Diabetes Readmission Prediction

Rayees Ahamed
Springboard Data Science Student (2021-2022)

The Problem

- Diabetes is a chronic health disease caused due to excess glucose levels
- From 1998 to 2008, US hospitals seen high number of diabetes patient's readmitted within 30 days

"What are the factors leading to high readmission of diabetes patients to hospital within 30 days of discharge?"

Our Goals

- Identifying important attributes that caused diabetes patients' getting readmitted

- Understanding the relationship between different attributes

- Key medicines used in treatment that helped avoid hospital readmission

- Building a prediction model to identify whether a patient will get readmitted or not

Data Exploration – Time at hospital

Target variable - readmitted
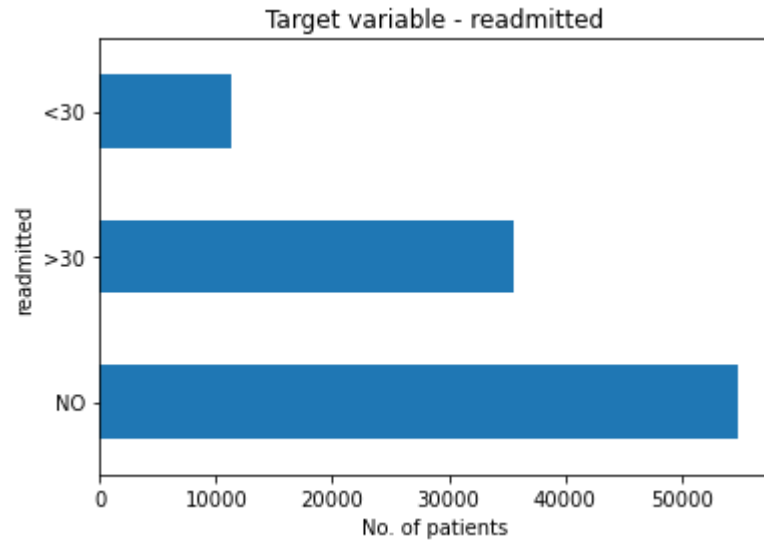
Time at hospital

Distribution of readmitted patients

- 54% - did not get readmitted

- 35% - readmitted after 30 days

- 11% - readmitted within 30 days
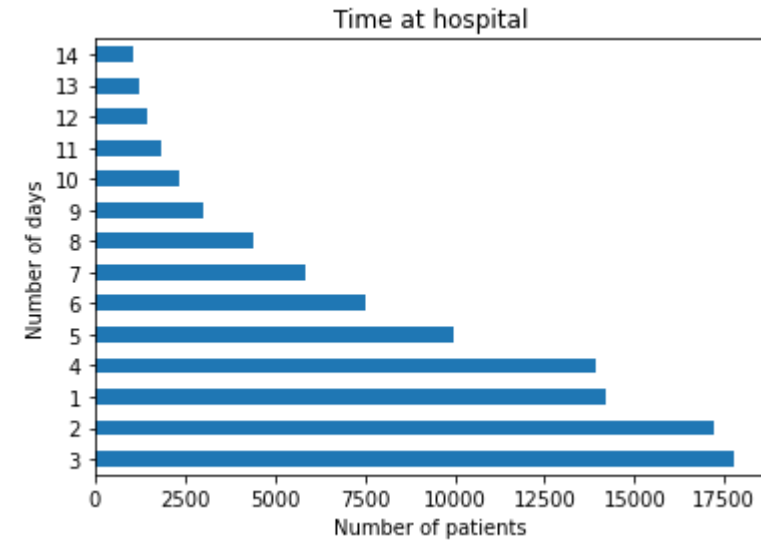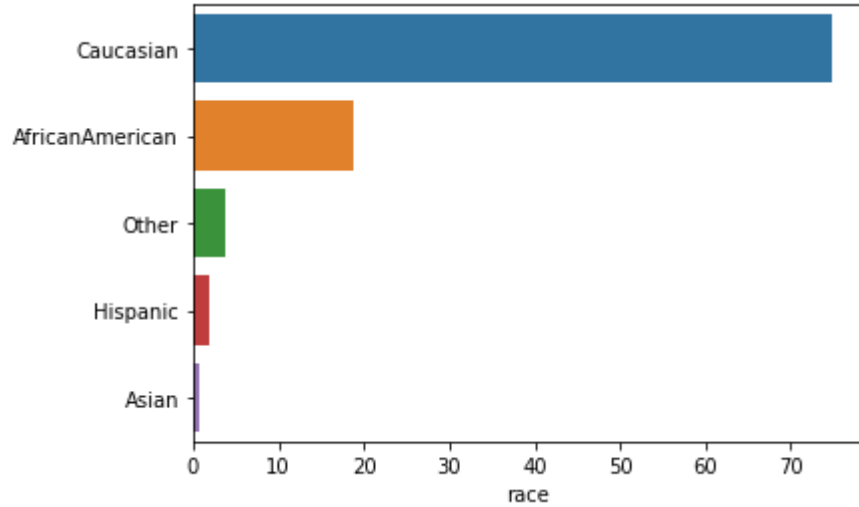
Time spent at hospital by patients

