Effective Heart Disease Prediction using IBM Auto AI Service

Dr.P. Mohana Priya
SASTRA Deemed University, Thanjavur – 613 401, India

1 INTRODUCTION

1.1 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 mortality by heart failure. In this project, machine learning model is build using Auto AI service and node-red web application where we can showcase the prediction of heart failure. The complete implementation procedures have been deployed using IBM Project Build – A - Thon in Smart Internz platform [1]

a. PURPOSE

The purpose of this project is to predict the heart failures based on the various parameters of the patient approaching a doctor. This project will be really useful for smart health care applications and heart failure issues can be predicted.

2 LITERATURE SURVEY

2.1 Existing Problem

Heart disease can be predicted using hybrid machine learning approaches [2] that are used for making decisions and predictions from the large quantity of data, datamining techniques by incorporating neural networks and rule mining procedures [3,4,5,6], Intelligent optimization algorithms and statistical analysis of the family history.

2.2 Proposed Solution

In this project, IBM Auto AI service incorporated with random forest classifier to predict whether the patient health is at risk because of heart failures. The proposed solution results with patient is either at risk (or) not risk. The proposed solution consist of modules such as creation of IBM services for creating IBM Watson studio and node-red cloud foundary application, building machine learning model, integration node-red application with the Watson studio project to showcase prediction results in User Interface.

3. THEORETICAL ANALYSIS

The patient dataset is downloaded in the form of Comma Separated Values (.CSV) extension which has attributes such as Average Heart Beats per minute, Palpitations per day, Cholesterol, Body Mass Index (BMI), Age, Sex, Family history, Smoking history and Exercise in mins per week. The dataset is feeded as an input to the Watson studio project named Heart Failure Prediction Model (HFPM)

followed by adding new auto ai experiment to predict whether the patient health is at risk. Once model building is completed, project deployment needs to be done by saving the model at the suitable location, organization and space. Cache generation is done based on the two pipelines such as ch-pipeline and pr-pipeline, on successful completion of the pipelines, deployment needs to be promoted. Node-red service needs to be created as the cloud foundary application where the Application Programming Interface (API) key and scoring end point is generated. Node-red service can be accessible by invoking visit app URL, necessary node-red plugins will be loaded thereby node-red editor can be opened preferring not recommended for user login credentials. Once the node-red editor is turned on, .json node flow file needs to be imported and configuration details needs to be replaced both for pre-token in which existing variable API key with the API key created for the project. In the http request node, scoring end point needs to be replaced for the newly created project. Node-red dashboard needs to be invoked where by web application is enabled in which the user needs to enter inputs for the said attributes. The prediction model results accuracy along with patient health risk information.

