

COMPARISON OF HEART DISEASE PREDICTION MODELS USING MACHINE LEARNING

THANPISIT NAKHAM | RODSATHON MALILA | KRITTIKA RATTANAWANNAPAN | NATTAWAT KEAMUSTIEN | CHALERMNAT NONTAPA
Data Science Research Center, Faculty Of Science, Chiang Mai University, Chiang Mai 50200, Thailand
Email : nattawat_keamust@cmu.ac.th

Abstract

Due to the increasing number of heart disease patients around the world. This makes heart disease a matter of concern. According to the World Health Organization (WHO), heart disease is the number 1 cause of death in the world, while in Thailand a patient who died from heart disease and stroke. An average of 7 people per hour, or 58,681 people per year, and the death rate from heart disease tends to increase every year. In this study, we studied factors affecting heart disease. The team therefore wanted to compare models for predicting heart disease using the function of a Decision Tree, Random Forest and XGBoost. By using the existing dataset to analyze the relationship through the heatmap of each variable, there are a total of 18 variables to bring variables that are related to heart disease into further analysis. Then the data set was divided into 2 sets Trainset number 1460 and Test set number 500, which were then used for data visualization, and predict the value using the three models. The results of the study found that the variables that were related to heart disease in the heatmap were smoking, stroke, Diabetic, Skin Cancer, Asthma and the best fit model was the decision tree has an accuracy of 78%

Introduction

According to statistics from the World Health Organization (WHO), cardiovascular disease is the number one cause of death in the world.

In addition, the death rate from cardiovascular disease in Thailand is increasing every year. Some congenital diseases can also affect heart disease. Therefore, the researchers intend to study the factors affecting heart disease by applying the work of Random forest.

