Heart Disease Prediction Model Report

BY OLOMO AYOBAMI

Content Layout

- Aim & Overview of the Prediction Model
- Data Exploration
- Model Development & Analysis
- Important Features & Business Approach

Aim & Overview of the Prediction Model

- The goal of this project is to develop a machine learning model that will detect if a patient has heart disease or not. Also, test the model with a heart disease dataset developed by Medical Practitioners in University Hospitals, Medical Centers and Cardiology Clinics in Cleveland.

- This prediction model will be used by the hospital cardiologists. It will aid the cardiologists in reconfirming a patient has heart disease or not based on their symptoms and most crucial test results.

- It will also be used by Clinical Laboratory Scientist and Nurses who work in heart specialist hospitals, Teaching and General Hospitals to determine the heart disease status of their patients.

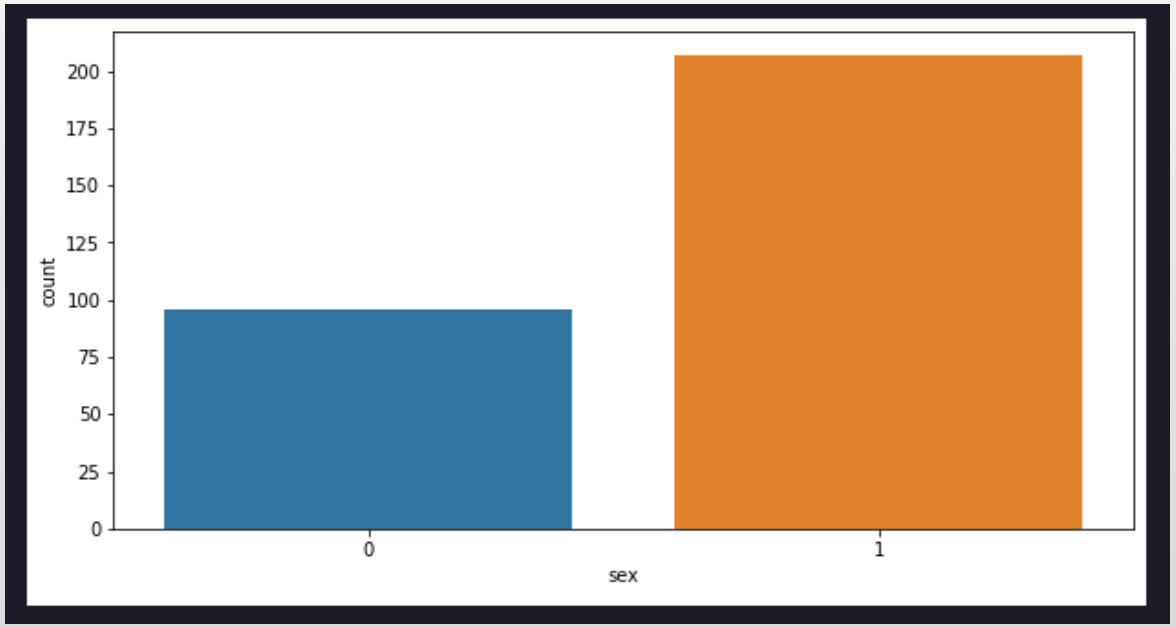
- 13 attributes were used to determine the heart condition of each patient in the dataset.
- 13 attributes were used to build the heart disease prediction model.
- The 13 attributes are the independent variables, and the target value is the dependent variable.
- Independent Variables = clinical tests, symptoms and patients biometrics.
- Dependent Variable = Heart Disease determination.

- Our Machine Learning System is a Supervised Learning Model.