3.1BLOCKDIAGRAM

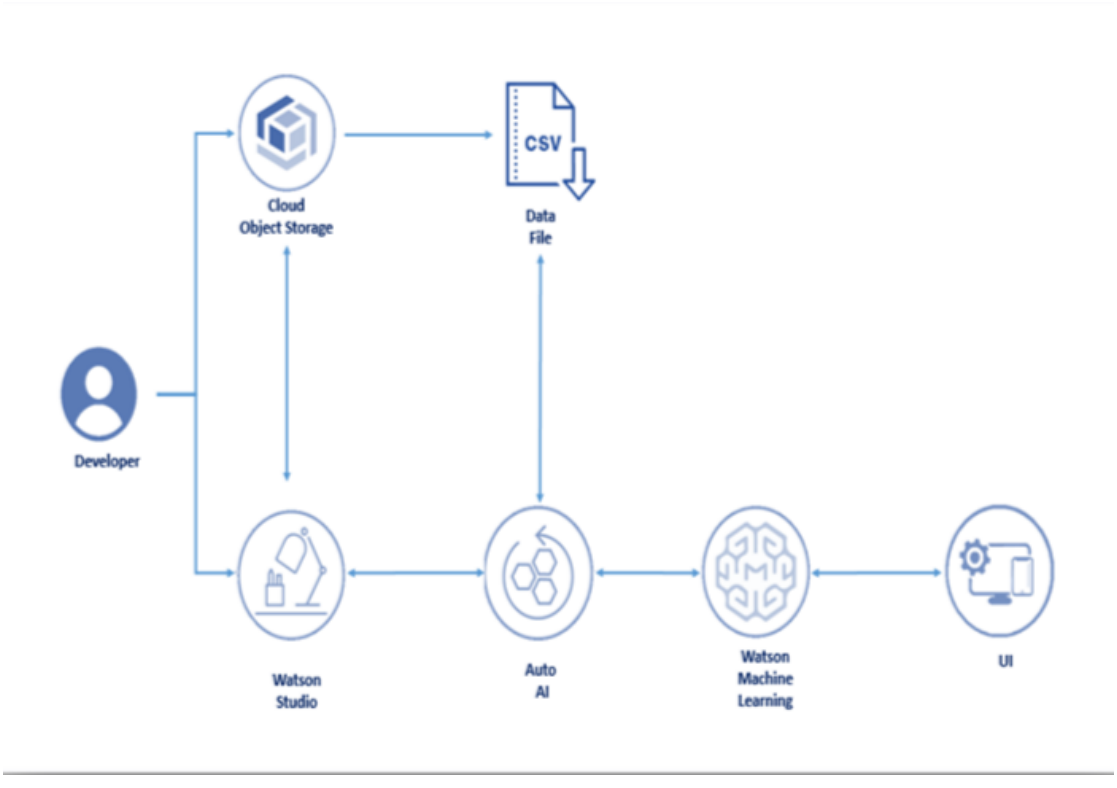


Fig 1 : Block diagram of the Proposed AI

3.2 HARDWARE / SOFTWARE DESIGNING

3.2.1 Hardware Requirements

- DELL Inspiron
- Intel Core i3 10th GEN

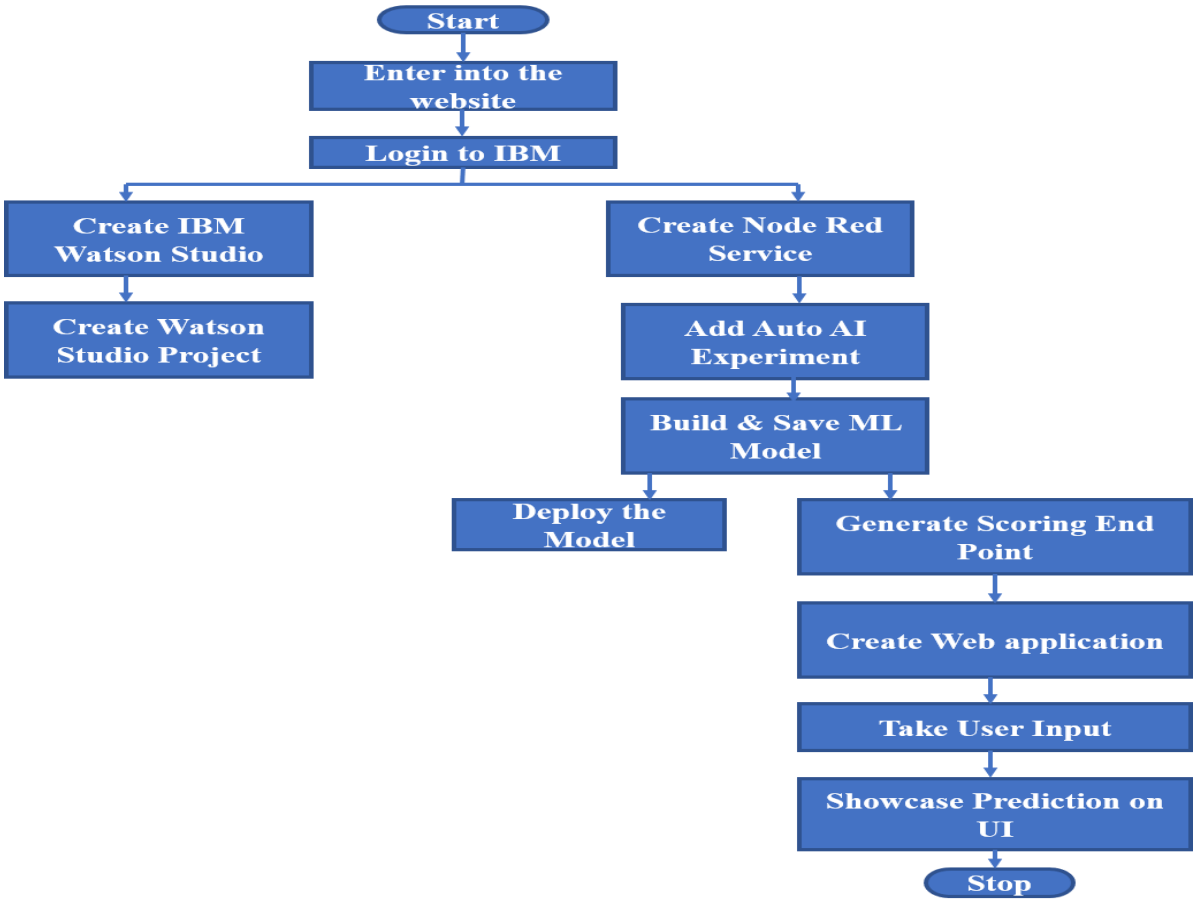
3.2.2 Software Requirements

- IBM Academic Initiative Account
- IBM Cloud Account
- IBM Watson Studio
- Auto AI
- IBM Web Server
- Node-Red Editor
- Node-Red Web Application
- User Interface

4 EXPERIMENTAL INVESTIGATIONS

Data manipulation and data analysis have been done for various set of attributes available in the patient data.csv file. Prediction with its result is given in section 5. The analyzed data of the different metrics include cross validation score accuracy, average precision, F1 score, recall, Receiver Operating Curve (ROC), precision, log loss etc. Average of all the said metrics are 0.867, 0.45, 0.67, 0.495, 0.8, 0.68, 0.39 and 0.7825.