Objectives

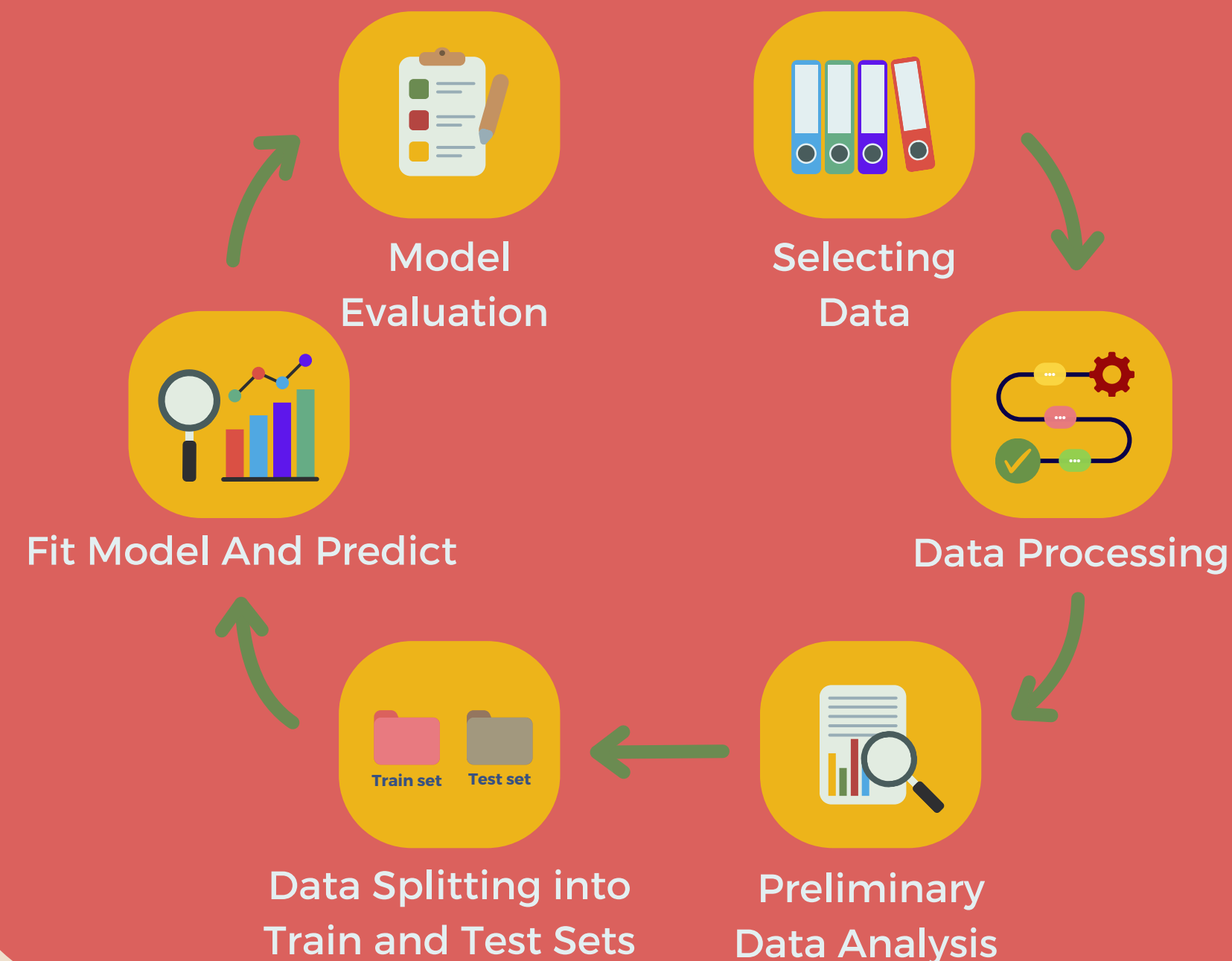


To study factors affecting
heart disease

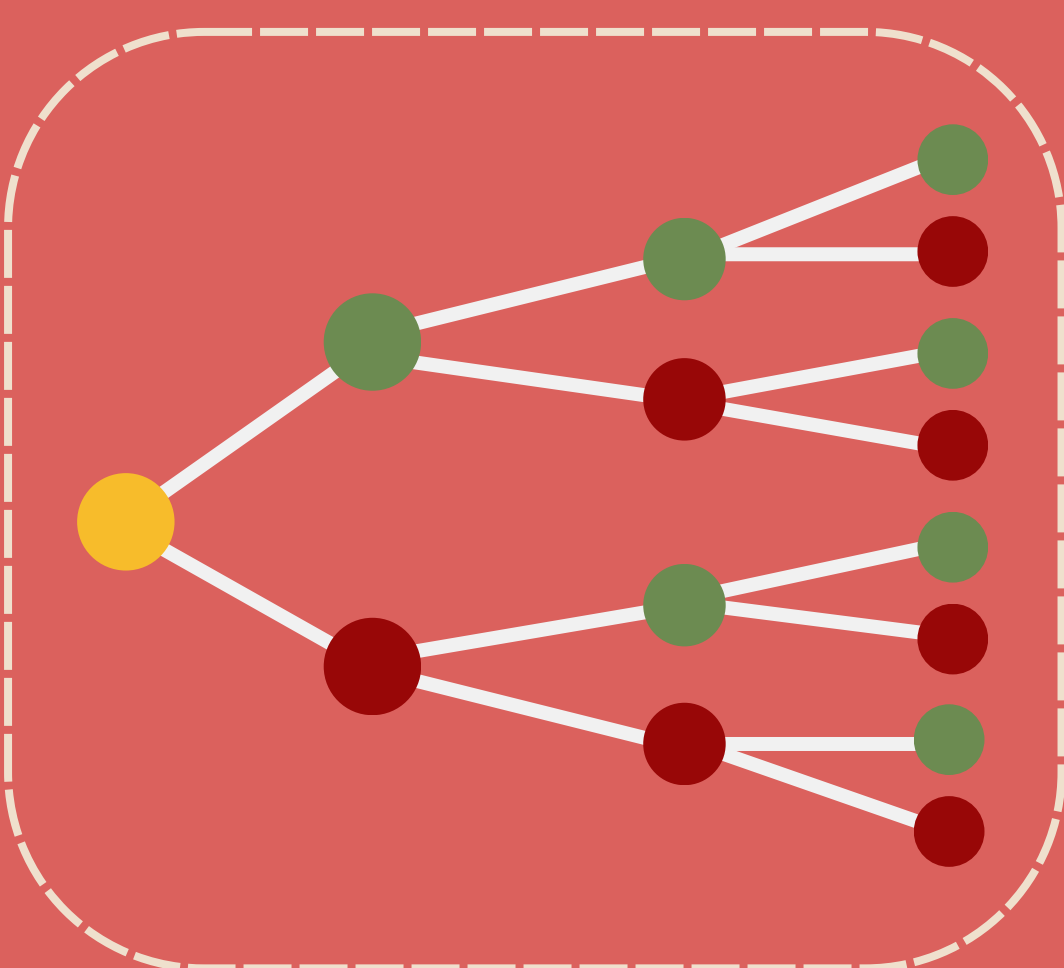


To compare models for predicting heart disease
using the function of a Decision Tree, Random
Forest and XGBoost

