
Hospital Readmission Prediction

Phase 3 Project



AVI PATEL

Objective

This project seeks to create a model that classifies whether a patient is likely to be readmitted to the hospital within 30 days of discharge given the health conditions. This information will help Hospitals to determine individual patients' risk of returning in the hospital in this time period.

Business Problem

Hospitals in USA spent over \$41 billion on patients who got readmitted within 30 days of discharge. Being able to determine factors that lead to higher readmission in such patients, and correspondingly being able to predict which patients will get readmitted can help hospitals save millions of dollars while improving quality of care.

The Data

The data was obtained from [Kaggle](#). The data contains attributes such as patient number, race, gender, age, admission type, time in hospital, number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

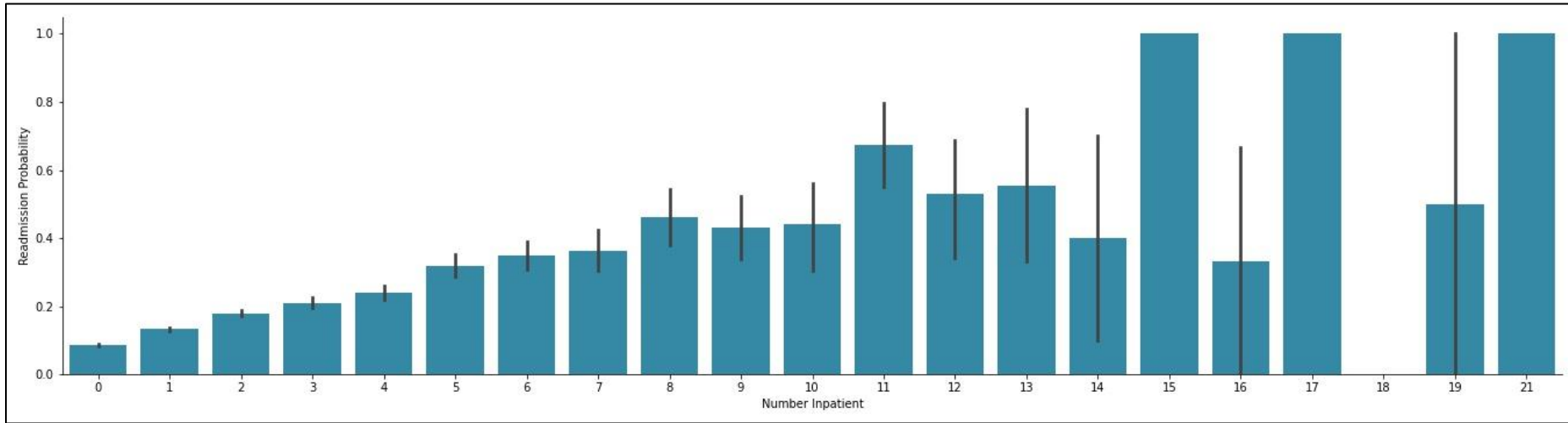


The dataset suffers from lack of class separation, presenting a challenge for generating predictions and class imbalance of approximately 90% - to - 10%

Out of 101766 Patients record, 99340 records were used for the analysis.

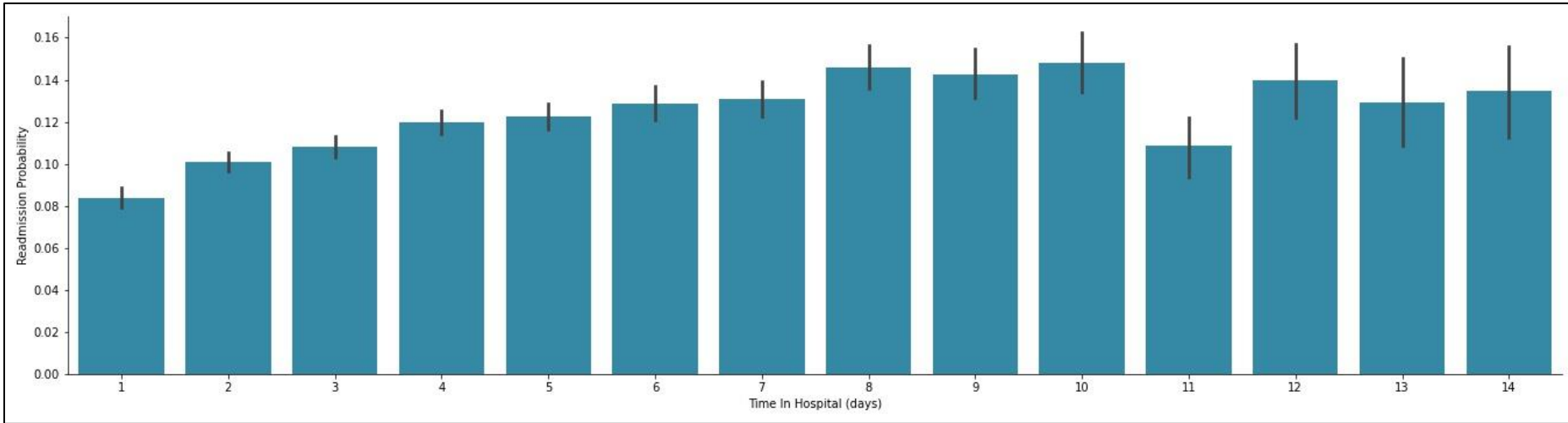
Reason to not use those 2400 records (2.3%) was because of the missing values and the decision was made to drop those records and not to use available techniques to fill the missing values as this is a medical dataset.