5 FLOW CHART



6 RESULTS

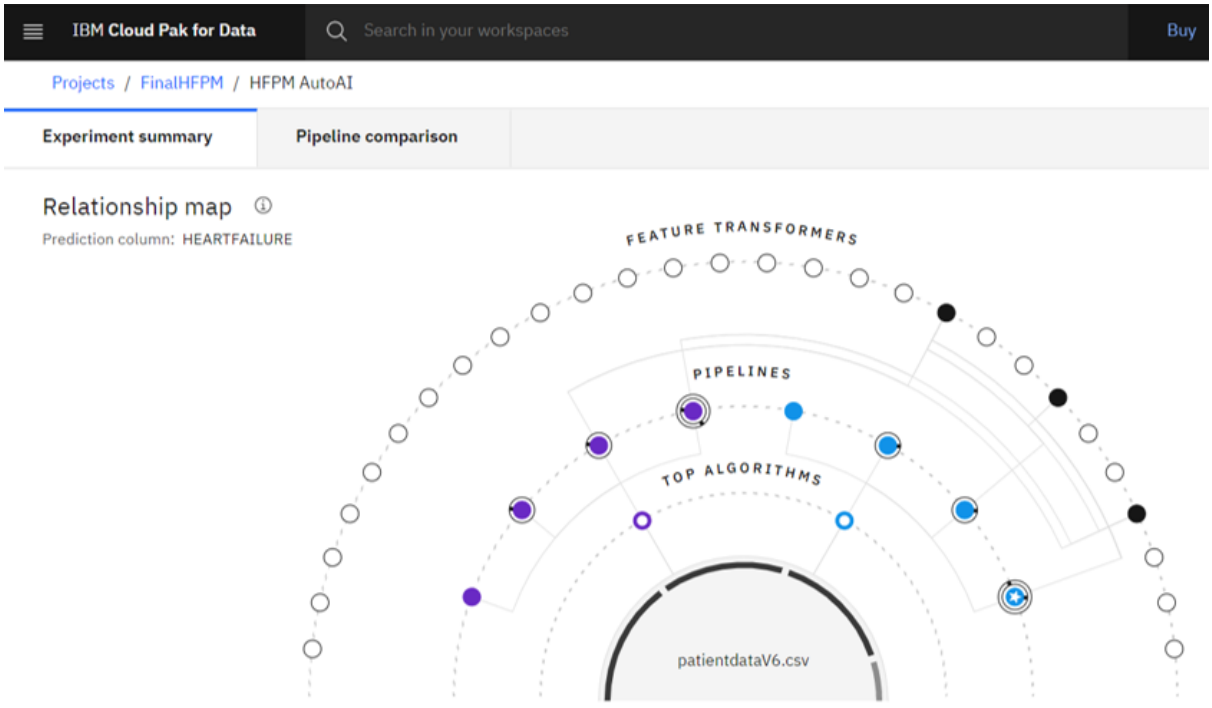


Fig 3 : Feature Transformations

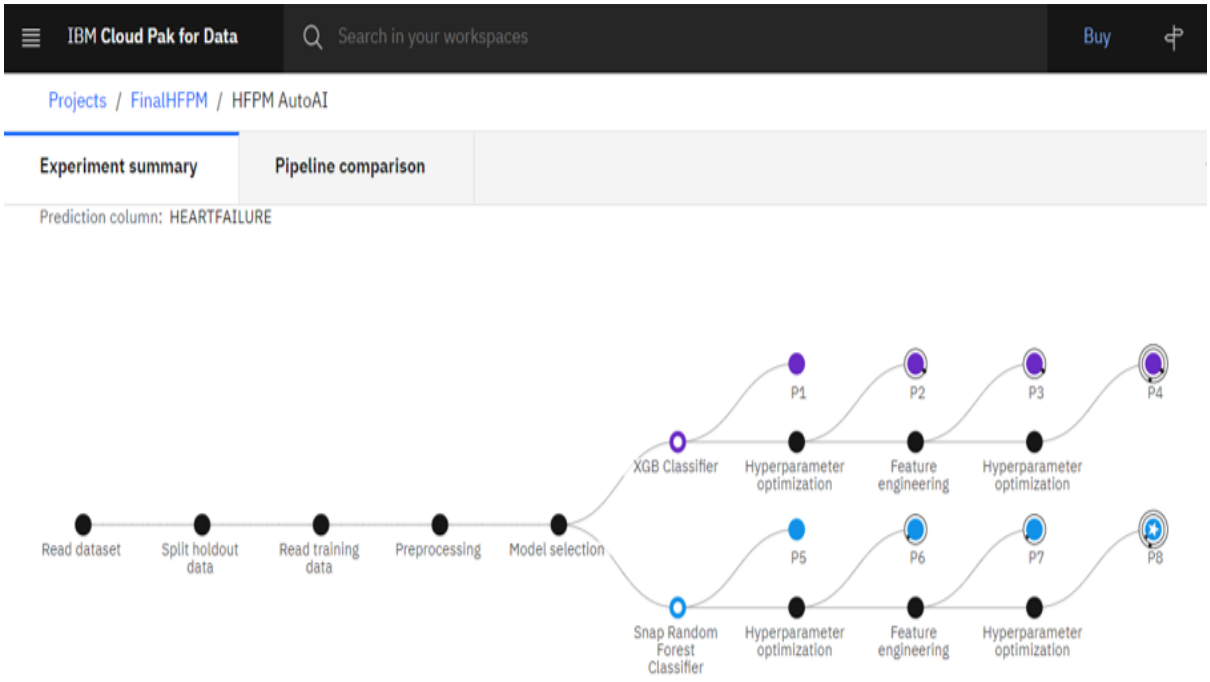


Fig 4 : Pipeline

Projects / FinalHFPM / HFPM AutoAI

Experiment summary | Pipeline comparison | ★ Rank by: Accuracy

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	00:01:17
	2	Pipeline 3	XGB Classifier	0.873	HPO-1 FE	00:00:43
	3	Pipeline 4	XGB Classifier	0.873	HPO-1 FE HPO-2	00:01:30
	4	Pipeline 7	Snap Random Forest Classifier	0.872	HPO-1 FE	00:00:54
	5	Pipeline 6	Snap Random Forest Classifier	0.869	HPO-1	00:00:09
	6	Pipeline 1	XGB Classifier	0.869	*None*	00:00:01
	7	Pipeline 2	XGB Classifier	0.869	HPO-1	00:00:13
	8	Pipeline 5	Snap Random Forest Classifier	0.861	*None*	00:00:01

