1. INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications. This project aims to predict future Heart Disease by analysing data of patients which classifies whether they have heart disease or not using machine-learning algorithms.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm.

1.1 Overview

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using three classification algorithms namely Naïve Bayes, Decision Tree, and Random Forest are used at different levels of evaluations.

Machine Learning is used across many ranges around the world. The healthcare industry is no exclusion. Machine Learning can play an essential role in predicting presence/absence of locomotors disorders, Heart diseases and more. Such information, if predicted well in advance, can provide important intuitions to doctors who can then adapt their diagnosis and dealing per patient basis. We work on predicting possible Heart Diseases in people using Machine Learning algorithms.

2 LITERATURE SURVEY

2.1 Existing problem

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate chance of heartdisease in human. Early detection of cardiac diseases can decrease the mortality rate and overallcomplications. However, it is not possible to monitor patients every day in all cases accurately andconsultation of a patient for 24 hours by a doctor is not available since it requires more sapience,time and expe

rtise. Since we have a good amount of data in today's world, we can use various

machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can beused for health diagnosis in medicinal data

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented

with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says “Prevention is better than cure”, early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

[1]Purushottam ,et ,al proposed a paper “Efficient Heart Disease Prediction System” using hill climbing and decision tree algorithms .They used Cleveland dataset and preprocessing of data is performed before using classification algorithms. The Knowledge Extraction is done based on Evolutionary Learning (KEEL), an open-source data mining tool that fills the missing values in the data set.A decision tree follows top-down order.

[2] Santhana Krishnan. J ,et ,al proposed a paper “Prediction of Heart Disease Using Machine Learning Algorithms” using decision tree and Naive Bayes algorithm for prediction of heart disease. In decision tree algorithm the tree is built using certain conditions which gives True or False decisions. The algorithms like SVM, KNN are results based on vertical or horizontal split conditions depends on dependent variables. Dataset splits in 70% training and 30% testing by using some methods. This algorithm gives 91% accuracy. The second algorithm is Naive Bayes, which is used for classification. It can handle complicated, nonlinear, dependent data so it is found suitable for heart disease dataset as this dataset is also complicated, dependent and nonlinear in nature. This algorithm gives an 87% accuracy.

[3] Sonam Nikhar et al proposed paper “ Prediction of Heart Disease Using Machine Learning Algorithms” their research gives point to point explanation of Naïve Bayes and decision tree classifier that are used especially in the prediction of Heart Disease.

[4] Aditi Gavhane et al proposed a paper “Prediction of Heart Disease Using Machine Learning”, in which training and testing of dataset is performed by using neural network algorithm multi-layer perceptron. In this algorithm there will be one input layer and one output layer and one or more layers are hidden layers between these two input and output layers.

[5] Avinash Golande et al, proposed “Heart Disease Prediction Using Effective Machine Learning Techniques” in which few data mining techniques are used that support the doctors to differentiate the heart disease. Usually utilized methodologies are k-nearest neighbour, Decision tree and Naïve Bayes. Other unique characterization-based strategies utilized are packing calculation, Part thickness, consecutive negligible streamlining and neural systems, straight Kernel selfarranging guide and SVM (Bolster Vector Machine).

[8] Senthil Kumar Mohan et al, proposed “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” in which their main objective is to improve exactness in cardiovascular problems. The algorithms used are KNN, LR, SVM, NN to produce an improved

exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with linear model(HRFLM).

[10] Aakash Chauhan et al, proposed “Heart Disease Prediction using Evolutionary Rule Learning”. Data is directly retrieved from electronic records that reduce the manual tasks. The amount of services are decreased and shown major number of rules helps within the best prediction of heart disease. Frequent pattern growth association mining is performed on patient’s dataset to generate strong association.

2.2 Proposed solution

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

1.) Collection of Dataset
2.) Selection of attributes
3.) Data Pre-Processing
4.) Disease Prediction

Collection of dataset

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model.

Collection of dataset

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model.

Selection of attributes

Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, exang, etc are selected for the prediction. The Correlation matrix is used for attribute selection for this model.

