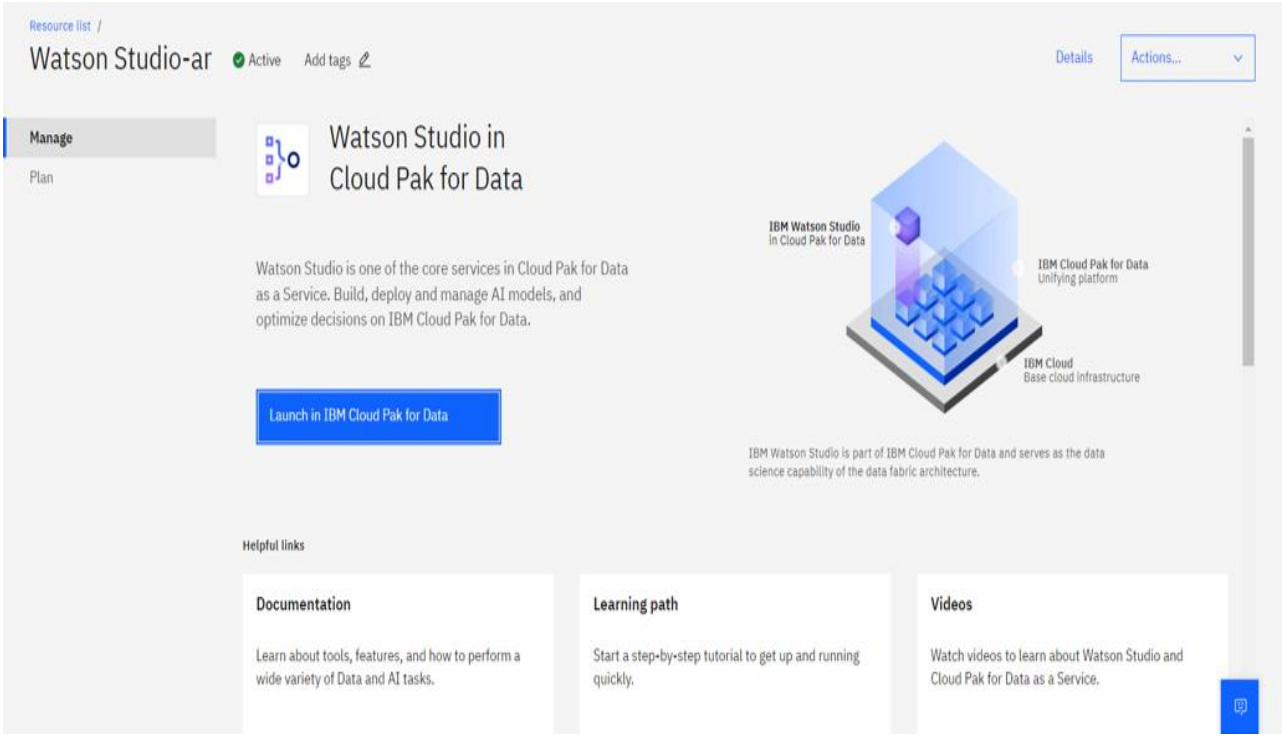
Effective Heart Disease Prediction Using IBM Auto AI Service