Number Of Inpatient visits



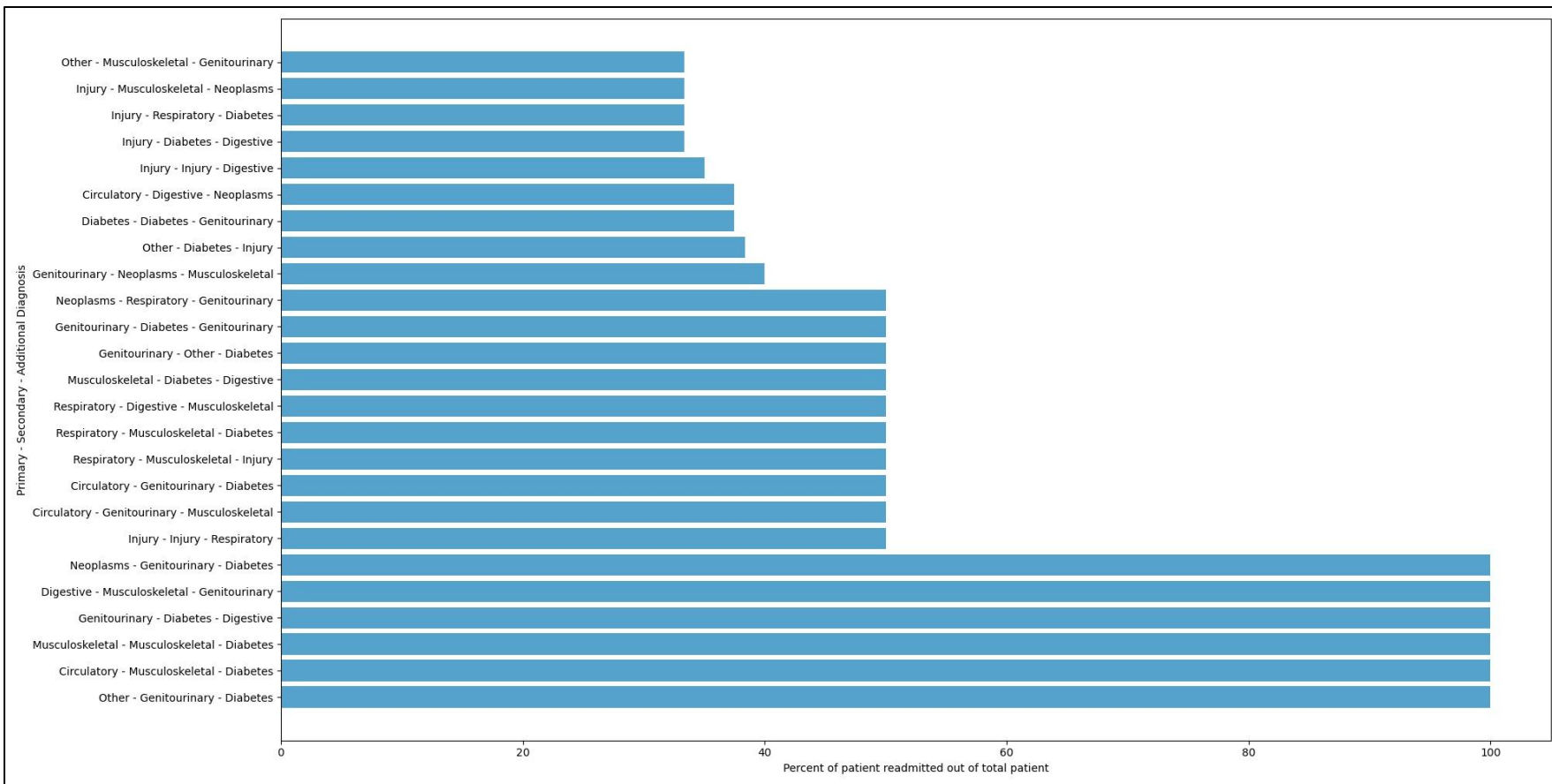
It was found that if the patient had more Inpatients stays in hospital over a past year the probability of readmission within 30 days was higher.

Time In Hospital (Days between admission and discharge)



It was found that if the patient had spent more time in hospital (considering serious illness) the probability of readmission within 30 days was higher.

Diabetic Encounter in the system during Diagnosis.



It was Found that the Patients that were encountered Diabetic during diagnosis had higher probability of readmission within 30 days.

The primary diagnosis is the root cause of the visit. The Secondary diagnosis/diagnoses, are the other conditions that were either present on admission & directly affect the care given for this visit

Results

4 classification models were performed to determine best fit:
Logistic Regression, Decision Tree, Random Forest, and XGBoost.

The **Random Forest** Model is reported to be the best model for prediction of the Readmission of the patient with given medical information with an emphasis **Recall** in an effort to minimize false predictions of no readmittance.

RECALL : 63.00%

ACCURACY : 62.25%

AUC : 67.40%

THANK YOU



<https://github.com/avithekkc>

AVI PATEL