Pre-processing of Data

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format.

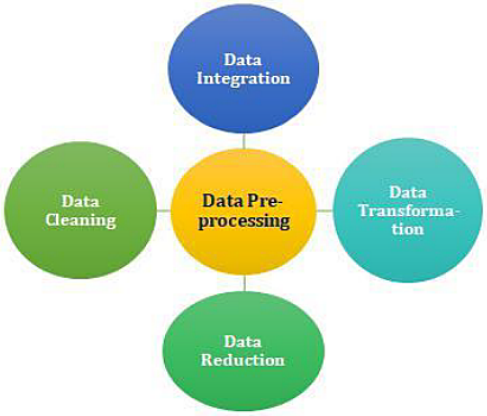


Figure: Data Pre-processing

Prediction of Disease

Various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression, Ada-boost, XG-boost are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction.

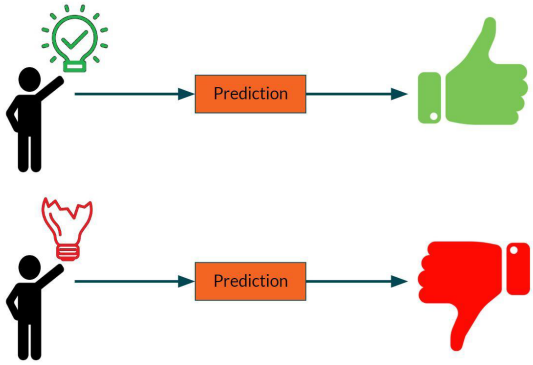


Figure: Prediction of Disease

3 THEORITICAL ANALYSIS

3.1 Block diagram

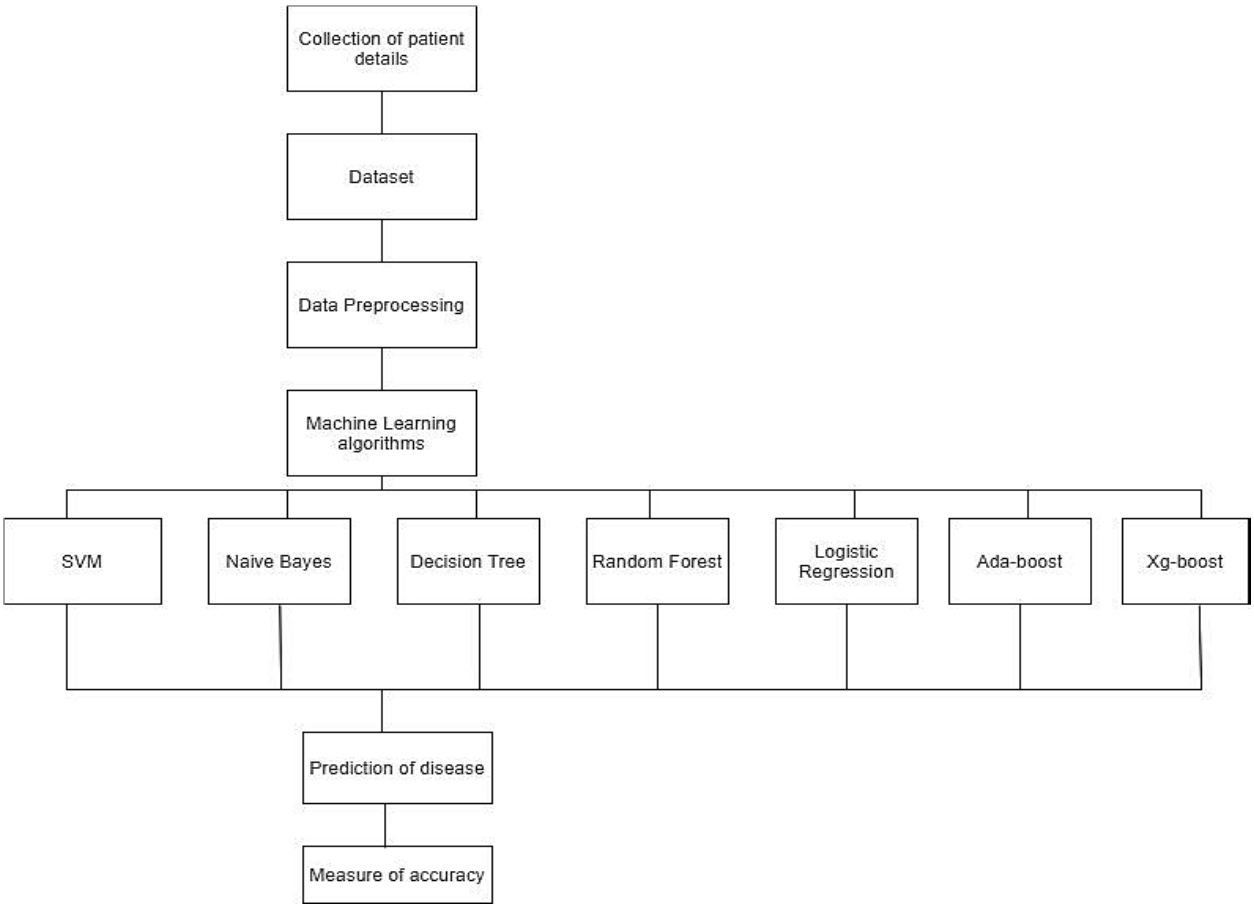


Fig. System Architecture

3.2 Hardware / Software designing

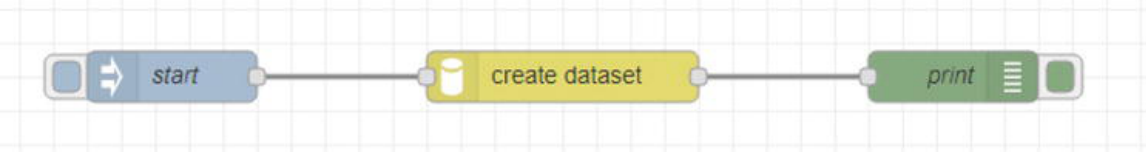
Services and Software:
Machine Learning Service
Watson Studio service
Node-Red service
Cloudant Service
Storage: Cloud Object Storage

4 EXPERIMENTAL INVESTIGATIONS

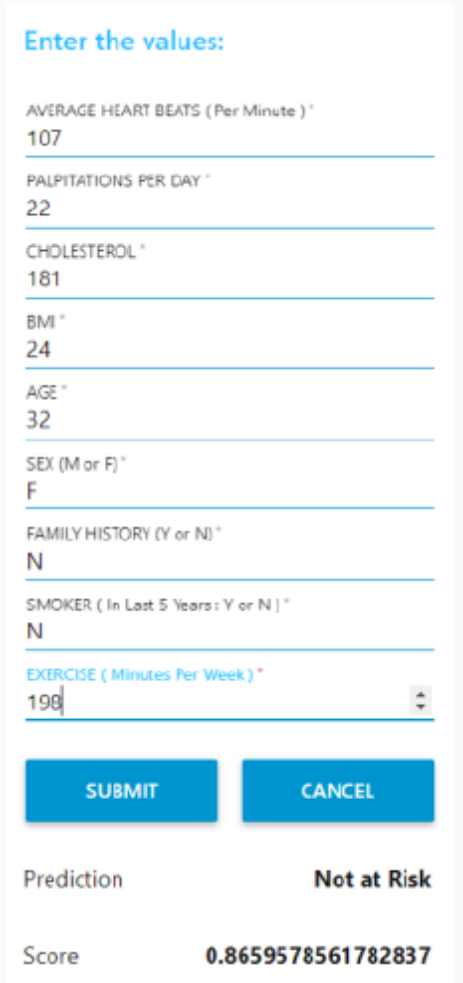
- Machine learning (ML) package for node-RED

This node-red-contrib-machine learning module for Node-RED contains a set of nodes offering machine learning functionalities. Such nodes have a python core that takes advantage of common ML libraries such as SciKit-Learn and Tenserflow. Classification and outlier detection can be performed using this package. performed using this package.
These flows create a dataset, train a model, and then evaluate it. Models, after training, can be used in real scenarios to make predictions. Flows and test datasets are available in the “test”

folder. We need to make sure that the paths specified inside nodes' configurations are correct before trying to execute the program. "node-red" can be run from the folder ".node-red/node-modules/node-red-contrib-machine-learning" and the paths will be automatically correct.