Introduction

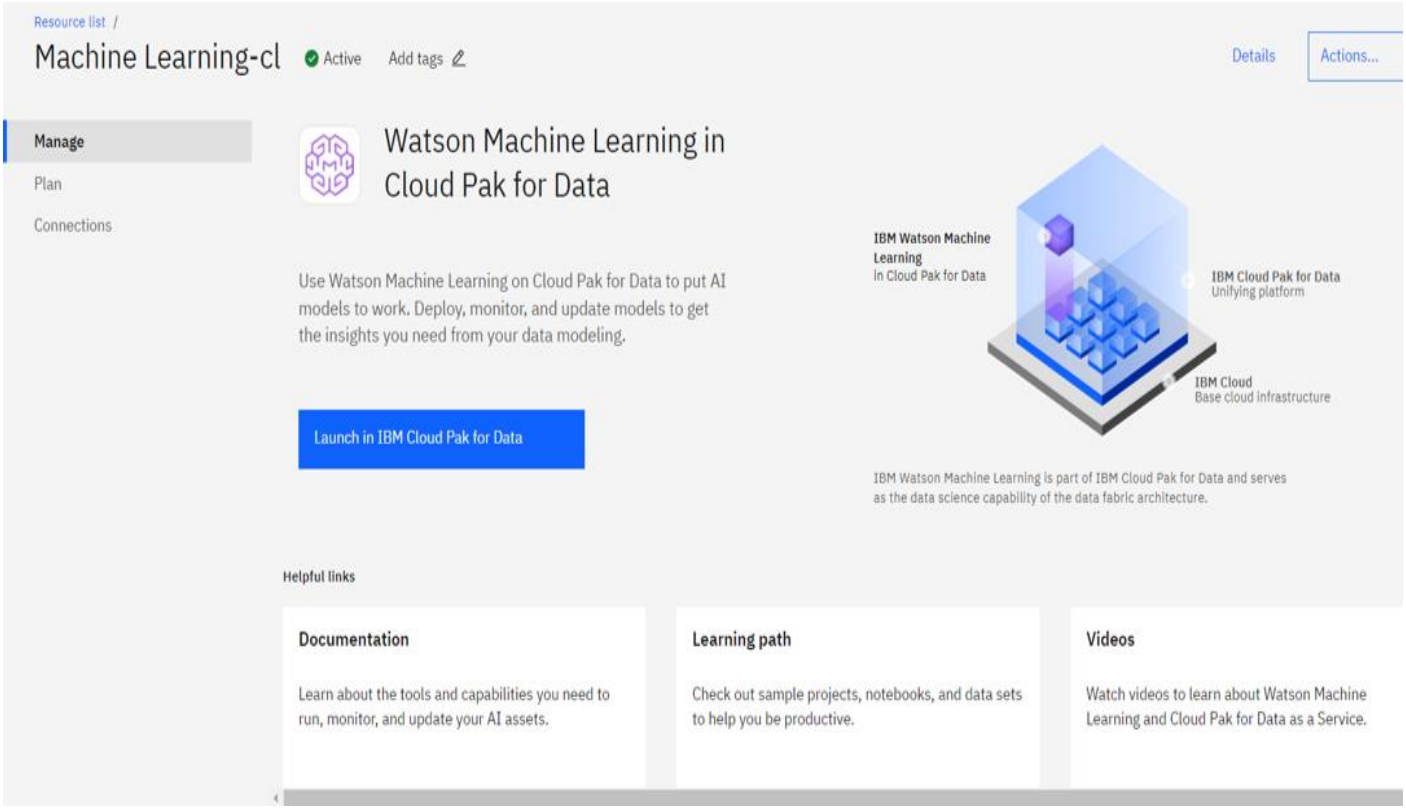
- Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide.
- Heart failure is a common event caused by CVDs
- In this project, a classification model is built using IBM Auto AI for the prediction of heart failure.
- For achieving this objective, a heart disease prediction dataset is used which contains 9 features such as Avg heartbeats per min, cholesterol, bmi, etc

