Project Title:

Effective Heart Disease Prediction Using IBM Auto AI Service

1 INTRODUCTION

1.1 Overview

Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases.

1.2 Purpose

Cardiovascular diseases (CVDs) are the number one 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 the dataset uses in this project contains 9 features that can be used to predict mortality by heart failure. In this project, a prediction model using IBM Auto AI service is built and a web server application also built to showcase the prediction of heart failure. The goal of this effective heart disease prediction project is to determine whether a patient should be diagnosed with heart disease or not, which is a binary outcome, so the positive result = 1, the patient will be diagnosed with heart disease and the negative result = 0, the patient will not be diagnosed with heart disease.

2 LITERATURESURVEY

2.1 Existing Problem

It is very important to take into account the prediction of risk level of heart disease for healthcare industry in order to ease the medical treatment for the patients. Data science

classification techniques are used in a number of applications like healthcare analytics, customer analytics, marketing analytics, water quality analytics, textile production analytics, manufacturing analytics and textile waste analytics etc.,

There are various heart disease prediction models are available based on data mining techniques such as regression, clustering, association rule and classification techniques such as decision tree, naïve Bayes, random forest, artificial neural network etc.,

Even though there are lot of prediction models and ensemble techniques available, there is no single infrastructure or framework existing to execute all the above techniques altogether. Therefore, for developing the effective and best heart disease prediction model, lot of efforts are needed to incorporate everything.

Ramalingam et al.,[1] provided the survey about the Heart disease prediction using machine learning techniques.They discussed about algorithms,techniques and performance of various models such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models.

Mohan et al., [2] also proposed effective heart disease prediction using hybrid machine learning techniques.Their method aimed at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. Their prediction model is introduced with different combinations of features and several known classification techniques. They produced an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

Various research workers like Rajdhan et al.,[4],Patel et al.,[5], Singh & Kumar[6], Jagtap et al.,[7],Khourdifi& Bahaj [8],Gavhane A[9] and Jindal et al., have provided different solutions using different machine learning techniques for developing effective heart disease prediction model.

2.2 Proposed Solution

The proposed solution is to develop a prediction model using IBM Auto AI service. Under Auto AI, there are various machine learning techniques are available. Using pipeline concept, various prediction models with different machine learning techniques are developed and also found best prediction model among them. Then finally,with the best prediction model, a web server application is also built to showcase the prediction of heart failure using node RED service.

