

PATIENT READMITTANCE ANALYSIS

Instructor: Prof. Daniel Acuna

Problem and Objective

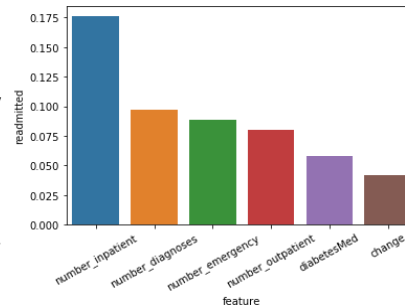
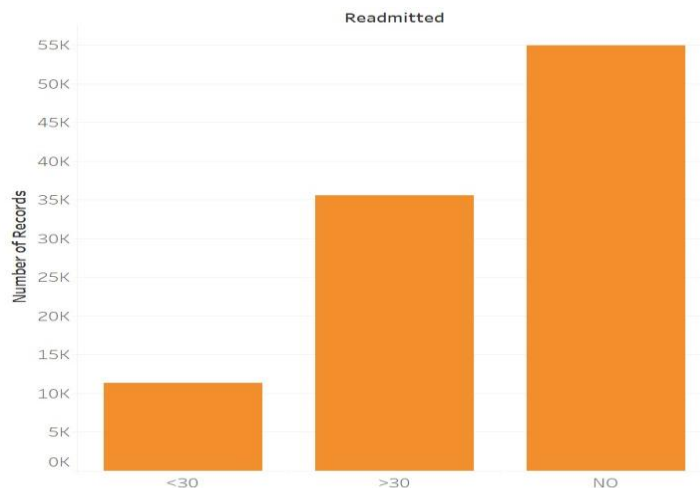
Diabetes is the underlying cause of many health issues. Our main objective is to determine if a patient will be readmitted or not by considering the medicines, diagnosis and various other features. To accomplish this we have decided to use two algorithms: Logistic Regression and Random Forest. This system can be used by healthcare providers to determine if a patient needs a follow-up appointment in the immediate future.

Goals

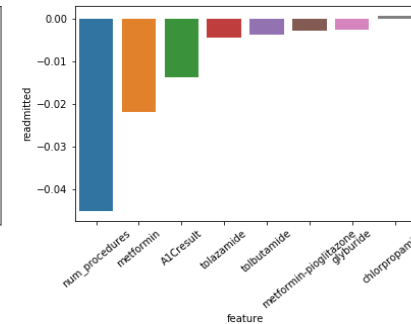
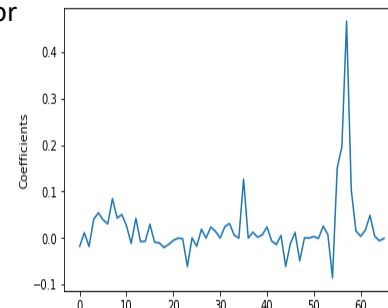
1. Predict if a patient will be readmitted or not.
2. Predict if a patient will be readmitted in less than 30 days or more than 30 days.

Data Description

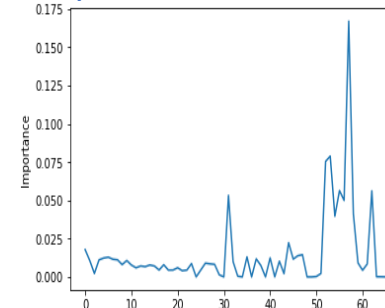
This data consists of 49 columns that can be used to determine the readmittance of the patient, twenty-four of which are various diabetic medicines and the remaining are the tests and columns containing demographic data. The data is in csv format and consists of 101,767 records.



Logistic Regression Coefficient Distribution



Random Forest Feature Importance



Model Evaluation

Logistic Models	AUC	Random Forest Models	AUC
LR with all features	0.654	RF with all features	0.660
LR with all features & reg. parameters	0.652	Multiclass RF	0.572
LR with important features	0.64	RF with important features	0.642

Best Models	Precision	Recall	AUC
Random Forest	0.62	0.36	0.660
Logistic Regression	0.63	0.31	0.654

We found that the Random Forest with all features included was the best model.

Top 10 Features of Both Models

Logistic Regression

Features	Coefficient
number_inpatient	0.466
number_emergency	0.196
number_diagnosis	0.151
diabetesMed	0.126
number_outpatient	0.103
Diabetes	0.084
Circulatory	0.054
Diabetes3	0.050
time_in_hospital	0.048
Diabetes2	0.042

Random Forest

Features	Importance
number_inpatient	0.167
num_medications	0.078
num_lab_procedures	0.075
number_diagnosis	0.056
time_in_hospital	0.055
age	0.053
number_emergency	0.049
number_outpatient	0.041
num_procedures	0.039
insulin	0.022

Conclusion

From the models used and feature importance that was obtained, the top features were found to be number of in patient visit, number of emergency visits, number of medications and number of diagnosis on which the readmission of a patient depends. From both the models, number of in patients was the most important one and it makes sense as patients administered by a doctor for a certain disease have a greater chance of revisiting and thus, readmission in a hospital.

We can also infer that higher number of diagnosis results in a higher chance of readmittance.

Data Source

<https://archive.ics.uci.edu/ml/datasets/diabetes>