- On average each patient spent 4.5 days

- 72% - spent below 6 days

- 20% - spent above 6 days

- Maximum stay – 14 days

Key questions to asked:

- How large group of people did not get readmitted?

- Why some patients alone staying longer time in hospital?

Data Exploration – Patient demographics

**Patients from different races in this dataset**

- High number of patients are from Caucasian & African American races

- Male and female had equal distribution in both readmitted & not readmitted group

- Patients aging from 45 to 85 are more vulnerable to diabetes and 65 to 85 aged patients are riskier to get readmitted within 30 days

Gender distribution
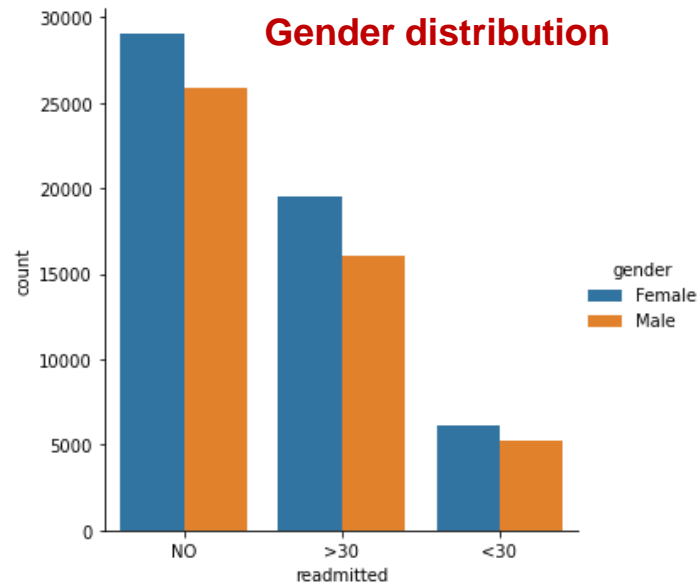
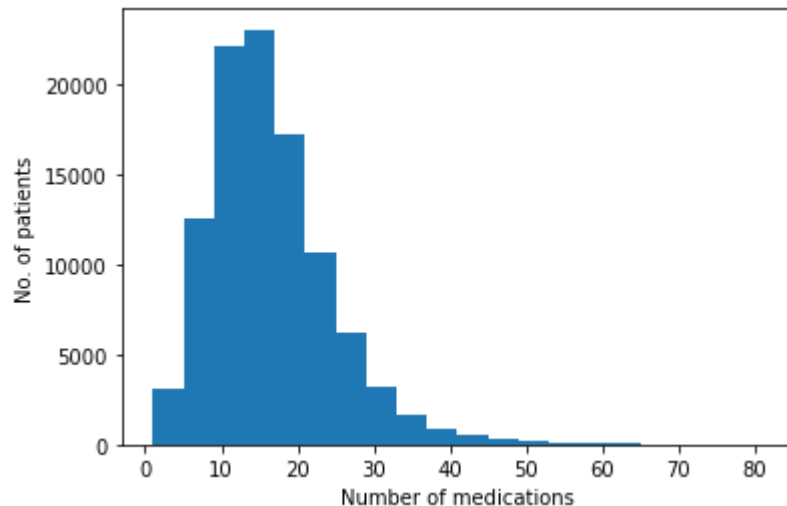**Age distribution of readmitted patients**
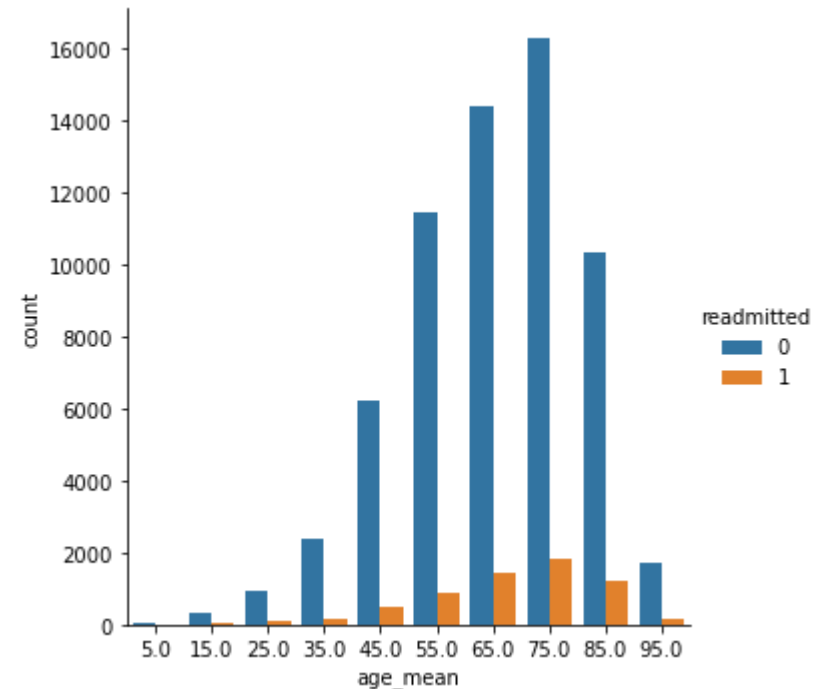
Data Exploration – Change of medications