Services Used

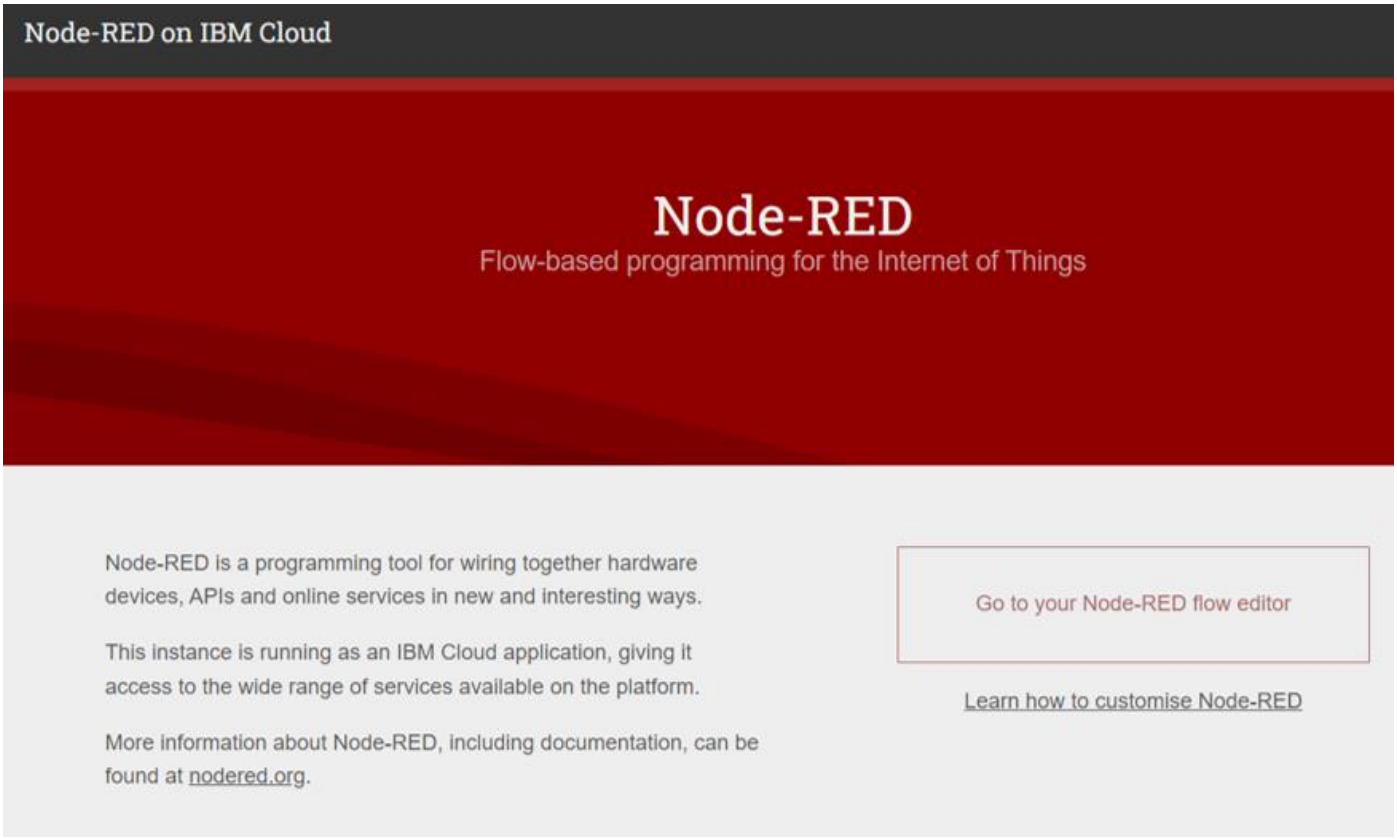
- IBM Watson Studio



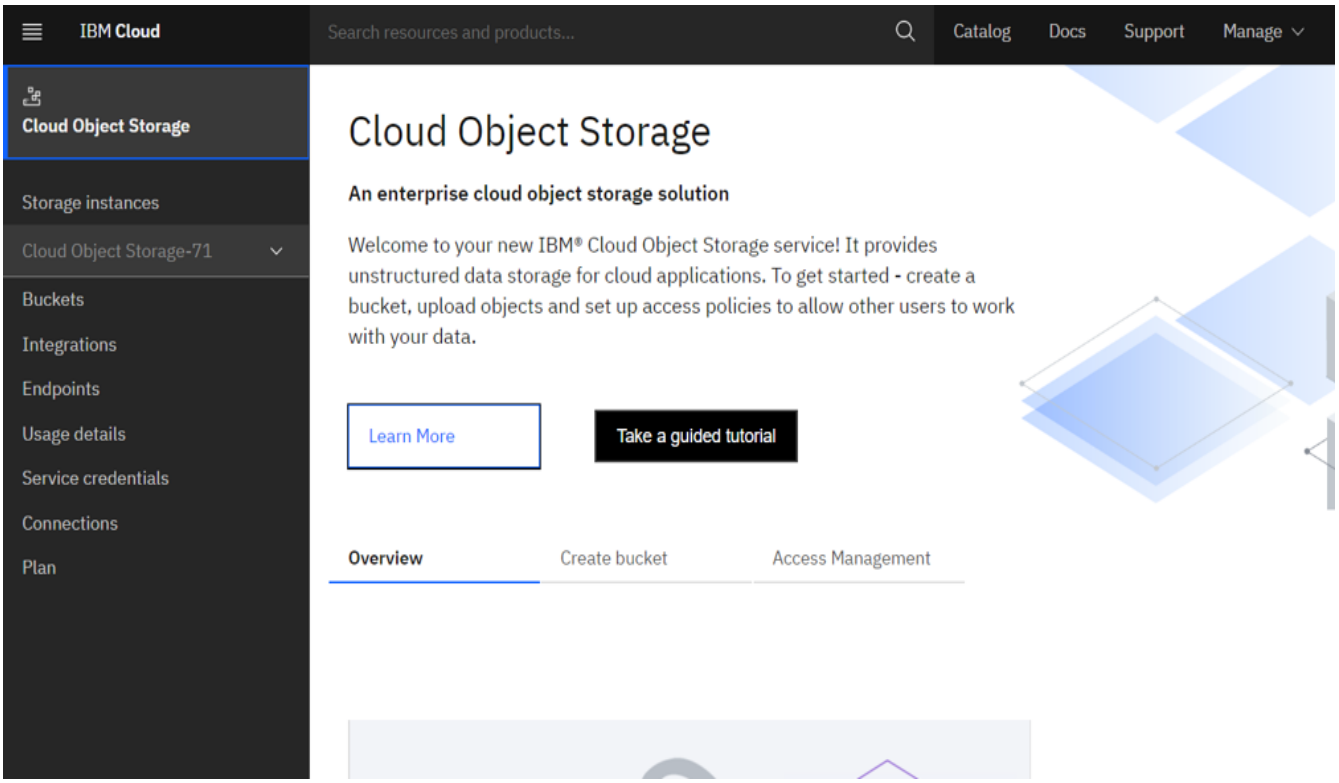
- IBM Watson Machine Learning



- Node-RED

