth dates, gender, and death indicators.

OMR.csv

(Output Measurement Records): Provides longitudinal patient data, including measurements such as height, weight, BMI, and blood pressure, recorded over time to track health trends and assess treatment effectiveness.

Diagnoses_ICD.csv

Description: Lists International Classification of Diseases (ICD) diagnosis codes assigned to patients during their hospital stay.

Services.csv

Description: Provides information about clinical services and procedures that patients received during their hospital stay, including billing-related data.

Labevents.csv

Description: Contains laboratory test results for patients, including test identifiers, results, timestamps, and units of measurement.