Cardiovascular Disease Prediction

Prabagini Thambinathan

Overview of cardiovascular disease

- Term cardiovascular disease is a group of disorders of the heart and blood vessels.
- It includes heart attack, stroke, heart rhythm problems, heart infections.
- Cardiovascular disease is the leading cause of death among men and women in the United States.
- Cardiovascular diseases are the leading cause of death globally, taking an estimated 17.9 million lives each year.

Business Problem

- Major cause for cardiovascular disease are coronary artery disease, lack of exercise, diabetes, overweight, smoking, excessive use of alcohol, high blood pressure, and stress.
- Project predicts whether a person has cardiovascular disease or not using different classification model.

Data

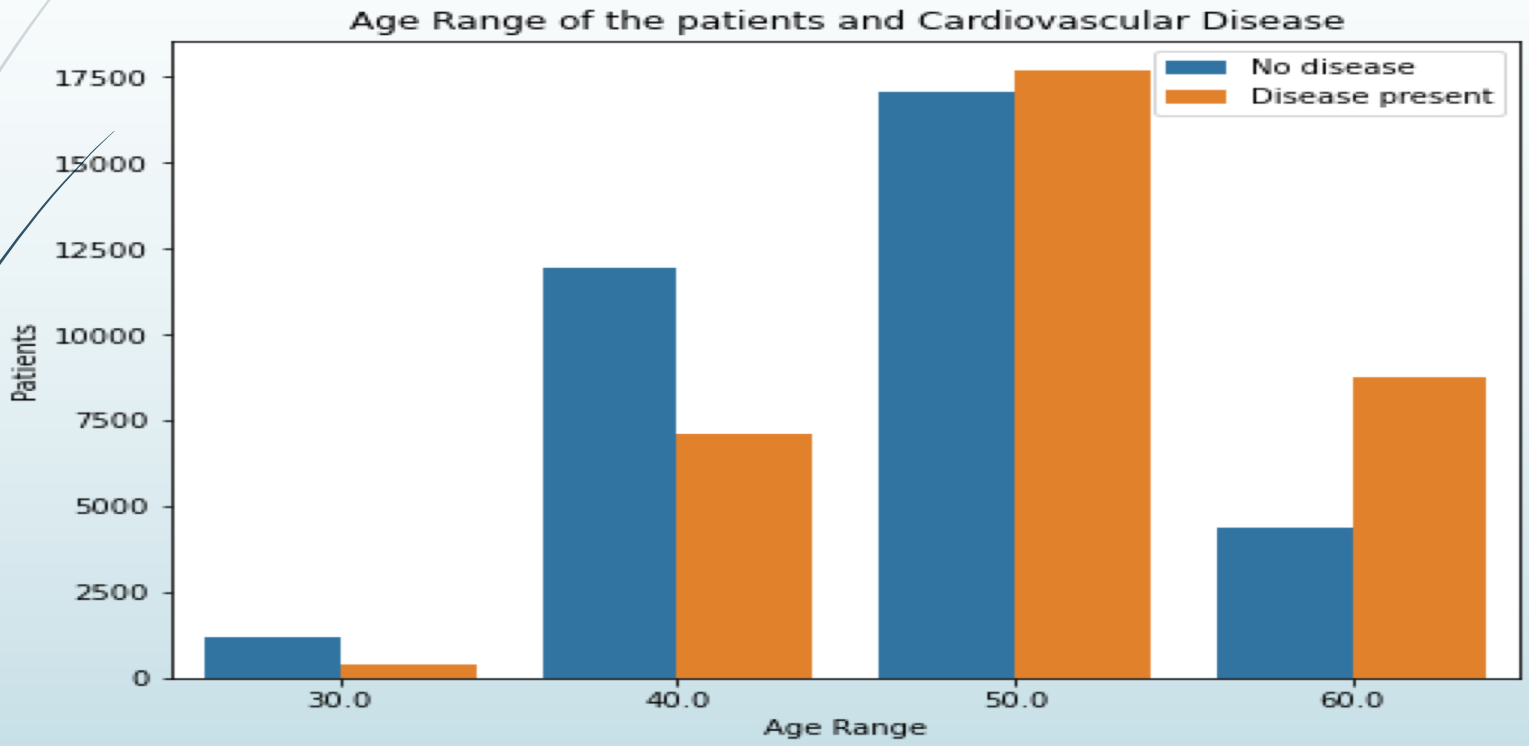
- Data used for this project was obtained from Kaggle.
- There are more than 70,000 entries.
- Data contains 12 features and 1 target.
- Three types of input features – objective(factual information), examination(results of medical examination), subjective(information given by the patient).

