|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | |  | **INTRODUCTION** | |  |  |  |
| * 1. Overview | | | | | | | |
|  | |  | * 1. Purpose | | | | |
| **2** | |  | **LITERATURE SURVEY** | |  |  |  |
|  | |  | 2.1 Existing problem | | |  |  |
|  | |  | 2.2 Proposed solution | | |  |  |
| **3** | |  | **THEORITICAL ANALYSIS** | |  |  |  |
|  | |  | 3.1 Block diagram | |  |  |  |
|  | |  | 3.2 Hardware / Software designing | | |  |  |
| **4** | |  | **EXPERIMENTAL INVESTIGATIONS** | | |  |  |
| **5** | |  | **FLOWCHART** | |  |  |  |
| **6** | |  | **RESULT** |  |  |  |  |
| **7** | |  | **ADVANTAGES & DISADVANTAGES** | | |  |  |
| **8** | |  | **APPLICATIONS** | |  |  |  |
| **9** | |  | **CONCLUSION** | |  |  |  |
| **10** | |  | **FUTURE SCOPE** | | **`** |  |  |
| **11** | |  | **BIBILOGRAPHY** | |  |  |  |
|  | |  | **APPENDIX** | |  |  |  |
|  |  | | A. Source code | |  |  |  |
|  |  | |  |  |  |  |  |

1. INTRODUCTION
   1. Overview

The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still "information rich" but "knowledge poor". There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. Number of experiment has been already conducted to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and sometime Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The second conclusion is that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

**Project Description:**

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

* 1. Purpose

There is an urgent need for quick and accurate diagnosis methods to combat the disease. Appropriate computer based information and/or decision support systems can help in accomplishing clinical tests at a decreased cost and inconveniences.

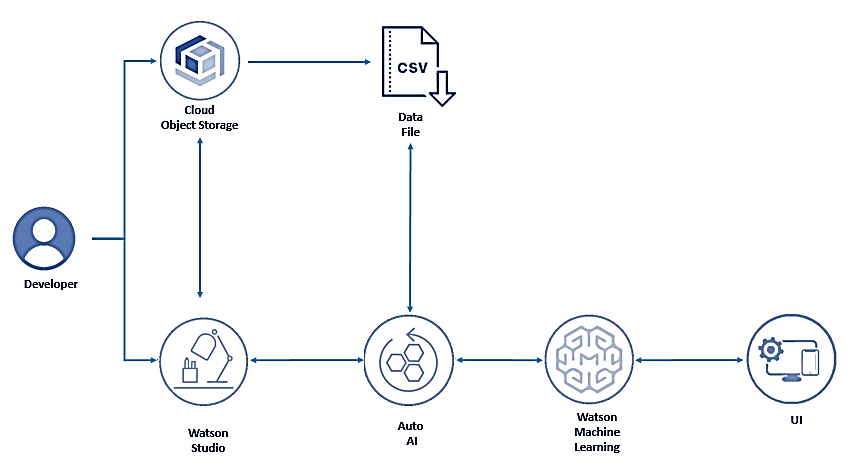
1. LITERATURE SURVEY
   1. Existing problem

It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason. The Heart Disease Prediction application is an end user support and online consultation project.

* 1. Proposed solution

Here, we propose a web application that allows users to get instant guidance on their heart disease through an intelligent system online. The application allows user to share their heart related issues. It then processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient’s details. Based on result, the can contact doctor accordingly for further treatment.  The system can be used for a heart disease consulting online.

1. THEORITICAL ANALYSIS
   1. Block diagram



* 1. Hardware / Software designing

**Required Environment:**  
IBM Watson Studio, IBM Machine Learning, IBM NodeRED, Data

**Cloud Object Storage**

Cloud object storage is a format for storing unstructured data in the cloud. Object storage is considered a good fit for the cloud because it is elastic, flexible and it can more easily scale into multiple petabytes to support unlimited data growth. The architecture stores and manages data as objects compared to block storage, which handles data as blocks, and logical volumes and file storage which store data in hierarchical files.

The object storage software design includes a globally unique identifier for each object along with rich, customizable metadata. The metadata is separated to enable other capabilities such as application- and user-specific data for indexing, interfaces that can be directly programmed by the application, a global namespace and more flexible data management policies.

An object identifier is an address tied to the object, which enables the object to be found over a distributed system. Objects may be spread across multiple data centers located in different parts of the world. The object storage-based data can be found without the user knowing the specific physical location of the data.

Object storage, along with the metadata, can be accessed directly via application program interfaces (APIs), HTTP and HTTPS. That differs from block storage volumes, which only can be accessed when they are attached to an operating system.

IBM Cloud Object Storage is designed to support exponential data growth and cloud-native workloads. With built-in high-speed file transfer capabilities, cross-region offerings and integrated services, IBM Cloud Object Storage can help you securely leverage your data.

**IBM Watson Studio**

IBM Watson Studio is an integrated environment designed to make it easy to develop, train, manage models, and deploy AI-powered applications and is a SaaS solution delivered on the IBM Cloud. It is evolving Data Science Experience on IBM Cloud with lot of new features to build AI applications. IBM Watson® Machine Learning helps data scientists and developers accelerate AI and machine-learning deployment. With its open, extensible model operation, Watson Machine Learning helps businesses simplify and harness AI at scale across any cloud.