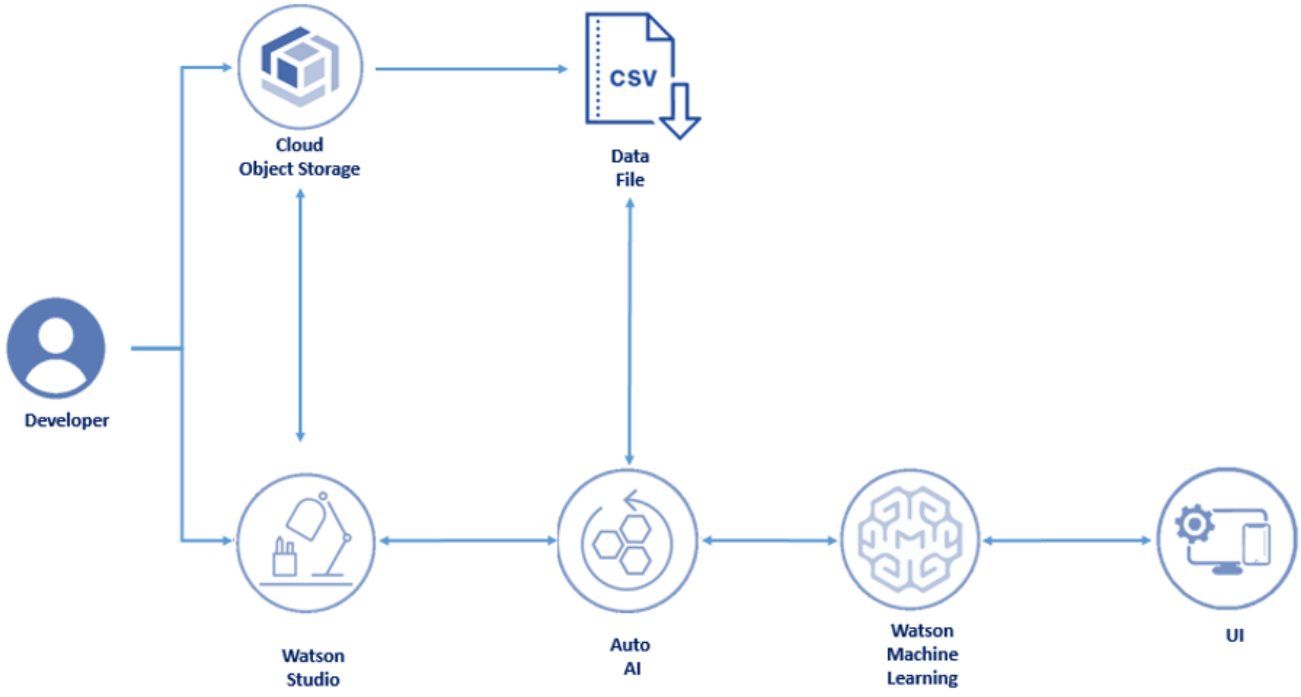


- IBM Cloud Object Storage



Technical Architecture

Given below is the technical architecture of the project.



