Predicting whether there is Heart Disease or Not Using Machine Learning

Saswat Panda

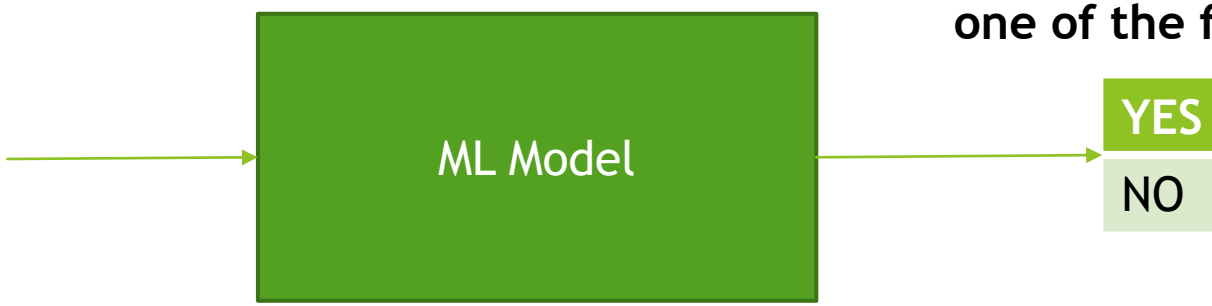
Introduction

- According to WHO, cardiovascular diseases (CVDs) is the top one killer over the world. There are seventeen million people died from it every year, especially heart disease.
- If we can evaluate the risk of every patient who probably has heart disease, that is, not only patients but also everyone can do something earlier to keep illness away.
- The dataset used is a real data of patients including important factors that might be responsible for heart disease.

ML Model

Input

Examinations of patients' body such as maximum heart rate and blood pressure.



Output:
one of the following

- YES
- NO

Predicted yes if there is heart disease

Comparison

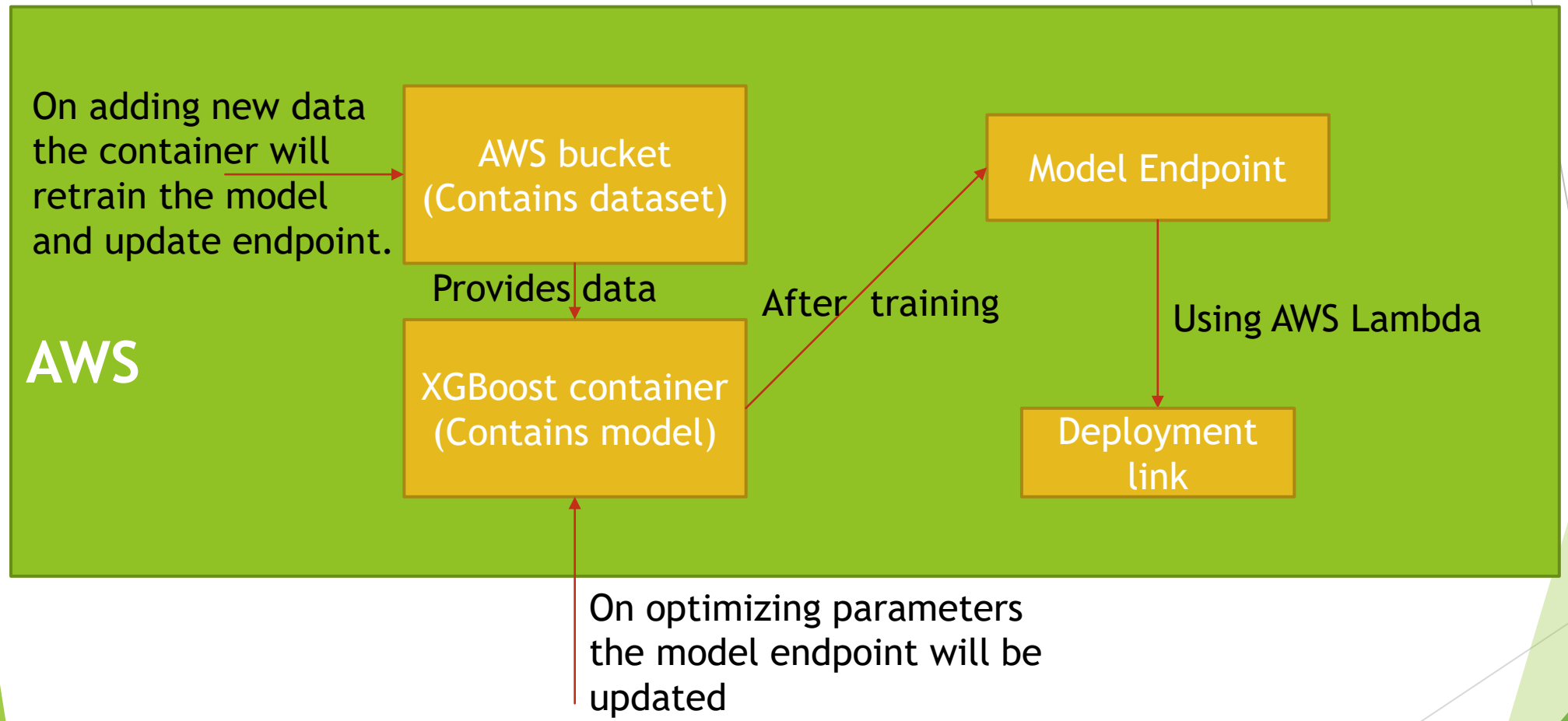
Classifier Name	Cross validation accuracy (%)
Decision-Tree	81.35
KNN	65.68
Random Forest / XGBoost	82.63
AdaBoost	75.13
Stochastic Gradient Descent	64.45

From the table it can be observed that the best model is Random forest/XGB classifier. The test accuracy obtained on using Random forest/XGB is 84%. The accuracies for classifying whether there is a heart disease or not are 85% and 81% percent respectively.

Most Important features

Feature name	Importance (%)
The number of major vessels	18.14
Chest pain experienced	18.13
Depression induced by exercise relative to rest	15.33
Thalassemia	13.24
Maximum heart rate achieved	12.59

ML Model deployment



The above cycle provides automated workflow, where you can add new training data or optimize model parameter without disturbing deployment.