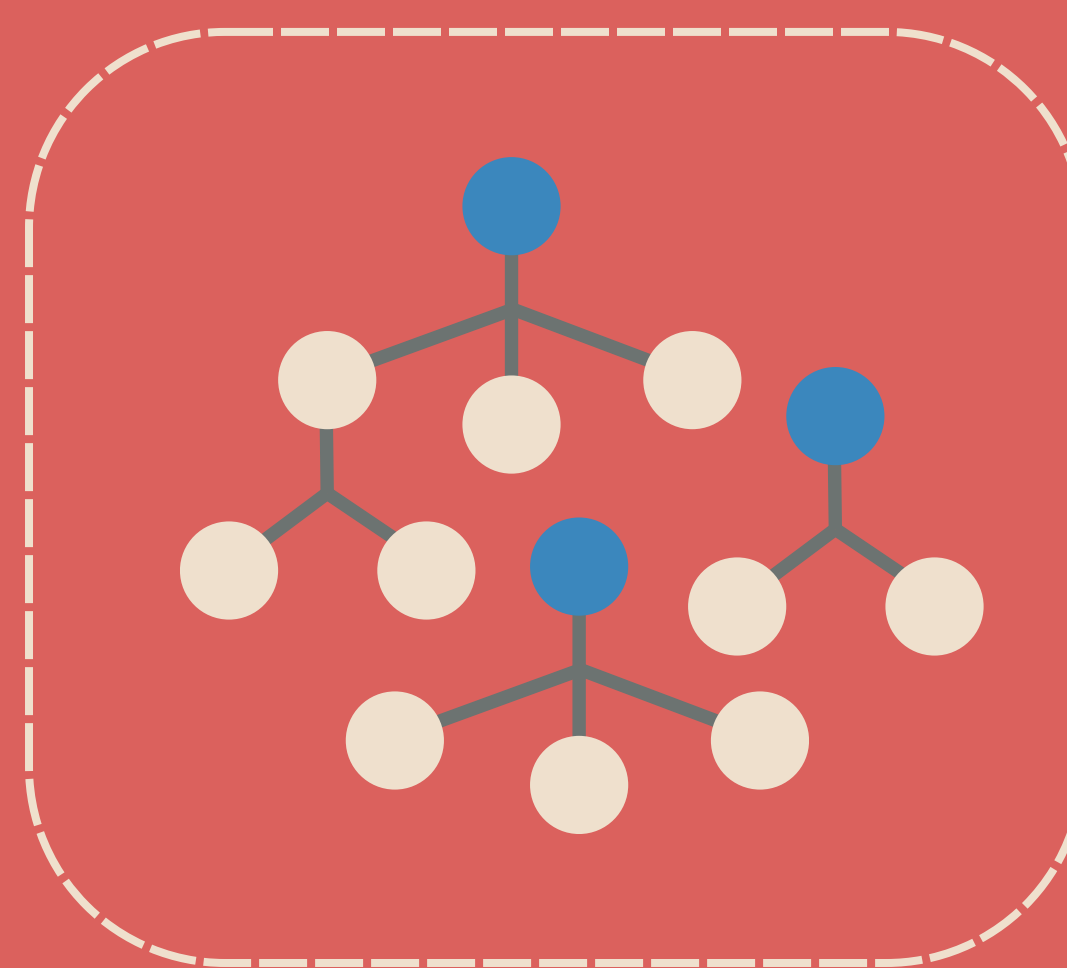
Medthology



