

Effective Heart Disease Prediction Using IBM Auto AI Service

Dr. Baljeet Kaur

1. INTRODUCTION

a. Overview

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. Heart failure is a common event caused by CVDs and this dataset contains 9 features that can be used to predict deathrate by heart failure.

b. Purpose

In this project, you need to build a model using Auto AI and build a web application where we can showcase the prediction of heart failure.

2. LITERATURE SURVEY

a. Existing problem

Existing approaches use Machine Learning. Data Scientist has to apply multiple algorithms and compare the accuracy of different algorithms to identify the best solution.

The given problem of heart disease prediction contains the data of various patients and contains information about the average number of heart beats per minute, a number of pulse petitions per day, cholesterol level of the patient, BMI, Age, Sex, Family History, either the person is smoker for last 5 years, how much time a person gives for exercise for week in minutes. Based on this given data as an input, the data has information about either the person had a heart failure or not. This information can be used to predict that the person will have a heart failure or not based on the other input parameters provided. Here the target variable is categorical in nature. It is a problem of classification, which can be solved by using Logistic regression or decision trees or by using random forest algorithm. Multiple solutions are possible for this given problem. Data scientist has to apply all these algorithms on the given data and then compare the accuracy of the algorithms to finalize the final machine learning algorithm for the given problem.

b. Proposed solution

The auto AI solution provided by IBM cloud environment, where we can create a service in Watson studio and provide the given data as an input, the auto AI service in Watson Studio will suggest your multiple machine learning algorithms which are suitable for the given data. The data scientist no need to worry about execution of multiple algorithms at their end, they can concentrate on the solution without spending more time on preprocessing of the data. The auto AI solution, AI experiment helps to create a model, Deploy the ML model as a webserver, Integrating Model and Node-RED Service and Build an Application using Node-RED which takes inputs from the user and showcases the prediction on UI.

3. THEORITICAL ANALYSIS

a. Block diagram

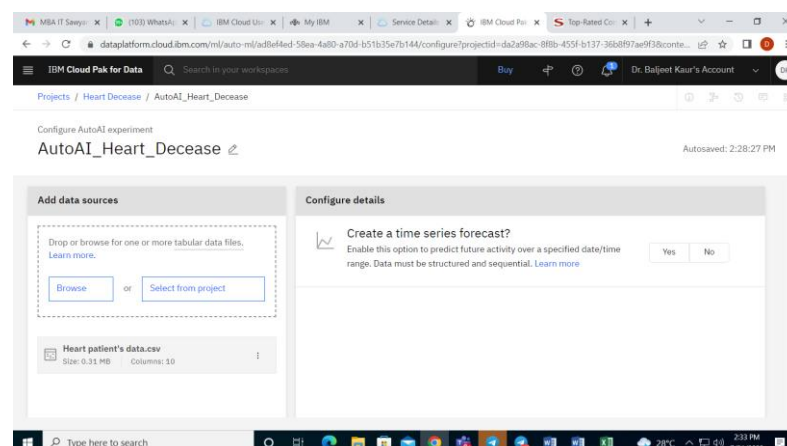


b. Hardware / Software designing

IBM Cloud Account with Cloud Object Storage service (COS)
Watson studio
Node-RED
Machine Learning service (ML)

4. **EXPERIMENTAL INVESTIGATIONS**

Creation of Auto AI Experiment for Heart Decease:



Importing Data Set and Setting up Experiment

The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes the IBM Cloud Pak for Data logo, a search bar, and user account information. The main content area is titled 'Configure AutoAI experiment' and 'AutoAI_Heart_Decease'. On the left, there's a 'Browse' button and a 'Select from project' button. Below them, a file named 'Heart patient's data.csv' is listed with a size of 0.31 MB and 10 columns. On the right, the 'What do you want to predict?' section shows 'HEARTFAILURE' as the prediction column. Below this, the 'Prediction column: HEARTFAILURE' is confirmed, and the 'CUH remaining' is 9.88 CUH. The 'PREDICTION TYPE' is set to 'Binary Classification', the 'POSITIVE CLASS' is 'Y', and it's 'OPTIMIZED FOR Accuracy & run time'. At the bottom, there's an 'Experiment settings' button and a 'Run experiment' button.

Execution of Experiment:

The screenshot shows the IBM Cloud Pak for Data interface during the execution of an experiment. The top navigation bar is the same as the previous screenshot. The main content area is titled 'Experiment settings'. On the left, there's a sidebar with 'Prediction', 'Data source', and 'Runtime' tabs. The 'Prediction' tab is selected, showing 'Prediction settings'. Under 'Prediction type', there are four options: 'Binary classification' (selected), 'Multiclass classification', 'Regression', and 'Time series forecast'. The 'Binary classification' option is described as 'Classify data into categories. Choose this if your prediction column contains two distinct categories'. Below the settings, there's a 'Cancel' button and a 'Save settings' button. The bottom part of the screenshot shows the 'Experiment summary' page. It has tabs for 'Experiment summary' and 'Pipeline comparison'. The 'Experiment summary' tab is selected, showing a 'Relationship map' and a 'Progress map'. The 'Relationship map' shows a diagram of the data flow. The 'Progress map' shows a diagram of the experiment progress. A notification box says 'Your AutoAI run has started' and 'Based on the size of the training data and configuration of your run, this run could take some time to complete.' The 'Submitted' section shows 'HEART PATIENT'S DA...' and 'The run will begin momentarily. Time elapsed: 0 seconds'.

IBM Cloud Pak for Data | Search in your workspaces | Buy | Dr. Baljeet Kaur's Account

Projects / Heart Decease / AutoAI_Heart_Decease

Experiment summary | Pipeline comparison | ★ Rank by: Accuracy (Optimized) | Cross validation score

Progress map

Prediction column: HEARTFAILURE

Read dataset → Split holdout data → Read training data → Preprocessing → Model selection

Model selection branches into two paths:

- Path 1: Selected algorithm → Hyperparameter optimization (P1) → Feature engineering (P2) → Hyperparameter optimization (P3) → P4
- Path 2: Selected algorithm → Hyperparameter optimization (P5) → Feature engineering (P6) → Hyperparameter optimization (P7) → P8

Relationship map

Swap view

Preprocessing

HEART PATIENT'S DA...

