

Effective Heart Disease Prediction Using IBM Auto AI Service

1. Introduction

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. People with cardiovascular disease or who are at high cardiovascular need early detection and management wherein a machine learning model can be of great help. Healthcare industries generate enormous amount of data on CVDs that accommodates hidden knowledge or pattern for decision making. The huge volume of such data may be used by supervised learning methods to make decision which is more accurate than intuition.

In this project, we train most sophisticated binary classification algorithms using Auto AI and choose the best performing method. We also build a web application where we can showcase the prediction of heart failure using the selected best performing method.

2. Proposed Solution:

The code pattern of the project can be thought of as two distinct parts:

- i) A predictive model is built using AutoAI on IBM Cloud Pak for Data. The model is then deployed to the Watson Machine Learning service.
- ii) An application is built using Node-RED which takes inputs from the user and showcases the prediction of heart failure on UI, via an API call to the Watson Machine Learning service.

2.1 Prerequisites

- * An IBM Cloud Account (<https://cloud.ibm.com>)
- * An account on IBM Cloud Pak for Data (<https://dataplatfom.cloud.ibm.com/>).

2.2 Architecture

The project is created using Watson Studio (<https://www.ibm.com/cloud/watson-studio>). A model is created with AutoAI by uploading data. The data is backed up and stored on Cloud Object Storage. The model is deployed using the Watson Machine Learning service. A [Node.js](<https://nodejs.org/>) web app is deployed on IBM Cloud. It calls the predictive model hosted on the Watson Machine Learning service. A user visits the web app, enters their information, and the predictive model returns a response. Figure 1 depicts the whole

process. The project flow is shown step by step with screen shots. The dataset used to train the model, contains 9 features, that is used to predict mortality by heart failure.