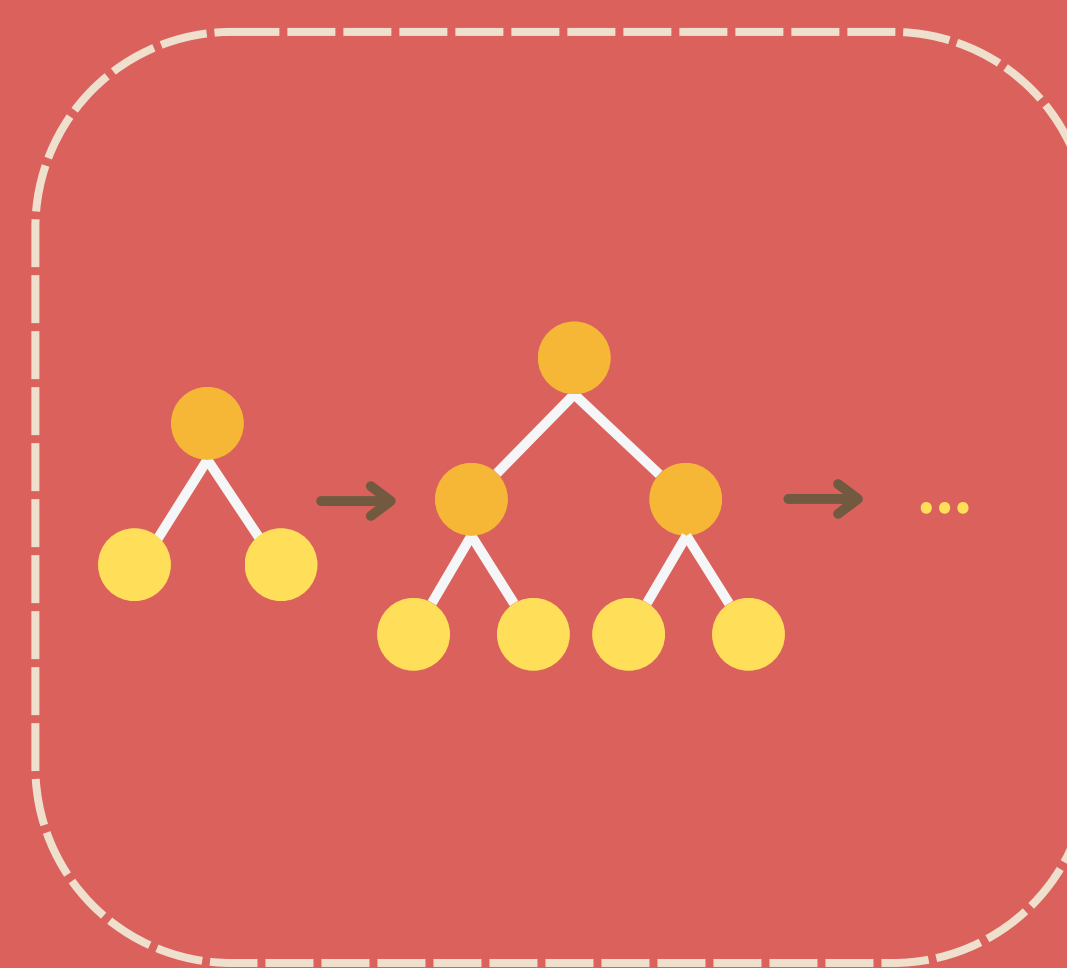
Models



Decision Tree



Random Forest