Setting default preprocessor parameters

Time elapsed: ~18 seconds

IBM Cloud Pak for Data | Search in your workspaces | Buy | Dr. Baljeet Kaur's Account

Projects / Heart Decease / AutoAI_Heart_Decease

Experiment summary | Pipeline comparison | ★ Rank by: Accuracy (Optimized) | Cross validation score

Progress map

Prediction column: HEARTFAILURE

Read dataset → Split holdout data → Read training data → Preprocessing → Model selection

Model selection branches into two paths:

- Path 1: XGB Classifier → Hyperparameter optimization (P1) → Feature engineering (P2) → Hyperparameter optimization (P3) → P4
- Path 2: Random Forest Classifier → Hyperparameter optimization (P5) → Feature engineering (P6) → Hyperparameter optimization (P7) → P8

Relationship map

Swap view

Experiment completed

8 PIPELINES GENERATED

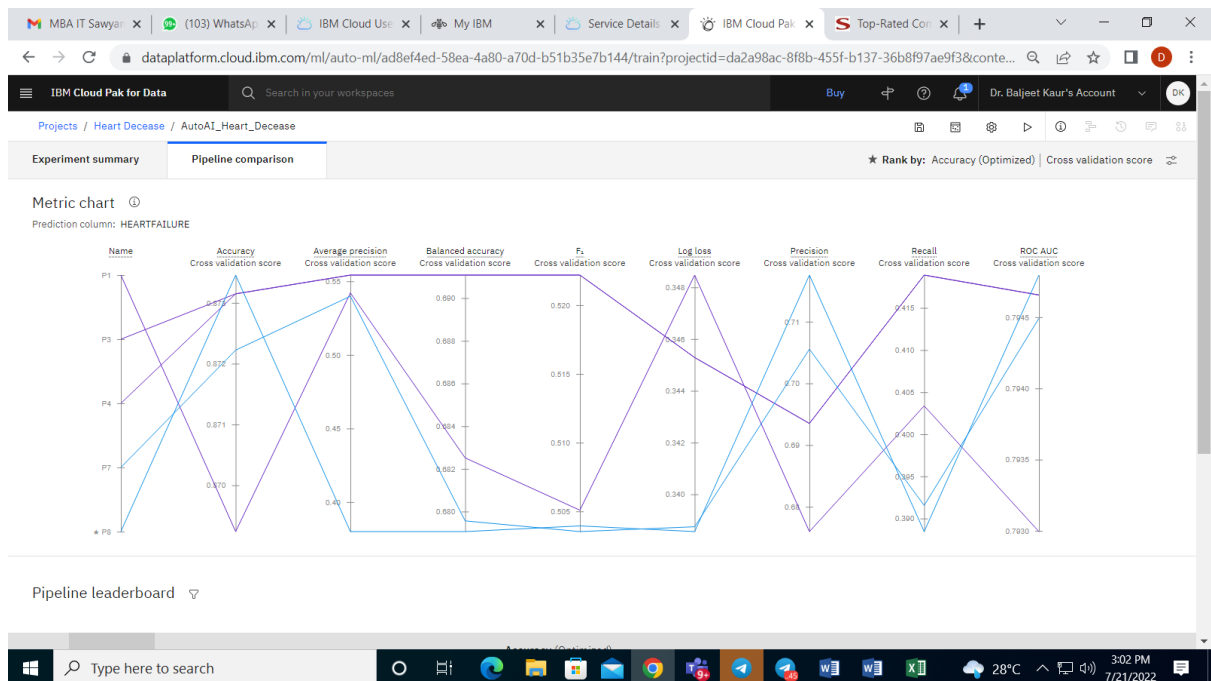
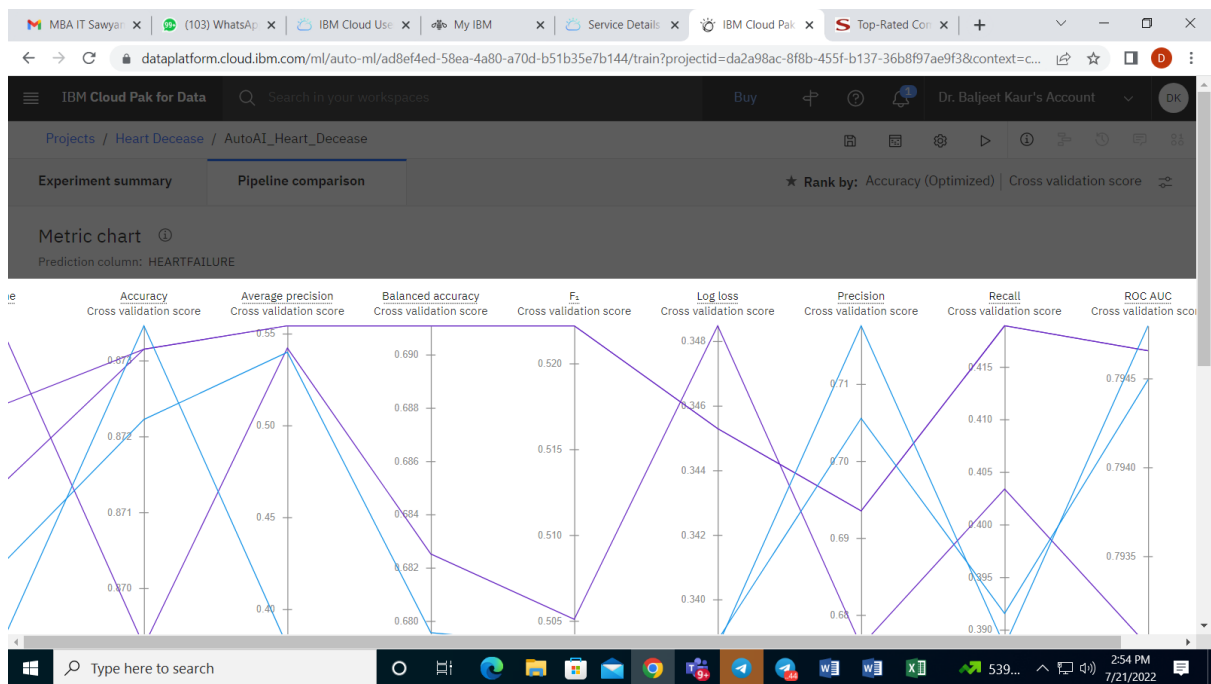
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 5 minutes

[View log](#) [Save code](#)

Pipeline leaderboard

Model pipeline:



Comparison of Results:

IBM Cloud Pak for Data

Search in your workspaces

Buy

Dr. Baljeet Kaur's Account

Projects / Heart Disease / AutoAI_Heart_Decease

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	00:00:37
2	Pipeline 4	XGB Classifier	0.873	HPO-1 FE HPO-2	00:01:21
3	Pipeline 1	XGB Classifier	0.869	None	00:00:01
4	Pipeline 2	XGB Classifier	0.869	HPO-1	00:00:11
5	Pipeline 7	Random Forest Classifier	0.862	HPO-1 FE	00:00:40
6	Pipeline 8	Random Forest Classifier	0.862	HPO-1 FE HPO-2	00:01:19
7	Pipeline 5	Random Forest Classifier	0.858	None	00:00:01

Type here to search

28°C 3:41 PM 7/21/2022

Saving the Model:

IBM Cloud Pak for Data

Search in your workspaces

Buy

Dr. Baljeet Kaur's Account

Save as

Select asset type

Model

Create a Watson Machine Learning model asset that you can test with new data, deploy to generate predictions, and trace lineage activity.