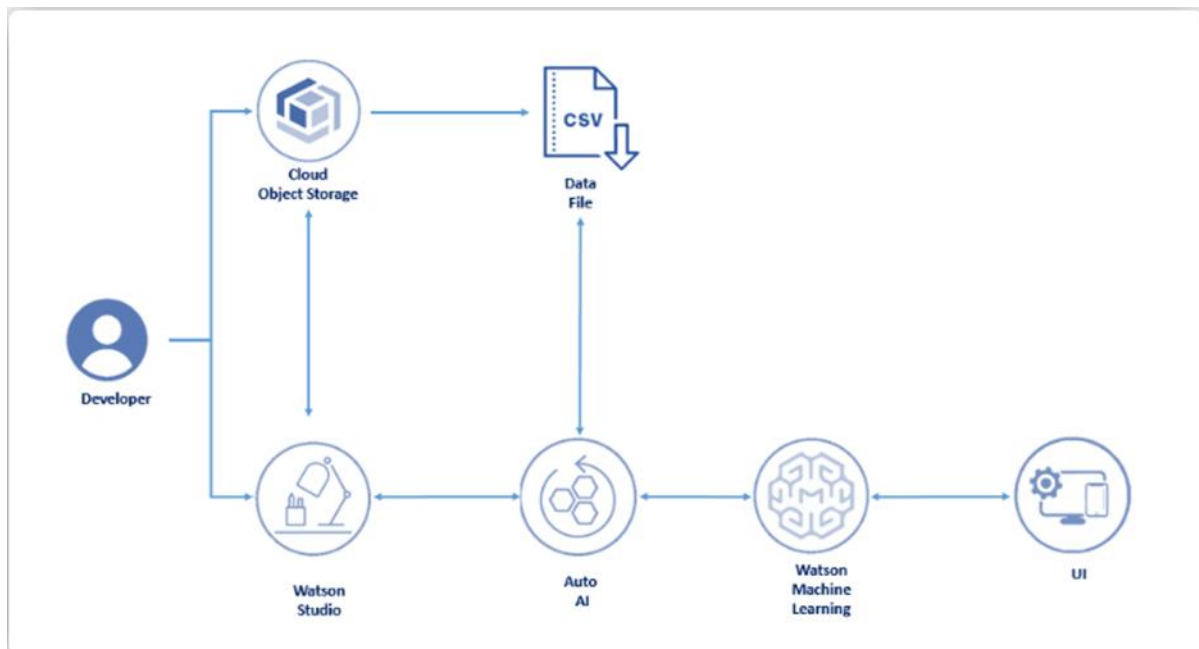


Fig 1: Architecture of the Model

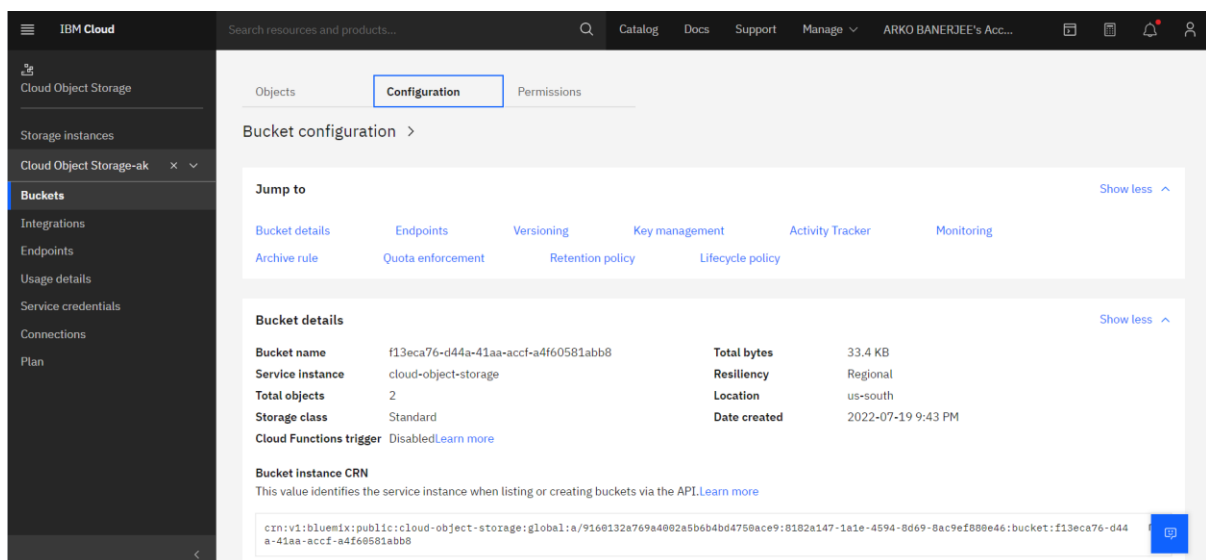


Fig 2: The data is backed up and stored on Cloud Object Storage.

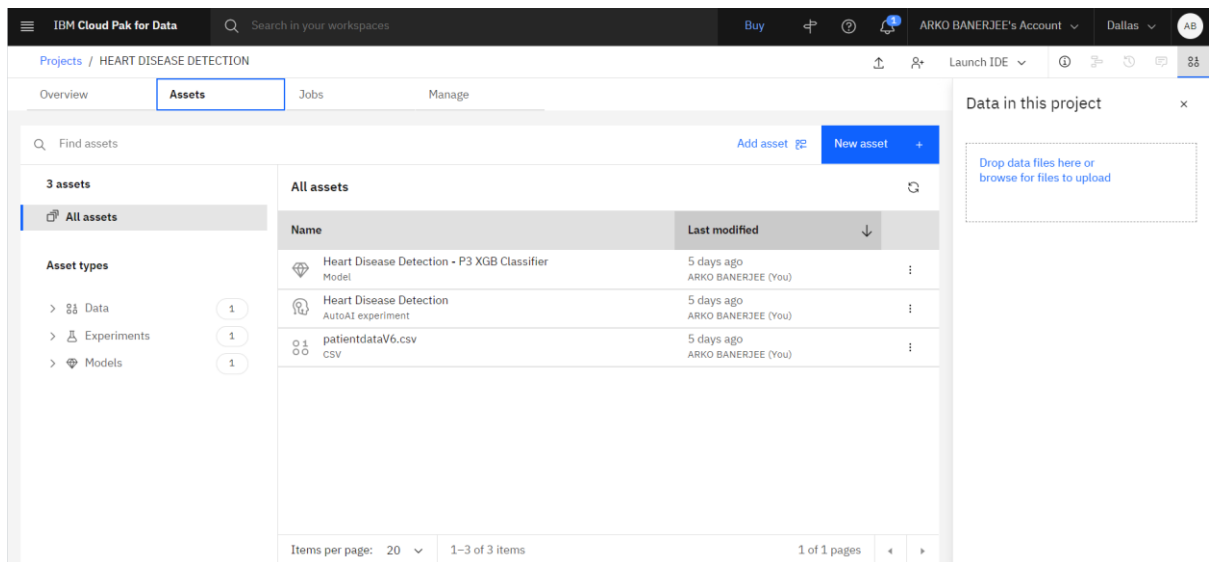


Fig 3: A model is created with AutoAI by uploading data.