Methodology

- OSEMN(Obtain-Scrub-Explore-Model-Interpret) framework is used in this cardiovascular disease prediction.
- Dropping Id column, changing age column from days to years.
- New BMI(body mass index) column.
- Systolic or diastolic pressure readings that were negative, low or high.
- Height and weight that seemed too high or too low.

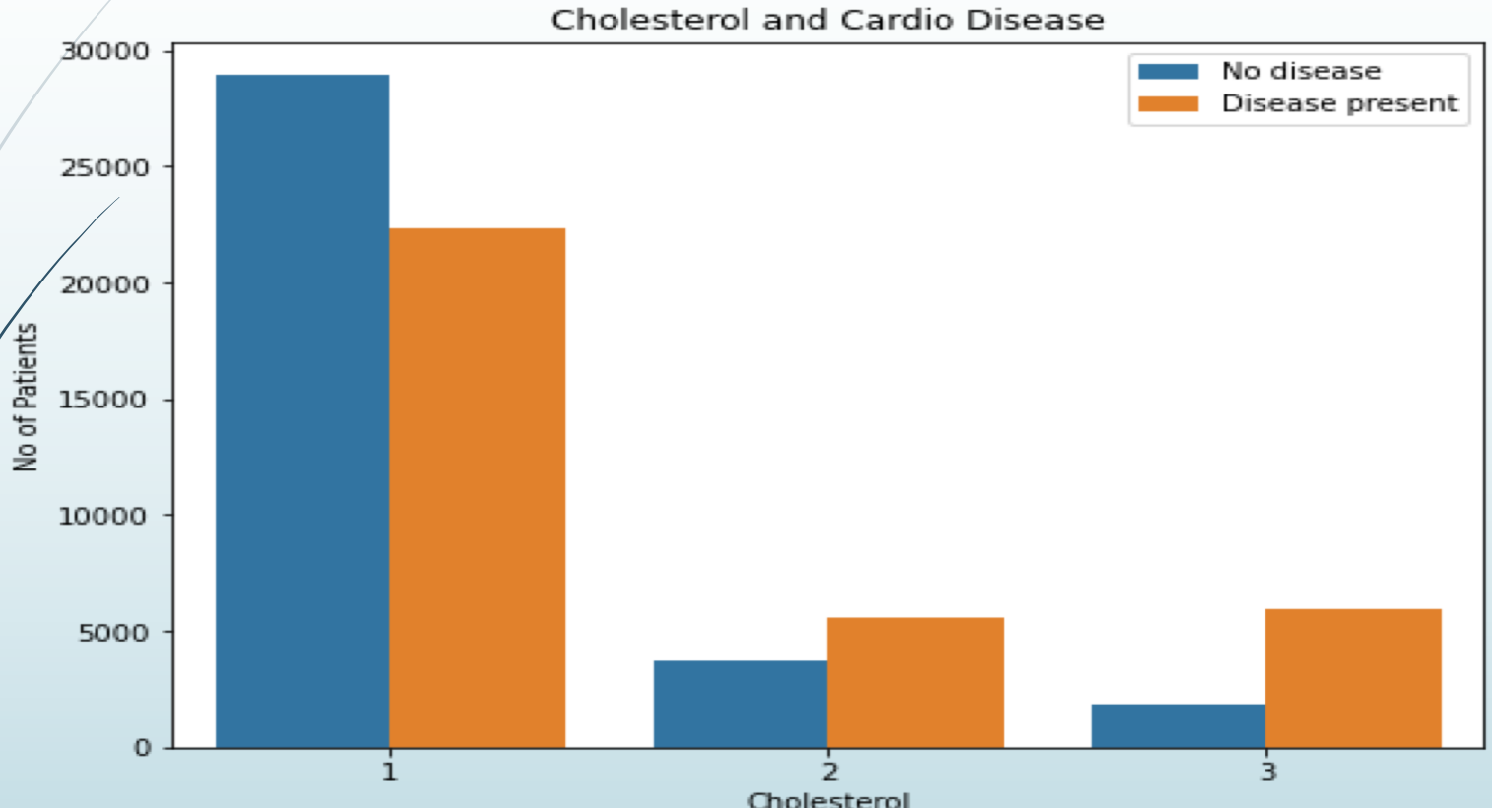
Age Analysis

Age ranging from 50 – 60 has more number of cardiovascular disease.



