Diabetes Readmission Analysis

Abdulla Elkhadrawy, Afm Khan, Theodore Jeliazkov, Daniel Lines, Andrew Qian, Rahul Ramarao

Introduction

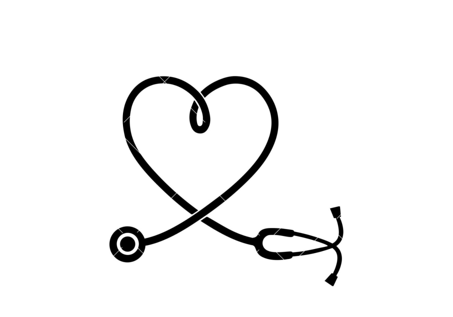
Overview

- Dataset covers 10 years (1999-2008) of clinical records from 130 US hospitals
- Contains 101,766 instances of hospitalized patients diagnosed with diabetes
- 47 features including patient details, hospital metrics, lab tests, medications

Introduction

Importance

- Evaluating readmission risk
- Improving diabetes management and glycemic control in hospitals
- Reducing healthcare costs and complications from suboptimal care



Goal

Develop a predictive model to identify high-risk patients for better diabetes management during hospital stays.

A1C Testing

The A1C blood test is used to diagnose diabetes and monitor the efficacy of diabetes treatment. It provides a measure of average blood glucose levels over the past 3 months.

- Lower than 5.7%: normal
- Between 5.7% and 6.5%: prediabetes
- Greater than 7%: elevated to high risk

A1C and Estimated Average Glucose Levels

	A1C Percentage	Estimated Average Glucose (EAG)	
In-range	< 5.7%	< 117 mg/dL	6.5 mmol/L
Prediabetes	5.7-6.4%	117-137 mg/dL	6.5-7.6 mmol/L
Diabetes	> 6.4%	> 137 mg/dL	> 7.6 mmol/L
Increased risk of complications	6.5%	140 mg/dL	7.8 mmol/L
	7.0%	154 mg/dL	8.6 mmol/L
	7.5%	169 mg/dL	9.4 mmol/L
	8.0%	183 mg/dL	10.1 mmol/L
	8.5%	197 mg/dL	10.9 mmol/L
	9.0%	212 mg/dL	11.8 mmol/L
	9.5%	226 mg/dL	12.6 mmol/L
	10%	240 mg/dL	13.4 mmol/L

Data Focus

Selected for individuals labelled with A1C test results over 7 and over 8, emphasizing patients at high risk of complications and readmission

Clinical Significance

Variables

- Insulin
- Age Group
- Weight

Variable Relationships

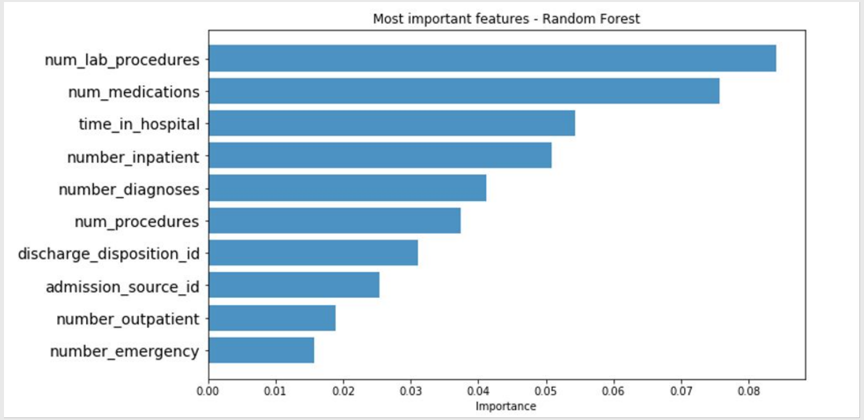
- Higher Age Groups → Weight Gain
- Excess body weight → Type 2 Diabetes
- Excess body fat → lower insulin sensitivity

