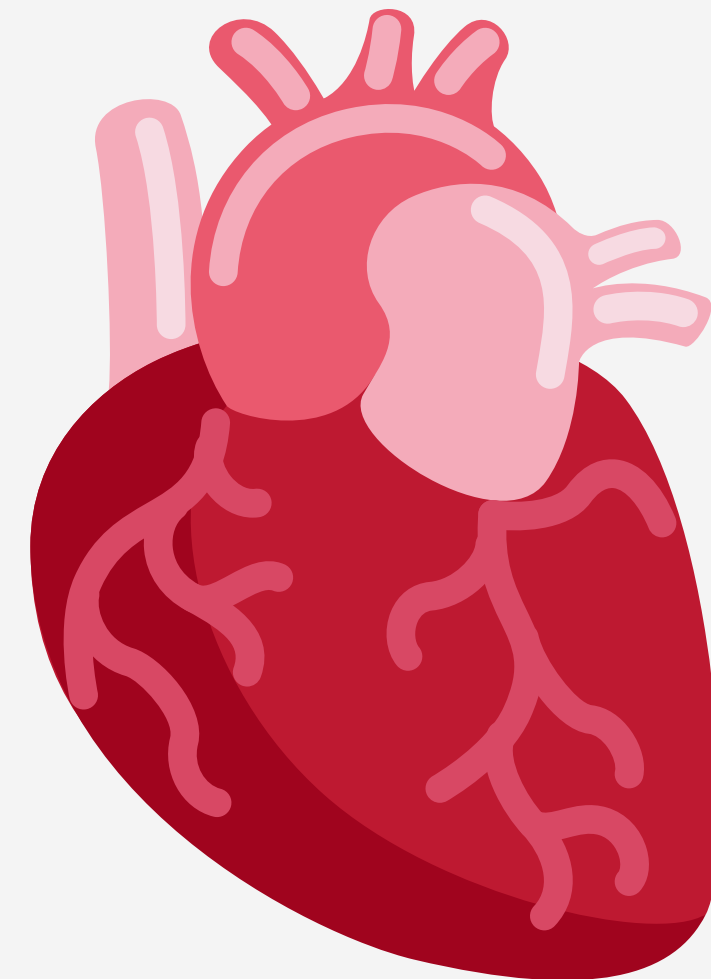


Cardiovascular Disease Prediction



Presentation by Shahd Aljabarti



Problem Statement



Data



Model



Results



Conclusion

Presentation Outline



Problem Statement

The cardiovascular diseases lead the cause of death across the world as the statement of World Health Organization. In 2019, 32% of the global death were due to the heart attacks and strokes.

This project aims to explore the features that affect in cardiovascular disease occurrence and build a model to predict whether a person has a heart disease or not.



Data Understanding

Source of the dataset is Kaggle website where the dataset was extracted in excel format then converted to csv format. It has 70,000 records and 12 features.

Predictors

Age in days	Systolic blood pressure	Smoking
Height	Diastolic blood pressure	Alcohol intake
Weight	Cholesterol	Physical activity
Gender	Glucose	

Target



Cardiovascular
disease
presence

Data Understanding

Source of the dataset is Kaggle website where the dataset was extracted in excel format then converted to csv format. It has 70,000 records and 12 features.

