

PREDICTION OF DIABETES READMISSIONS

Aisha Kamara, Data Scientist
October 4th, 2021

PRESENTATION OVERVIEW

01

Introduction

Brief background on Diabetes and readmission rates in the US.

02

Problem Statement

Problem statement introduction.

03

Data Cleaning

Cleaning of the dataset

04

EDA Findings

EDA results

05

Modeling Results

Models used and results

06

Conclusion

Findings and recommendations

INTRODUCTION



3.3 Million

Readmitted Patients



\$41 Billion

Hospital Costs

DIABETES

Diabetes is a heterogeneous group of diseases that, through various mechanisms, cause hyperglycemia, commonly known as high blood sugar, a buildup of glucose in a person's bloodstream at dangerously high levels due to a person's body not being able to produce enough insulin in order to regulate glucose within the bloodstream.



PROBLEM STATEMENT

DATASET OVERVIEW

- The data obtained represents 100,000+ unique inpatient diabetes medical visits over 10 years (1999–2008) of clinical care at 130 hospitals and integrated delivery networks in the United States.
- The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, diagnosis, number of medications, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

PROBLEM STATEMENT

Problem 1

What factors are the strongest indicators of hospital readmission for a diabetic patient?

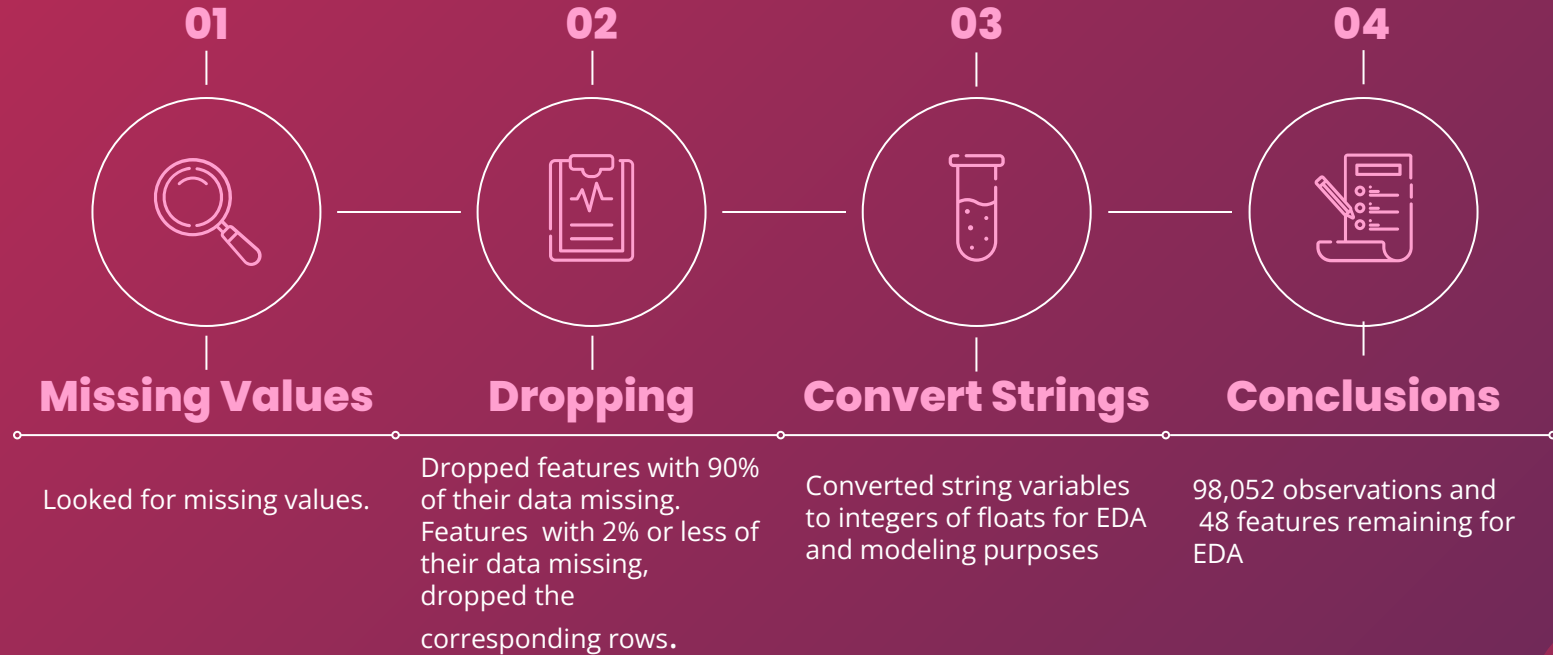
Problem 2

How well can I predict hospital readmission with "limited" features in this dataset?

