Heart Disease Prediction

Azure ML

Date: 21st November 2021
Created by: Varun Bhavnani
No. of Pages: 6

Document Details

Document Name	Document Owner	Date of release	Version
Heart Disease Prediction using Azure ML	Varun Bhavnani	21st November 2021	V1.0

Contents

1. Problem Description

1.1. Problem Statement

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict possible heart disease.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidemia, or already established disease) need early detection and management wherein a machine learning model can be of great help.

1.2. Business requirements

Following are the business requirements

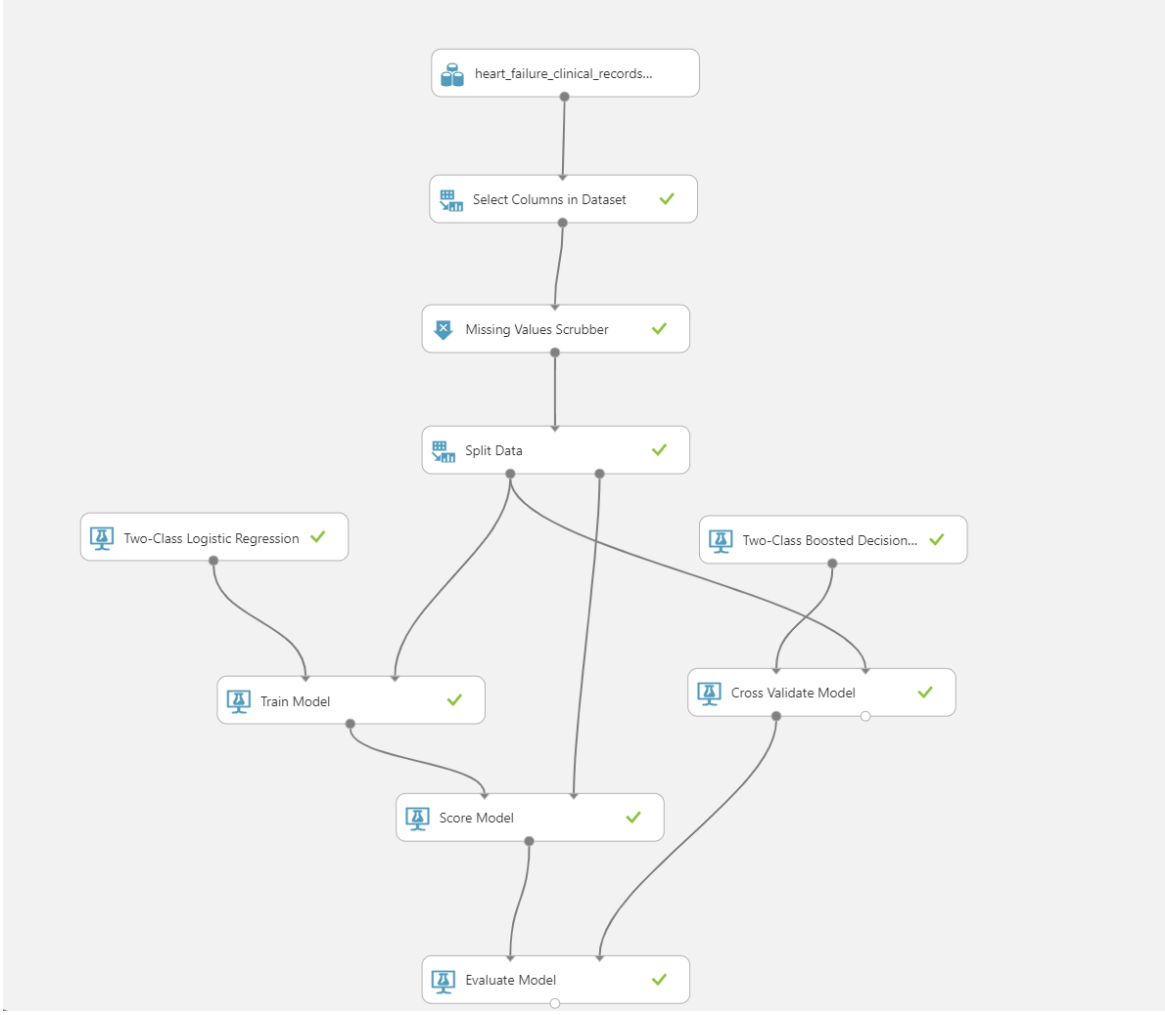
1. Perform exploratory data analysis and draw out key insights
2. Build predictive model:
    a. Build two-class decision tree and evaluate the accuracy of the model
    b. Build two-class logistics regression and evaluate the accuracy of the model
3. Deploy the model and create the web server
4. Usrr Interface: Develop a user interface that will take the inputs from the end user and predict the heart disease in less than a minute