**IBM Machine Learning**

Watson Machine Learning provides capabilities to help you:

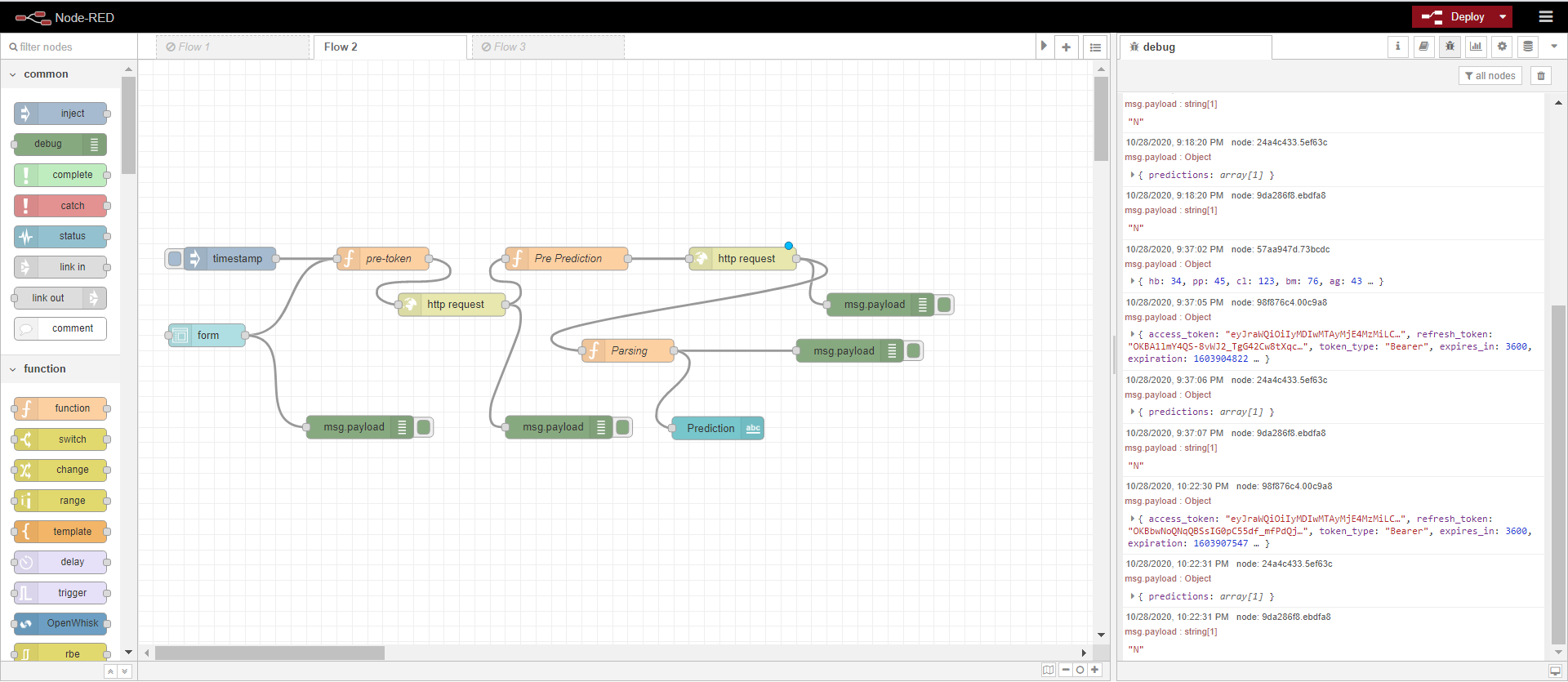
* Deploy models built with IBM Watson Studio and open source tools.
* Dynamically retrain models
* Automatically generate APIs to build AI-powered applications
* Streamline model management and deployment end-to-end with an easy-to-use interface

1. EXPERIMENTAL INVESTIGATIONS

A most general ways of experimental investigation stages are given as follows.

1. Project Idea
2. Conduct Background Research
3. Compose a Hypothesis
4. Design your Experiment
5. Collect Data
6. Analyse Data and Draw Conclusions

In this IBM Watson NodeRED application the experimentation can be done at the flow level of NodeRED by selecting debug mode.