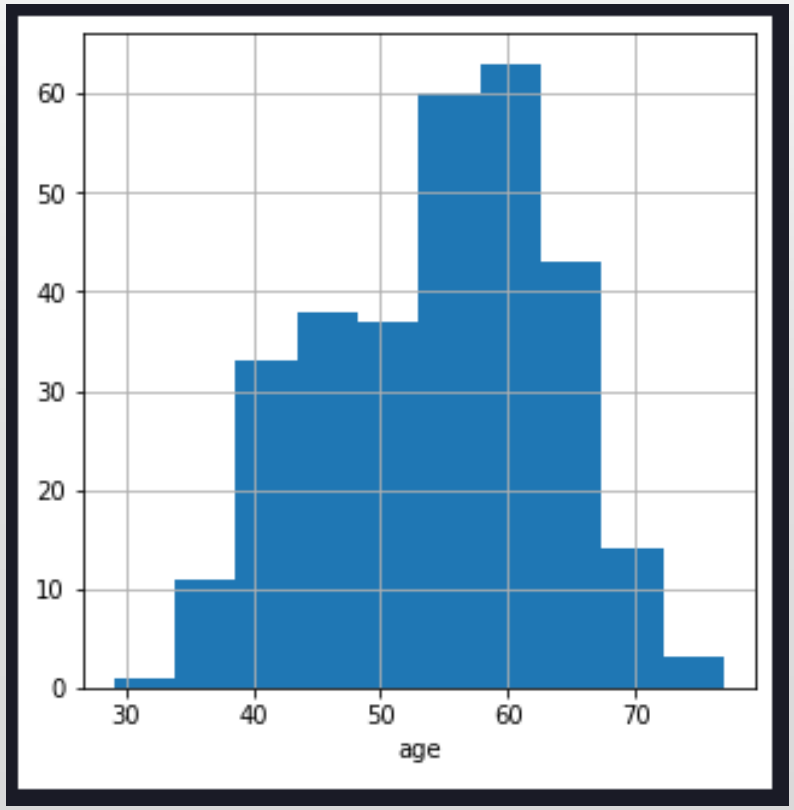
Data Exploration

Gender Results

1 = FEMALE ; 0 = MALE



Age Distribution

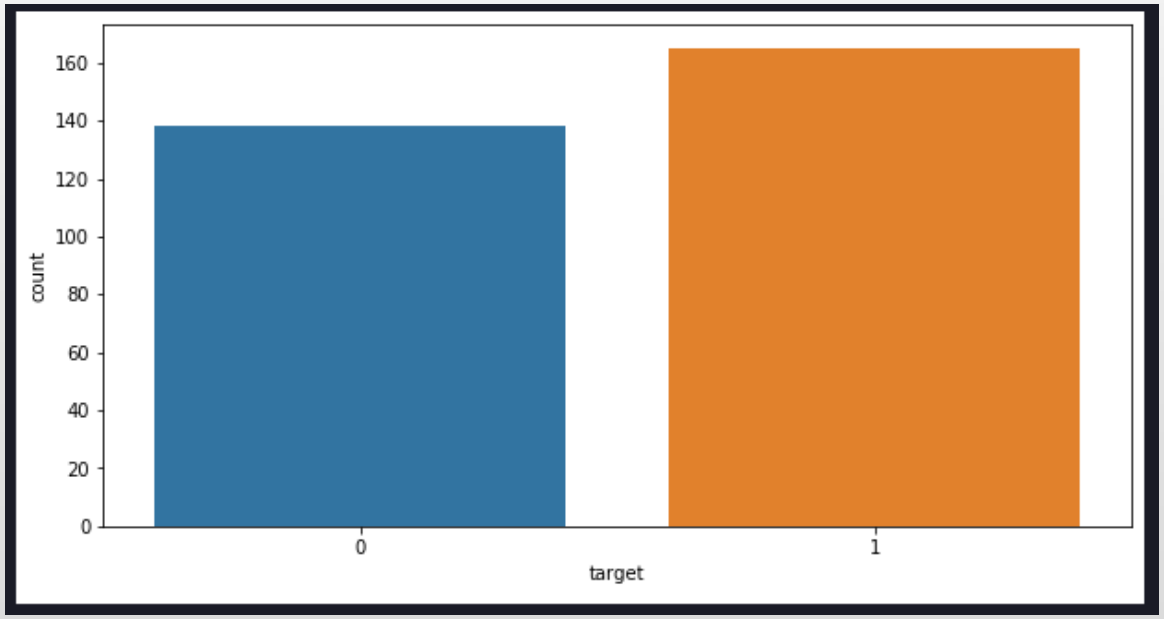


Data Exploration

Target Variable from Dataset

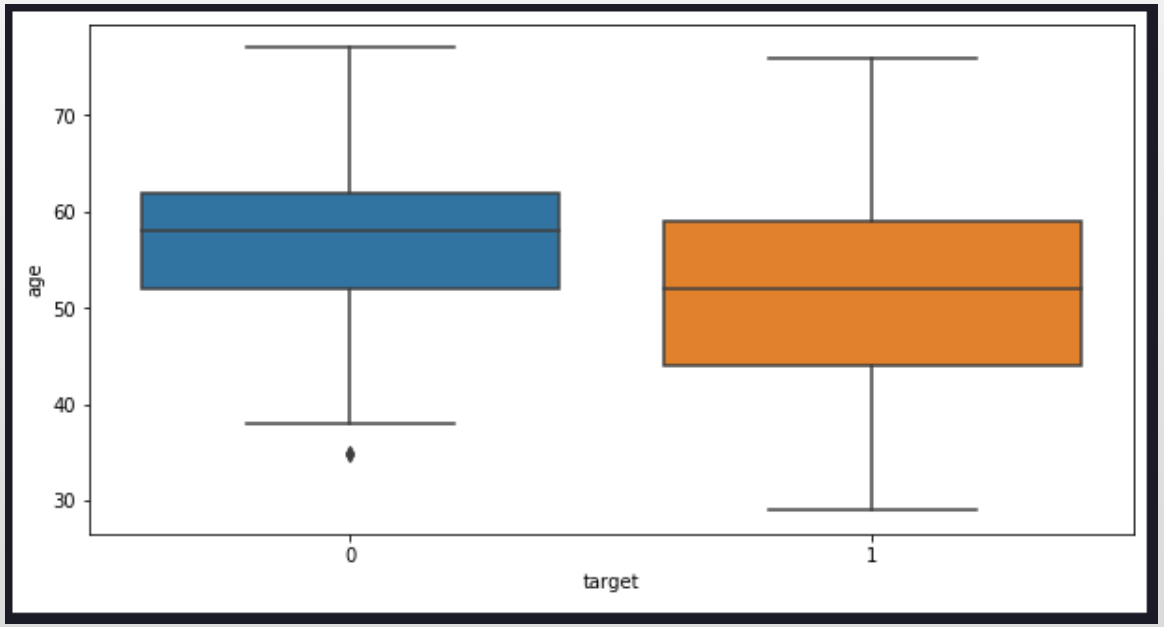
Heart Disease; 1 = Positive

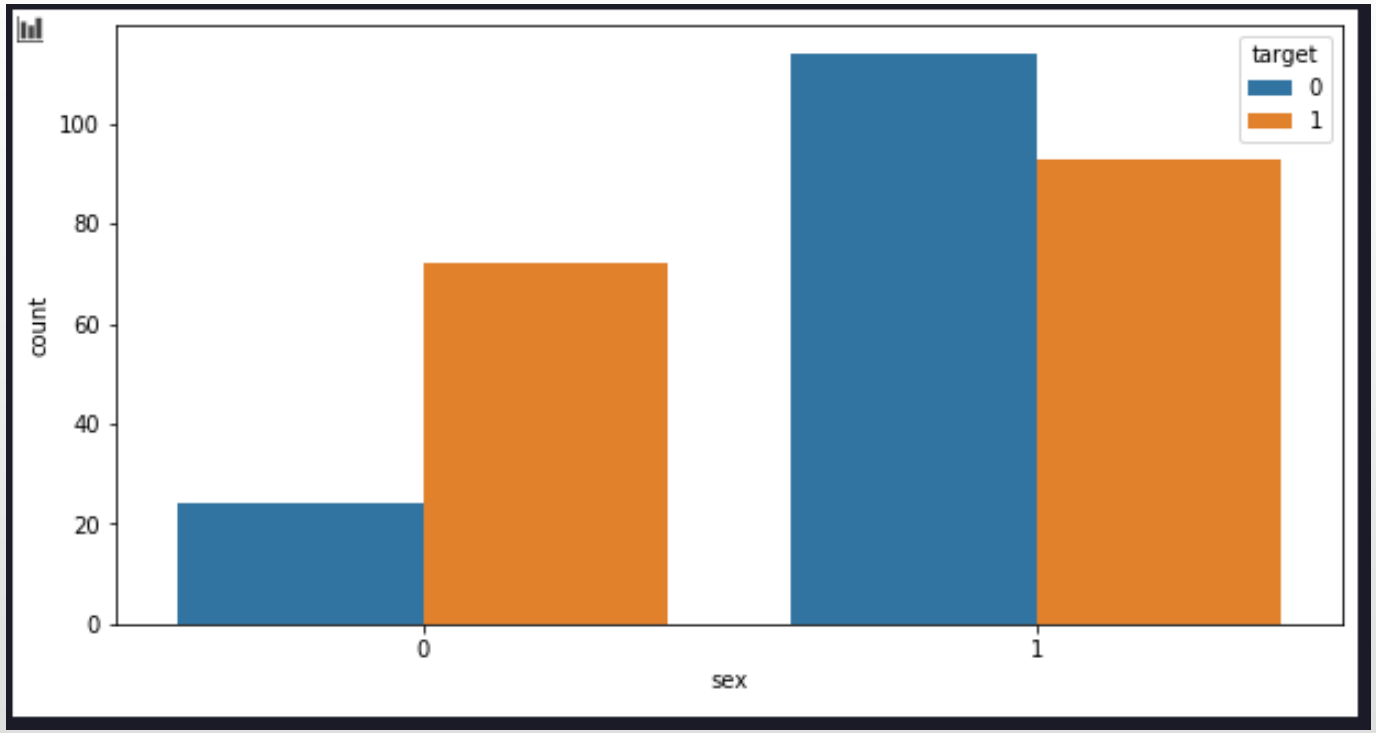
0 = Negative



Target Variable vs Age

Is heart disease age related?





Target Variable vs Gender

Is heart disease gender related?

Model Development & Analysis

ITERATION PROCESS & RESULTS

Iteration Process

- The reasons for the iterative process are as follows:
- To pick the model with the best performance.
- Also, to discover the most important features that help the model to make a highly accurate prediction, and the attributes the doctors should look out for during their decision making process.
- Lastly, to create a balanced model. That can true positives, true negatives, false positives and false negatives.