1.3. Data Description

1. AGE: Age of the individual
2. ANAEMIA: Anaemia is a deficiency in the number or quality of red blood cells in your body.
3. CREATININE_PHOSPHOKINASE: Creatine kinase or creatine phosphokinase is an enzyme chiefly found in the brain, skeletal muscles, and heart. An elevated level of creatine kinase is seen in heart attacks, when the heart muscle is damaged, or in conditions that produce damage to the skeletal muscles or brain.
4. DIABETES: People with diabetes are also more likely to have heart failure. Heart failure is a serious condition, but it doesn’t mean the heart has stopped beating; it means your heart can’t pump blood well. This can lead to swelling in your legs and fluid building up in your lungs, making it hard to breathe. Heart failure tends to get worse over time, but early diagnosis and treatment can help relieve symptoms and stop or delay the condition from getting worse.
5. EJECTION_FRACTION: A normal ejection fraction is more than 55%. This means that 55% of the total blood in the left ventricle is pumped out with each heartbeat. Heart failure with reduced ejection fraction happens when the muscle of the left ventricle is not pumping as well as normal. The ejection fraction is 40% or less.
6. HIGH_BLOOD_PRESSURE: If you have heart failure, there's a good chance you also have high blood pressure, or "hypertension." About two-thirds of people whose hearts can't pump enough blood because of the condition also have high BP or once did. Hypertension is a major risk factor for heart failure.