Based on the architecture, the flow of the project can be explained as follows:

1.Log in to IBM account
Firstly, log in to your IBM cloud account.
2. Create IBM Watson Studio, Cloud Object Storage and Node-RED Service
Search the catalog for IBM Watson Studio, Cloud Object Storage and Node-RED Service and create instances of the same.
3. Create a Watson Studio project
Create a machine learning service in IBM Watson for deploying your classification model.
4. Add Auto AI Experiment
5. Run the Auto AI Experiment to build an optimal Machine learning model on the given dataset
6. Save the model
7. Deploy the model on a web server and generate scoring End Point
8. Create a WEB application Using Node-RED to take user input
9. Showcase Prediction on UI

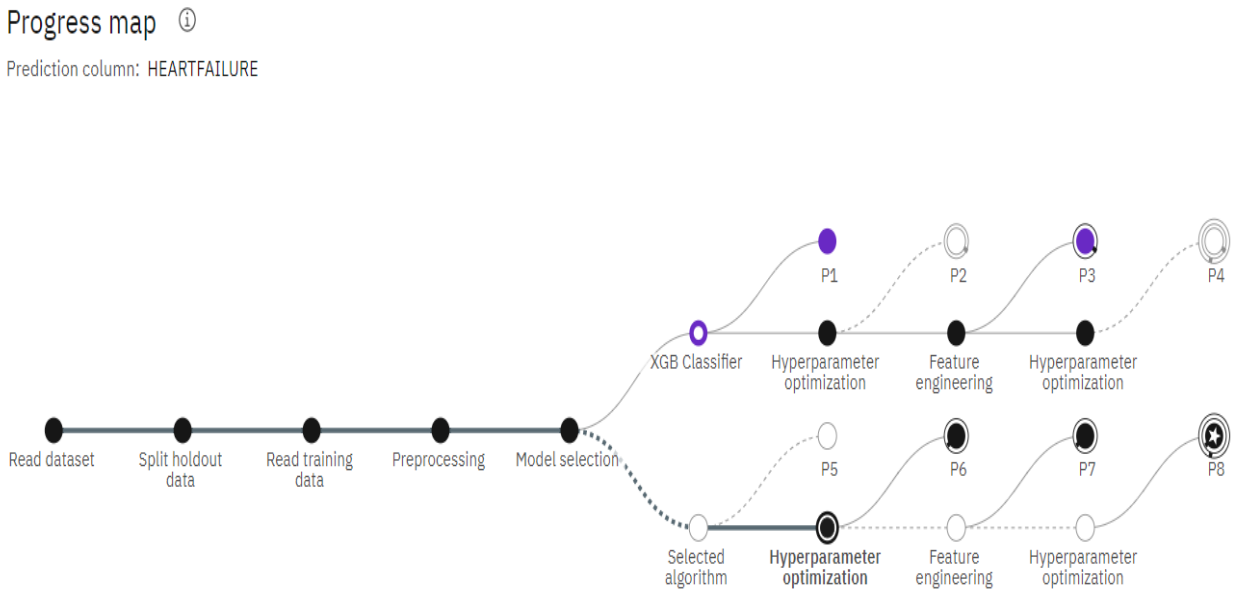
Result

The given dataset was used to create an Auto-AI experiment. A binary classification model was built which takes the various features from the dataset as input and classifies the given instance into either one of the two classes i.e. 'Y'(prone to heart disease) or 'N'(not prone to heart disease).
The Auto-AI experiment chose the following two algorithms for classification:

1. SNAP Random Forest classifier and
2. XGB classifier

The flow of the Auto-AI experiment is given as follows:

1. Read dataset
2. Split holdout data
3. Read training data
4. Preprocessig
5. Model Selection
6. From here, the flow proceeds to both the classification algorithms via dedicated pipelines and makes the predictions supported by hyperparameter optimization and feature engineering techniques.