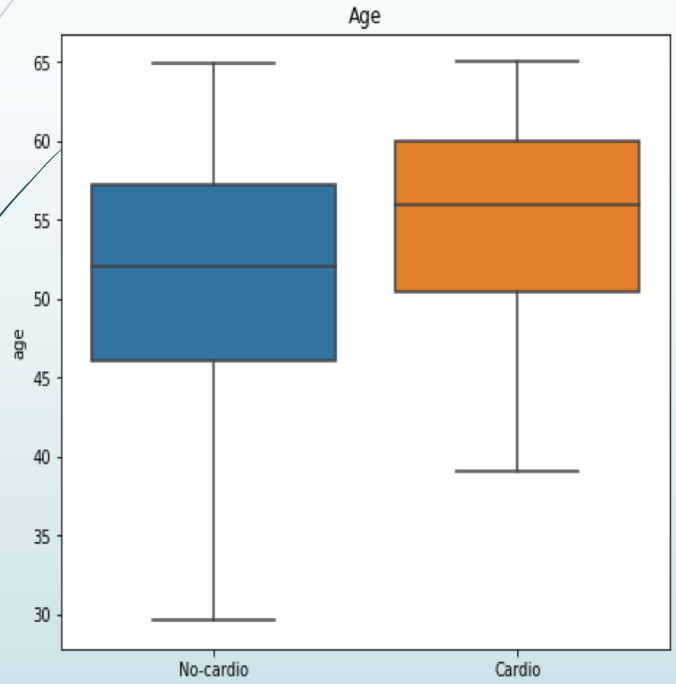
Cholesterol Analysis

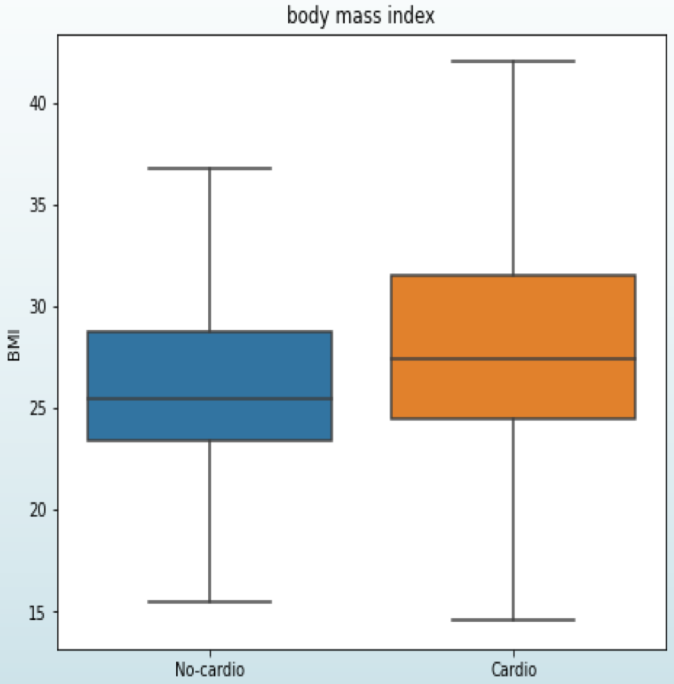
People with above normal and well above normal conditions are highly affected by cardiovascular disease.



BMI Analysis

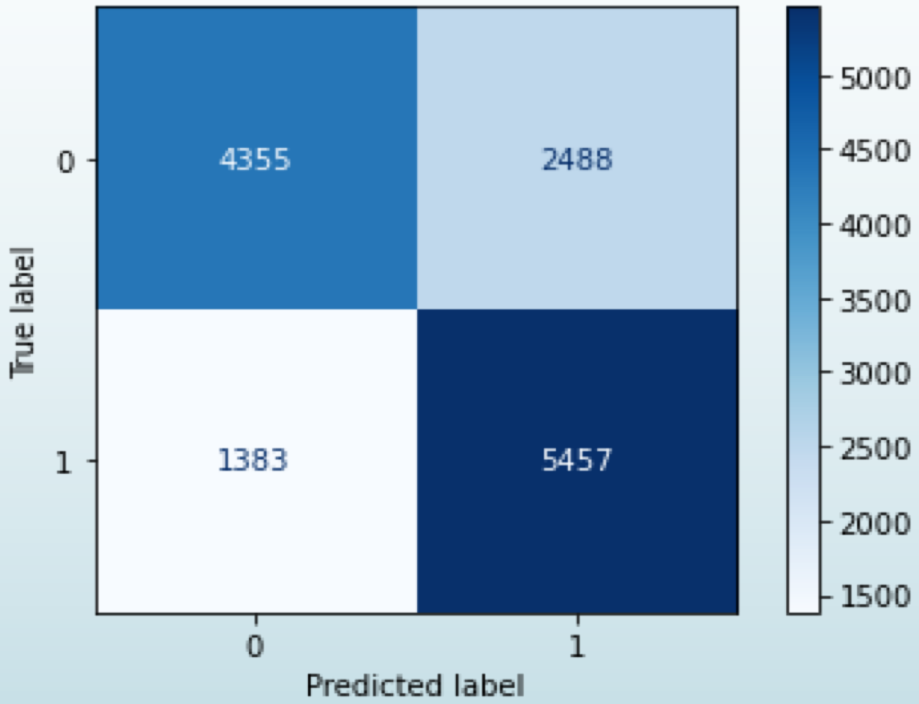
People with higher body mass index is affected with cardiovascular disease.





