Diabetes Readmission Within 30 Days Prediction

MANAL ALENIZI

#TF010

outline

Introduction

Problem statement

Methodology
Data preparation and feature engendering

Finding

Result

Introduction

Diabetes is a chronic disease associated with abnormal high levels of glucose in the blood. Diabetes makes many kinds of complications, which also leads to a high rate of repeated admission of patients with diabetes.

▶ Hospital readmission is a healthcare quality measure that helps in determining the level of quality of care.

Identify whether a hospitalized diabetic patient will be readmitted within 30 days will help to:

▶ Reducing early hospital readmissions is a policy priority aimed at improving healthcare quality.

▶ Reduce cost.

Problem statement

Identify the major factors that contribute to hospital readmissions

Compare accuracy of each model.

Data

Historical data for diabetes patients of clinical care hospitals.

The dataset included around 100,000 records and 50 features.

Methodology

Pre-process, analyze, visualize, and conduct supervised learning on dataset.

Classification algorithm used in order to classify whither patients will readmitted based on the features.

Random forest and Boosting algorithm.

Data preparation and Feature Engendering

- Fill missing , encoding, drop useless columns
- Ensured data classification were the same
- Generate new column to calculate total visits of patients
- Divided data into testing and validation

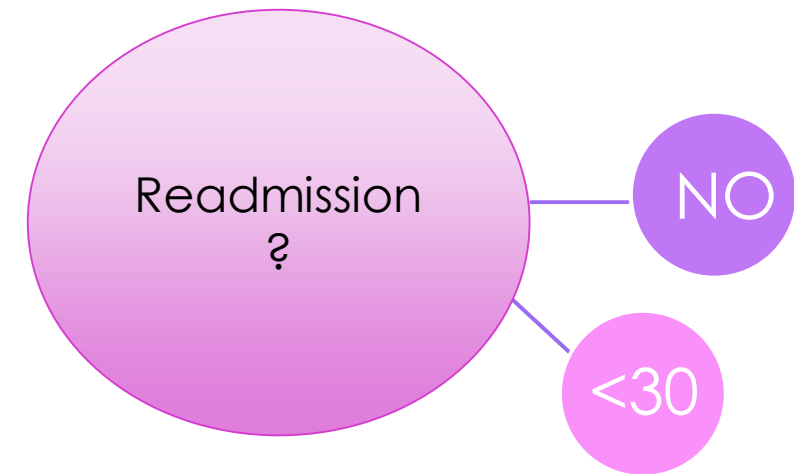# Finding
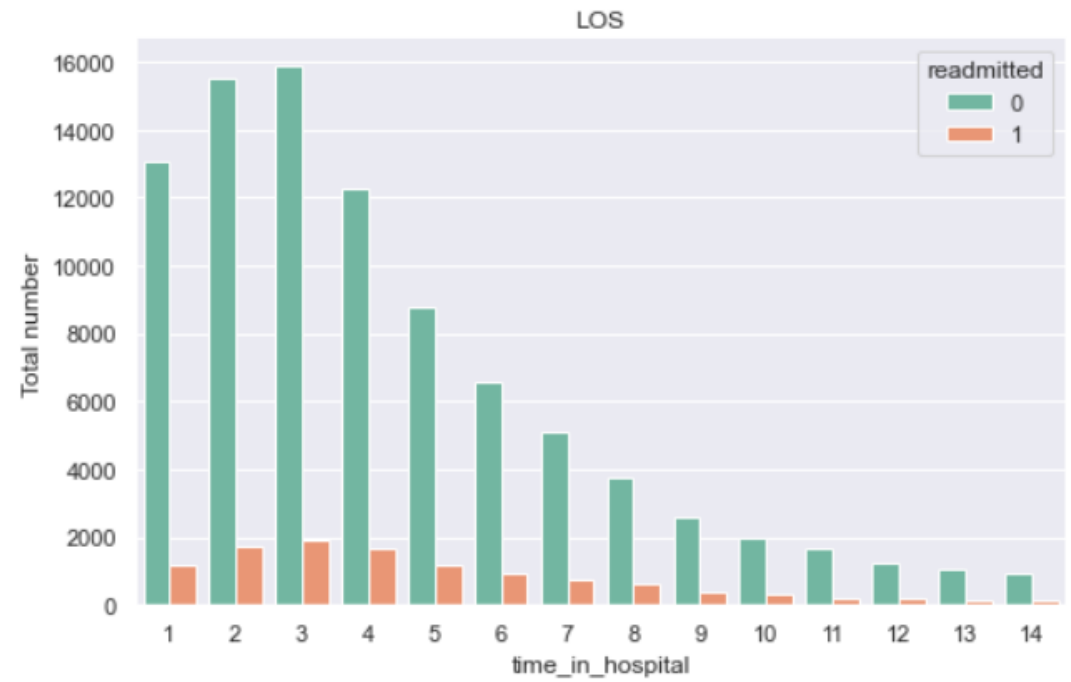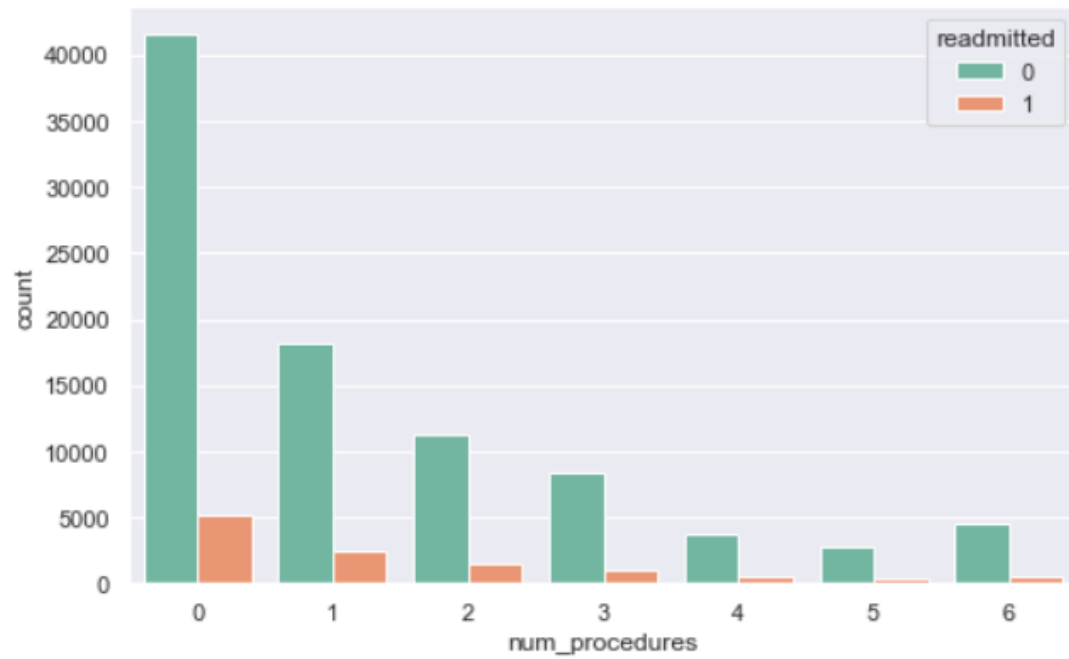
Prediction

Classification problem

Predicate whether diabetes patients will be readmitted to the hospital based on several factors.

Readmission ?

NO

<30

prediction

Random Classifier

```
Accuracy (overall correct predictions): 0.63
Auc: 0.63
Recall (all 1s predicted right): 0.58
Precision (confidence when predicting a 1): 0.64
Detail:
              precision    recall  f1-score   support

           0       0.62      0.68      0.65      1267
           1       0.64      0.58      0.61      1251

    accuracy                           0.63      2518
   macro avg       0.63      0.63      0.63      2518
weighted avg       0.63      0.63      0.63      2518
```
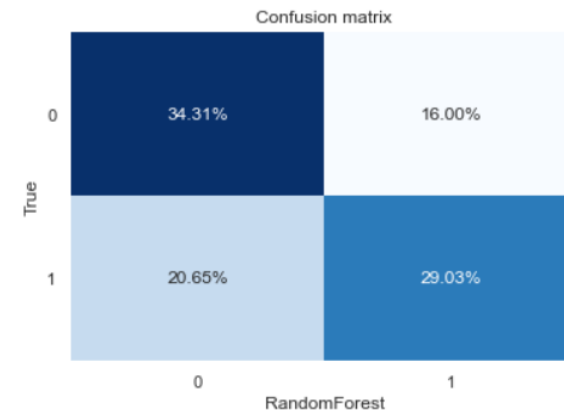