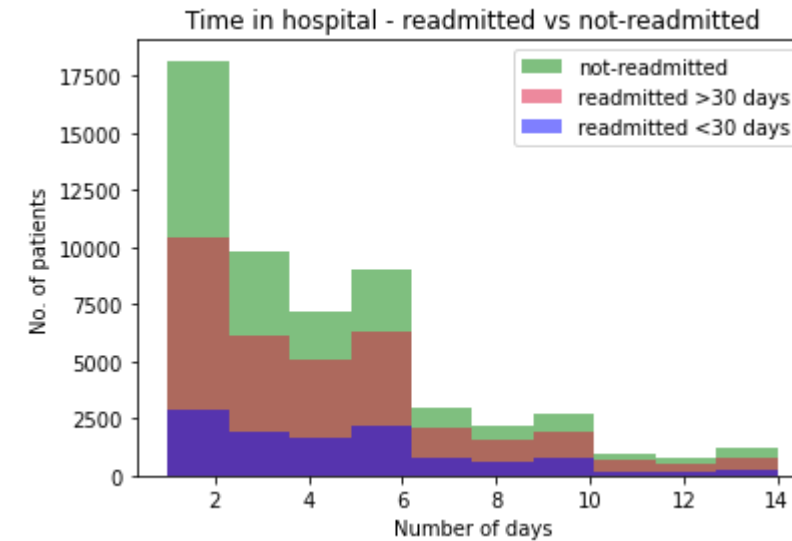**Did patients are suggested to change medications?**

- 60% - asked to change

- 40% - no change

Number of medications taken by patients

- Most patients taken from 2 to 40 medicines over the period

- Average number of medications taken 16-17 medicines per patient.