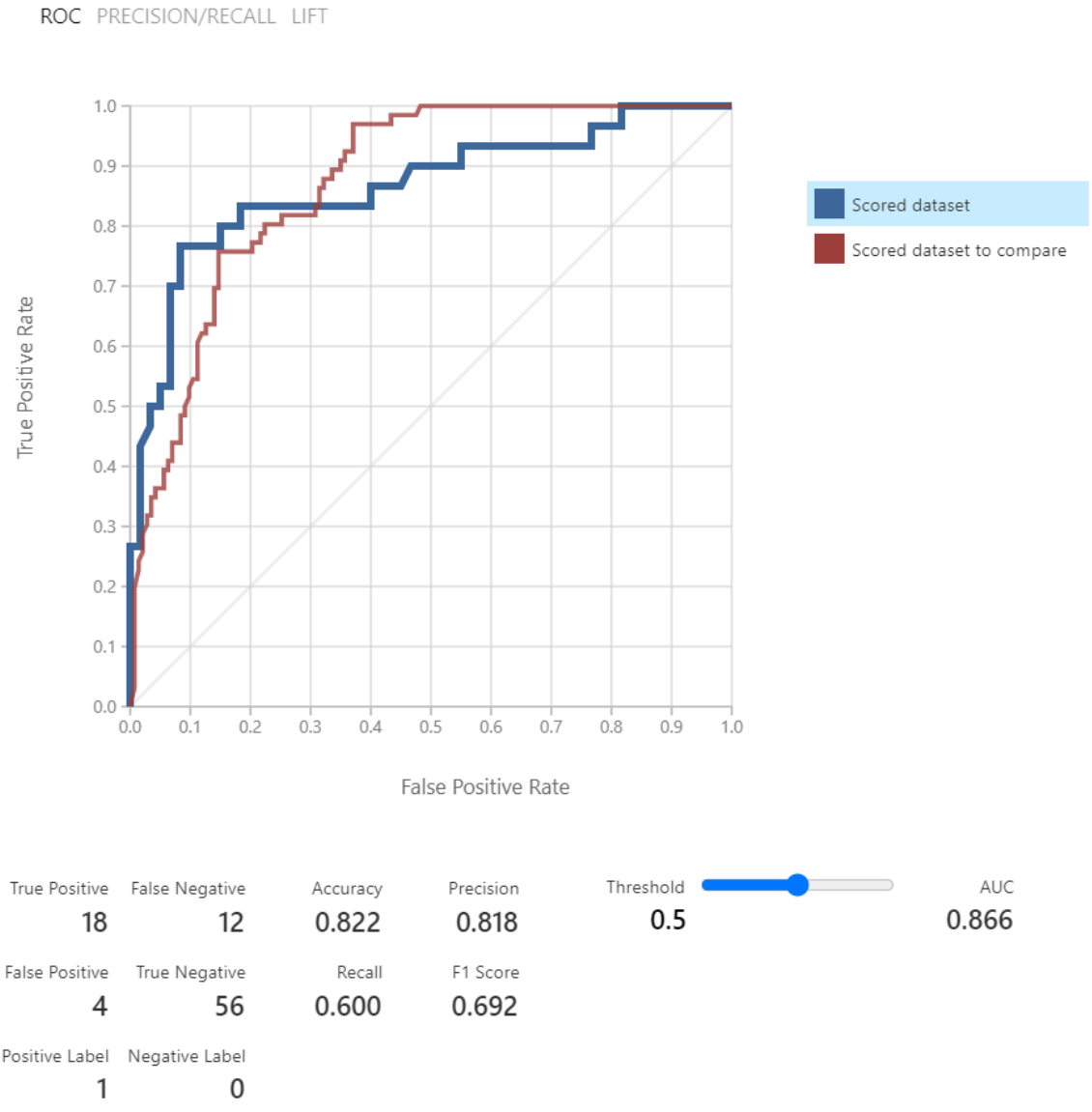
7. Platelets: Heart failure patients have increased whole blood aggregation,7 platelet-derived adhesion molecules. CHF patients also have higher mean platelet volume9 and soluble (and platelet-bound) P-selectin, regardless of the aetiology.
8. SERUM_CREATININE: Patients with severe heart failure, particularly those on large doses of diuretics for long periods, may have elevated BUN and creatinine levels indicative of renal insufficiency owing to chronic reductions of renal blood flow from reduced cardiac output.
9. SERUM_SODIUM: Hyponatremia or low serum sodium level is typically defined as a serum sodium concentration of <135 mEq/L and is one of the most common biochemical disorders featured in heart failure patients, with a prevalence close to 25% [2–4]. HF affects cardiac output by either decreasing heart rate or reducing the stroke volume
10. SEX: Gender of the individual
11. SMOKING: Somking habits of the inidividual

2. Azure ML Prediction Process Flow Diagram



Following are the steps we took in order to build the model

1. Selected the required columns
2. Cleaned the data set by using missing value scrubber
3. Split the data into 70% training data and 30% validation data
4. Trained the model using Two-class logistics regression & Two-class boosted tree
5. Extra step of cross validation was applied for two-class boosted tree
6. Scored and evaluated the model