Notebook

Create a notebook if you want to view the code that created this model pipeline or interact with the model programmatically.

Define details

Name

AutoAI_Heart_Decease - P3 XGB Classifier

Description (optional)

Enter description here

Tags

Add tags to make assets easier to find.

Add a tag

Cancel Create

Type here to search

28°C 3:41 PM 7/21/2022

IBM Cloud Pak for Data

Search in your workspaces

Buy ? Dr. Baljeet Kaur's Account

Projects / Heart Decease / AutoAI_Heart_Decease

Experiment summary Pipeline comparison Rank by: Accuracy

Pipeline leaderboard

Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★ 1	Pipeline 3	XGB Classifier	0.873	HPO-1 FE	00:00:37
2	Pipeline 4	XGB Classifier	0.873	HPO-1 FE HPO-2	00:01:21
3	Pipeline 1	XGB Classifier	0.869	None	00:00:01
4	Pipeline 2	XGB Classifier	0.869	HPO-1	00:00:11
5	Pipeline 7	Random Forest Classifier	0.862	HPO-1 FE	00:00:40
6	Pipeline 8	Random Forest Classifier	0.862	HPO-1 FE HPO-2	00:01:19
7	Pipeline 5	Random Forest Classifier	0.858	None	00:00:01

Saved model successfully. AutoAI_Heart_Decease - P3 XGB Classifier was successfully saved to Heart Decease. View in project

Save as

Type here to search

28°C 3:42 PM 7/21/2022

Promoting Model to deployment space:

IBM Cloud Pak for Data

Search in your workspaces

Buy ? Dr. Baljeet Kaur's Account

Models / f38a9a53-6a36-4b9d-abc1-7fd09fb928c2?projectid=da2a98ac-8f8b-455f-b137-36b8f97ae9f3&context=cpd...

Promote to space

Use a deployment space to organize supporting resources such as input data and environments; deploy models or functions to generate predictions or solutions; and view or edit deployment details.

New deployment space

Use a space to collect assets in one place to create, run, and manage deployments

Define details

Name: Heart_Decease_Deploy

Description (Optional):

Deployment space description:

Deployment space tags (optional):

Add a tag:

Select services

The space is being prepared...

The space "Heart_Decease_Deploy" is being created.

Step 1 of 1. Creating deployment space.

Close

Cancel Creating

Type here to search

28°C 3:45 PM 7/21/2022

IBM Cloud Pak for Data

Promote to space

Use a deployment space to organize supporting resources such as input data and environments; deploy models or functions to generate predictions or solutions; and view or edit deployment details.

Target space: Heart_Decease_Deploy

Tags (optional): Start typing to add tags

☐ Go to the model in the space after promoting it

Selected assets (1)

Asset name	Format
AutoAI_Heart_Decease - P3 XGB Classifier	Model

Select version

Promoting a version of an asset to a space creates a new asset in the space, with a new asset ID.

Current

Description (optional)

Description of assets

Cancel Promoting...

IBM Cloud Pak for Data

AutoAI_Heart_Decease - P3 XGB Classifier

Promote to deployment space

Successfully promoted AutoAI_Heart_Decease - P3 XGB Classifier to the associated deployment space. Go to the deployment space to prepare the assets for deployment.