Predictors

Age in years

BMI

Gender

blood pressure

Cholesterol

Glucose

Smoking

Alcohol intake

Physical activity

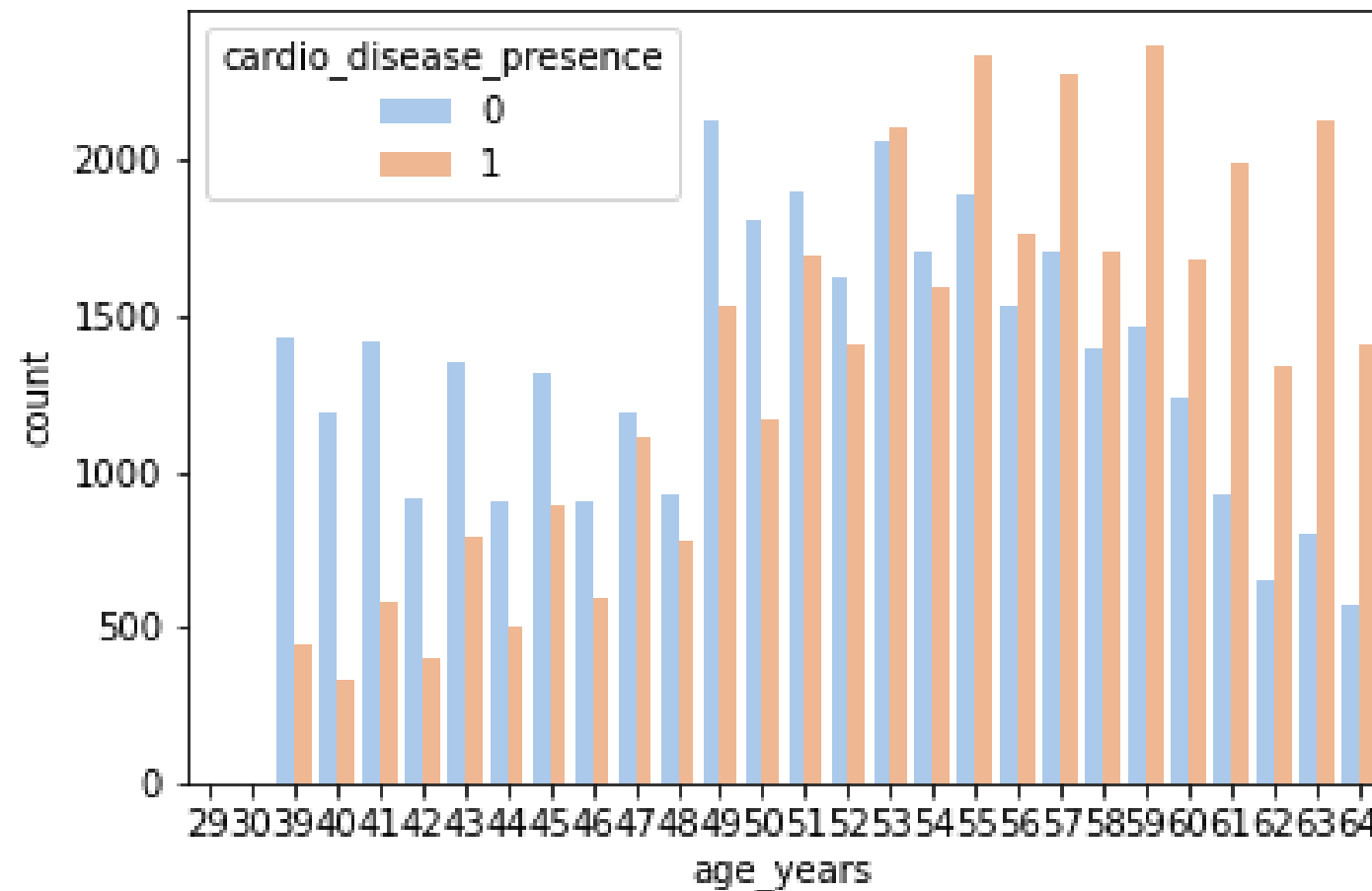
Target

Cardiovascular
disease
presence

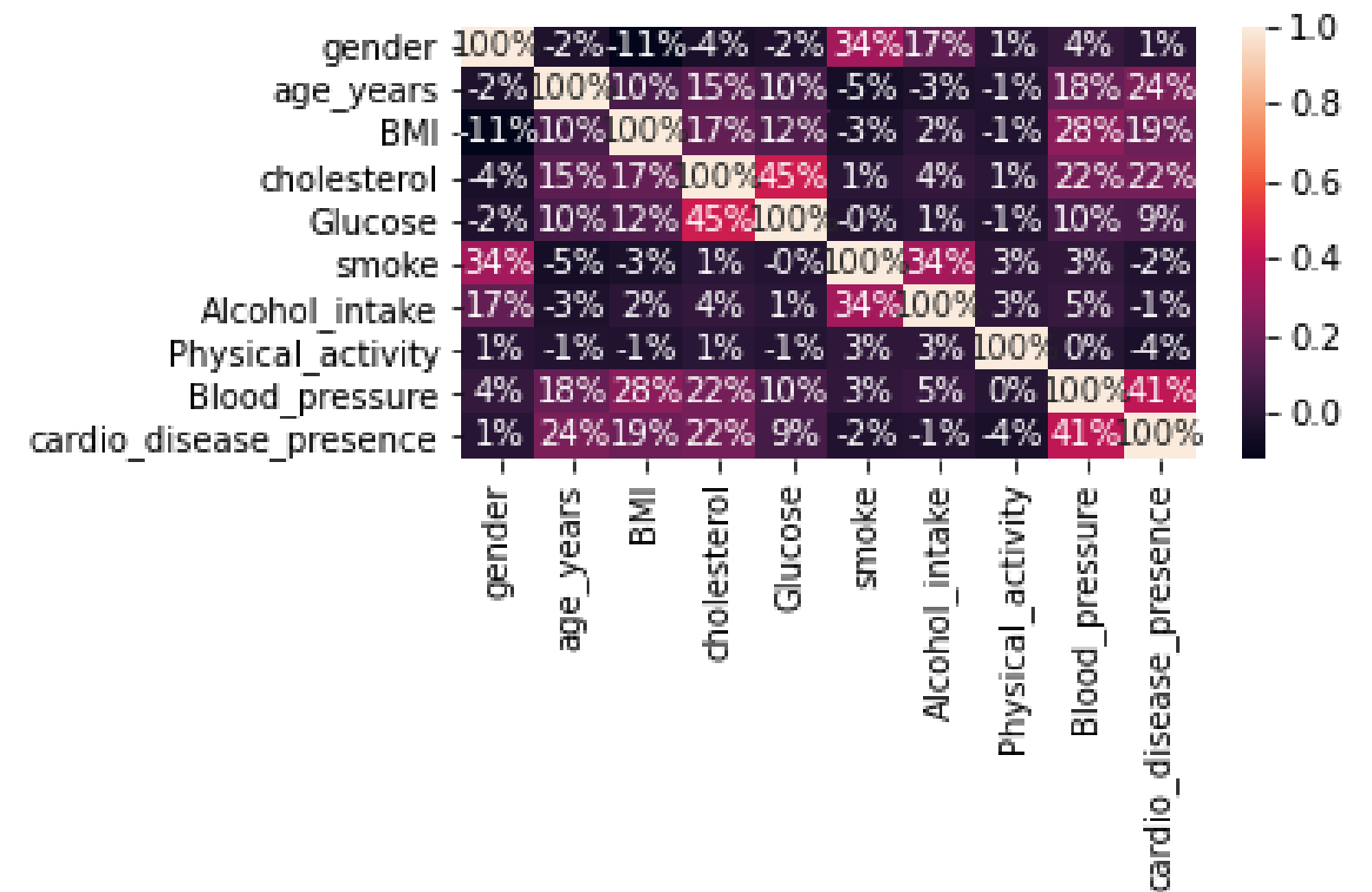
Data Understanding

Interesting plots!

Count Plot



Pairwise Correlation



Models

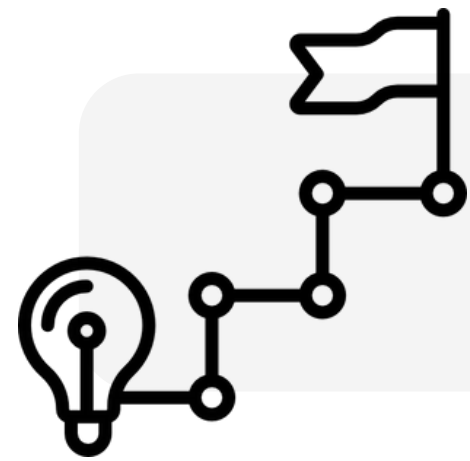
The following algorithms both were used before choosing logistic regression as the model with better performance.

Logistic Regression

The model was trained with scaled data.