DATA CLEANING

The background is a solid dark pink color. It is decorated with various geometric elements: several white-outlined circles of different sizes, some solid red circles, and thin white line segments. These elements are scattered across the frame, creating a modern, abstract aesthetic.

CLEANING PROCESS

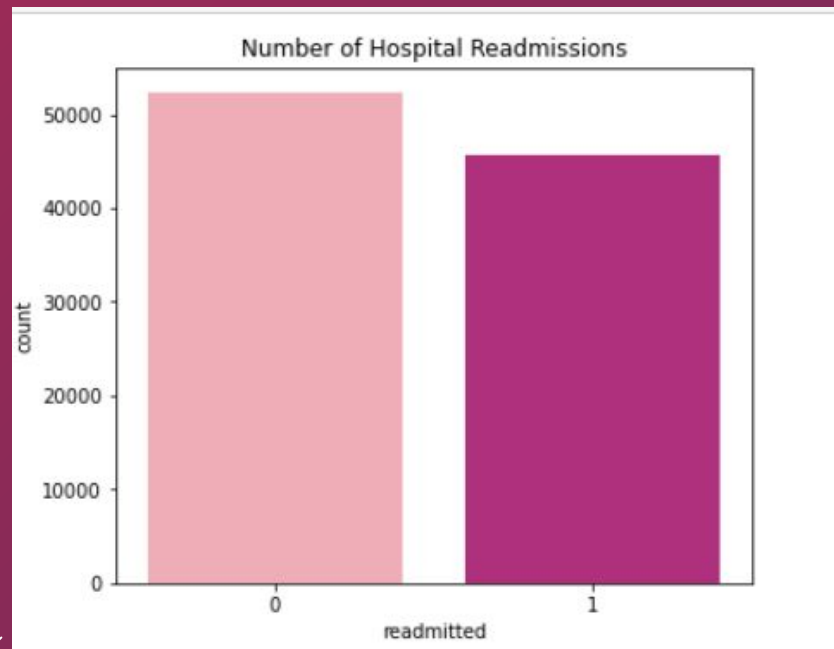


The background is a solid dark red color. It is decorated with various white geometric elements: thin circles of different sizes, some solid red circles, and several short, parallel white line segments. These elements are scattered across the frame, creating a modern, abstract aesthetic.

EDA

READMITTANCE RATES

- 98,052 Observations
- 47 Features
- 1 Target Feature
- 54% of patient had no record of readmission
- 46% of patients, were readmitted after discharge



Categorical Features

The background is a solid dark red color. It is decorated with various white geometric elements: thin circles of different sizes, some solid red circles, and several short, parallel white line segments. These elements are scattered across the frame, creating a modern, abstract aesthetic.

DEMOGRAPHICS



48%

Age

48 % of patients readmitted are between 60-80 years of age.



53%

Gender

53 % of readmitted patients are females.