Timestamp 3:45:57 PM

Created Jul 21, 2022 3:44 PM

Type wml-hybrid_0.1

Model ID f38a9a53-6a36-4b9d-abc1-7fd09fb928c2

Software specification hybrid_0.1

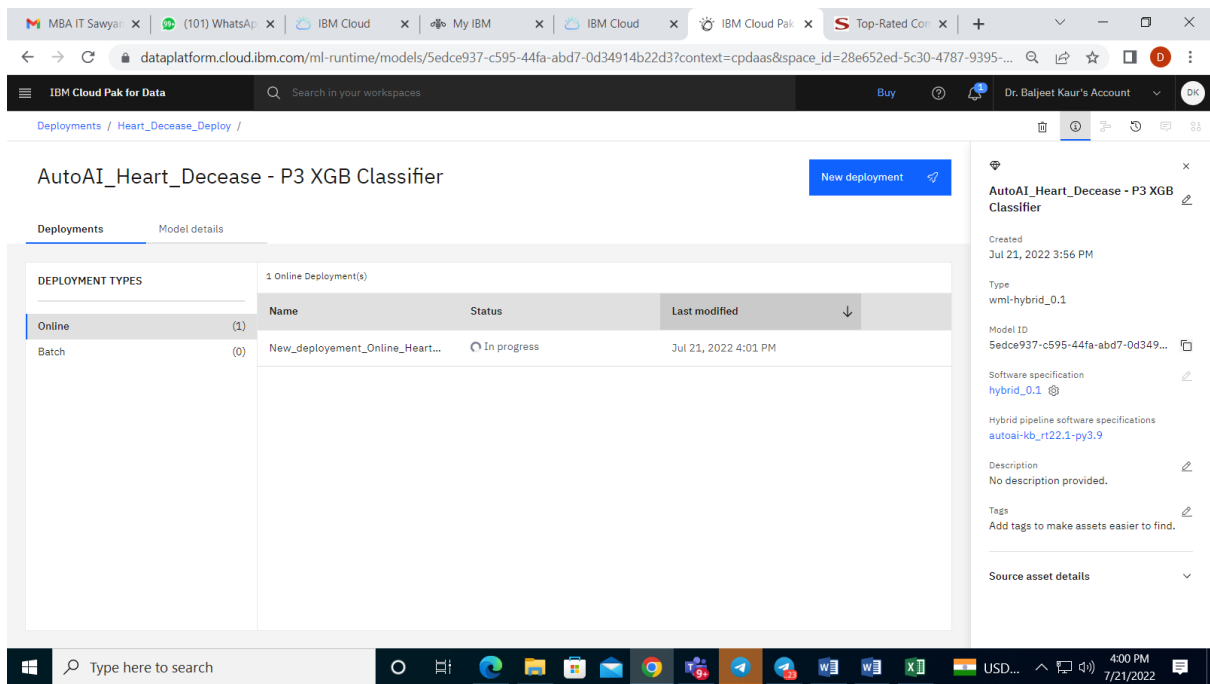
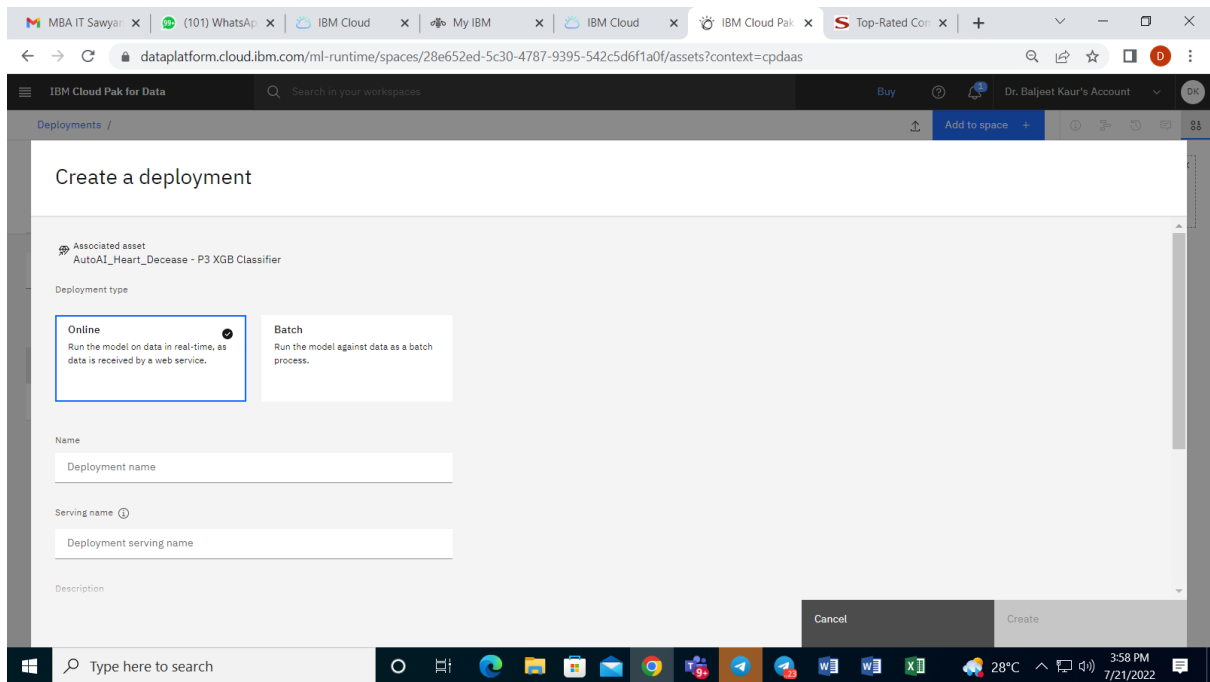
Hybrid pipeline software specifications autoai-kb_rt22.1-py3.9

Tags Add tags to make assets easier to find.

Feedback

Input Schema

Column	Type
AGE	"integer"
AVGHEARTBEATSPERMIN	"integer"
BMI	"integer"
CHOLESTEROL	"integer"
EXERCISEMINPERWEEK	"integer"
FAMILYHISTORY	"other"
PALPITATIONSPERDAY	"integer"
SEX	"other"
SMOKERLASTSYRS	"other"



AutoAI_Heart_Decease - P3 XGB Classifier test prediction

Prediction type: ☒ Table view ☐ JSON view

Binary classification

Prediction percentage

1 Record

Confidence level distribution

Amount of records

	Prediction	Confidence
1	N	93%
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

Error: Invalid input data.

Download

Creation of Node Red Service:

Resource list

Create resource +

Name	Group	Location	Product	Status	Tags
Filter by name or IP address... Filter by group or org. Filter... Filter... Filter... Filter...					
Devices (0)					
VPC infrastructure (0)					
Clusters (0)					
Container Registry (0)					
Satellite (0)					
Cloud Foundry apps (1)					
Node RED GZPYL 2022-07-12	SIUSICSR / node	Dallas	Node.js	Started	-
Cloud Foundry services (1)					
Services and software (1)					

apikey.json Show all

Node-RED interface showing a flow named "Flow 1" with two nodes: "Hello Node-RED!" and "msg.payload". The interface includes a left sidebar with node palettes (common, function), a top bar with tabs, and a right sidebar with flow information and a message box.

Flow 1

Nodes: inject, debug, complete, catch, status, link in, link call, link out, comment

Function nodes: apikey.json

Flow 1 info

Flow: "384c2ba73a1e9b3"

You can confirm your changes in the node edit tray with **ctrl-enter** or cancel them with **ctrl-escape**

Windows taskbar: 4:39 PM 7/21/2022

Node-RED interface showing the "User Settings" dialog box. The dialog displays a list of installed and available nodes, including "cn-dashboard-nodes", "node-red-dashboard", "feezal", and "node-red-contrib-dashboard-average-bars". The "debug" console on the right shows the output of the flow.

User Settings

View: Nodes Install

Palette

Keyboard

Search: dashboard

69 / 3954

node-red-dashboard

Web Components based Dashboard UI with WYSIWYG Editor

0.8.1 1 year, 7 months ago

node-red-contrib-dashboard-average-bars

debug

msg.payload: string[15]

"Hello Node-RED!"

7/21/2022, 4:57:31 PM node: f2f2649a.0d0d98

msg.payload: string[15]

"Hello Node-RED!"

7/21/2022, 4:57:36 PM node: f2f2649a.0d0d98

msg.payload: string[15]

"Hello Node-RED!"

7/21/2022, 4:57:48 PM node: f2f2649a.0d0d98

msg.payload: string[15]

"Hello Node-RED!"