Relative distribution of hospital time in readmitted patients

Time in hospital - readmitted vs not-readmitted

- Does change in medications help patients avoid readmission?
- Can it reduce the time each patient spent at hospital?
- What medicines specifically taken by patients who are not readmitted?
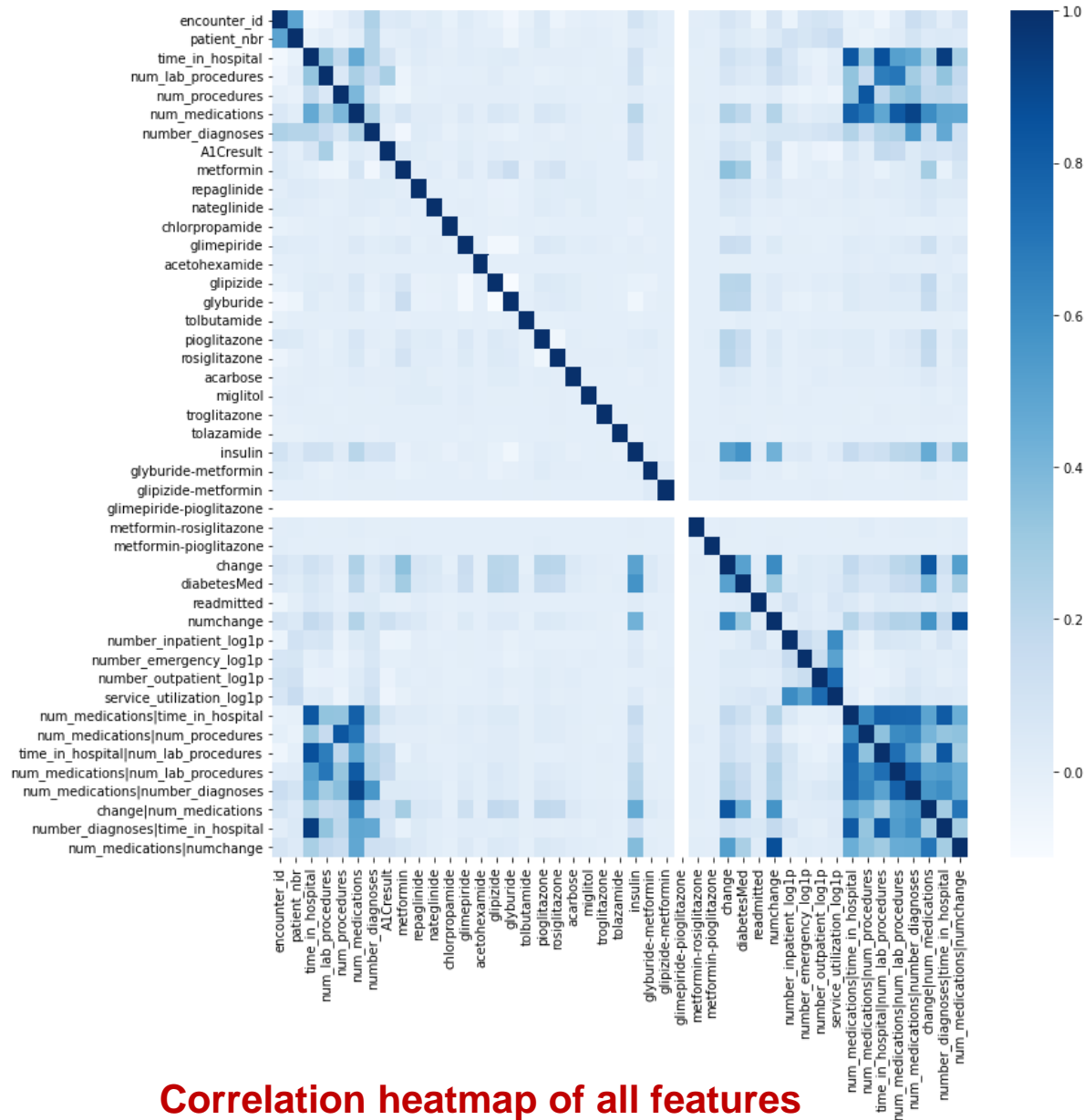
Distribution of medicines in readmitted patients
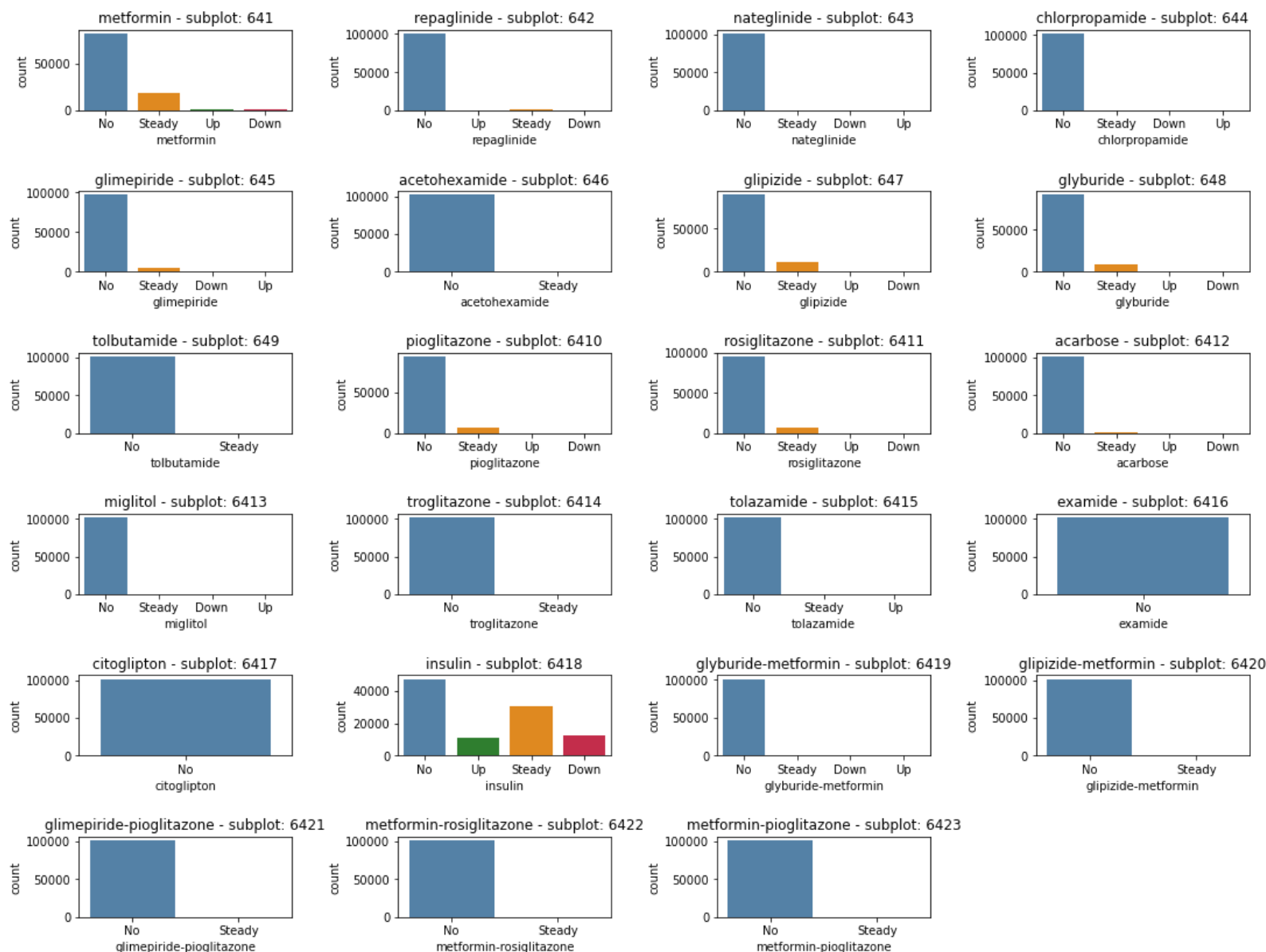
Selected medicines with significant change in their prescription levels:

- Insulin
- Metformin
- Glimepiride
- Glipizide
- Glyburide
- Pioglitazone
- Rosiglitazone

Feature Engineering
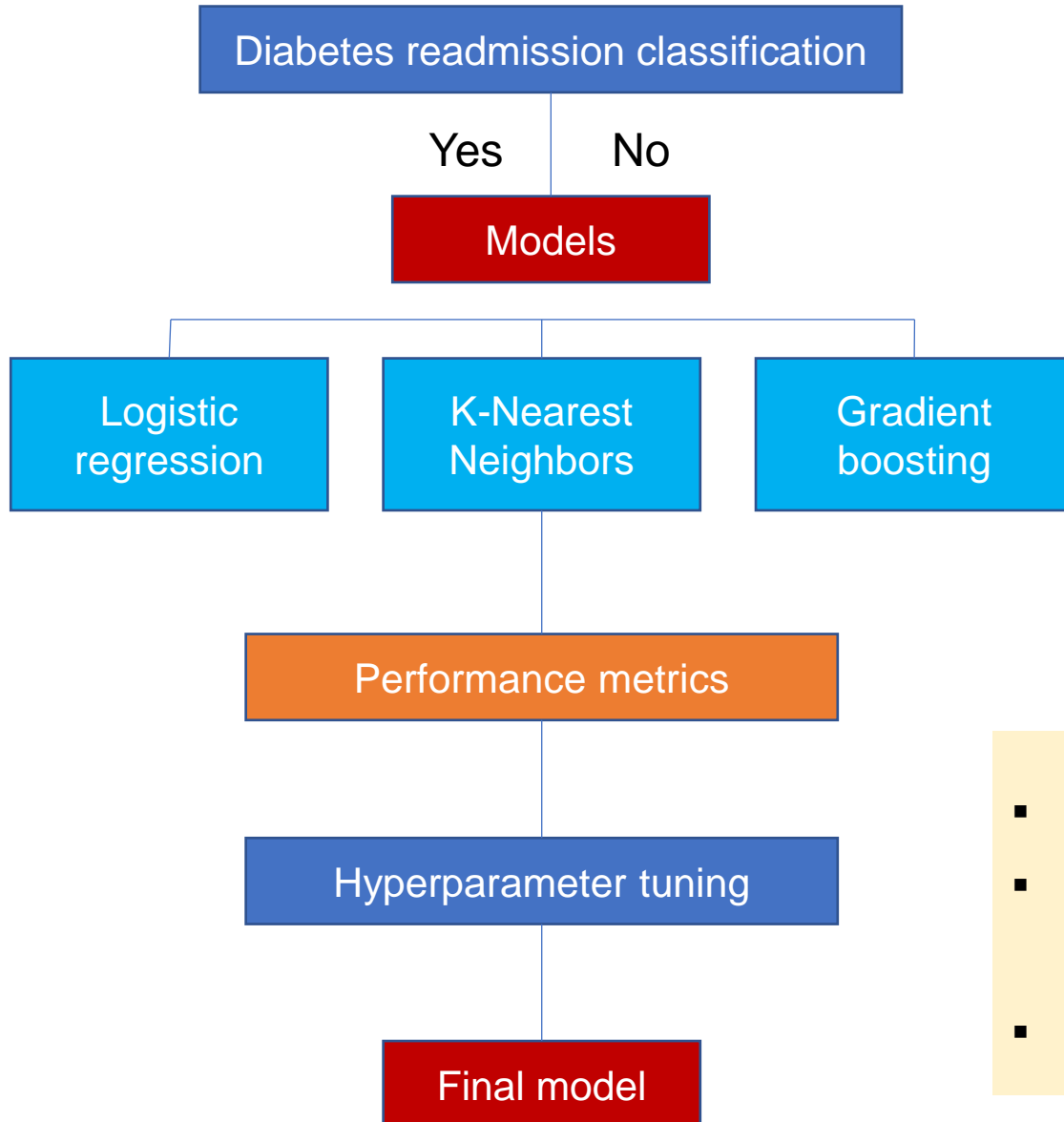

Correlation heatmap of all features

Interaction between variables and new features identified :

- Number of medications | time in hospital

- Number of medications | number of lab procedures

- Number of diagnoses | number of medications

- Age | number of diagnoses

- Number of medications | change of medicines
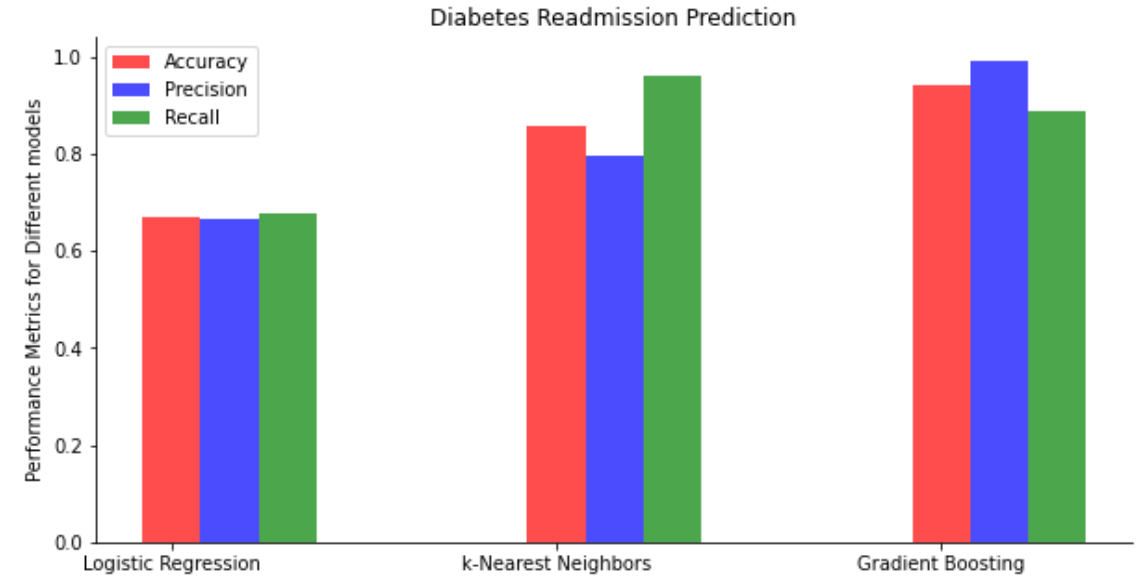
Observations from heatmap:

- Time in hospital & number of medications - positively correlated

- Number of procedures & number of medications - positively correlated

- Age & number of emergencies - negatively correlated

Building the prediction model


```
Diabetes readmission classification
        Yes          No
            Models
Logistic      K-Nearest      Gradient
regression    Neighbors      boosting
        Performance metrics
        Hyperparameter tuning
            Final model
```