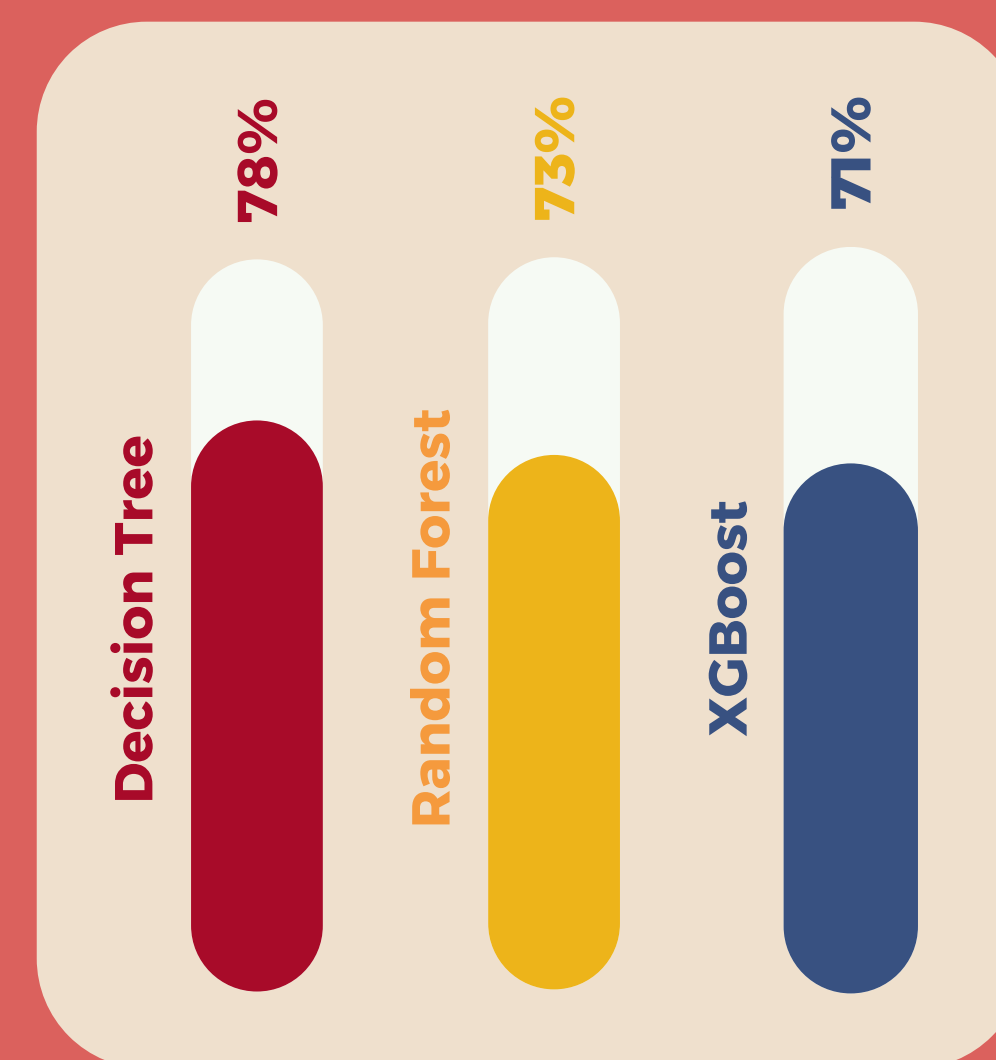