patient_nbr	race	gender	age	**weight**
80041266	Caucasian	Female	[80-90)	[50-75)
103586670	Caucasian	Male	[70-80)	[100-125)
67852377	?	Male	[50-60)	[75-100)
81133947	Caucasian	Male	[50-60)	[75-100)
107804349	Caucasian	Male	[70-80)	[50-75)
24514578	?	Female	[30-40)	[50-75)
74612331	Caucasian	Female	[70-80)	[75-100)
5236596	Caucasian	Male	[0-10)	[0-25)
84952215	Caucasian	Male	[60-70)	[75-100)
82870227	Caucasian	Female	[50-60)	[75-100)
103008159	Caucasian	Male	[30-40)	[75-100)
107536383	Caucasian	Male	[70-80)	[75-100)
85220019	Caucasian	Female	[60-70)	[0-25)
21053511	Caucasian	Male	[60-70)	[25-50)
76773537	Caucasian	Female	[80-90)	[50-75)
47503989	Caucasian	Female	[50-60)	[150-175)
91560816	Caucasian	Male	[70-80)	[0-25)
5884920	Caucasian	Female	[10-20)	[50-75)
110849184	Caucasian	Female	[80-90)	[25-50)
28481058	Caucasian	Male	[60-70)	[75-100)
78657957	Caucasian	Male	[50-60)	[100-125)
10975032	Caucasian	Female	[50-60)	[100-125)
6363873	Caucasian	Female	[70-80)	[50-75)
88846479	AfricanAmerican	Female	[40-50)	[50-75)
65086533	Caucasian	Male	[50-60)	[50-75)
61504731	Caucasian	Male	[70-80)	[50-75)

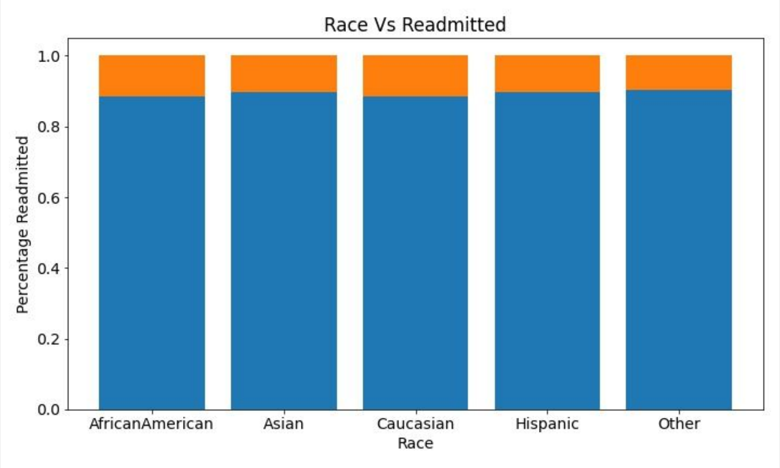
Clinical Significance

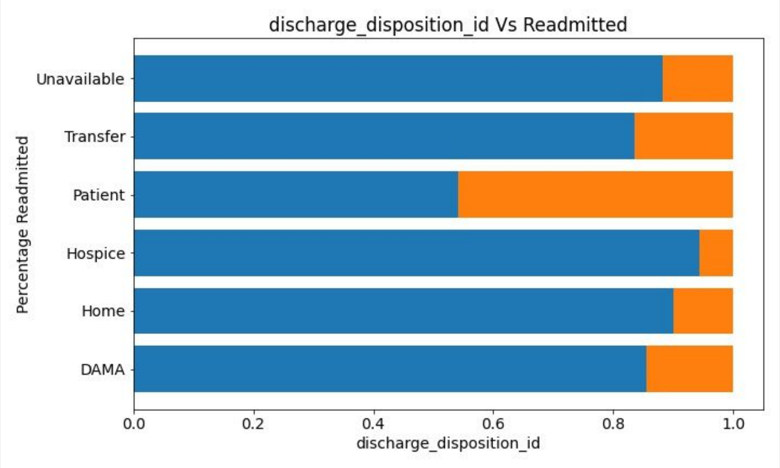
Number of Lab Procedures

- Positive Correlation: Patients have more severe diabetes and need extra care and face increased risk of readmission