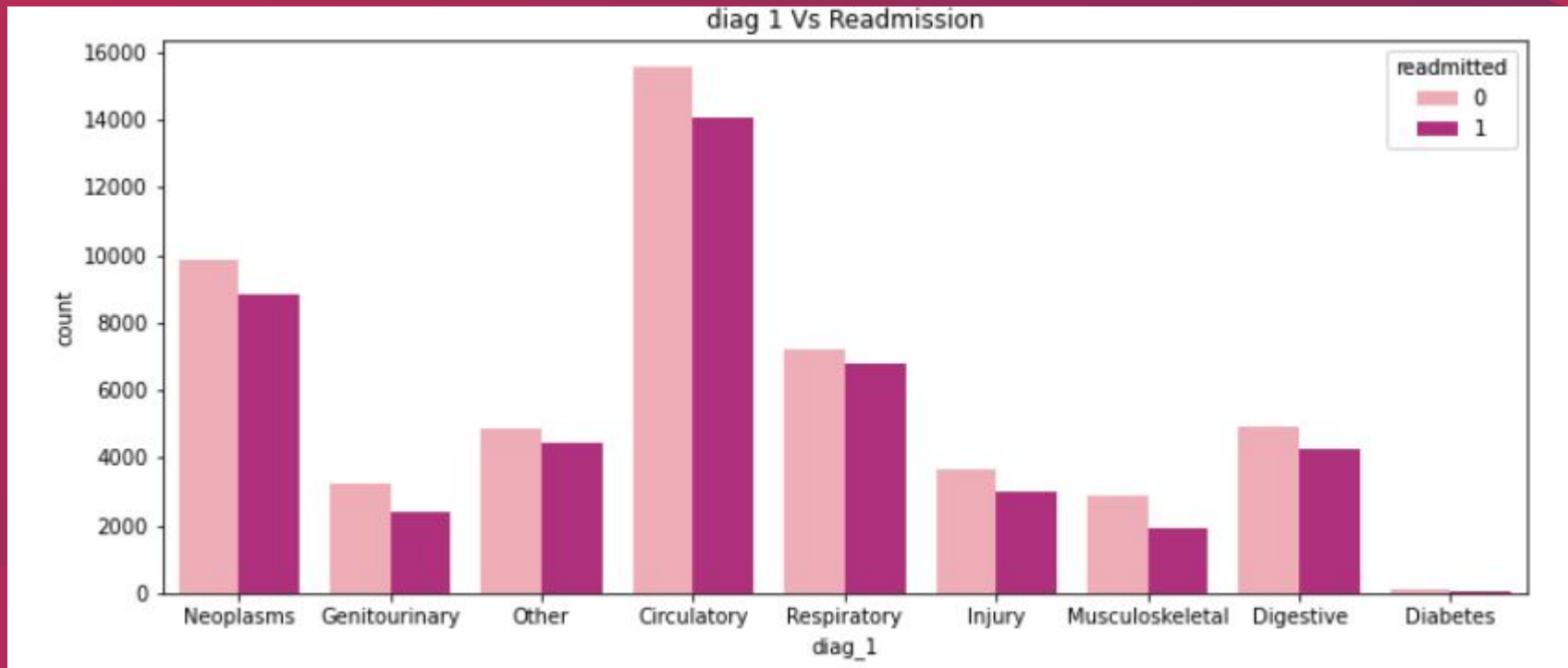


76%

Race

76% of readmitted patients are Caucasian.

DIAGNOSIS TYPE



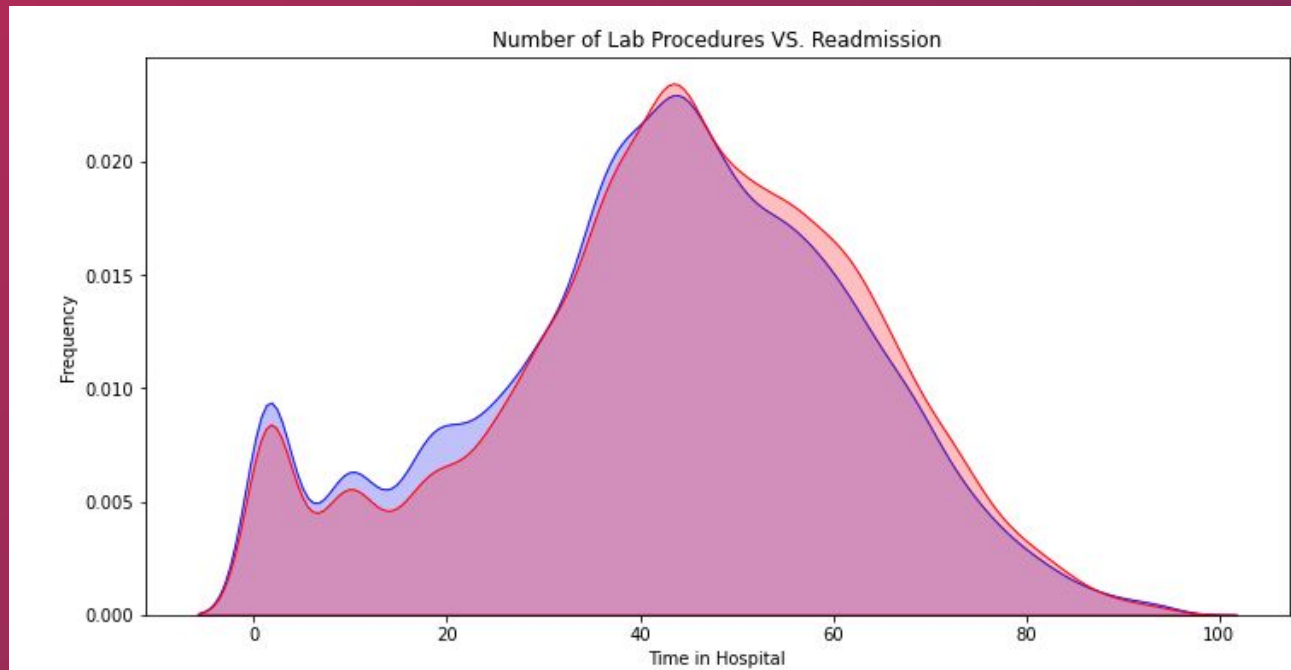
The background is a solid dark red color. It is decorated with various white geometric elements: thin circles, larger semi-transparent circles, and short line segments scattered across the frame. The text is centered in a bold, white, sans-serif font.