Comparison of performance metrics between models

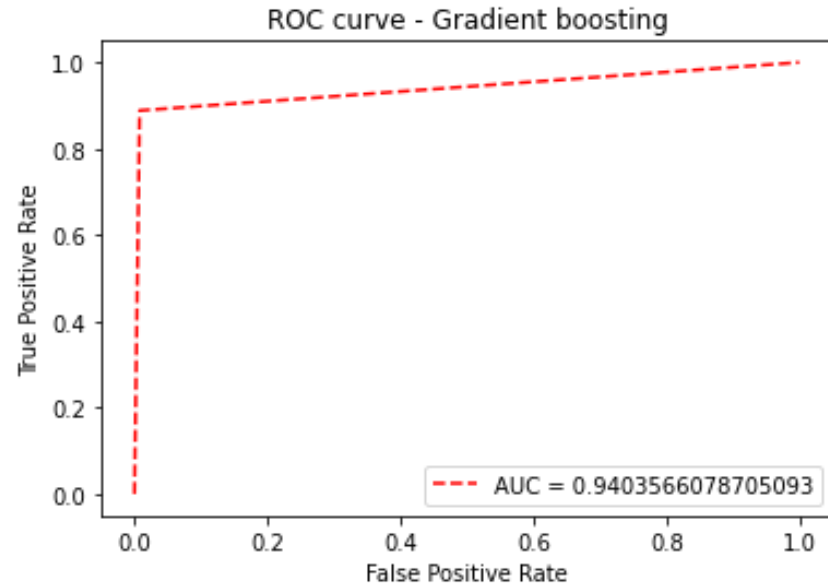Diabetes Readmission Prediction

- **Gradient boosting classifier** topped in accuracy and precision scores
- Compared to k-NN classifier, it is faster and computationally less expensive.
- We selected gradient boosting classifier as our best model