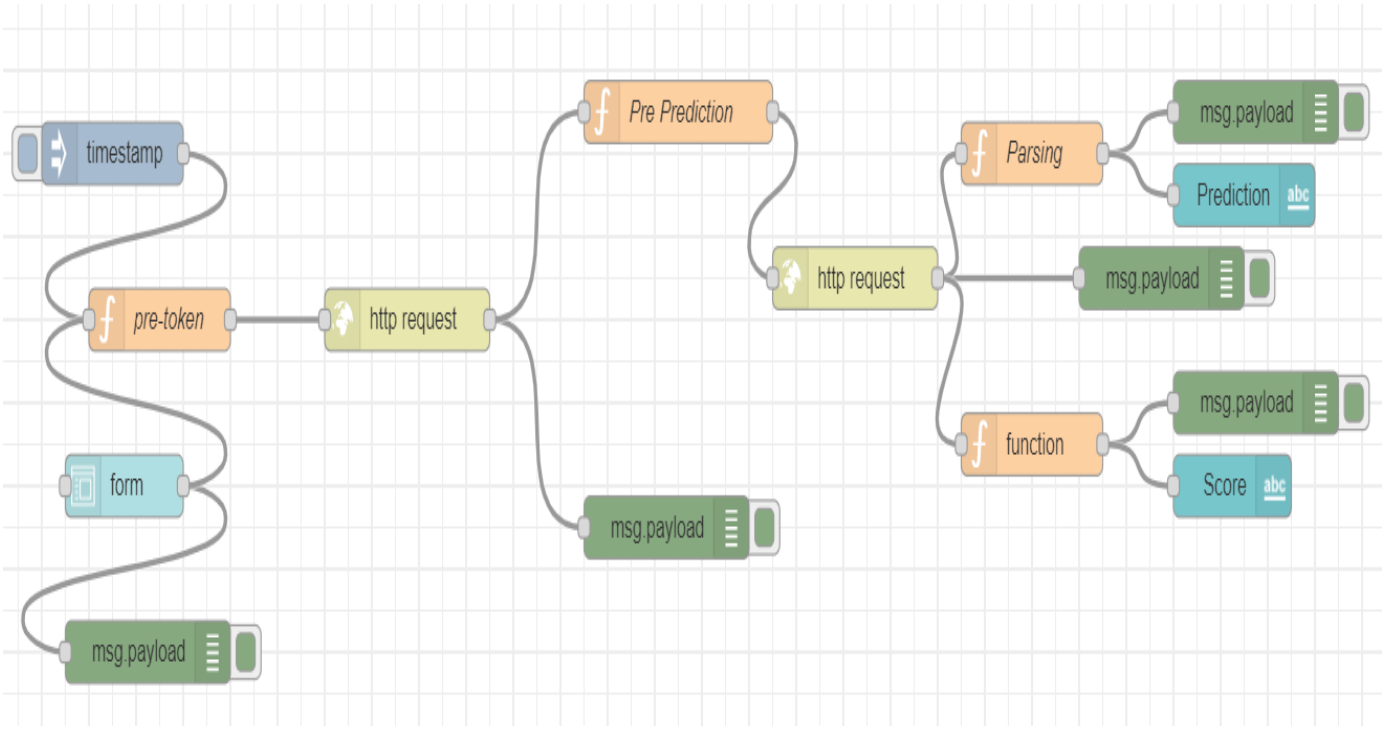


After the models are built, the SNAP random forest algorithm along with two rounds of hyperparameter optimization and feature engineering gives the best accuracy i.e. 87.3%. Hence, this model is selected for further deployment as a service.