Random Forest

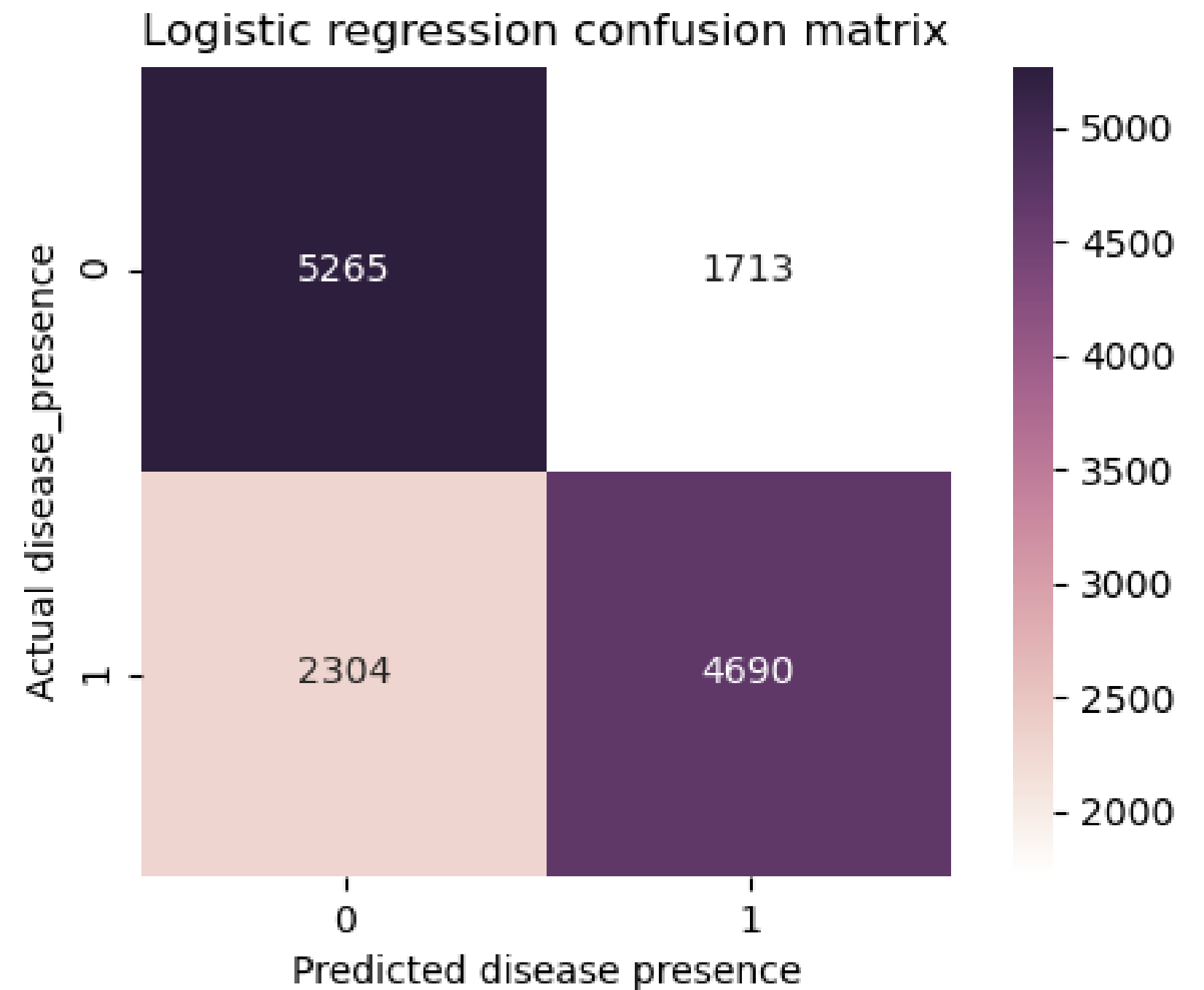
100 trees were built before taking the maximum voting.

