PREDICTING PATIENT READMISSIONS IN HEALTHCARE

BUS 9430 - Business Analytics Project Management

TEAM MediForecast Consultants -
Rupa Poddar, Sahithi Arnika Modadugu,Sujasna Tamang,
Mustafa Ekinci.

Data Source & Information

Data Source - Health Facts database -Cerner Corporation

- Dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals.
- All the data is from inpatient encounters
- Diabetic encounter data, where any kind of diabetes was entered to the system as a diagnosis.
- 55 attributes & 100,000 instances
- Attributes: Patient Number, Race, Gender, Age, Admission Type, Time in Hospital, Medical Specialty of Admitting Physician, Number of Lab Tests Performed, HbA1c Test Result, Diagnosis, Number of Medications, Diabetic Medications, Number of Outpatient Visits, Number of Inpatient Visits, Number of Emergency Visits in the Year Before the Hospitalization.

Data Cleaning

- Attributes removed : >40% missing values

  - Weight, Payer Code and Medical Specialty

- Outliers removed using interquartile range method

  - Number of medications, Number of outpatient visits, Number of inpatient visits, Number of emergency visits

- Changed attributes to binary variables

  Readmission
  0: Indicates that the patient was not readmitted to the hospital after discharge.
  1: Indicates that the patient was readmitted to the hospital within 30 days or after 30 days of discharge.

  Medication Change, Diabetic Medication
  0: If there is no change in medication / If the patient is not taking diabetic medication
  1: If there is a change in medication / If the patient is taking diabetic medication

Research Questions

- Predictive modeling for readmission risk assessment: Can we accurately predict patient readmission based on attributes such as medical history, admission details, and treatment plans? How effective are predictive modeling techniques (e.g., machine learning algorithms) in forecasting readmission risk?

- Segmentation of patients based on risk factors: Can we segment patients into different risk groups based on their medical profiles and demographic characteristics? Are there distinct patterns or clusters of patients with higher or lower readmission risks?

Feature Selection For Predicting Patient Readmission

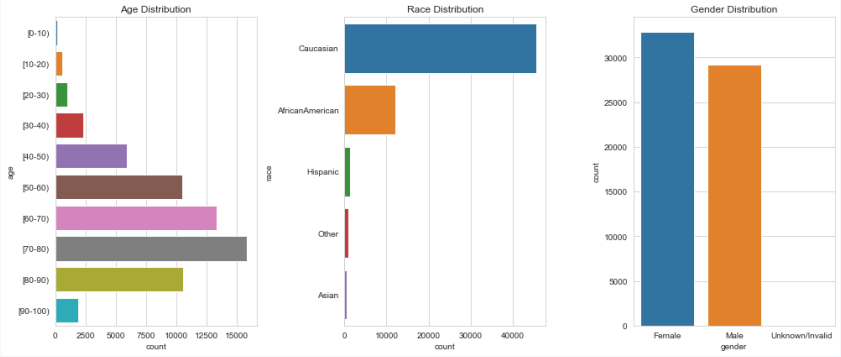
For Predictive Modeling:

- Evaluate impact of healthcare encounters (outpatient, emergency visits, inpatient admission) and medication usage (e.g., metformin, insulin) on readmission risk.

For Patient Segmentation:

- Consider demographic factors (age, gender, race) alongside healthcare encounters and medication usage to identify distinct patterns or clusters associated with readmission risk.