Fig 5 : Pipeline Leader Board

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

Metric chart

Prediction column: HEARTFAILURE

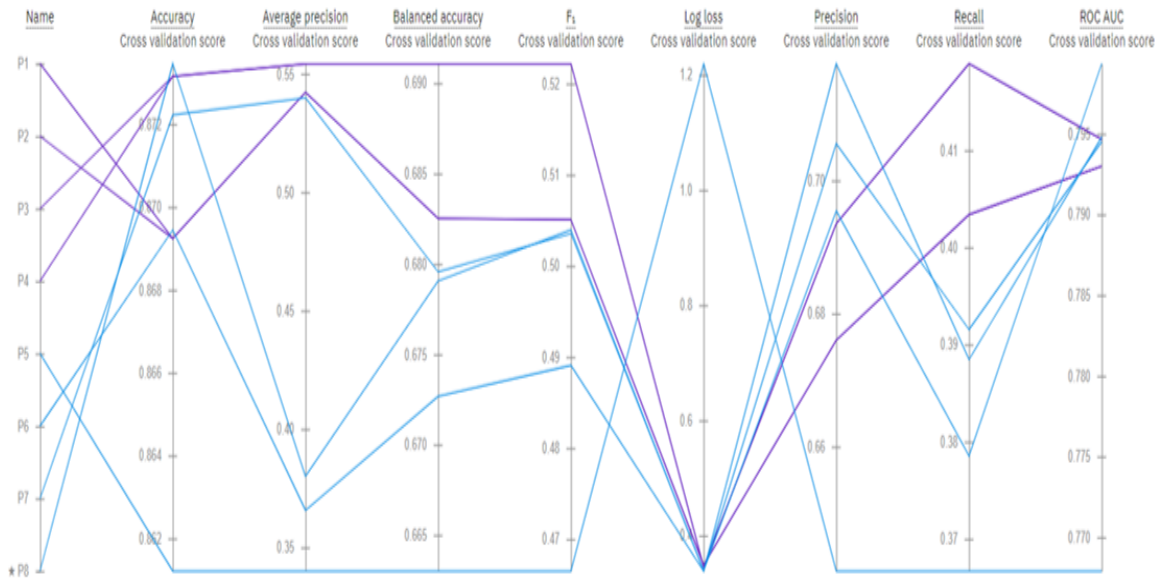


Fig 6 : Pipeline Comparison

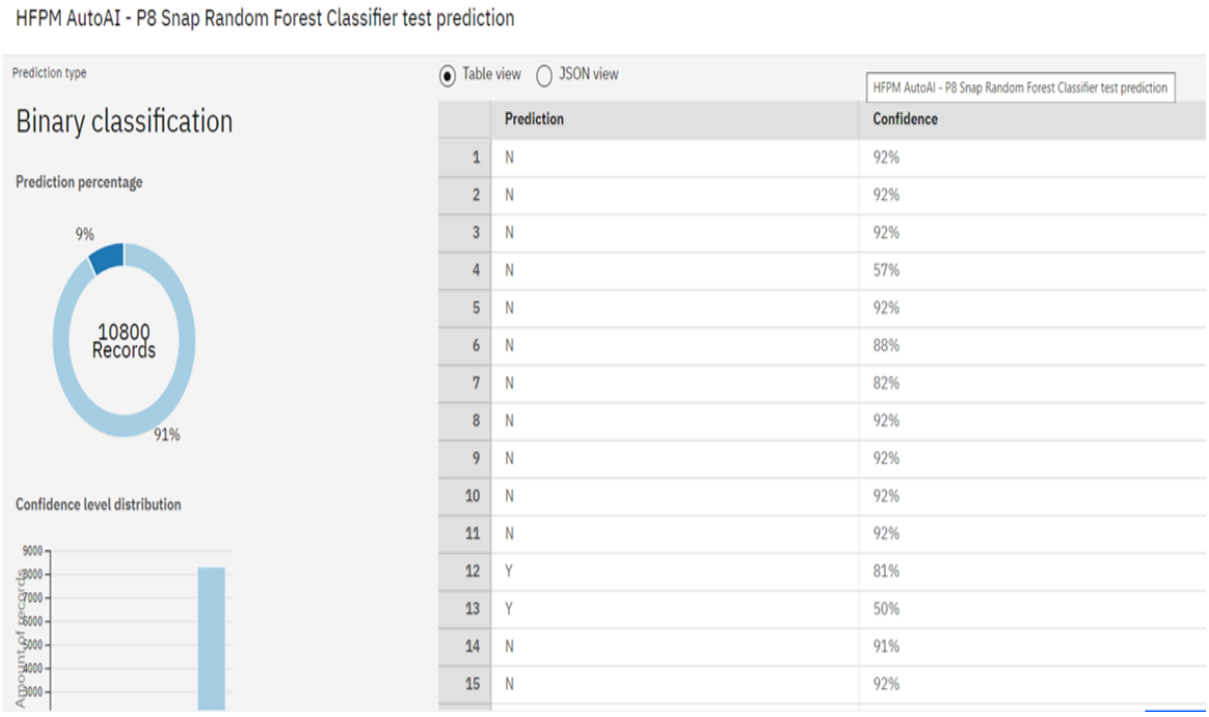


Fig 6 : Random Forest Classifier Prediction Table View

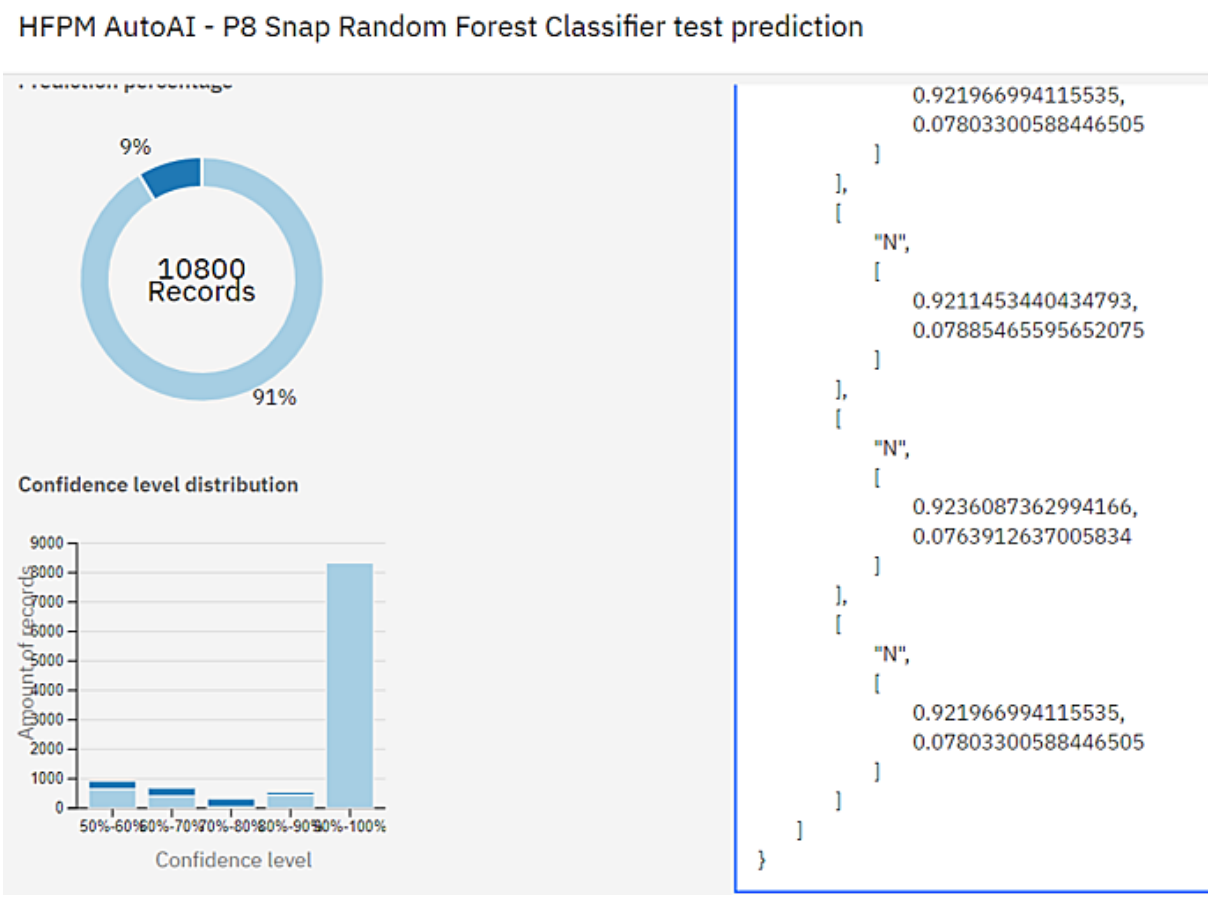


Fig 7 : Random Forest Classifier Prediction Json View

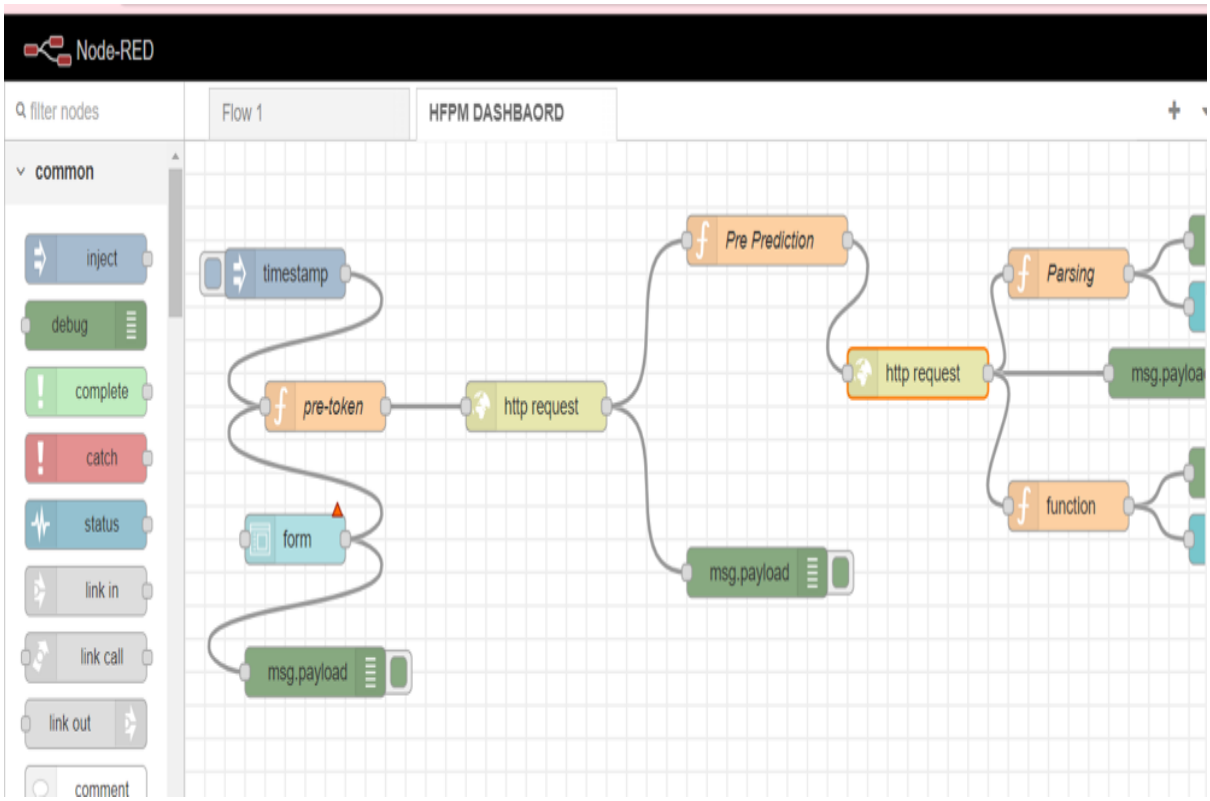


Fig 8 : Node - Red Editor

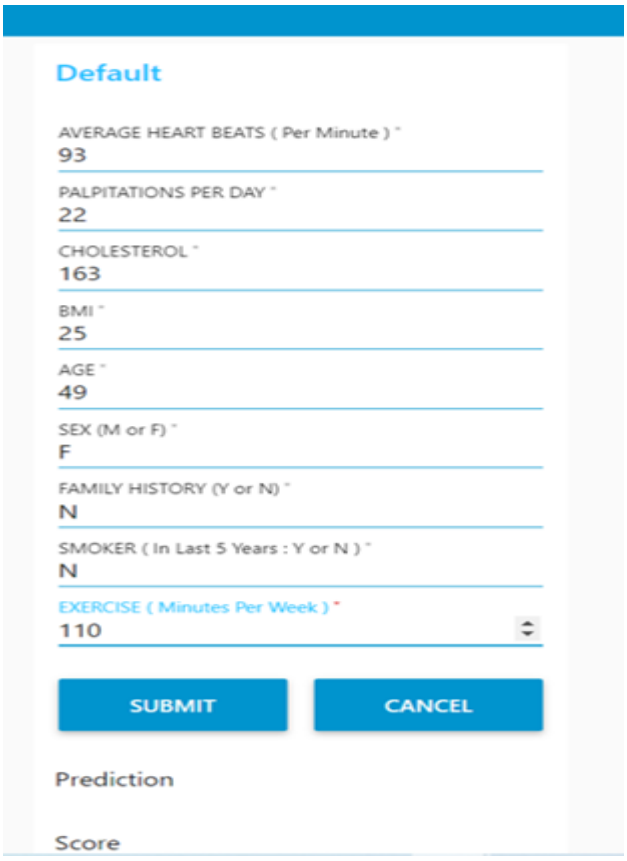


Fig 9 : Node Red UI

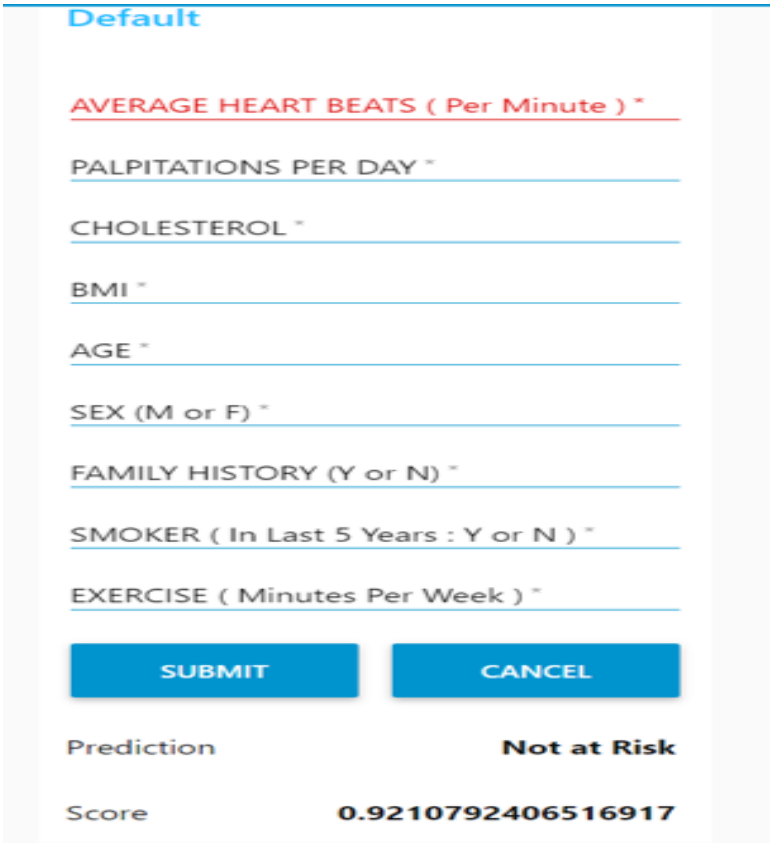


Fig 10 : Prediction Output 1

Fig 11 : Prediction Output 2

7 ADVANTAGES AND DISADVANTAGES

- IBM Auto AI service reduced the human intervention to implement feature engineering procedures.
- Prediction accuracy can be optimized in Auto AI service.
- Easy for comparative analysis of various machine learning algorithms whereby leader board shows the topmost algorithms with higher prediction accuracy.
- As auto AI service implements all phases of machine learning algorithm, those can't be visualized other than visualizations of the imported dataset.
- Difficult to access IBM cloud object storage, space deployment.
- Capacity Unit Hour (CUH) limitation for the user account.
- Needs internet connection until completion of the project work.

8 APPLICATIONS

- Smart Health Care
- Internet of Medical Things (IoMT)