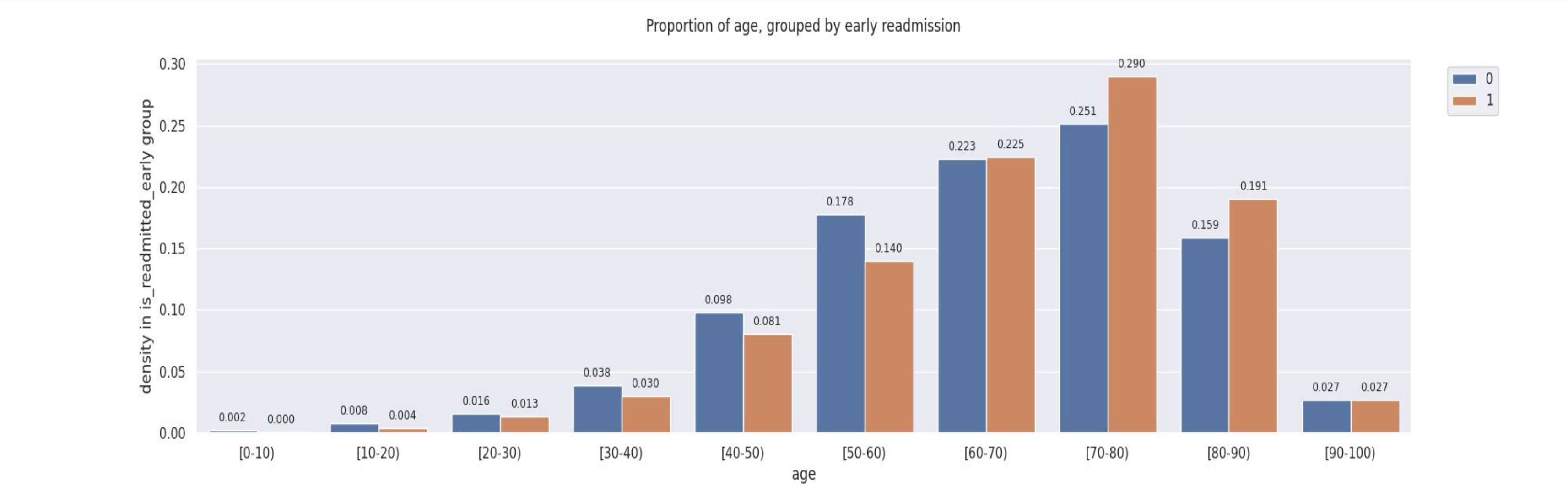


Exploratory Data Analysis

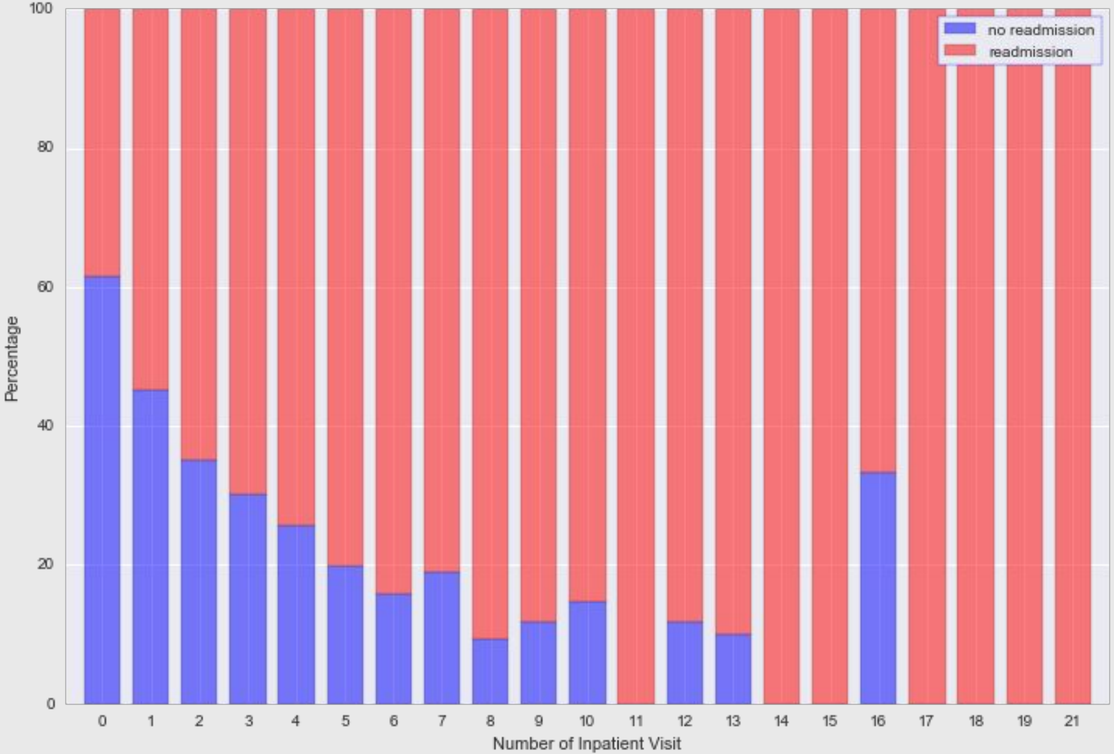




Exploratory Data Analysis



Exploratory Data Analysis



Logistic Regression: Data Preprocessing

-Not all data is quantitative
- Must encode categorical variables with integer values

Ordinal Encoding

Grades
A
B
C
D
Fail

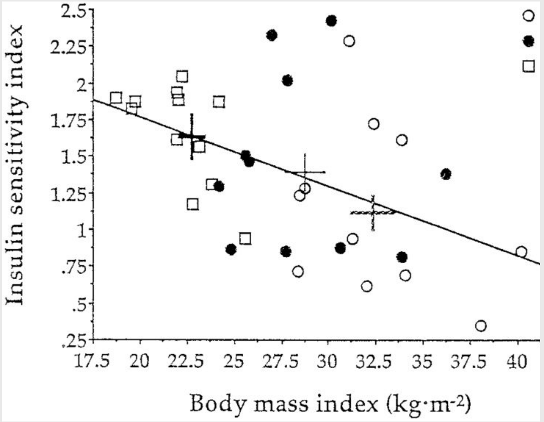
Grades	Encoded
A	4
B	3
C	2
D	1
Fail	0

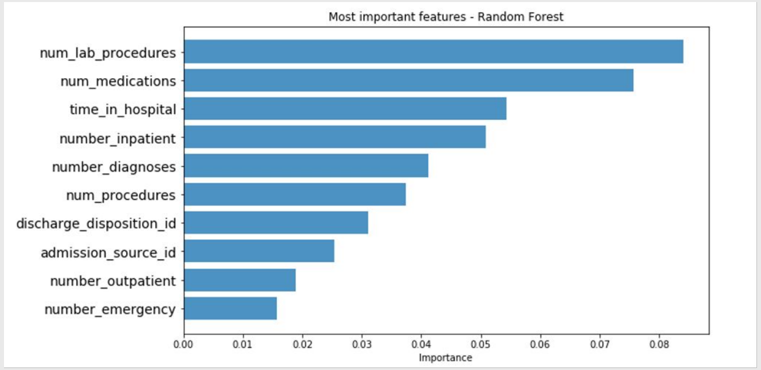
Logistic Regression + Interactions

- **Binary outcomes: situations where the response variable will be 0 or 1 (false or true)**
  - In logistic regression, binary outcomes are modeled using the logistic function
- **Interactions between highly correlated variables**
  - Interactions occur when the effect of one predictor variable on the outcome depends on the value of another predictor variable
- **Logistic Regression with Interactions**
  - This gives a better picture of how variables work together to predict outcomes, but it needs careful thinking, testing, and understanding because it can get a bit tricky to interpret the results.