Exploratory Data Analysis

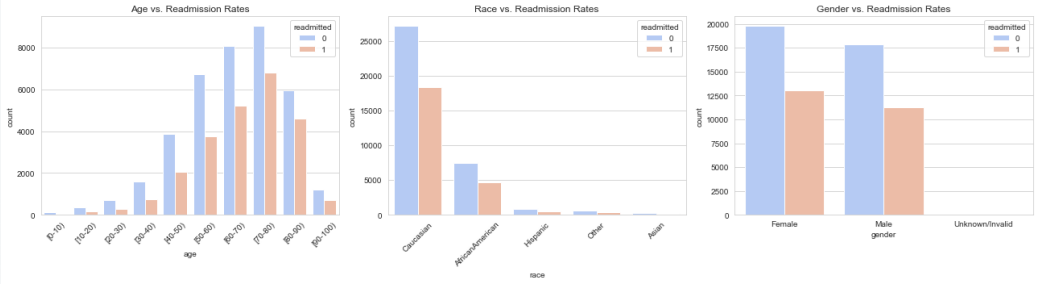


The Distribution of Key Demographic Variables

Age Distribution: The age distribution of the patient population.

Race Distribution: The distribution of patients by race.

Gender Distribution: The distribution of patients by gender.



The Relationship between Demographic Variables and Readmission Rates

Age vs. Readmission Rates: How readmission rates vary across different age groups?

Race vs. Readmission Rates: How patients of different racial backgrounds experience readmission?

Gender vs. Readmission Rates: The comparison of readmission rates between genders.

Exploratory Data Analysis

Variable	Number of encounters	% of the population	Readmitted	
			Number of encounters	% in group
HbA1c				
No test was performed	57,080	81.6%	5,342	9.4%
Result was high and the diabetic medication was changed	4,071	5.8%	361	8.9%
Result was high but the diabetic medication was not changed	2,196	3.1%	166	7.6%
Normal result of the test	6,637	9.5%	590	8.9%

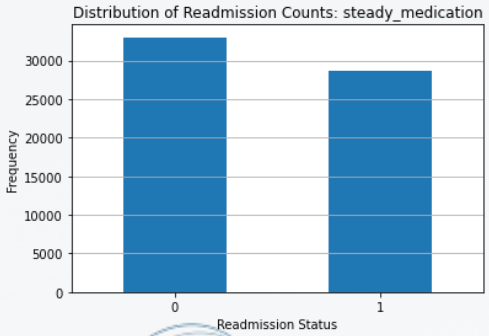
HbA1c (Blood test used to diagnose type 2 diabetes)
Test Results and Changes in Diabetic Medication

No HbA1c Test Performed
81.6% of the population, did not have an HbA1c test, 9.4% were readmitted.

High HbA1c Result with Medication Change
5.8% of the population, had high HbA1c results, 8.9% readmission rate.

High HbA1c Result with No Medication Change
3.1% of patients, had high HbA1c results, 7.6% readmission rate.

Normal HbA1c Results
9.5% of the patient, had normal HbA1c tests, 8.9% readmission rate.



The Frequencies of 'Readmitted' and 'Not Readmitted' Patients who had their medication dosages unchanged

Strengths

- **Comprehensiveness**: The dataset appears to include a wide range of variables, from basic demographics to detailed medical and medication histories, which are essential for an in-depth analysis of patient readmissions
- **Granularity**: Data such as the number of lab procedures, diagnoses, and medications indicate a high level of detail, which could allow for refined predictive modeling
- **Low Missing Data**: Several key features, such as encounter ID, patient number, and some health indicators, have a 0% missing rate, which suggests the data is relatively complete

Limitations

- **High Percentage of Missing Data for Certain Variables**: Some variables, such as payer code and medical specialty, have over 50% missing data, which could limit the analysis for these aspects
- **Data Skewness**: With a majority of the population not undergoing HbA1c tests, the dataset might be skewed towards patients with less monitoring, which could affect the model's predictions

Ethical Consideration

- **Confidentiality**: Given the sensitive nature of the clinical data, including HbA1c test results, admission types, and diagnoses, crucial to remove patient identifiers

- **Consent for Data Use**: Obtain consent from all patients regarding the use of their medical data

- **Compliance**: All data handling and analysis must adhere to healthcare regulations, such as HIPAA in US. Includes secure data storage, proper data encryption, and access control to protect information from unauthorized user

THANK YOU!