- Web application for government and private hospitals
- Mobile application for government and private hospitals

9 CONCLUSION

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. Predicting the health risk of the patient is essential for survival point of view. In this project random forest classifier is used to predict the health risk of a patient and it classifies with 87% accuracy and in node-red user interface, accuracy is tested for the different set of user inputs.

10 FUTURE SCOPE

The future research directions include prediction of heart disease using hybrid machine learning algorithms and the comparative analysis of various supervised and unsupervised machine learning algorithms with respect to prediction accuracy.

11 BIBLIOGRAPHY

[1] IBM Project – Build – A – Thon https://smartinternz.com/ibm-build-a-thon.

[2] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE access 7 (2019): 81542-81554.

[3] Singh, Poornima, Sanjay Singh, and Gayatri S. Pandi-Jain. "Effective heart disease prediction system using data mining techniques." International journal of nanomedicine 13, no. T-NANO 2014 Abstracts (2018): 121.

[4] Ibrahim SP, Sivabalakrishnan M. An enhanced weighted associative classification algorithm without preassigned weight based on ranking hubs. Int J Adv Comput Sci Appl. 10(10); 2019.

[5] Methaila A, Kansal P, Arya H, Kumar P. Early heart disease prediction using data mining techniques. Comput Sci Inf Technol J. 53–59; 2014.