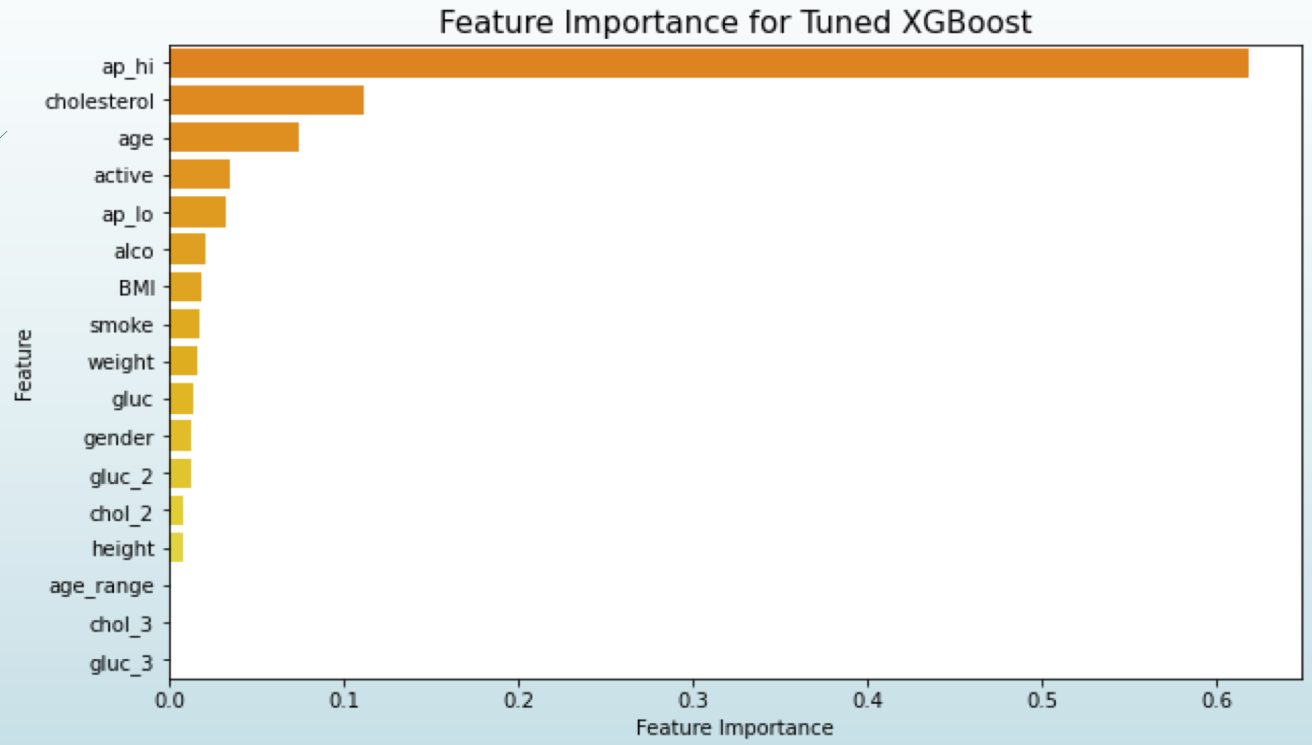
Model Result

1383 False Negatives : people in test group have CVD and were not diagnosed.
5457 True Positives : people in test group have CVD and were diagnosed correctly.



Important Features

Systolic - blood pressure predicts CVD
Cholesterol - above normal and well above normal indicates CVD
Age - getting older increases CVD



Conclusion

- Successfully built a model(XG Boost) that classifies true positives and false negatives with 74% f1 score.
- From the model, we can conclude that having a low blood pressure, low cholesterol, being young and being active are the best ways to avoid cardiovascular disease.

Next Step

- More data about cholesterol, like HDL and LDL.
- Information about diabetes, lung disease, anxiety/depression.
- Data on region, ethnicity, socioeconomic background.

Thank You

mail4prabagini@gmail.com