7/21/2022, 4:58:01 PM node: f2f2649a.0d0d98

msg.payload: string[15]

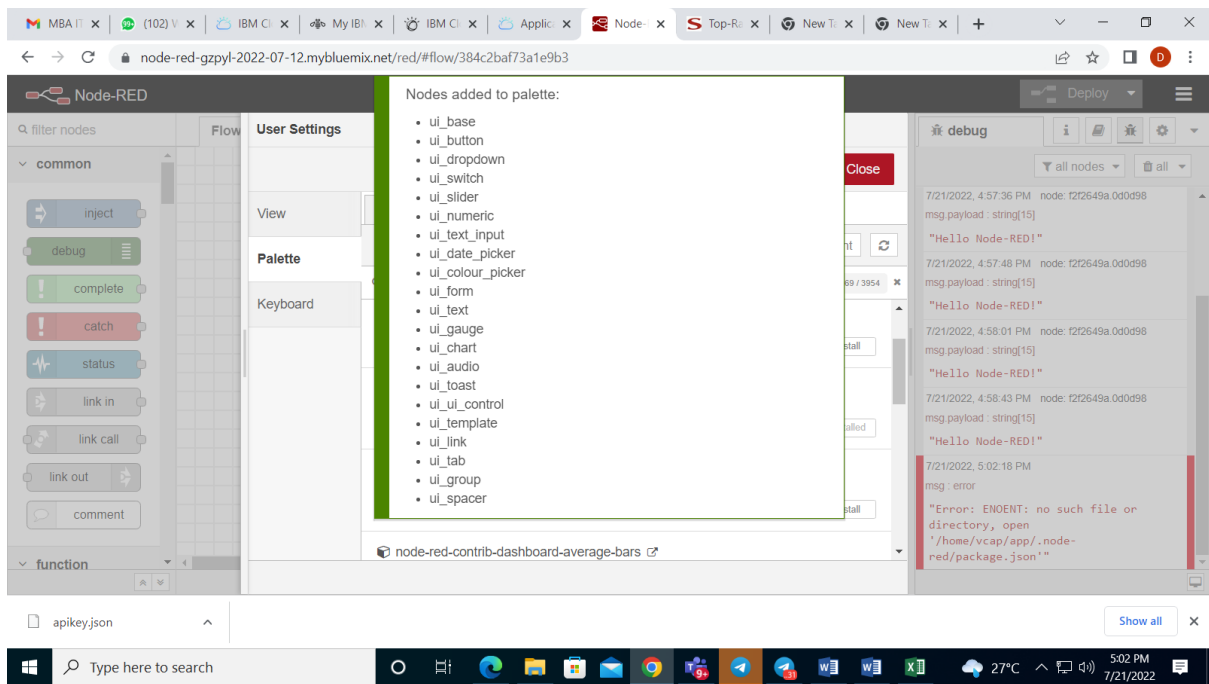
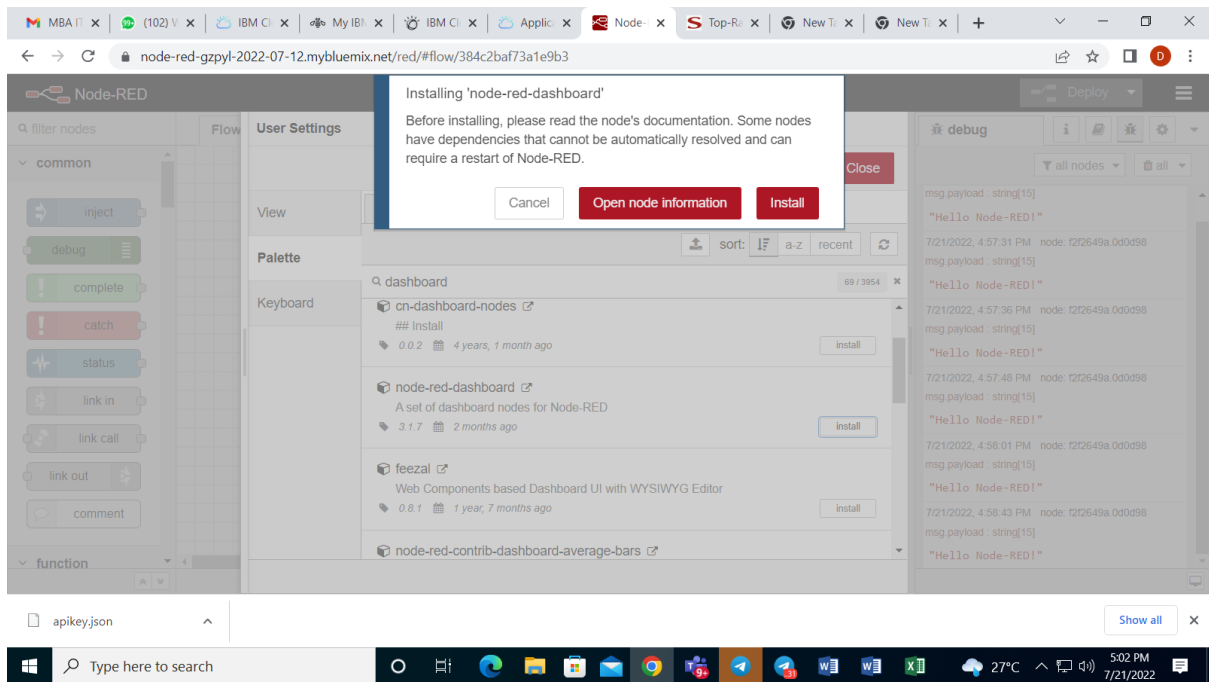
"Hello Node-RED!"

7/21/2022, 4:58:43 PM node: f2f2649a.0d0d98

msg.payload: string[15]

"Hello Node-RED!"

Windows taskbar: 5:01 PM 7/21/2022



Node-RED interface showing the "Edit dropdown node" configuration for a node named "dropdown". The configuration includes:

- Group: [Heart Disease Predictor] Effective Heart Disease Pri
- Size: auto
- Label: SMOKER LAST 5YRS
- Tooltip: optional tooltip
- Placeholder: Select option
- Options: Y (YES), N (NO)

The debug console shows the following messages:

```
7/21/2022, 4:57:36 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 4:57:48 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 4:58:01 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 4:58:43 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 5:02:18 PM
msg: error
"Error: ENOENT: no such file or
directory, open
'/home/vcap/app/.node-
red/package.json'"
```

Node-RED interface showing the "Flow 2" configuration. The flow includes the following nodes:

- form
- msg.payload
- SEX
- msg.payload
- FAMILY HISTORY
- msg.payload
- SMOKER LAST 5YRS
- msg.payload

The debug console shows the following messages:

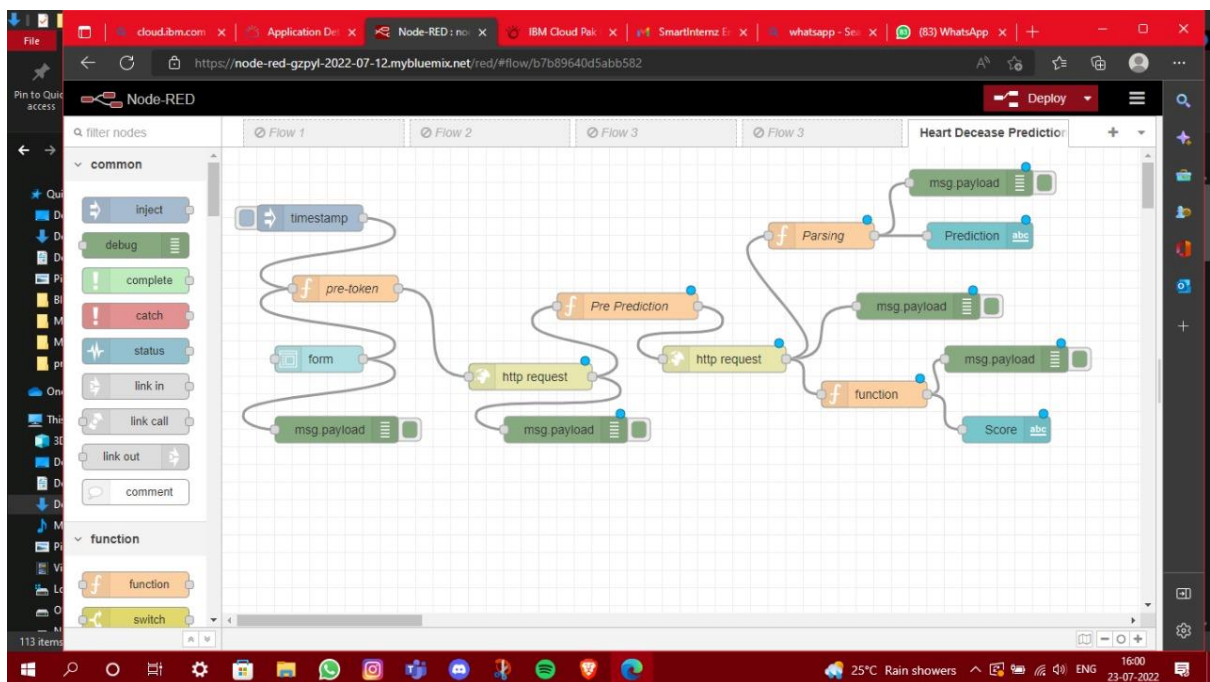
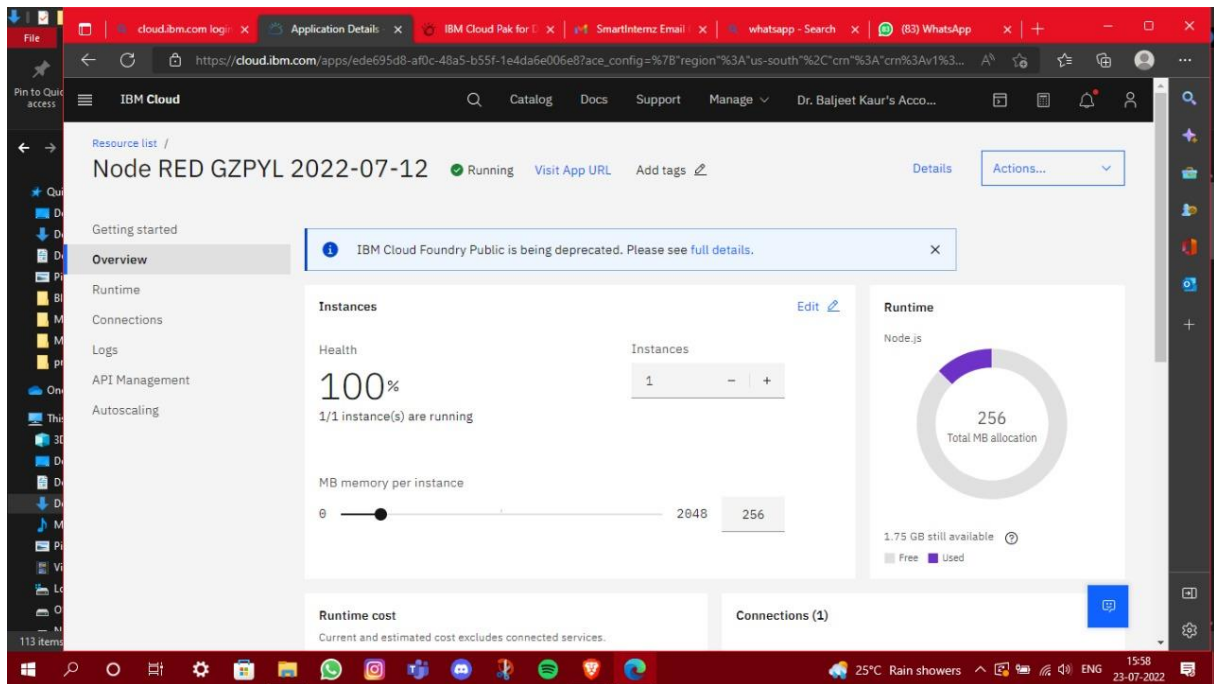
```
7/21/2022, 4:57:36 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 4:57:48 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

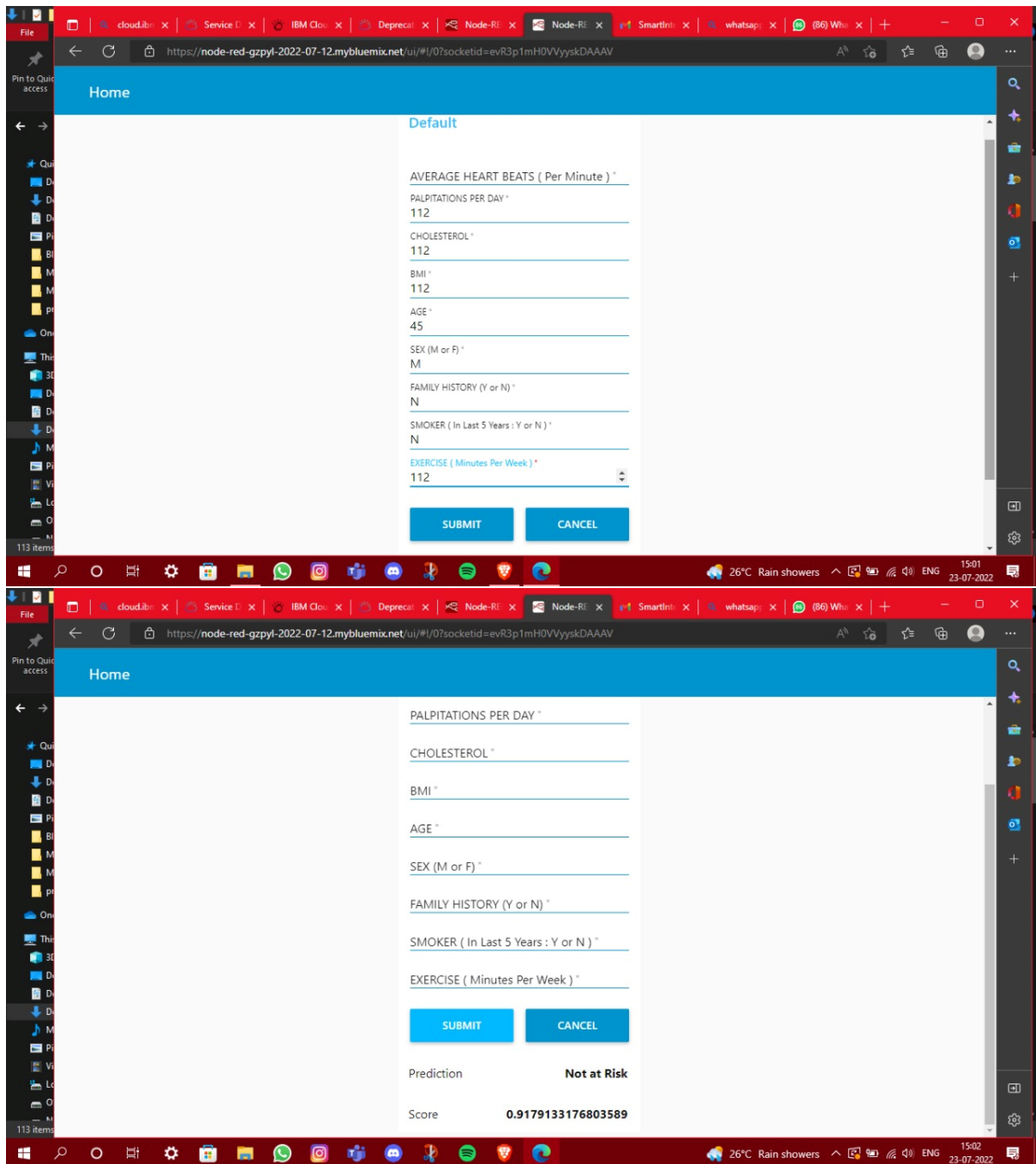
7/21/2022, 4:58:01 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 4:58:43 PM node: f2f2649a.0d0d98
msg.payload: string[15]
"Hello Node-RED!"

7/21/2022, 5:02:18 PM
msg: error
"Error: ENOENT: no such file or
directory, open
'/home/vcap/app/.node-
red/package.json'"
```