[6] Soni J, Ansari U, Sharma D, Soni S. Intelligent and effective heart disease prediction system using weighted associative classifiers. International Journal on Computer Science and Engineering. 2011;3(6):2385–92.

APPENDIX

A. SOURCE CODE


```
[
   {
      "id": "a4b58694.dd7a08",
      "type": "tab",
      "label": "HFPM DASHBAORD",
      "disabled": false,
      "info": ""
```

```
    },
    {
        "id": "ddaf6701.3ffd98",
        "type": "debug",
        "z": "a4b58694.dd7a08",
        "name": "",
        "active": true,
        "tosidebar": true,
        "console": false,
        "tostatus": false,
        "complete": "false",
        "statusVal": "",
        "statusType": "auto",
        "x": 130,
        "y": 320,
        "wires": []
    },
    {
        "id": "21769e48.4b87b2",
        "type": "function",
        "z": "a4b58694.dd7a08",
        "name": "pre-token",
                                                                    "func":
"global.set(\"hb\",msg.payload.hb)\nglobal.set(\"ppd\",msg.payload.ppd)\nglobal.set
(\"ch\",msg.payload.ch)\nglobal.set(\"bm\",msg.payload.bm)\nglobal.set(\"ag\",msg.
payload.ag)\nglobal.set(\"s\",msg.payload.s)\nglobal.set(\"fh\",msg.payload.fh)\nglo
bal.set(\"sm\",msg.payload.sm)\nglobal.set(\"em\",msg.payload.em)\n\nvar
apikey=\"3wJaOkK3HPyfdMylxVm9IN2Ox5RVlFHN1UrmnO_pZovx\";\nmsg.headers=