Confusion matrix

prediction

Boosting Classifier

```
Accuracy (overall correct predictions): 0.65
Auc: 0.65
Recall (all 1s predicted right): 0.59
Precision (confidence when predicting a 1): 0.67
Detail:
              precision    recall  f1-score   support

           0       0.64      0.72      0.67      1267
           1       0.67      0.59      0.63      1251

    accuracy                           0.65      2518
   macro avg       0.65      0.65      0.65      2518
weighted avg       0.65      0.65      0.65      2518
```
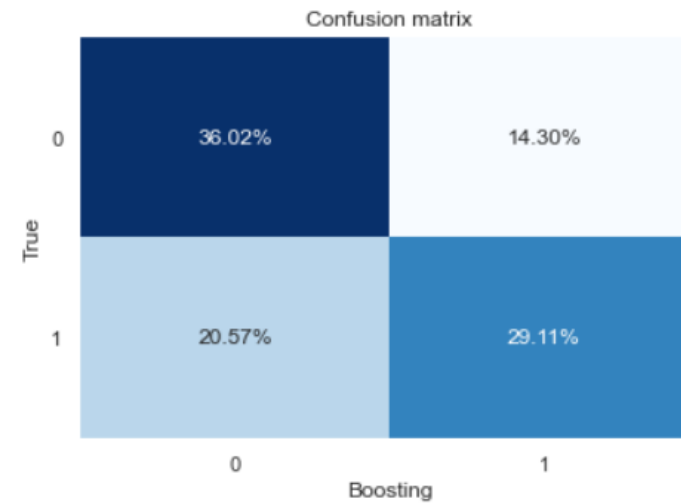

Confusion matrix

Result

Boosting Forest classification:

What are the top 3 factors that predict a diabetic readmission within 30 days?

▶ Number of medication.

▶ Number of lab procedures.

▶ Time in hospital.

If the length of a patients stay in the hospital is predictor of a readmission ? Yes

If the number of procedures a patient entered into the electronic medical record is indicator for readmission ? Yes
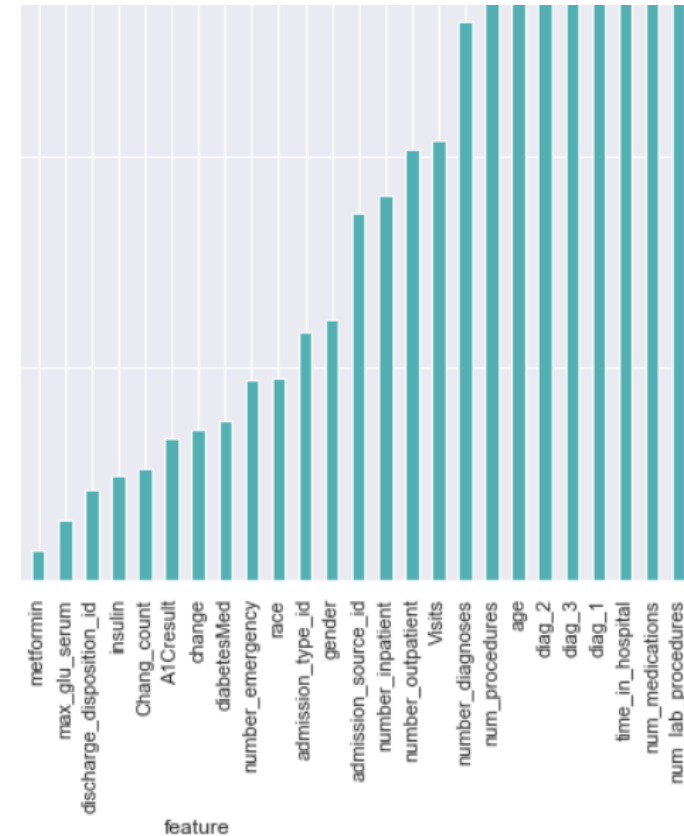
Future Work

- In the future, I'll try the following to improve the performance of our classifier:
- Generate new features.
- Neural network.