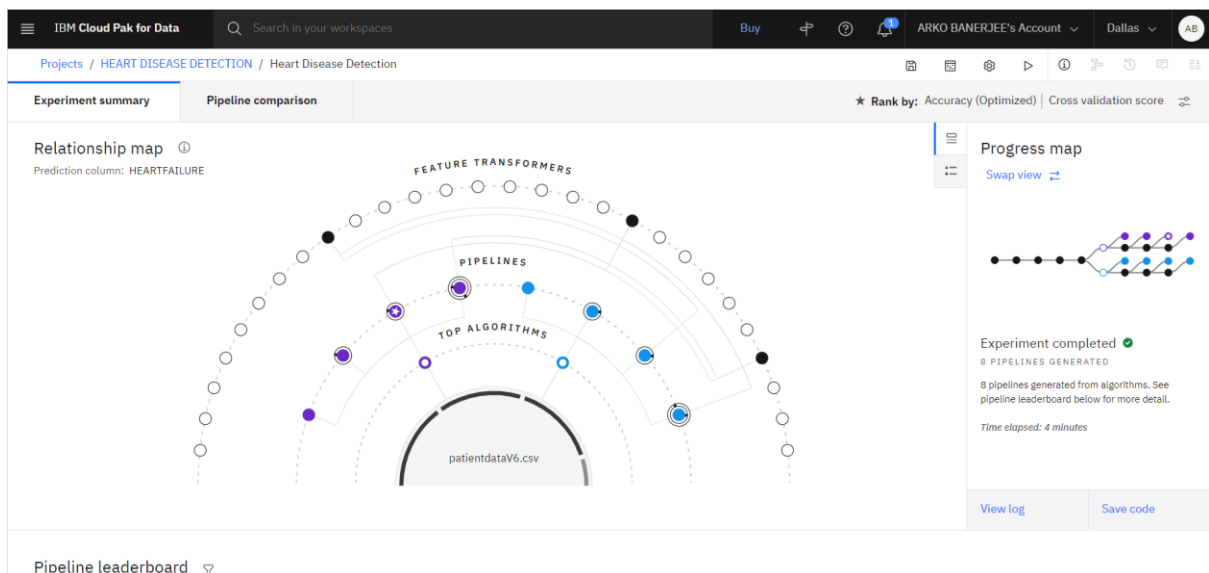


Fig 4: Binary Classification algorithms are applied by Auto AI

IBM Cloud Pak for Data

Search in your workspaces

Buy

ARKO BANERJEE's Account

Dallas

AB

Projects / HEART DISEASE DETECTION / Heart Disease Detection

Experiment summary

Pipeline comparison

★ Rank by: Accuracy (Optimized) | Cross validation score

Pipeline leaderboard

	Rank	↑	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1		Pipeline 3	XGB Classifier	0.873	HPO-1 FE	00:00:45
	2		Pipeline 4	XGB Classifier	0.873	HPO-1 FE HPO-2	00:01:36
	3		Pipeline 1	XGB Classifier	0.869	None	00:00:01
	4		Pipeline 2	XGB Classifier	0.869	HPO-1	00:00:13
	5		Pipeline 7	Extra Trees Classifier	0.858	HPO-1 FE	00:00:38
	6		Pipeline 8	Extra Trees Classifier	0.858	HPO-1 FE HPO-2	00:01:01
	7		Pipeline 5	Extra Trees Classifier	0.857	None	00:00:01

Fig 5: Binary Classification methods are sorted according to their performances.