{\"content-type\":\"application/x-www-form-
urlencoded\"}\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:grant-
type:apikey\",\"apikey\":apikey}\nreturn msg;\n",
        "outputs": 1,
        "noerr": 0,
        "initialize": "",
        "finalize": "",
        "x": 140,
        "y": 160,
        "wires": [
            [
                "1775fa.9dce3a06"
            ]
        ]
    },
    {
```

```
    "id": "1775fa.9dce3a06",
    "type": "http request",
    "z": "a4b58694.dd7a08",
    "name": "",
    "method": "POST",
    "ret": "obj",
    "paytoqs": "ignore",
    "url": "https://iam.cloud.ibm.com/identity/token",
    "tls": "",
    "persist": false,
    "proxy": "",
    "authType": "",
    "x": 350,
    "y": 160,
    "wires": [
      [
        "c1b55bc1.e78c38",
        "34bd01eb.09e15e"
      ]
    ]
  },
  {
    "id": "c1b55bc1.e78c38",
    "type": "debug",
    "z": "a4b58694.dd7a08",
    "name": "",
    "active": true,
    "tosidebar": true,
    "console": false,
    "tostatus": false,
    "complete": "false",
    "statusVal": "",
    "statusType": "auto",
    "x": 570,
    "y": 260,
    "wires": []
  },
  {
    "id": "34bd01eb.09e15e",
    "type": "function",
    "z": "a4b58694.dd7a08",
    "name": "Pre Prediction",
        "func": "var hb = global.get(\"hb\")\nvar ppd = global.get(\"ppd\")\nvar ch =
global.get(\"ch\")\nvar bm = global.get(\"bm\")\nvar ag = global.get(\"ag\")\nvar s =
```