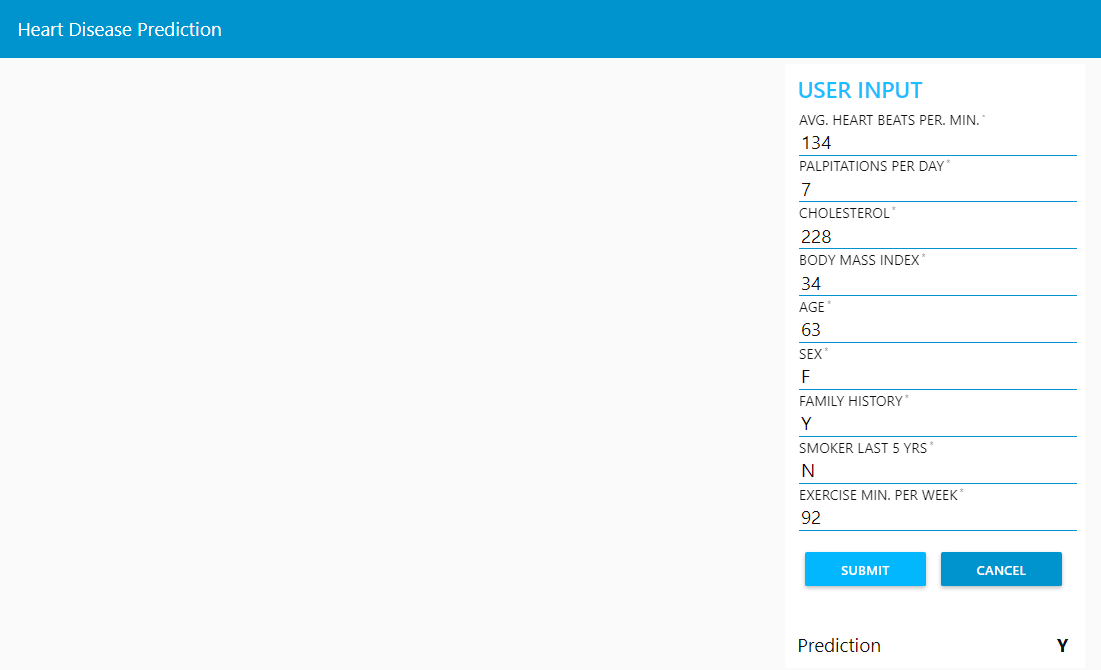
1. FLOWCHART



1. RESULT

The sample test data and one of the execution output is given below. The user has to enter heartbeat per minutes, palpitations per day, cholesterol level, Body mass index, age, sex, family history of heart disease, smoking habit, exercise per week data into the form prediction to assess the possibility of heart disease and predict an exact single value for heart failure.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AVG HEARTBEATS  PERMIN | PALPITATIONS  PERDAY | CHOLESTEROL | BMI | AGE | SEX | | FAMILY  HISTORY | SMOKER LAST5YRS | EXERCISE MINPERWEEK | HEART FAILURE |
| 93 | 22 | 163 | 25 | 49 | F | N | | N | 110 | N |
| 108 | 22 | 181 | 24 | 32 | F | N | | N | 192 | N |
| 86 | 0 | 239 | 20 | 60 | F | N | | N | 121 | N |
| 80 | 36 | 164 | 31 | 45 | F | Y | | N | 141 | Y |
| 66 | 36 | 185 | 23 | 39 | F | N | | N | 63 | N |
| 125 | 27 | 201 | 31 | 47 | M | N | | N | 13 | N |
| 83 | 27 | 169 | 20 | 71 | F | Y | | N | 124 | N |
| 107 | 31 | 199 | 32 | 55 | F | N | | N | 22 | N |
| 92 | 28 | 174 | 22 | 44 | F | N | | N | 107 | N |
| 84 | 12 | 206 | 25 | 50 | M | N | | N | 199 | N |
| 60 | 1 | 194 | 28 | 71 | M | N | | N | 27 | N |
| 134 | 7 | 228 | 34 | 63 | F | Y | | N | 92 | Y |



1. ADVANTAGES & DISADVANTAGES
   1. Advantages
      1. User can search for doctor’s help at any point of time.
      2. User can talk about their Heart Disease and get instant diagnosis.
      3. Doctors get more clients online.
      4. Very useful in case of emergency.
   2. Disadvantages
      1. Accuracy Issues: A computerized system alone does not ensure accuracy, and the warehouse data is only as good as the data entry that created it.
      2. The system is not fully automated; it needs data from user for full diagnosis.
2. APPLICATIONS

One of the most prominent application is to reduce the concerns of a patient at the earliest minutes. The application allows user to share their heart related issues. It then processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient’s details. Based on result, the can contact doctor accordingly for further treatment.  The system can be used for a heart disease consulting online.

1. CONCLUSION

Heart Disease is a fatal disease by its nature. This disease makes a life threatening complexities such as heart attack and death. The importance of Data Mining in the Medical Domain is realized and steps are taken to apply relevant techniques in the Disease Prediction. The observations from various work has led to the deployment of the proposed system architecture of this work using AutoML of Watson.

1. FUTURE SCOPE

The future scope of this project may include a most general model for various data sets. Also the custom models and hybrid performing models of the same could be a better choice of the future, to optimise outcome of selective data models.

1. BIBILOGRAPHY

[1] P. K. Thenmozhi 1, *International Journal of Engineering Research and General Science ,* vol. 2, no. 6, 2014.