Pipeline details

Pipeline 3

Rank

1

Accuracy (Optimized)

0.869 (Holdout)

Algorithm

XGB Classifier

Enhancements

HPO-1FE

Save as

Model viewer

Model information

Feature summary

Evaluation

Model evaluation

Confusion matrix

Precision recall

Feature name	Transformation	Feature importance
FAMILYHISTORY	None	100.00%
BMI	None	12.00%
SMOKERLAST5YRS	None	8.00%
PALPITATIONSPERDAY	None	7.00%
EXERCISEMINPERWEEK	None	6.00%
AVGHEARTBEATSPERMIN	None	6.00%
CHOLESTEROL	None	4.00%
NewFeature_0	sum(AVGHEARTBEATSPERMIN, PALPITATIONSPERDAY)	4.00%
NewFeature_2	sum(AVGHEARTBEATSPERMIN, AGE)	3.00%
NewFeature_5	sum(PALPITATIONSPERDAY, EXERCISEMINPERWEEK)	1.00%

Fig 6: Importance of features in the data is evaluated by the model

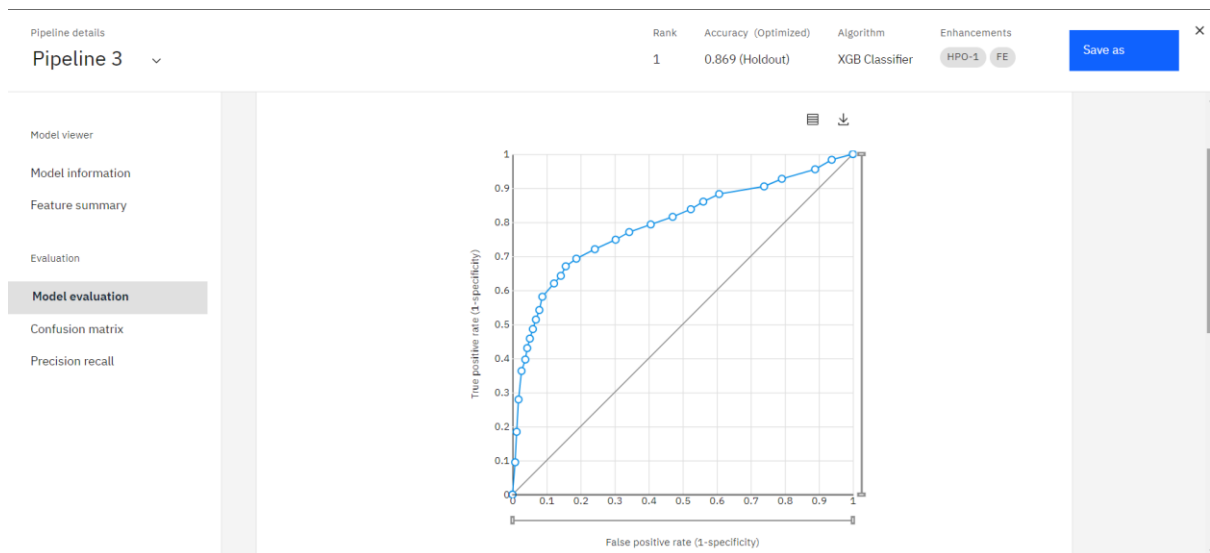