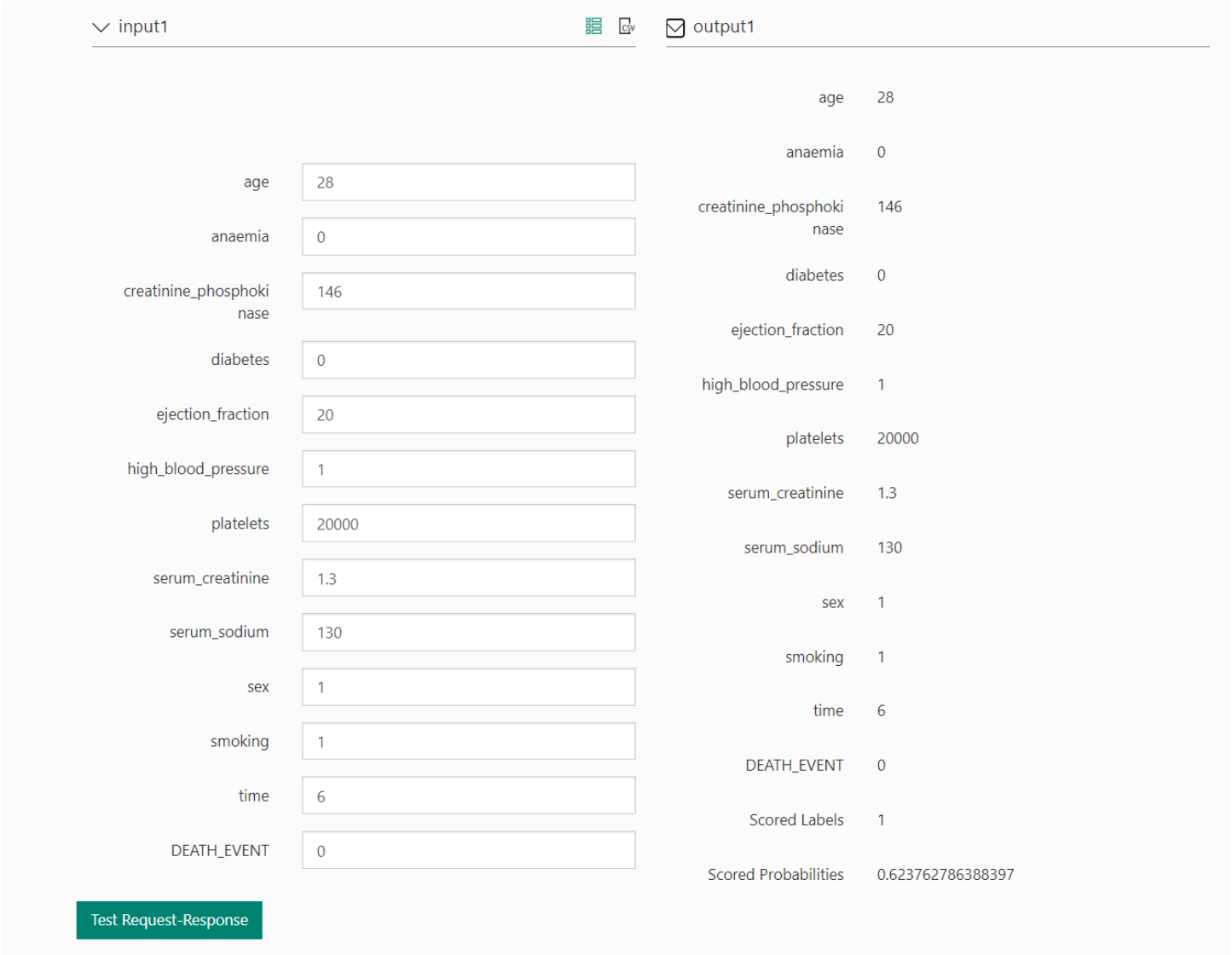
From the above graph, we can interpret that the area under the curve for

- Logistics Regression: 0.866
- Boosted tree: 0.873

Higher the value of AUC, better the model. From the values above, boosted tree is a better model

3. Model Deployment using Azure Web Server

- Built-in feature in Azure ML allowed us to deploy a web server that creates a user interface with the feature of providing the input and the result
- After building the model, we deployed the model as a web service which will take the variables as input and provide the heart disease prediction as the output. Please refer to the screenshots below



- End of Document -