1. APPENDIX
   1. Source code

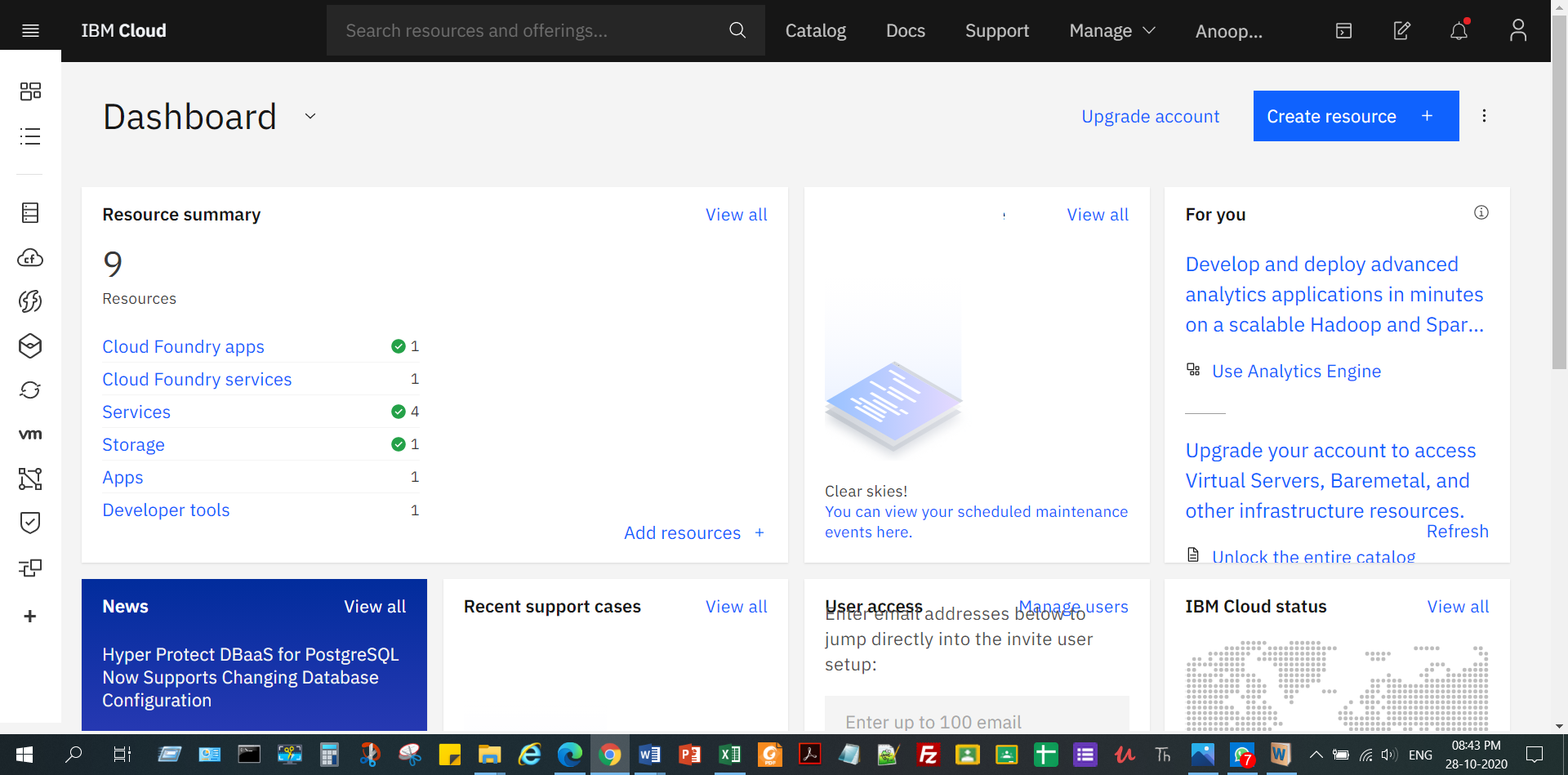
JSON – node red code

* [{"id":"204b0de7.c35dd2","type":"tab","label":"Flow 2","disabled":false,"info":""},{"id":"c62a9601.da3918","type":"ui\_form","z":"204b0de7.c35dd2","name":"","label":"","group":"c3393bbb.2feb28","order":2,"width":0,"height":0,"options":[{"label":"AVG. HEART BEATS PER. MIN.","value":"hb","type":"number","required":true,"rows":null},{"label":"PALPITATIONS PER DAY","value":"pp","type":"number","required":true,"rows":null},{"label":"CHOLESTEROL","value":"cl","type":"number","required":true,"rows":null},{"label":"BODY MASS INDEX","value":"bm","type":"number","required":true,"rows":null},{"label":"AGE","value":"ag","type":"number","required":true,"rows":null},{"label":"SEX","value":"sx","type":"text","required":true,"rows":null},{"label":"FAMILY HISTORY","value":"fh","type":"text","required":true,"rows":null},{"label":"SMOKER LAST 5 YRS","value":"sm","type":"text","required":true,"rows":null},{"label":"EXERCISE MIN. PER WEEK","value":"xr","type":"number","required":true,"rows":null}],"formValue":{"hb":"","pp":"","cl":"","bm":"","ag":"","sx":"","fh":"","sm":"","xr":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":90,"y":360,"wires":[["3edf7620.f8f9ca","57aa947d.73bcdc"]]},{"id":"57aa947d.73bcdc","type":"debug","z":"204b0de7.c35dd2","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","statusVal":"","statusType":"auto","x":290,"y":480,"wires":[]},{"id":"3edf7620.f8f9ca","type":"function","z":"204b0de7.c35dd2","name":"pre-token","func":"global.set(\"hb\",msg.payload.hb)\nglobal.set(\"pp\",msg.payload.pp)\nglobal.set(\"cl\",msg.payload.cl)\nglobal.set(\"bm\",msg.payload.bm)\nglobal.set(\"ag\",msg.payload.ag)\nglobal.set(\"sx\",msg.payload.sx)\nglobal.set(\"fh\",msg.payload.fh)\nglobal.set(\"sm\",msg.payload.sm)\nglobal.set(\"xr\",msg.payload.xr)\nvar apikey=\"0gc03RoZxh\_k2ip2k-8AlNkpcqcgN-54MUTjW1S-nvOx\";\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":320,"y":260,"wires":[["c7eb9d5d.74abc"]]},{"id":"c7eb9d5d.74abc","type":"http request","z":"204b0de7.c35dd2","name":"","method":"POST","ret":"obj","paytoqs":"ignore","url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":410,"y":320,"wires":[["98f876c4.00c9a8","54636c17.8cecd4"]]},{"id":"98f876c4.00c9a8","type":"debug","z":"204b0de7.c35dd2","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","statusVal":"","statusType":"auto","x":550,"y":480,"wires":[]},{"id":"54636c17.8cecd4","type":"function","z":"204b0de7.c35dd2","name":"Pre Prediction","func":"var hb = global.get(\"hb\")\nvar pp = global.get(\"pp\")\nvar cl = global.get(\"ad\")\nvar bm = global.get(\"bm\")\nvar ag = global.get(\"ag\")\nvar sx = global.get(\"sx\")\nvar fh = global.get(\"fh\")\nvar sm = global.get(\"sm\")\nvar xr = global.get(\"xr\")\nvar token=msg.payload.access\_token\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"Accept\":\"application/json\"}\nmsg.payload={\"input\_data\":[{\"fields\": [[\"AVG. HEART BEATS PER. MIN.\", \"PALPITATIONS PER DAY\", \"CHOLESTEROL\", \"BODY MASS INDEX\", \"AGE\", \"SEX\",\"FAMILY HISTORY\",\"SMOKER LAST 5 YRS\", \"EXERCISE MIN. PER WEEK\"]],\"values\": [[hb,pp,cl,bm,ag,sx,fh,sm,xr]]}]}\nreturn msg;","outputs":1,"noerr":0,"initialize":"","finalize":"","x":560,"y":260,"wires":[["69867b4a.d08904"]]},{"id":"69867b4a.d08904","type":"http request","z":"204b0de7.c35dd2","name":"","method":"POST","ret":"obj","paytoqs":"ignore","url":"https://us-south.ml.cloud.ibm.com/ml/v4/deployments/1275ea87-437d-48e2-ba83-8a8611ca0199/predictions?version=2020-09-01","tls":"","persist":false,"proxy":"","authType":"","x":790,"y":260,"wires":[["24a4c433.5ef63c","fd19cd4b.f6101"]]},{"id":"24a4c433.5ef63c","type":"debug","z":"204b0de7.c35dd2","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","statusVal":"","statusType":"auto","x":970,"y":320,"wires":[]},{"id":"c542f310.fb92f","type":"inject","z":"204b0de7.c35dd2","name":"","props":[{"p":"payload"},{"p":"topic","vt":"str"}],"repeat":"","crontab":"","once":false,"onceDelay":0.1,"topic":"","payload":"","payloadType":"date","x":120,"y":260,"wires":[["3edf7620.f8f9ca"]]},{"id":"9da286f8.ebdfa8","type":"debug","z":"204b0de7.c35dd2","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","statusVal":"","statusType":"auto","x":930,"y":380,"wires":[]},{"id":"fd19cd4b.f6101","type":"function","z":"204b0de7.c35dd2","name":"Parsing","func":"msg.payload=msg.payload.predictions[0].values[0][0]\nreturn msg;","outputs":1,"noerr":0,"initialize":"","finalize":"","x":640,"y":380,"wires":[["9da286f8.ebdfa8","9ac123f5.59ef2"]]},{"id":"9ac123f5.59ef2","type":"ui\_text","z":"204b0de7.c35dd2","group":"c3393bbb.2feb28","order":2,"width":0,"height":0,"name":"","label":"Prediction","format":"{{msg.payload}}","layout":"row-spread","x":757.6000366210938,"y":481.20001220703125,"wires":[]},{"id":"c3393bbb.2feb28","type":"ui\_group","z":"","name":"USER INPUT","tab":"8d99c5fb.2b9dd8","order":2,"disp":true,"width":"6","collapse":false},{"id":"8d99c5fb.2b9dd8","type":"ui\_tab","z":"","name":"Heart Disease Prediction","icon":"dashboard","disabled":false,"hidden":false}]