Model Training

- N = 288 patients (test on 20%)
- Using feature importance
- Using previous clinical knowledge to build model
  - Prevents superficial assumptions of your data
  - Prevents overfitting to data that might not be the highest quality



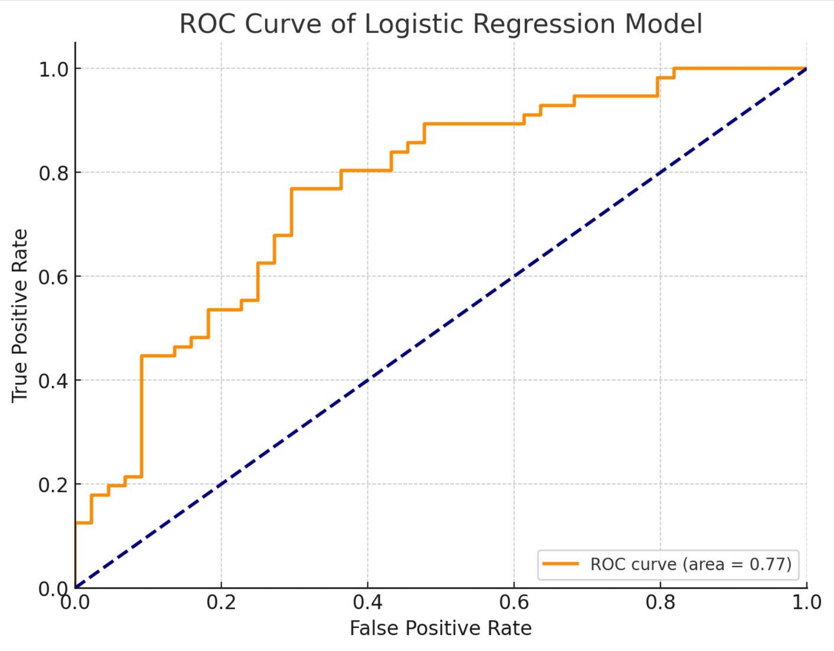


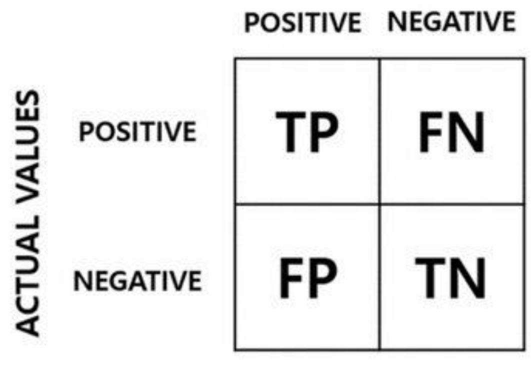
Model Formula

Logit(Readmission)= β0 + β1***Age** + β2***Weight** + β3***Insulin** + β4***Number_of_Labs** + β5***Time_in_Hospital** + β6***Discharge_ID** + β11***(Age*Insulin)**+ β12***(Insulin*Weight)**+ β12***(Age*Weight)**+ β7***Number In Patient Visit**

Model Performance

AUC = 0.77
Accuracy = 0.74
F1 Score = 0.77
Precision = 0.77
Recall = 0.77





$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Next Steps

- Utilize improved data for building stronger, more reliable models.
- Highlight the potential for physicians to incorporate these models as supplementary tools in patient care.
- Consider prolonging hospital stays and conducting additional tests for patients flagged by the model as having a high likelihood of readmission.