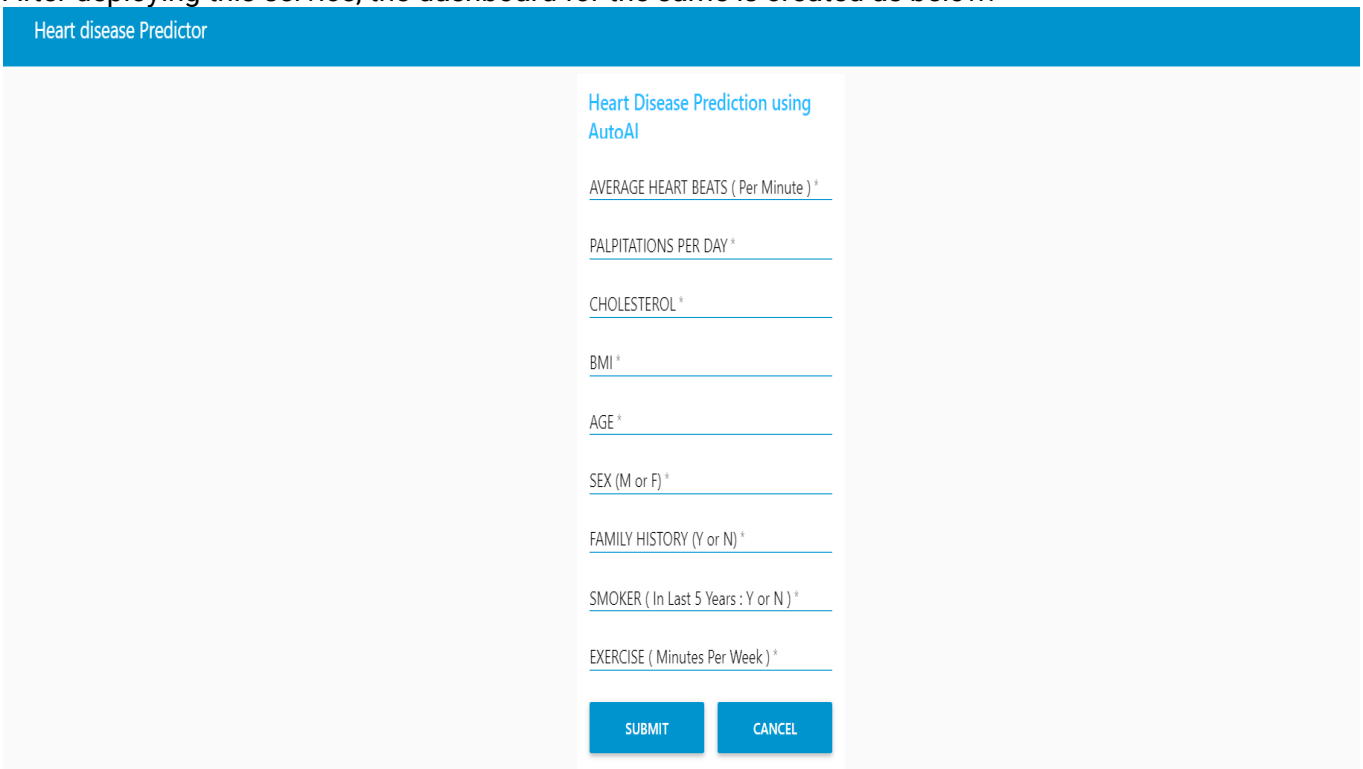
Pipeline leaderboard

	Rank ↑	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 8	Snap Random Forest Classifier	0.873	HPO-1 FE HPO-2	00:01:39
	2	Pipeline 3	XGB Classifier	0.873	HPO-1 FE	00:00:52
	3	Pipeline 7	Snap Random Forest Classifier	0.872	HPO-1 FE	00:01:10
	4	Pipeline 6	Snap Random Forest Classifier	0.869	HPO-1	00:00:11
	5	Pipeline 1	XGB Classifier	0.869	*None*	00:00:01

After that, a json flow for the web service is created using node-RED.



After deploying this service, the dashboard for the same is created as below:



Hence, according to the various parameters inserted by the user, the prediction is done by the model and is displayed accordingly.