Final model evaluation

ROC curve - Gradient boosting

Final model - Gradient Boosting Classifier
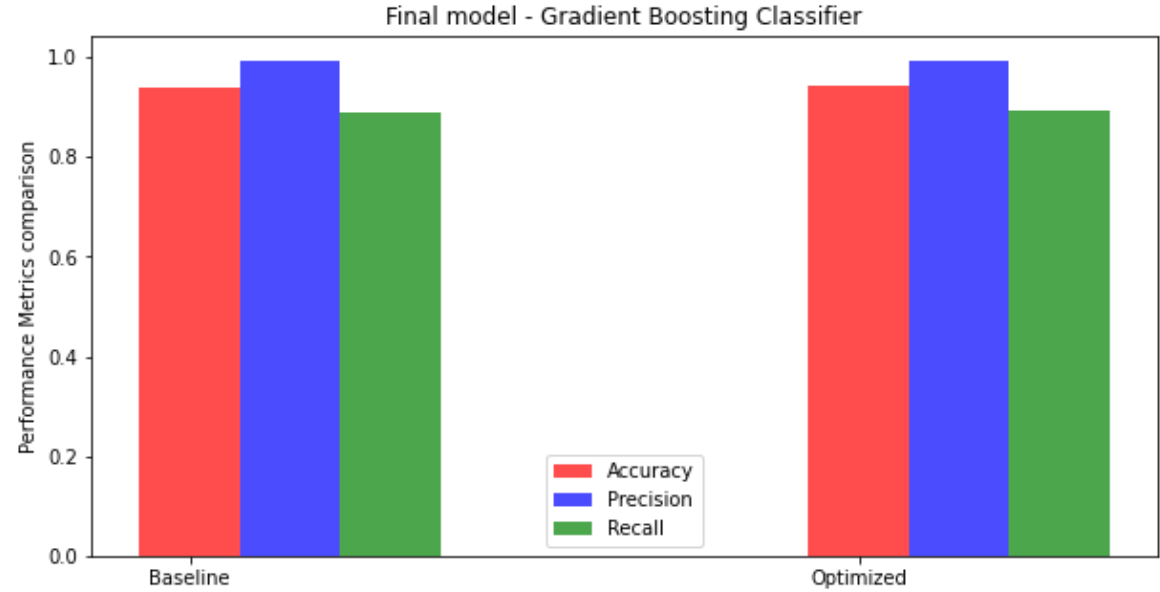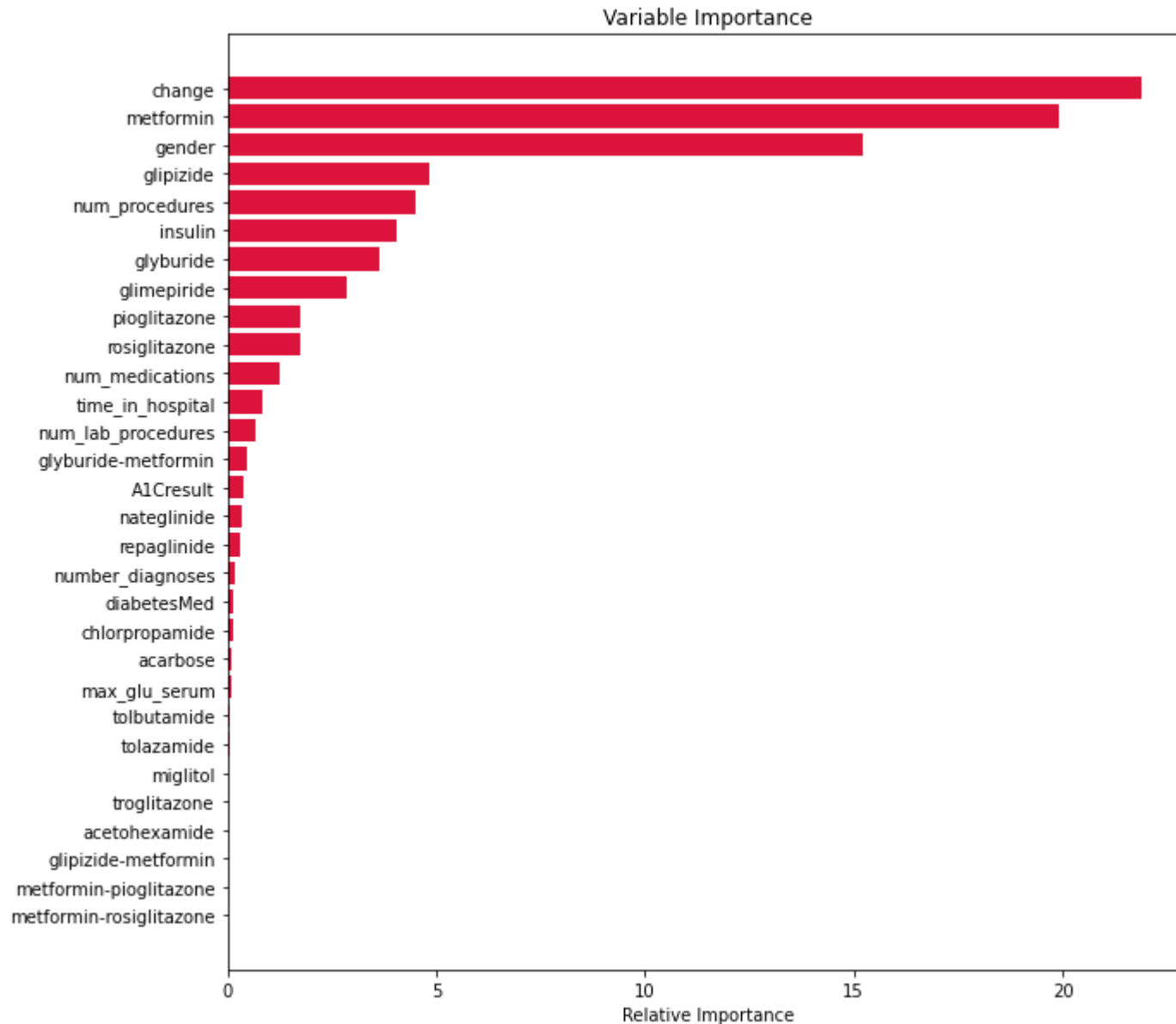
Hyperparameter tuning and best parameters:

- Randomized Search CV
- Learning rate: 1
- Max depth: 2
- Number of estimators: 400
- Cross validation folds: 5
- Tuned model ROC-AUC score: 0.94

Best Model: Gradient Boosting Classifier

Final model performance metrics:

- Accuracy – 94%
- Precision – 99%
- Recall – 89%

Features importance for this dataset

Variable Importance

Most important features of this dataset:

- Change in medications
- Gender
- Metformin
- Glipizide
- Insulin
- Glyburide
- Glimepiride

Summary

- Gradient boosting classifier – best model for prediction diabetes patients' hospital readmission

- Model performance scores were above 90% (i.e., Accuracy of 94%, precision of 99%)

- Medicines played a significant role in predicting the right class of patients

Limitations:

- Class imbalance of the dataset should be taken care prior to deployment

- Prediction also listed gender as important feature, hence more data can be added to evaluate this fact

- Weight column – an important feature had 97% missing values

  - Weight information of the patients can relate to their obesity and other disease status

  - Obesity linked diseases also may be a reason for their frequent readmission

Thank you

Acknowledgements

Springboard

Dipanjan Sarkar
Lead Data Scientist, SIT Academy.
Google Developer Expert - ML