Experimental result:
This model basically shows the results in the form of prediction rate with accuracy as shown in below diagram.



5 FLOWCHART

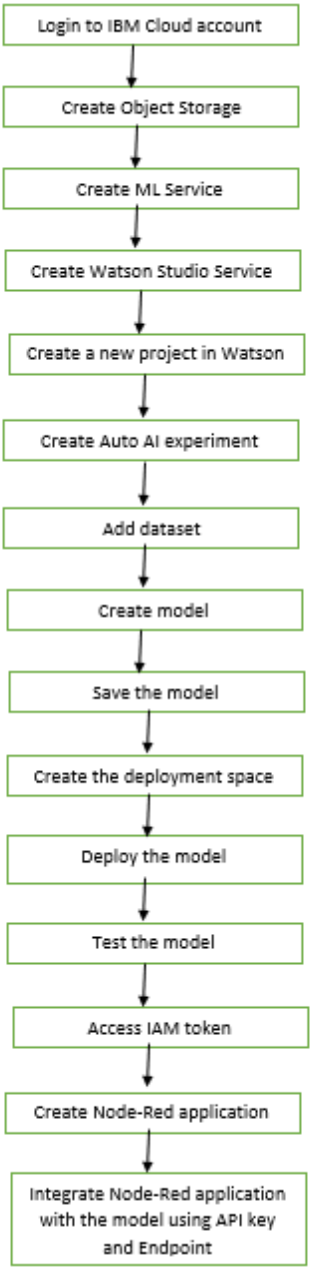


Fig. a

7 ADVANTAGES & DISADVANTAGE

- Higher prediction rate
- simple process
- If heart disease is predicted as at risk so countermeasures can be applied.

8 APPLICATIONS

- It can be used in health monitoring systems

9 CONCLUSION AND FUTURE SCOPE

The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes inhigh risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. This project resolved the feature selection i.e. backward elimination and RFECV behind the models and successfully predict the heart disease, with 85% accuracy. The model used was XGB regression.

Further for its enhancement, we can train on models and predict the types of cardiovascular diseases providing recommendations to the users, and also use more enhanced models.

10 BIBILOGRAPHY

1] Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43

[2] Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-8.

[3] Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334-43.

[4] Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies, 6(1), 637-9.

[5] Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp. 259-64). IEEE. ICCRDA 2020 IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012072 IOP Publishing doi:10.1088/1757-899X/1022/1/012072 9

[6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.

[7] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.

[8] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE. International Journal of Management, Technology And Engineering
Volume XI, Issue VI, JUNE/2021
ISSN NO : 2249-7455 Page No: 34

[9] Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). Inpatient deaths from acute myocardial infarction, 1982-92: analysis of data in the Nottingham heart attack register. BMJ, 315(7101), 159-64.

[10] Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women. International journal of epidemiologyy, 18(2), 361-7.

[11] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60). IEEE.

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)\nglobal.set(\"sm\",msg.payload.sm)\nglobal.set(\"em\",msg.payload.em)\n\nvar apikey=\"eeeSlOfL__xoU6-nRZ4uzYg485-5-2O0eBFrkcI8cAmx\";\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\",\"EXERCISEMINPERWEEK\"],\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/435304c8-dc84-42f8-b68a-e22dafbd6c21/predictions?version=2022-07-20",
    "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"
        ]
    ]
},
{
    "id": "ecc83872.c85c08",
    "type": "ui_text",
    "z": "a4b58694.dd7a08",
    "group": "80c2766c.ace4c8",
    "order": 2,
    "width": 0,
    "height": 0,
    "name": "",
    "label": "Prediction",
    "format": "{{msg.payload}}",
    "layout": "row-spread",
    "x": 1060,
    "y": 100,
    "wires": []
},
{
    "id": "df6f8738.fbce68",
    "type": "ui_form",
    "z": "a4b58694.dd7a08",
    "name": "",
    "label": "",
    "group": "80c2766c.ace4c8",
    "order": 1,
    "width": 0,
    "height": 0,
    "options": [
        {
            "label": "AVERAGE HEART BEATS ( Per Minute )",
            "value": "hb",
            "type": "number",
            "required": true,
            "rows": null
        },
        {
            "label": " PALPITATIONS PER DAY",
            "value": "ppd",
            "type": "number",
            "required": true,
            "rows": null
```