Output Screenshots:

Heart Disease Prediction using AutoAI

AVERAGE HEART BEATS (Per Minute) *
93

PALPITATIONS PER DAY *
22

CHOLESTEROL *
163

BMI *
25

AGE *
49

SEX (M or F) *
F

FAMILY HISTORY (Y or N) *
N

SMOKER (In Last 5 Years : Y or N) *
N

EXERCISE (Minutes Per Week) *
110

SUBMIT CANCEL

Prediction **Not at Risk**

Score **0.9210792406516917**

Heart Disease Prediction using AutoAI

AVERAGE HEART BEATS (Per Minute) *
134

PALPITATIONS PER DAY *
7

CHOLESTEROL *
228

BMI *
34

AGE *
63

SEX (M or F) *
F

FAMILY HISTORY (Y or N) *
Y

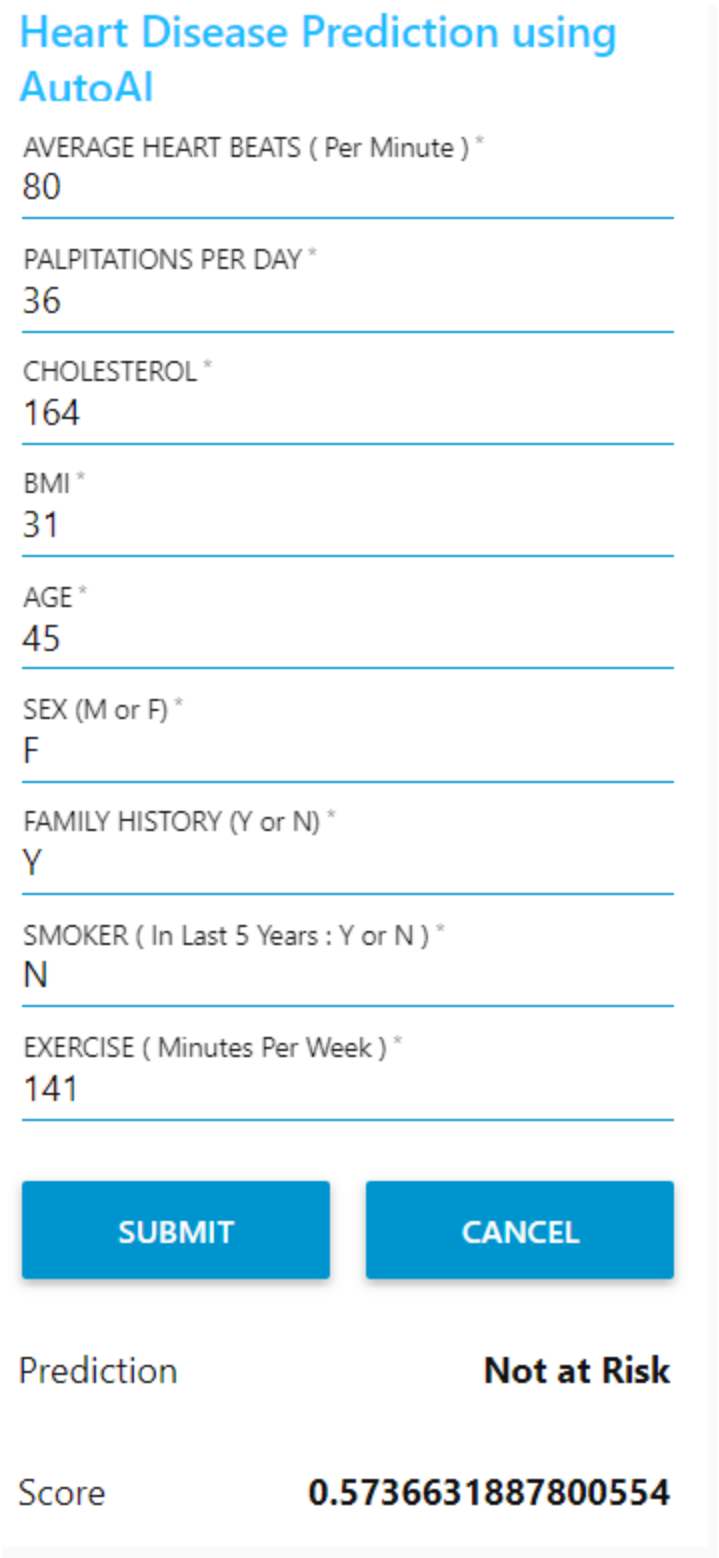
SMOKER (In Last 5 Years : Y or N) *
N

EXERCISE (Minutes Per Week) *
92

SUBMIT CANCEL

Prediction **At Risk**

Score **0.8132361145580516**



Conclusion:

Thus, using the various IBM cloud services and Auto-AI, the classification model is successfully built and deployed for the effective heart disease prediction.