global.get(\"s\")\nvar fh = global.get(\"fh\")\nvar sm = global.get(\"sm\")\nvar em = global.get(\"em\")\n\nvar
token=msg.payload.access_token\nmsg.headers={'Content-Type':
'application/json',\"Authorization\":\"Bearer
\"+token,\"Accept\":\"application/json\"}\nmsg.payload={\"input_data\":[{\"fields\":[\"
AVGHEARTBEATSPERMIN\",\"PALPITATIONSPERDAY\",\"CHOLESTEROL\",\n
\"BMI\",\"AGE\",\"SEX\",\"FAMILYHISTORY\",\"SMOKERLAST5YRS\",\"EXERCISEMINP
ERWEEK\"],\n \"values\":[[hb,ppd,ch,bm,ag,s,fh,sm,em]]}]}\nreturn msg;",
    "outputs": 1,
    "noerr": 0,
    "initialize": "",
    "finalize": "",
    "x": 580,
    "y": 60,
    "wires": [
      [
        "119647ca.2b3888"
      ]
    ]
  },
  {
    "id": "119647ca.2b3888",
    "type": "http request",
    "z": "a4b58694.dd7a08",
    "name": "",
    "method": "POST",
    "ret": "obj",
    "paytoqs": "ignore",
    "url": "https://us-south.ml.cloud.ibm.com/ml/v4/deployments/fe2957d8-ff61-
4842-b54a-4b4f381eb7a2/predictions?version=2020-09-01",
    "tls": "",
    "persist": false,
    "proxy": "",
    "authType": "",
    "x": 730,
    "y": 140,
    "wires": [
      [
        "750758a0.e310c8",
        "3cbb3720.1d3b78",
        "7a3f7170.c6936"
      ]
    ]
  },

```
{
    "id": "750758a0.e310c8",
    "type": "debug",
    "z": "a4b58694.dd7a08",
    "name": "",
    "active": true,
    "tosidebar": true,
    "console": false,
    "tostatus": false,
    "complete": "payload",
    "targetType": "msg",
    "statusVal": "",
    "statusType": "auto",
    "x": 990,
    "y": 140,
    "wires": []
},
{
    "id": "5a3e627e.04601c",
    "type": "inject",
    "z": "a4b58694.dd7a08",
    "name": "",
    "props": [
        {
            "p": "payload"
        },
        {
            "p": "topic",
            "vt": "str"
        }
    ],
    "repeat": "",
    "crontab": "",
    "once": false,
    "onceDelay": 0.1,
    "topic": "",
    "payload": "",
    "payloadType": "date",
    "x": 100,
    "y": 80,
    "wires": [
        [
            "21769e48.4b87b2"
        ]
```

```
      ]
    },
    {
      "id": "e929f78c.388bb8",
      "type": "debug",
      "z": "a4b58694.dd7a08",
      "name": "",
      "active": true,
      "tosidebar": true,
      "console": false,
      "tostatus": false,
      "complete": "payload",
      "targetType": "msg",
      "statusVal": "",
      "statusType": "auto",
      "x": 1070,
      "y": 60,
      "wires": []
    },
    {
      "id": "3cbb3720.1d3b78",
      "type": "function",
      "z": "a4b58694.dd7a08",
      "name": "Parsing",
                "func": "if (msg.payload.predictions[0].values[0][0] == 'N')\n        {\n
msg.probablity = msg.payload.predictions[0].values[0][1][0];\n          msg.payload =
'Not at Risk';\n          }\nelse \n          {\n                    msg.probability =
msg.payload.predictions[0].values[0][1][1];\n                    msg.payload = 'At Risk';\n
}\nreturn msg;",
      "outputs": 1,
      "noerr": 0,
      "initialize": "",
      "finalize": "",
      "x": 880,
      "y": 80,
      "wires": [
        [
          "e929f78c.388bb8",
          "ecc83872.c85c08"
        ]
      ]
    },
```