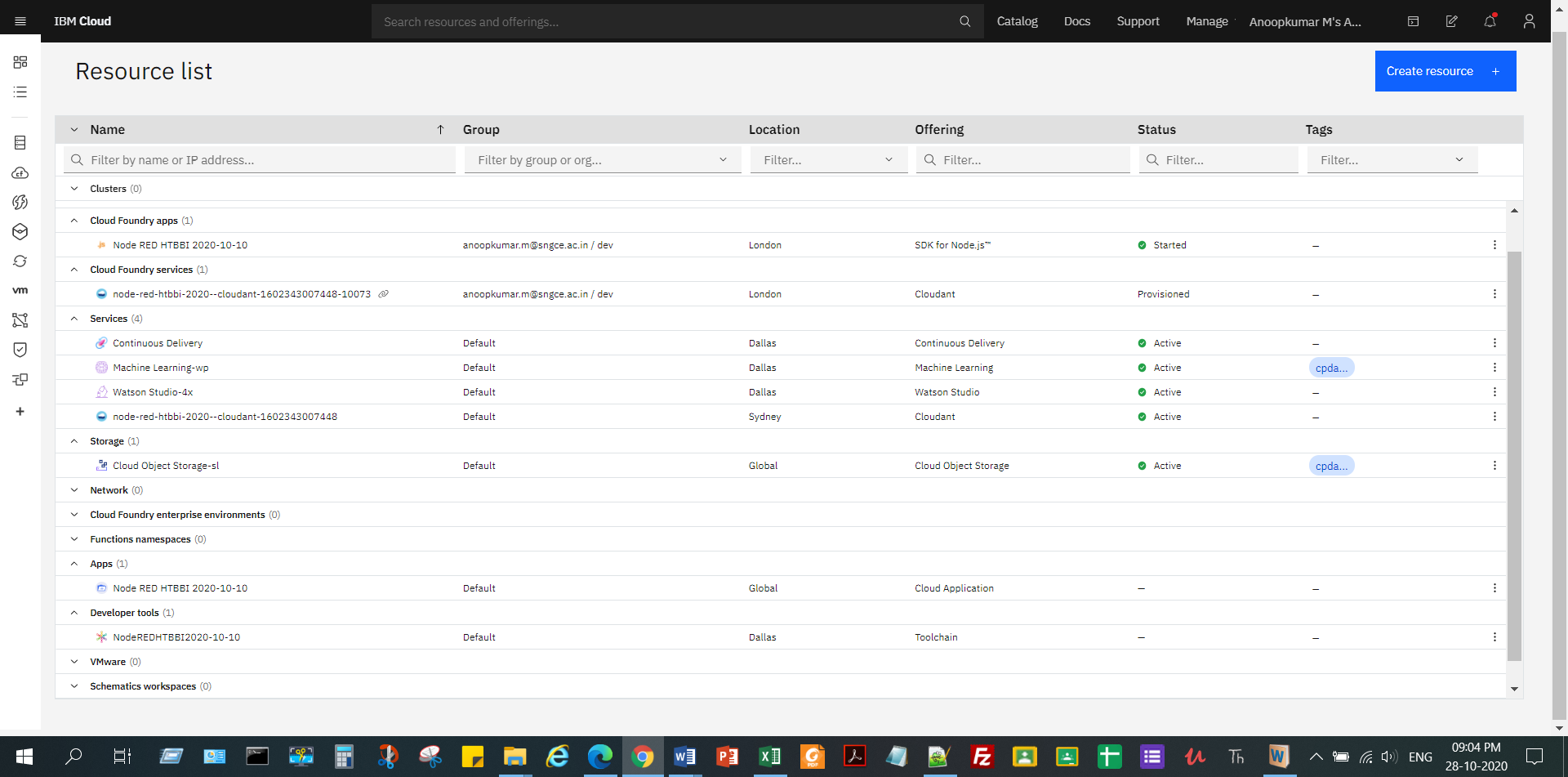
1. NodeRED URL :