```
},
{
  "label": "CHOLESTEROL",
  "value": "ch",
  "type": "number",
  "required": true,
  "rows": null
},
{
  "label": "BMI",
  "value": "bm",
  "type": "number",
  "required": true,
  "rows": null
},
{
  "label": "AGE",
  "value": "ag",
  "type": "number",
  "required": true,
  "rows": null
},
{
  "label": "SEX (M or F)",
  "value": "s",
  "type": "text",
  "required": true,
  "rows": null
},
{
  "label": "FAMILY HISTORY (Y or N)",
  "value": "fh",
  "type": "text",
  "required": true,
  "rows": null
},
{
  "label": "SMOKER ( In Last 5 Years : Y or N )",
  "value": "sm",
  "type": "text",
  "required": true,
  "rows": null
},
{
  "label": "EXERCISE ( Minutes Per Week )",
```

```
            "value": "em",
            "type": "number",
            "required": true,
            "rows": null
        }
    ],
    "formValue": {
        "hb": "",
        "ppd": "",
        "ch": "",
        "bm": "",
        "ag": "",
        "s": "",
        "fh": "",
        "sm": "",
        "em": ""
    },
    "payload": "",
    "submit": "Submit",
    "cancel": "Cancel",
    "topic": "",
    "x": 110,
    "y": 240,
    "wires": [
        [
            "ddaf6701.3ffd98",
            "21769e48.4b87b2"
        ]
    ]
},
{
    "id": "addbc880.c08138",
    "type": "ui_text",
    "z": "a4b58694.dd7a08",
    "group": "80c2766c.ace4c8",
    "order": 2,
    "width": 0,
    "height": 0,
    "name": "",
    "label": "Score",
    "format": "{{msg.payload}}",
    "layout": "row-spread",
    "x": 1050,
    "y": 240,
    "wires": []
```

```
    },
    {
        "id": "7871bbfa.db2964",
        "type": "debug",
        "z": "a4b58694.dd7a08",
        "name": "",
        "active": true,
        "tosidebar": true,
        "console": false,
        "tostatus": false,
        "complete": "payload",
        "targetType": "msg",
        "statusVal": "",
        "statusType": "auto",
        "x": 1070,
        "y": 200,
        "wires": []
    },
    {
        "id": "7a3f7170.c6936",
        "type": "function",
        "z": "a4b58694.dd7a08",
        "name": "",
            "func": "if (msg.payload.predictions[0].values[0][0] == 'N')\n      {\n          msg.payload =
msg.payload.predictions[0].values[0][1][0]\n          \n      }\nelse \n      {\n          msg.paylaod =
msg.payload.predictions[0].values[0][1][1]\n    }\nreturn msg;",
        "outputs": 1,
        "noerr": 0,
        "initialize": "",
        "finalize": "",
        "x": 880,
        "y": 220,
        "wires": [
            [
                "7871bbfa.db2964",
                "addbc880.c08138"
            ]
        ]
    },
    {
        "id": "80c2766c.ace4c8",
        "type": "ui_group",
        "z": "",
        "name": "Default",
        "tab": "547fcc14.7f90b4",
```

```
        "order": 1,
        "disp": true,
        "width": "6",
        "collapse": false
    },
    {
        "id": "547fcc14.7f90b4",
        "type": "ui_tab",
        "z": "a4b58694.dd7a08",
        "name": "Home",
        "icon": "dashboard",
        "disabled": false,
        "hidden": false
    }
]
```