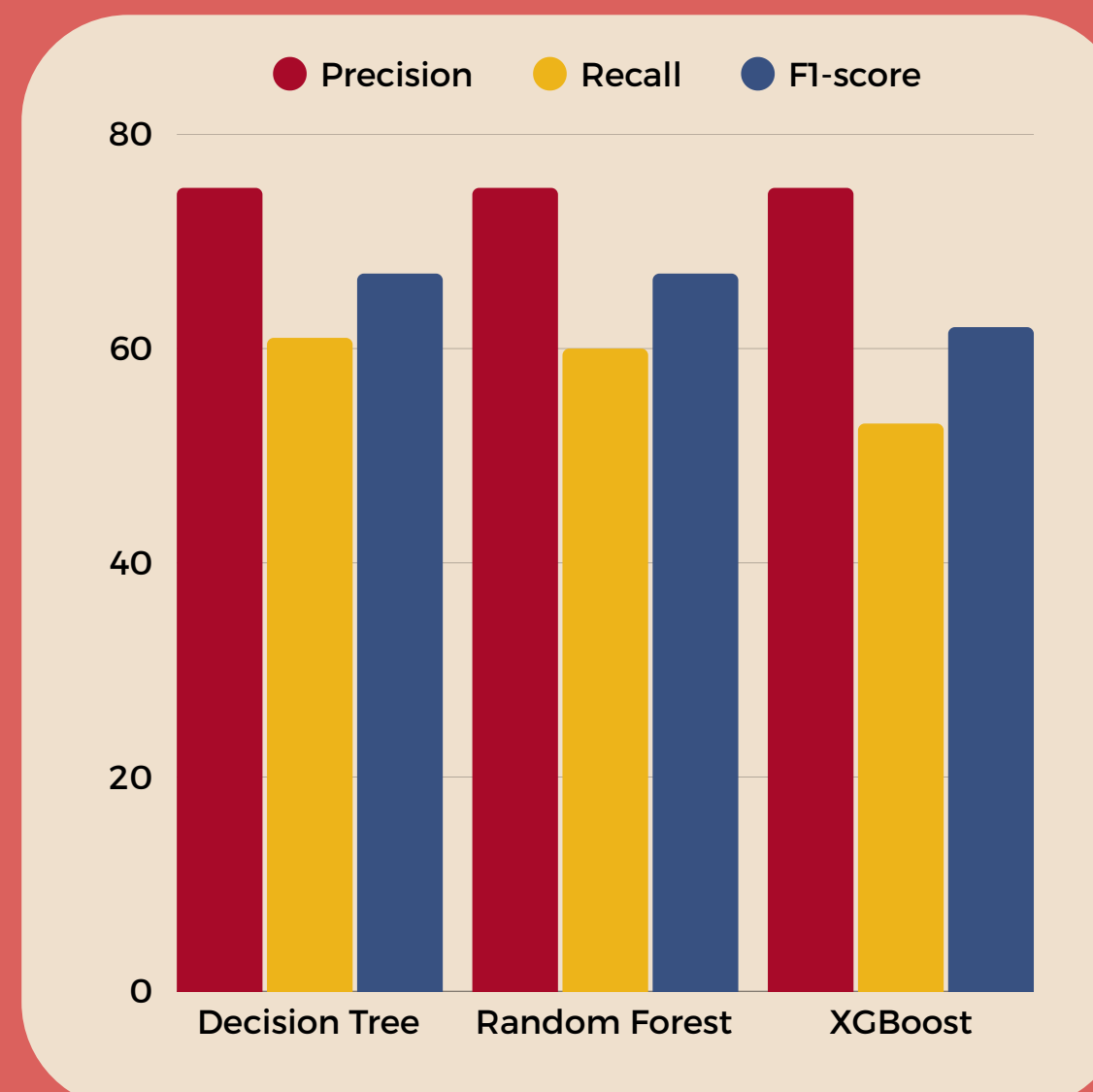
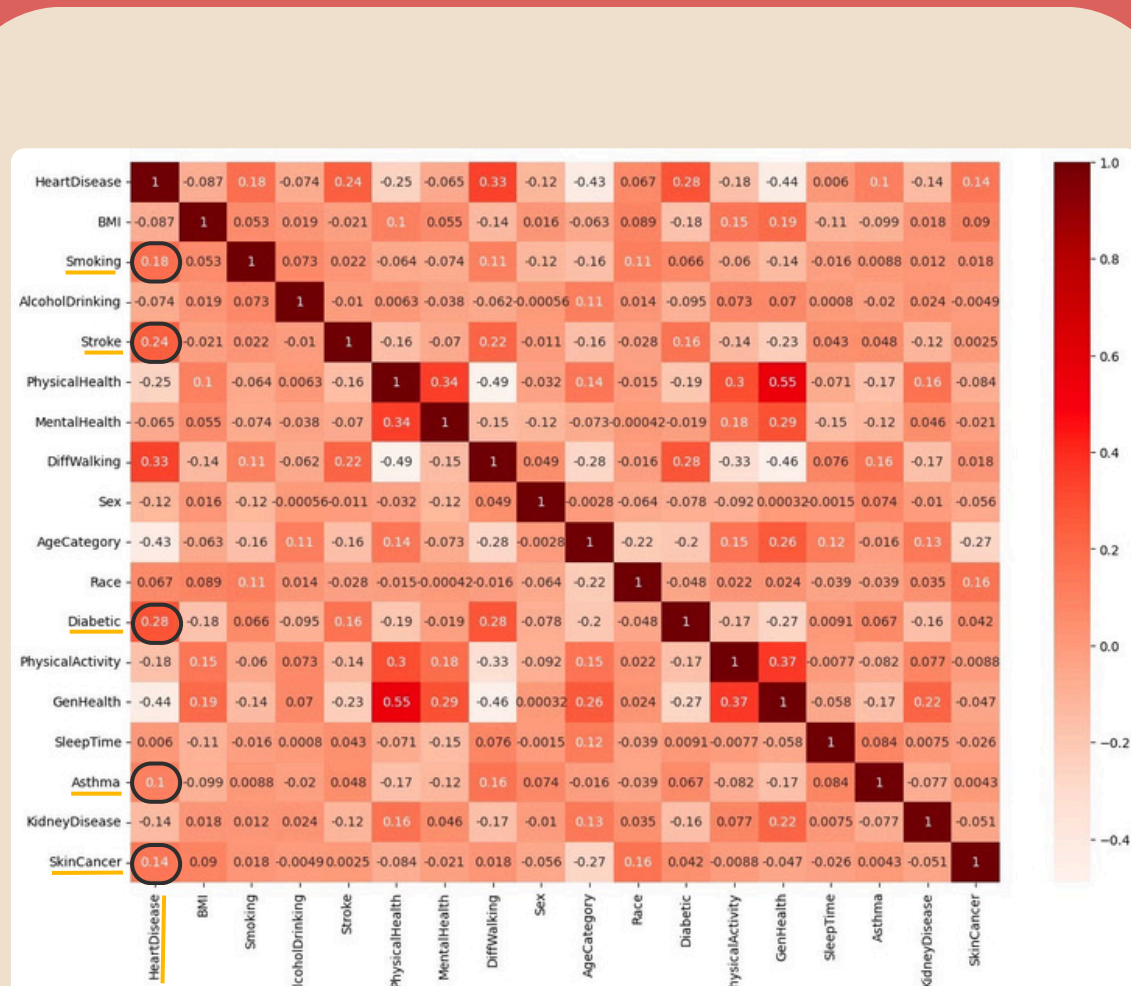