https://node-red-htbbi-2020-10-10.eu-gb.mybluemix.net/ui/#!/0?socketid=opXwUpYQ4jFd2-Q\_AAAB

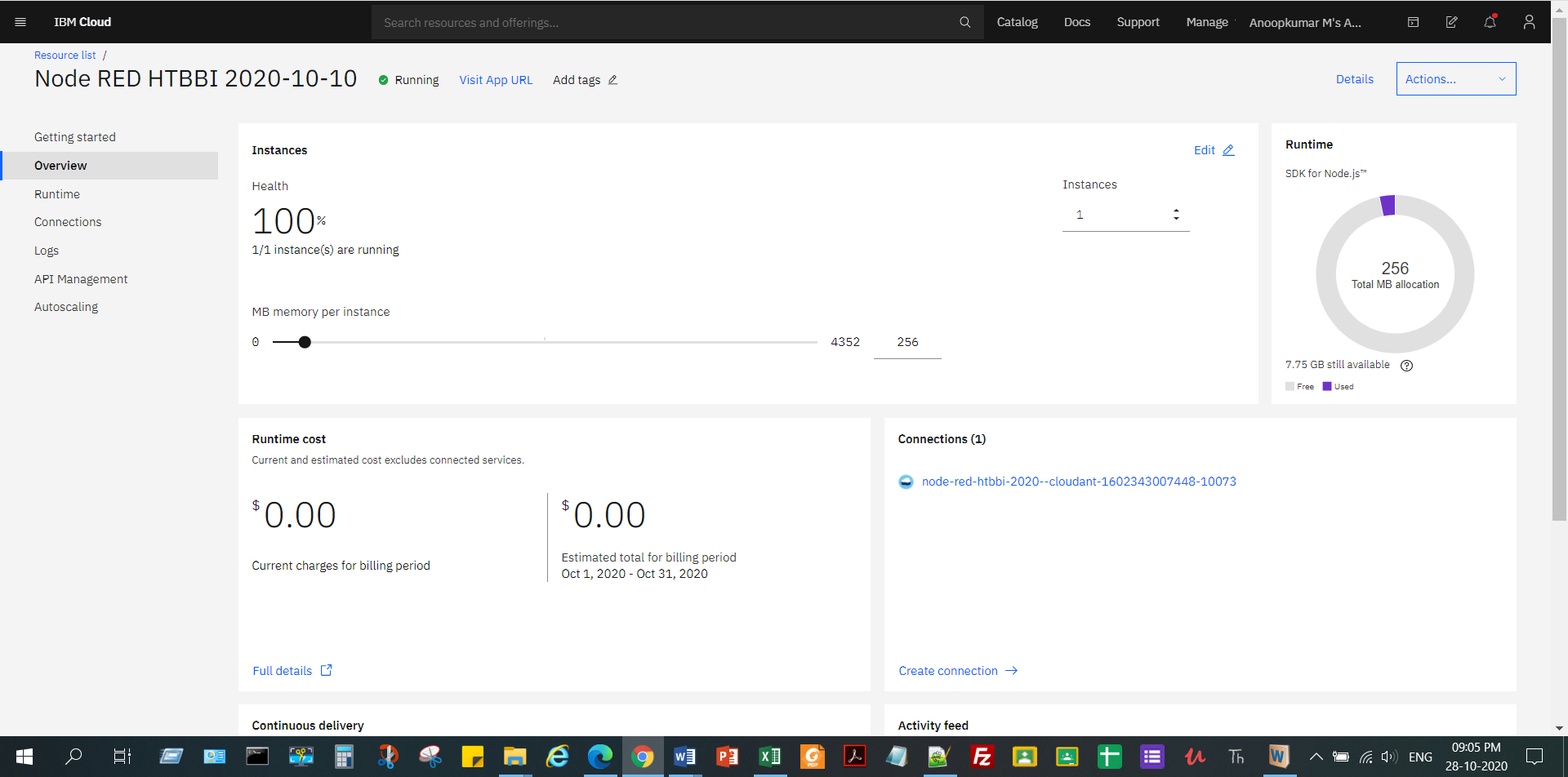
Dashboard



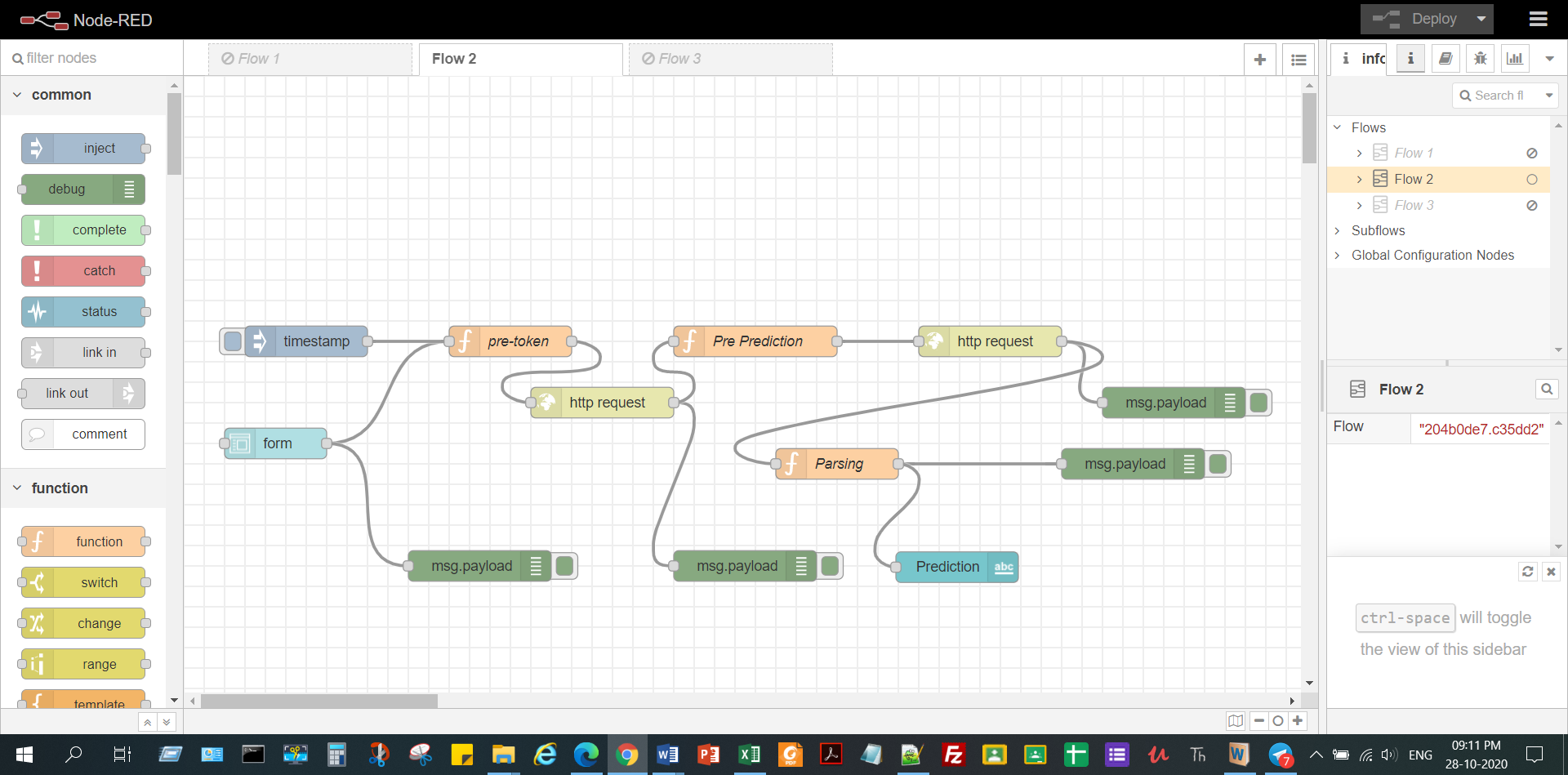