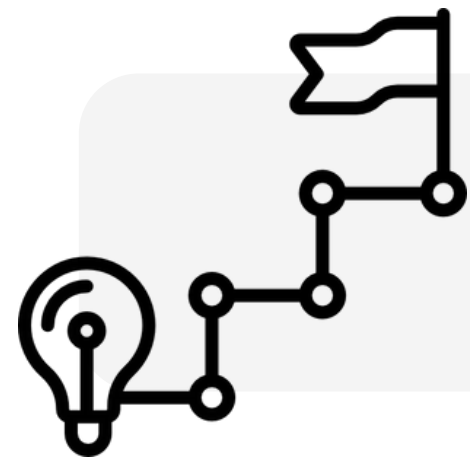


Results

- Logistic Regression

- Recall 67.06%
- Precision 73.25%
- Cross Validation 71.72%
- ROC AUC 77.91%



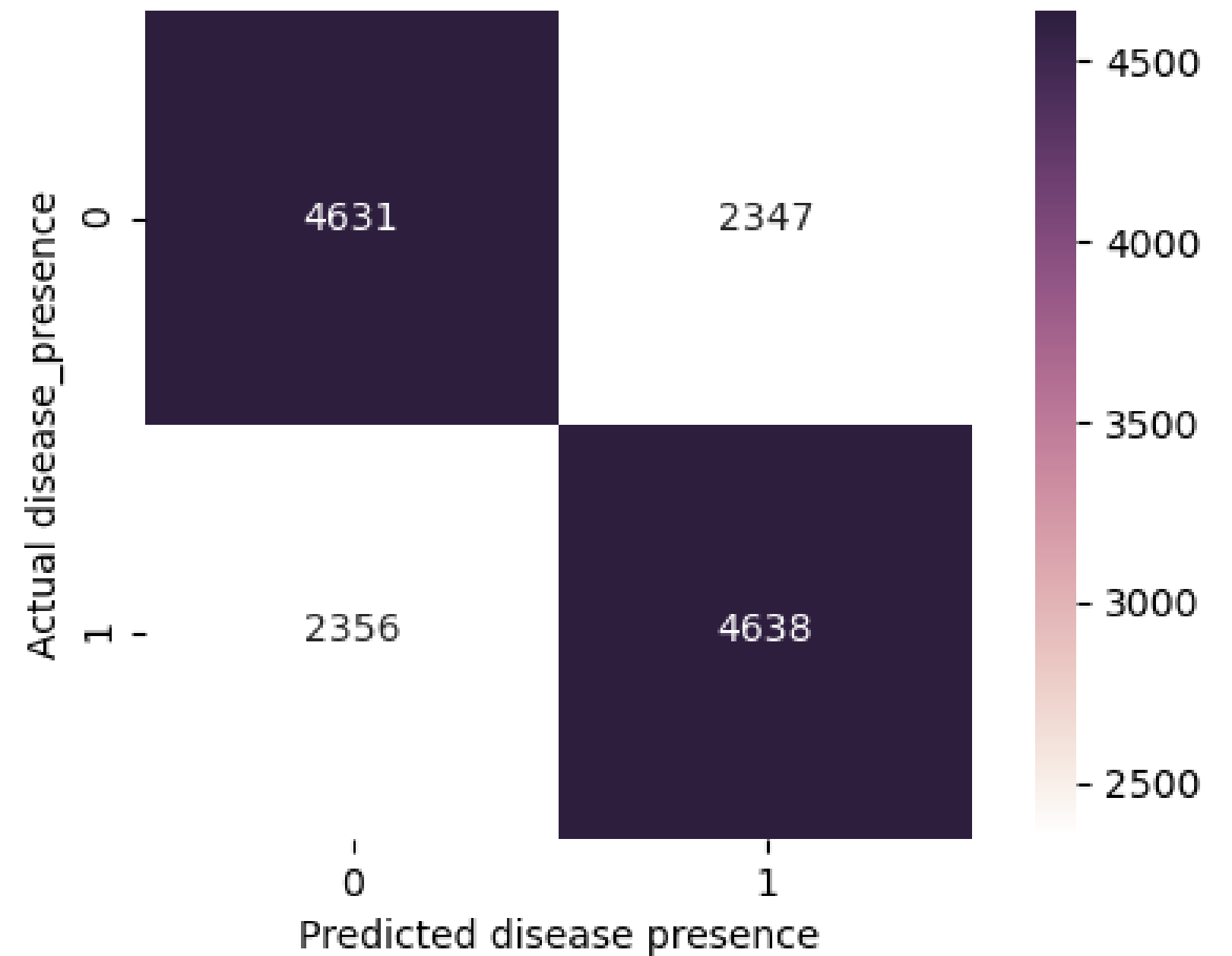


Results

- Random Forest

- Recall 66.51%
- Precision 66.6%
- Cross Validation 66.08%
- ROC AUC 59.24%

Random Forest confusion matrix




Conclusion

- In this case, the logistic regression has better performance than random forest classifier.
- The more the blood pressure, BMI, age & cholesterol level increase the greater your risk for developing heart disease or having a heart attack
- After I studied this topic I should advise others to pay attention to their weights, their life styles since the most of bad habits has high impact on the blood pressure so that will directly affects their hearts health.



Thank you!

Any question?



Have a
great day
ahead.