Continuous Features

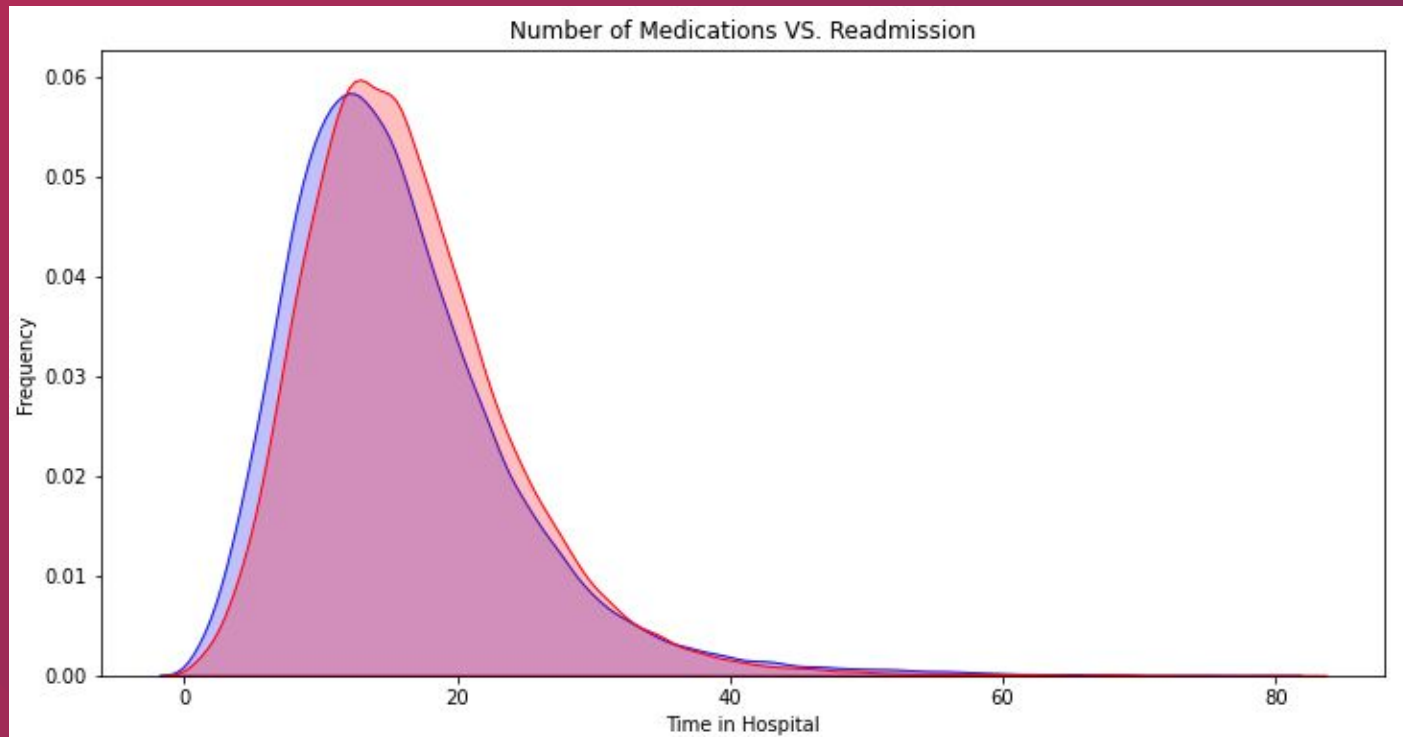
CONTINUOUS FEATURES

- 8 continuous features
- Outliers
 - Number of Lab Procedures
 - Number of Medication