XGBoost

Factors



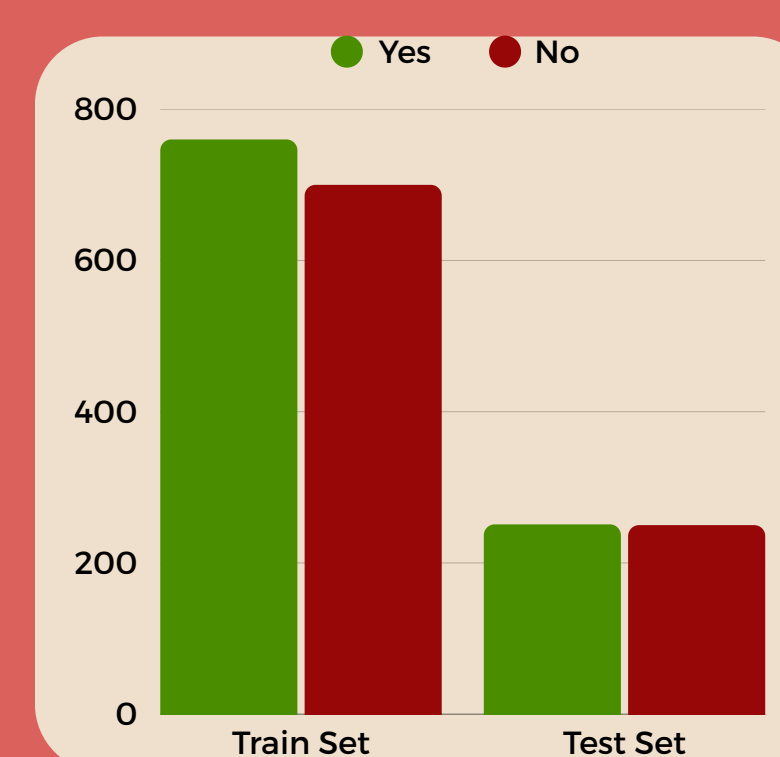
Result



Conclusion and Discussion

From the heatmap, the variables that are related to heart disease are smoking, stroke, diabetes, diff walking and skin cancer. When trained in the model using Random forest, the results are 275 people with heart disease and 509 people without heart disease. According to the test set that is used to predict, there are 309 people with heart disease and 475 people without heart disease. Therefore, this model has an accuracy value of 78%.

The student's application of Random Forest to examine the correlation between factors influencing heart disease can be leveraged for medical purposes. Additionally, it can also be used to predict financial investment risks and various other applications.



	No	Yes
No	120 TN	78 FP
Yes	41 FN	210 TP

Reference

Paolohospital. (2023). Heart disease. Why do so many Thai people have it? Check your heart health to help reduce your risk. Retrieved Sep 9, 2023, from <https://www.paolohospital.com/en-US/phahol/Article/Details/Heart/Cardiovascular-Disease>
Witchapong, D. (2018). Delving into Random Forest!!! – Part 2 of “Get to know Decision Tree, Random Forest, and XGBoost!!!” Retrieved Sep 10, 2023, from <https://medium.com/@witchapongdaroontham/เจาะลึก-random-forest-part-2-of-รู้ลึก-decision-tree-random-forest-และ-xgboost-79b9f41a1c1c>