Resource List and Services



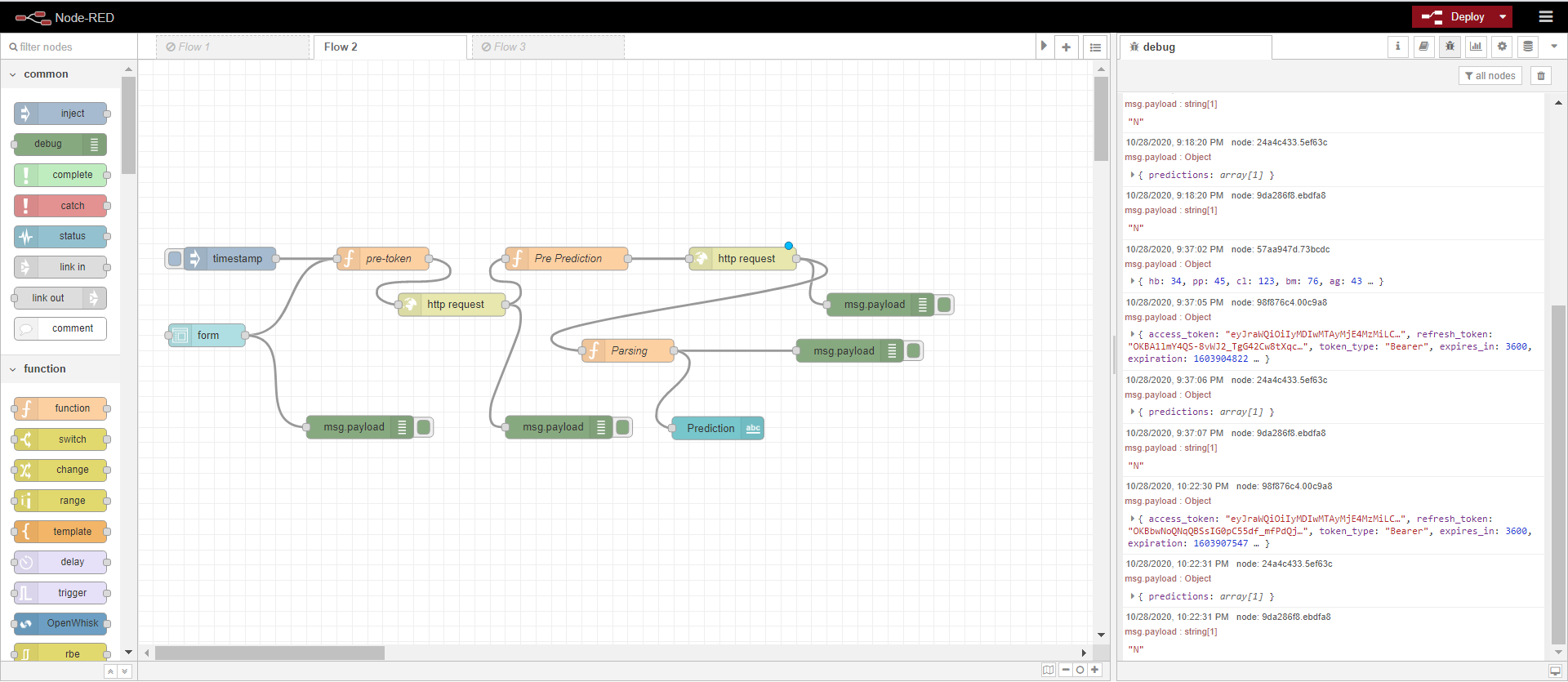
Node RED



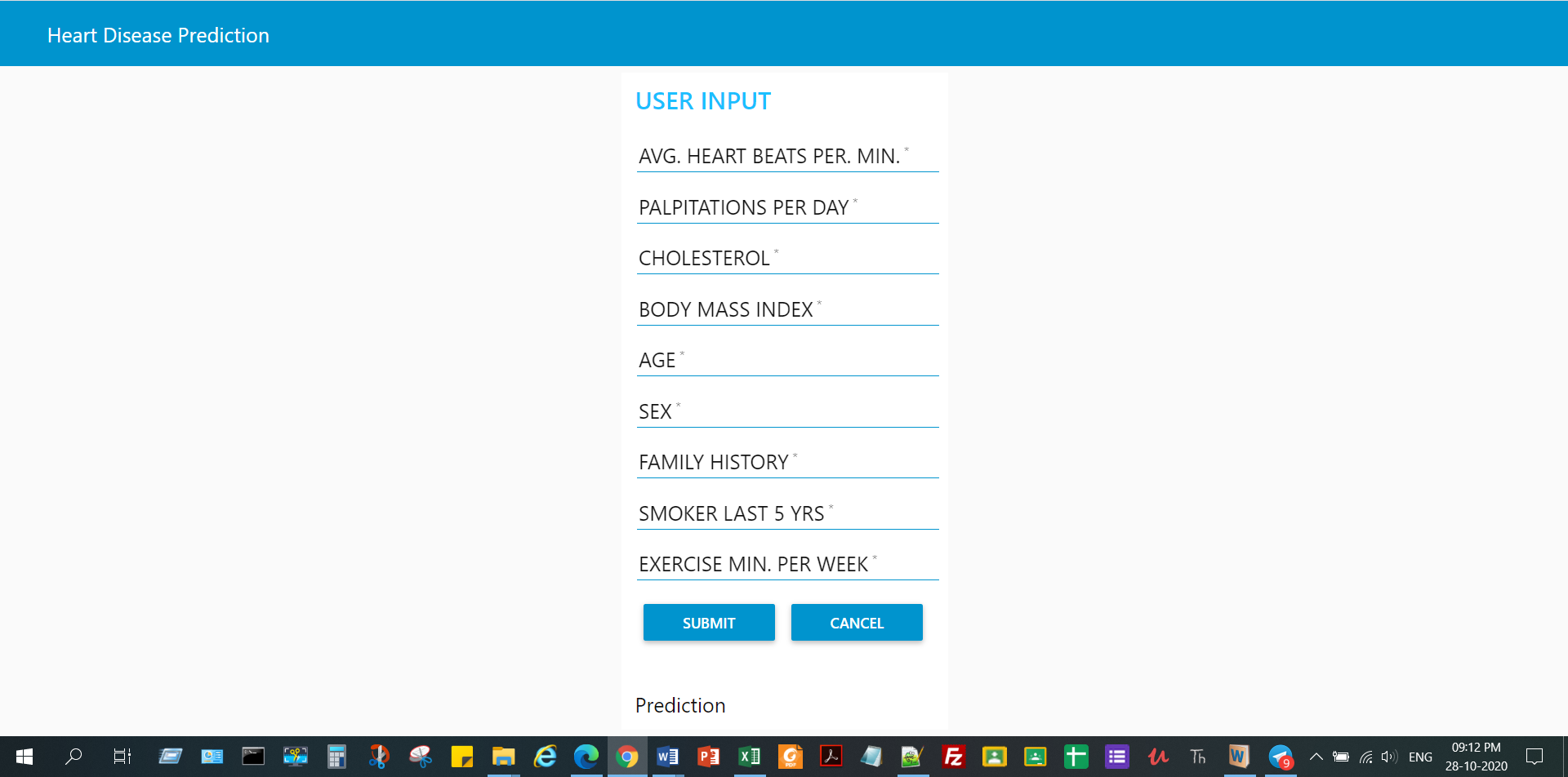
Node RED Flow Editor



Flow Experimentation



Heart Disease Prediction



Prediction Result