Number of Lab Procedures



Number of Medications



MODELING

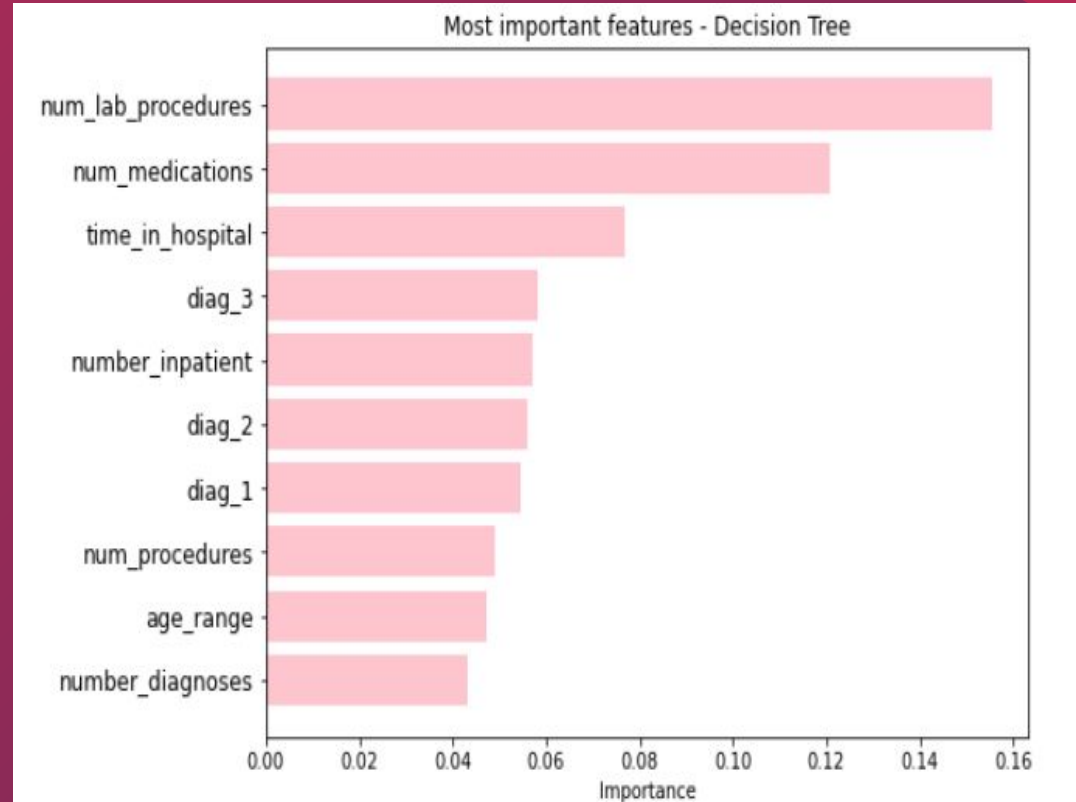
Logistic Regression, KNN, Random Forest Classifier

BASIC MODEL RESULTS

- Random Forest Classifier
- **Training Score:** 0.9999
- **Testing Score:** 0.6207
- **Accuracy:** 0.6194
- **Recall Score:** 0.5134

MOST IMPORTANT FEATURES

- Decision Tree Classifier
- Most important feature option
- Limited to 10 features out of the 50 in the initial dataset
- Limited 3 classification models to just these 10 features.



MOST IMPORTANT FEATURES

RESULTS

- Logistic Regression model
 - **Training Score:** 0.6090
 - **Testing Score:** 0.6093
 - **Accuracy:** 0.6083
 - Lowest recall score
-
- KNN Model
 - **Recall Score:** 0.4990

PRINCIPAL COMPONENT ANALYSIS

RESULTS

- Overall scores were lower than the scores computed by the limited features variation of each model.
- Logistic Regression function, sensitivity score of 0.5414

GRID SEARCH FINE TUNING

RESULTS

- Produced the best scores
- Random Forest model
- **Training Score:** 0.9999
- **Testing Score:** 0.6218
- **Accuracy:** 0.6181
- **Recall Score:** 0.5151

CONCLUSION

CONCLUSION

- Ten major features are found to have high impact on diabetes patient readmission.
- Although not the best scores, still beneficial for medical practitioners to pay attention to these features
- Using Grid search for each of our classification models produced the best accuracy and sensitivity scores.
- The best model, Random Forest Classifier, provided an accuracy score of 0.62 and a sensitivity score of 0.52.
- Attempt to use other models moving forward.
- Including new data such as family history may be helpful in increasing primary diagnosis rates and effectively decrease readmission rates.

THANKS

Do you have any questions?

akamara@xyz.com

+1 240 333 4456

xyzconsulting.com



CREDITS: This presentation template was created
by **Slidesgo**, including icons by **Flaticon** and
infographics & images by **Freepik**
Please keep this slide for attribution