Fig 7: ROC



Fig 8: Confusion Matrix

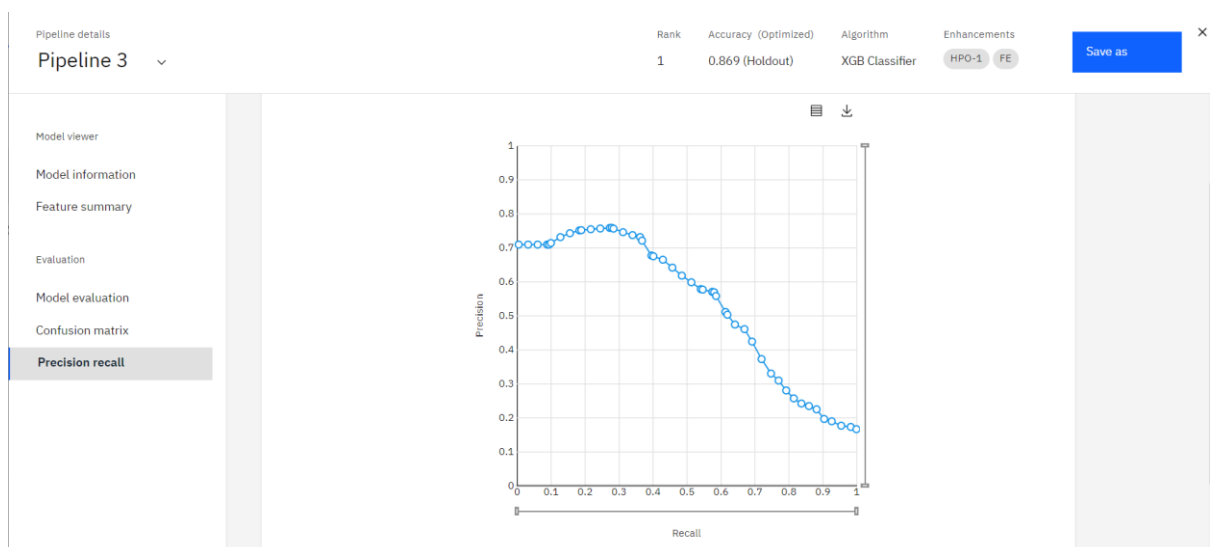


Fig 9: Precision Recall

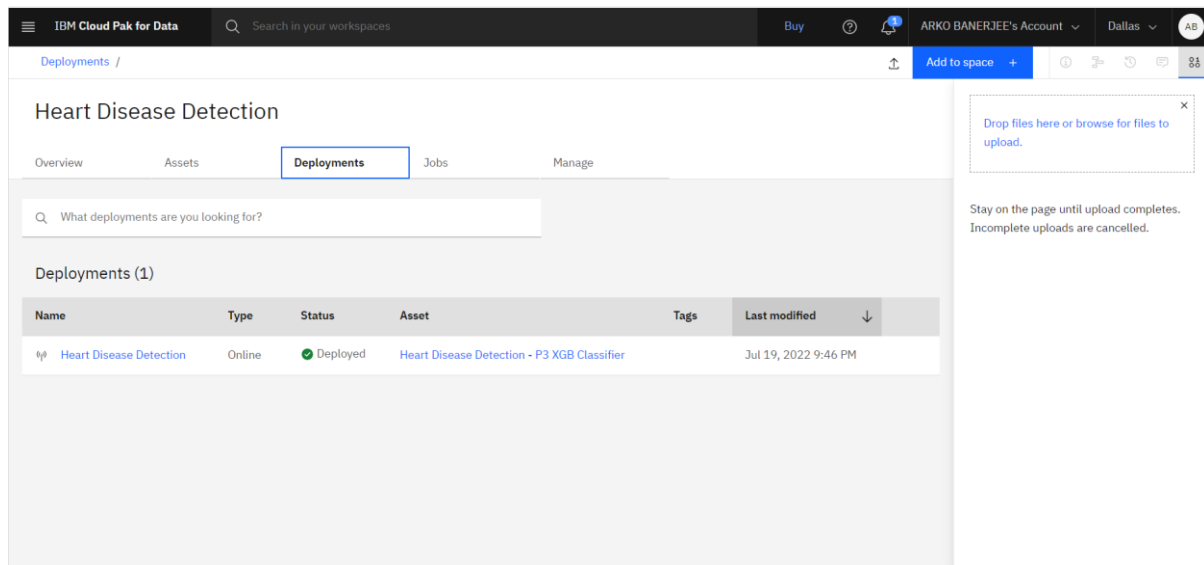


Fig 10: Deployment of Model

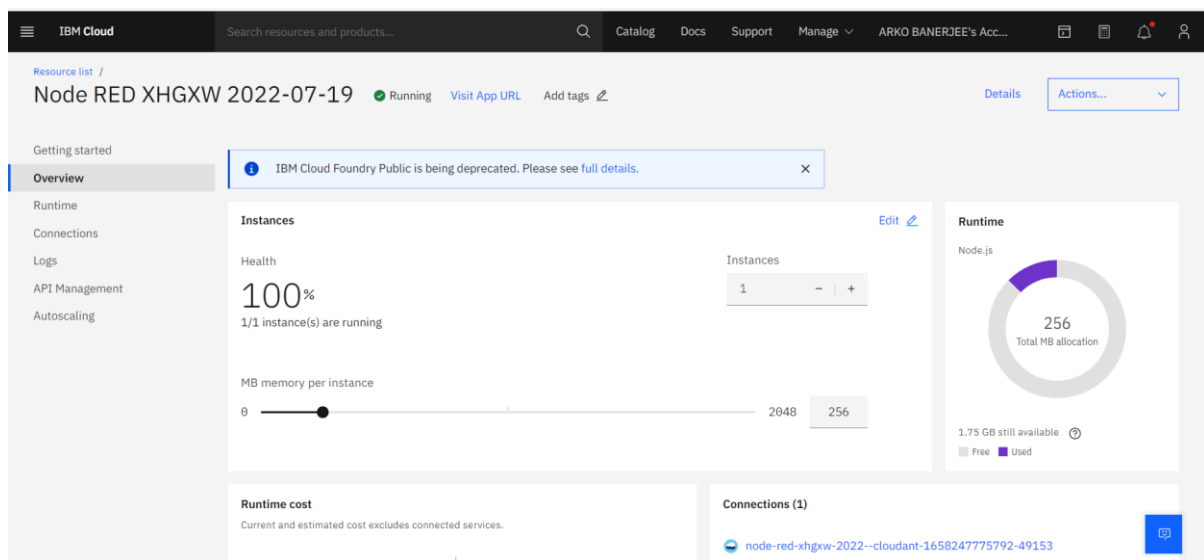


Fig 11: Node RED app creation

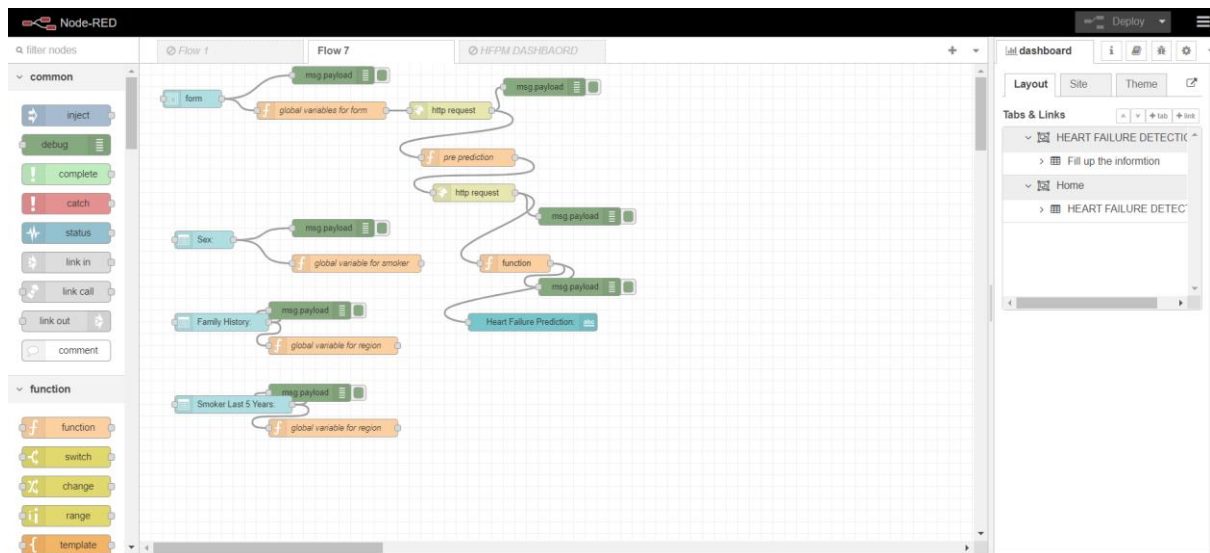


Fig 12: Node RED dashboard

The image shows the 'HEART FAILURE DETECTION' form. The form is titled 'HEART FAILURE DETECTION' and has a 'Fill up the information' section. The form contains the following fields: 'Sex' (Select option), 'Family History' (Select option), 'Smoker Last 5' (Select option), 'Avg Heart Beats per Mins *' (text input), 'Palpitations per Day *' (text input), 'Cholesterol *' (text input), 'BMI *' (text input), 'Age *' (text input), and 'Exercise Min per Week *' (text input). There are 'SUBMIT' and 'CANCEL' buttons. Below the buttons is a 'Heart Failure Prediction:' label.

Fig 13: Node RED UI

HEART FAILURE DETECTION

Fill up the information

Sex: Male

Family History: Yes

Smoker Last 5: No

Avg Heart Beats per Mins *
80

Palpitations per Day *
20

Cholesterol *
120

BMI *
35

Age *
43

Exercise Min per Week *
100

SUBMITCANCEL

Heart Failure Prediction:

Fig 14: User Input values into Node RED UI

HEART FAILURE DETECTION

Fill up the information

Sex: Male

Family History: Yes

Smoker Last 5: No

Avg Heart Beats per Mins *

Palpitations per Day *

Cholesterol *

BMI *

Age *

Exercise Min per Week *

SUBMITCANCEL

Heart Failure Prediction: **N**

Fig 15: Heart Failure Prediction in Node RED UI against user's provided input