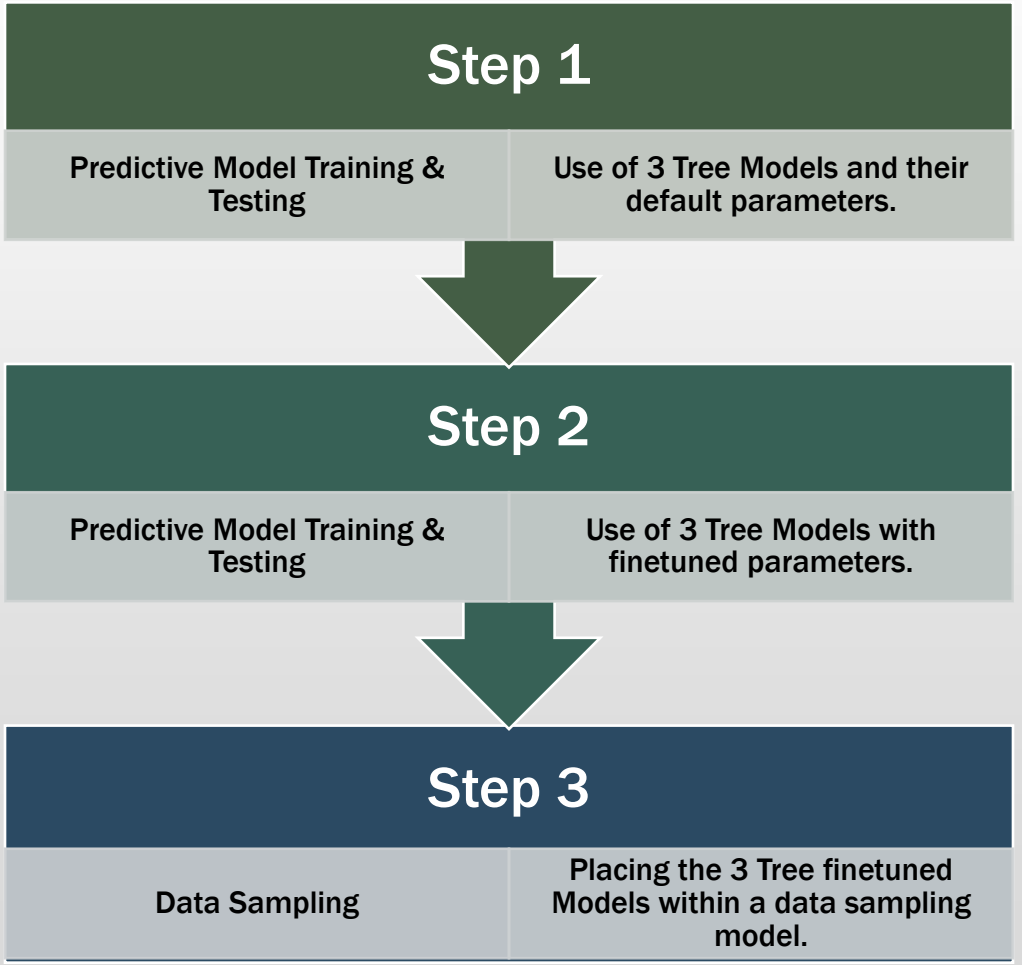


Table of Models Result

	A	B	C	D	E	F	G	H	I
1					Get scores for the target you are more interested in				
2	Runs	Algorithm	Data Sampling	Parameters	Accuracy	Precision	Recall	F-Score	AUC
3	1	Decision Tree	Normal	Default Parameters	0.69	0.64	0.73	0.68	
4	2	Random Forest	Normal	Default Parameters	0.8	0.73	0.9	0.8	
5	3	GBM	Normal	Default Parameters	0.78	0.71	0.85	0.78	
6	4	Decision Tree	Normal	entropy, depth=12, leaf =10	0.79	0.72	0.88	0.79	
7	5	Random Forest	Normal	estimators=1000, depth=6, leaf=10, jobs= -1	0.8	0.73	0.9	0.8	0.91
8	6	GBM	Normal	deviance, learning_rate=0.001, estimators=1000,depth=7, leaf=10	0.78	0.71	0.88	0.78	
9	7	Decision Tree	Combination	entropy, depth=12, leaf =10	0.86	1	0.71	0.83	
10	8	Random Forest	Combination	estimators=1000, depth=6, leaf=10, jobs= -1	1	1	1	1	
11	9	GBM	Combination	deviance, learning_rate=0.001, estimators=1000,depth=7, leaf=10	0.93	1	0.86	0.92	
12	10	Random Forest	Ensemble	estimators=1000, depth=6, leaf=10, jobs= -1	1	1	1	1	

- The best model for each sampling method is highlighted in pink.
- The overall best model is highlighted in blue. It is the best model because it has the best accuracy score, it is a balanced model and it can accurately predict false negatives, true positives and has an impressive f1-score. This implies the model will be able to predict when a person has heart disease and when and vice versa.
- I choose recall as the most important score.
- It is detrimental to have false negatives for a heart disease condition.

	feature	importance_score
5	thalach	0.28898
0	age	0.13592
7	oldpeak	0.09682
8	ca	0.08888
13	thal_2	0.08795

Important Features & Business Approach

- These 4 attributes/ features are important tests that should be carried out on patients suspected for having heart disease.
- Also the results of these 4 tests should be critically analyzed when carried out, in order to avoid mistakes in the diagnosis of patients.
- Also age has a large role to play in heart disease conditions.

Thank You!!!

For your audience.