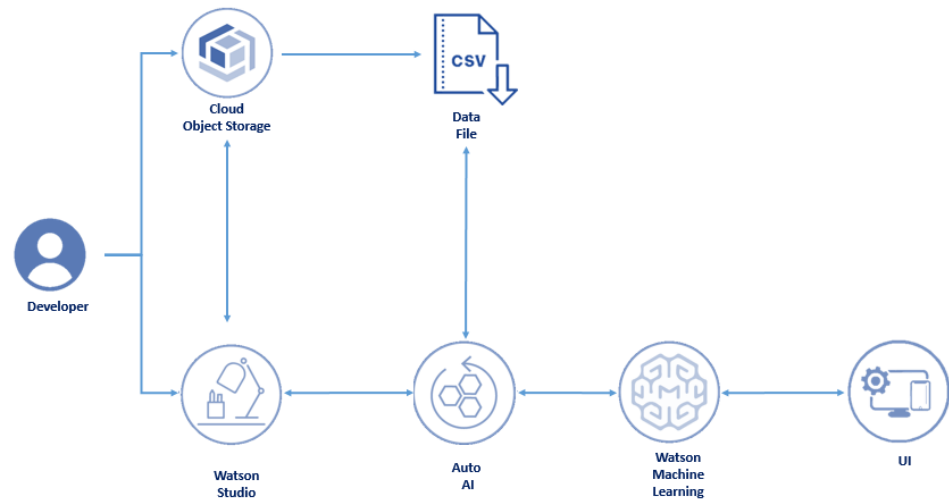


Launching the service Dashboard:

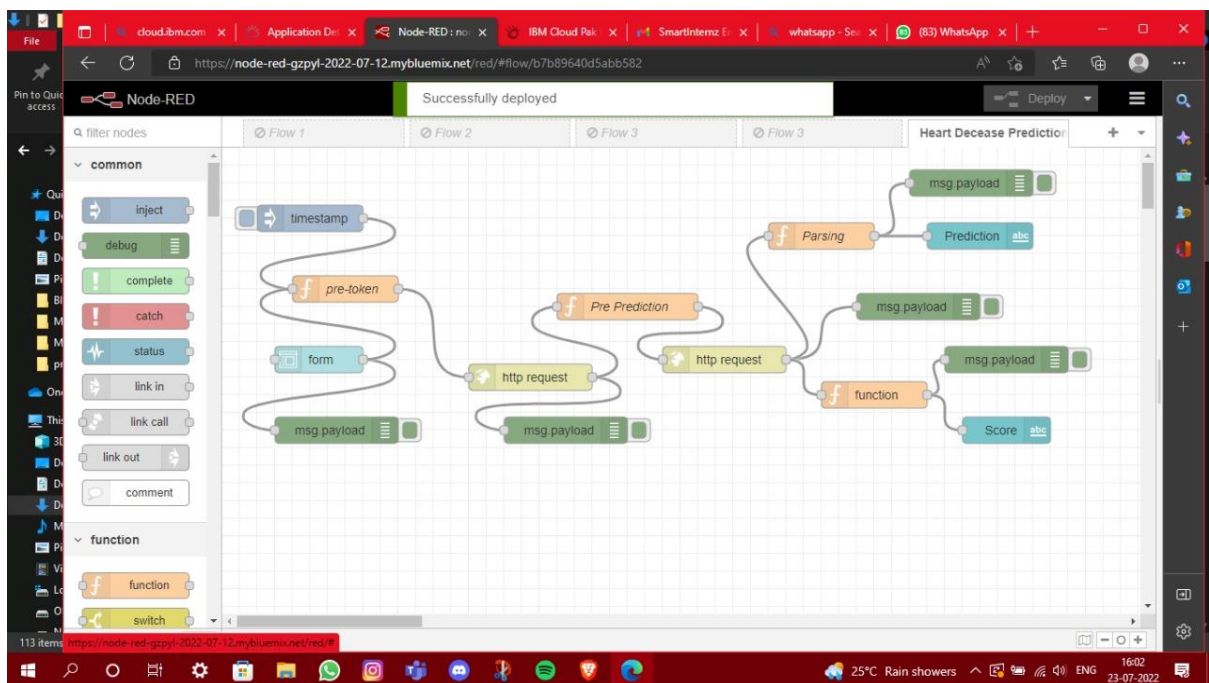


5. FLOWCHART

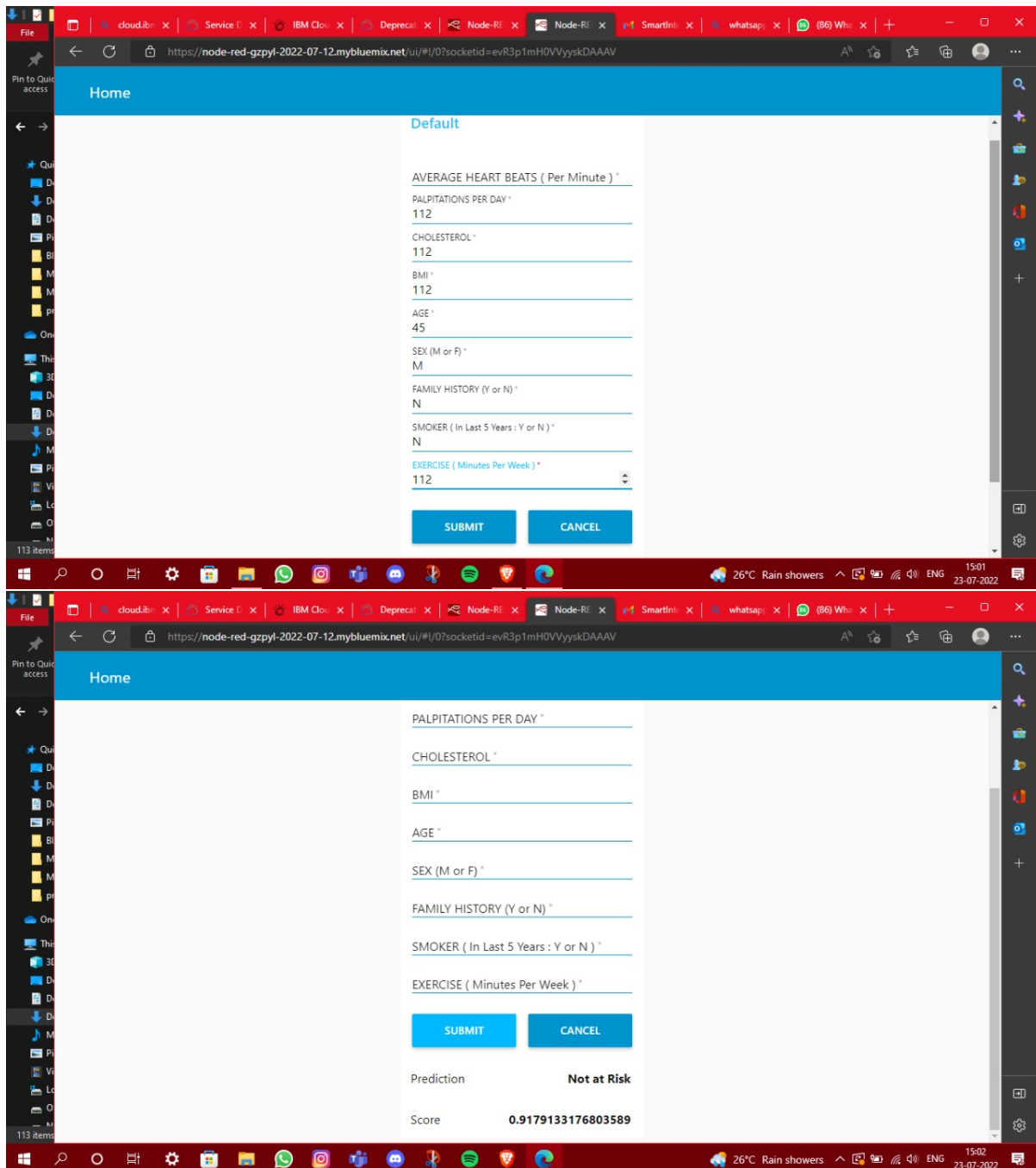
Diagram showing the control flow of the solution



6. RESULT



Launching the service Dashboard:



7. ADVANTAGES & DISADVANTAGES

Advantages:

- IBM Cloud provide end to end solution from taking data as input and providing prediction as output
- Organizations can concentrate of solution part of the problem that developing the model.

Disadvantages;

- a. Don't see any disadvantages as such.
- b. A lot of customization is possible for setting up the experiment, so don't find any disadvantages.

8. APPLICATIONS

This is a wonderful solution for businesses which are looking for end to end solutions. Following are few examples:

- a. Healthcare is a big market. Right from the timely diagnosis of the disease to smart health care solutions can use these services.
- b. Market Segmentation
- c. smart banking services
- d. Recommender systems

9. CONCLUSION

Heart failure is a common event caused by CVDs and this dataset contains 9 features that were used to predict mortality by heart failure. The Model was created using Auto AI Service. XGBoost provides was identified as the best algorithm to identify heart failure and the algorithm was used to create the model. Auto AI used different hyper parameters to improve the performance of XGBoost. In this project, we created a model using Auto AI and build a web application which was launched with the help of Dashboard for the prediction of heart failure.

10. FUTURE SCOPE

One of the significant difficulties that India as a nation faces is that we have limited infrastructure for medical healthcare which result in delay in healthcare service and identifying a critical disease at early stage. If the simple dashboards are available for people to understand the critical stage of a disease or to diagnose a disease will be really good social service.

The Similar application can be scaled to multiple case studies in Financial/ Marketing/ e-learning domains.

11. BIBLIOGRAPHY

IBM Smartinternz Workshop attended from 6th July to 12th July 2022. Course material and tutorial was referred in solving assignment.

Day 1 Link- https://www.youtube.com/watch?v=Zj-DFJpyT_A

Day 2 Link- <https://youtu.be/4CMldzb701M>

Day 3 Link- https://youtu.be/PSi_1DUr5og

Day 4 Link- <https://youtu.be/Huc9wGetngA>

Day 5 Link- <https://youtu.be/R22RqlGvtzE>

Day 6 Link- <https://youtu.be/59QeXVdK8rY>