3 THEORETICAL ANALYSIS

3.1 Block Diagram

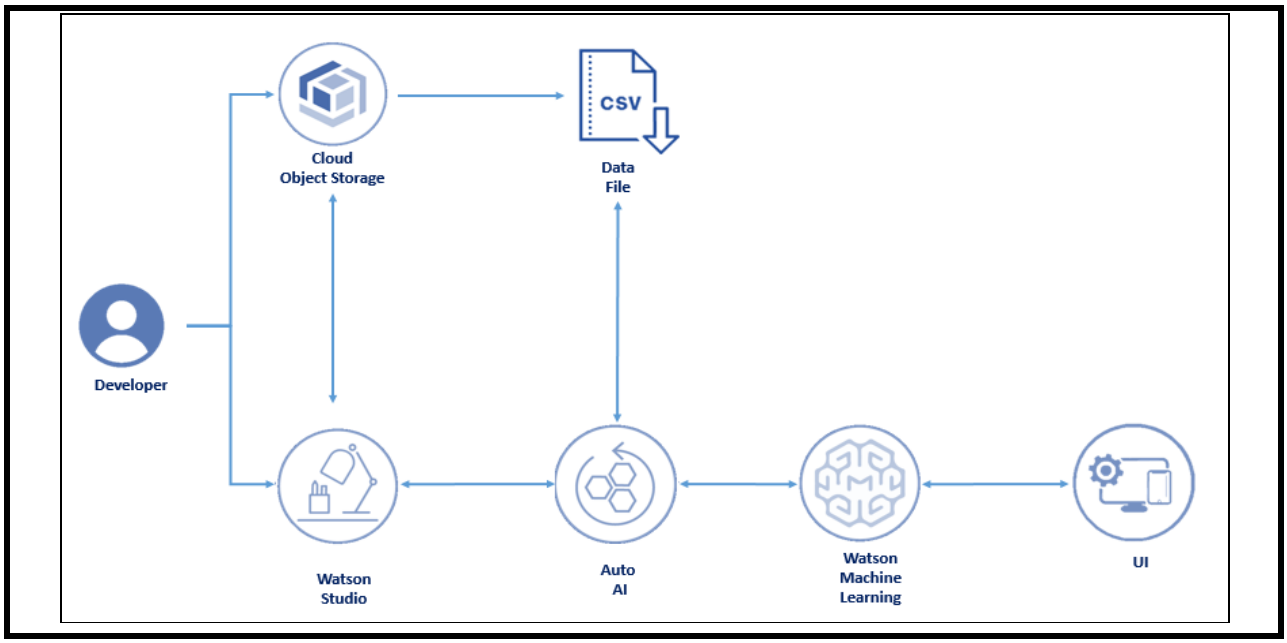


Figure.1 Technical Architecture of Effective Heart Disease Prediction Model

Figure.1 depicts the technical architecture of the proposed effective heart disease prediction model using IBM Auto AI service. The IBM cloud provides SaaS(Software-As-A-Service).Using this service, the developer creates the IBM cloud account to acquire the services provided by the IBM cloud. Initially,the developer creates cloud object storage service and Watson studio services. After that create Node RED service for deploying web server application.Then create Watson machine learning service and create Auto AI experiment with the patient dataset to build a machine learning model.Now the model is ready to deploy as web server and generate scoring end point.Then using Node RED service , create User Interface for accessing the web server application by the user and create web application to take user input and display prediction result on User Interface.

3.2 Hardware / Software designing

a. Hardware Designing

- PC/Laptop
- Operating System - Windows 7 / Windows 10 Pro
- 64-bit operating system, x64-based processor
- 4 GB / 8 GB RAM

b. Software Designing

- Google Chrome browser / Any other browser
- Internet Connection with optimum bandwidth
- SmartInternz account
- SmartInternz MailID credentials
- IBM Academic Initiative Account
- IBM Cloud Account

4 EXPERIMENTAL INVESTIGATIONS

For this project, the dataset named patientdataV6.csv is used. This dataset contains 10 attributes and 10,800 patient's sample records. Out of 10 attributes, 9 attributes are conditional /independent attributes and 1 is decision/dependant attribute. The conditional attributes are AVGHEARTBEATSPERMIN,FAMILYHISTORY,PALPITATIONSPERDAY,CHOLESTEROL, BMI, AGE,SEX,SMOKERLAST5YRS and EXERCISEMINPERWEEK.The decision attribute is HEART FAILURE. Out of 10,800 sample values of the dataset, one group consisted of 90% of the sample values (9720 sample records) of the dataset using for the learning (training purpose) and another group consisted of 10% of the sample values(1080 sample records) of the dataset for testing purposes.

Before the model development, the data preprocessing procedures such as missing value analysis, smoothing noisy data and data standardization were applied on the patient data set to produce reliable data.Then using cross validation technique, the dataset is split into training dataset and test dataset. The training dataset is used for model development purpose and 10% of the dataset is used for model validation purpose.Then the different prediction models are developed and best prediction model, snap random forest classifier is selected using IBM Auto AI service.

The AutoAI graphical tool in Watson Studio automatically analyses patient dataset and generates candidate model pipelines customized for predictive modelling problem. These model pipelines are created iteratively as AutoAI analyses the patient dataset and discovers data transformations, algorithms, and parameter settings that work best for problem setting. Results are displayed on a leader board, showing the automatically generated model pipelines ranked according to the given problem optimization objective.Then this prediction model is validated using test data. The accuracy of the best prediction model obtained is 87.3%. Once the pipeline creation is complete, view and compare the ranked pipelines in a leader board.The Saved model from the action menu for the pipeline with the highest accuracy or low error rate has chosen. This saves the pipeline as a Machine Learning asset in this project. A notification is received that the link to view the saved model in this project.

Then this model is deployed for ready to use. The API key and the relevant endpoint URL

is generated. The user interface is created using IBM node RED service with Node RED flow editor. The Node RED flow in json format is also deployed and associated with prediction model already created. Then invoke the Input User Interface screen with dashboard.Now the patient input details are given through UI and the predicted output is displayed with score.The flow work of this project is clearly depicted in Figure.2.

5 FLOWCHART

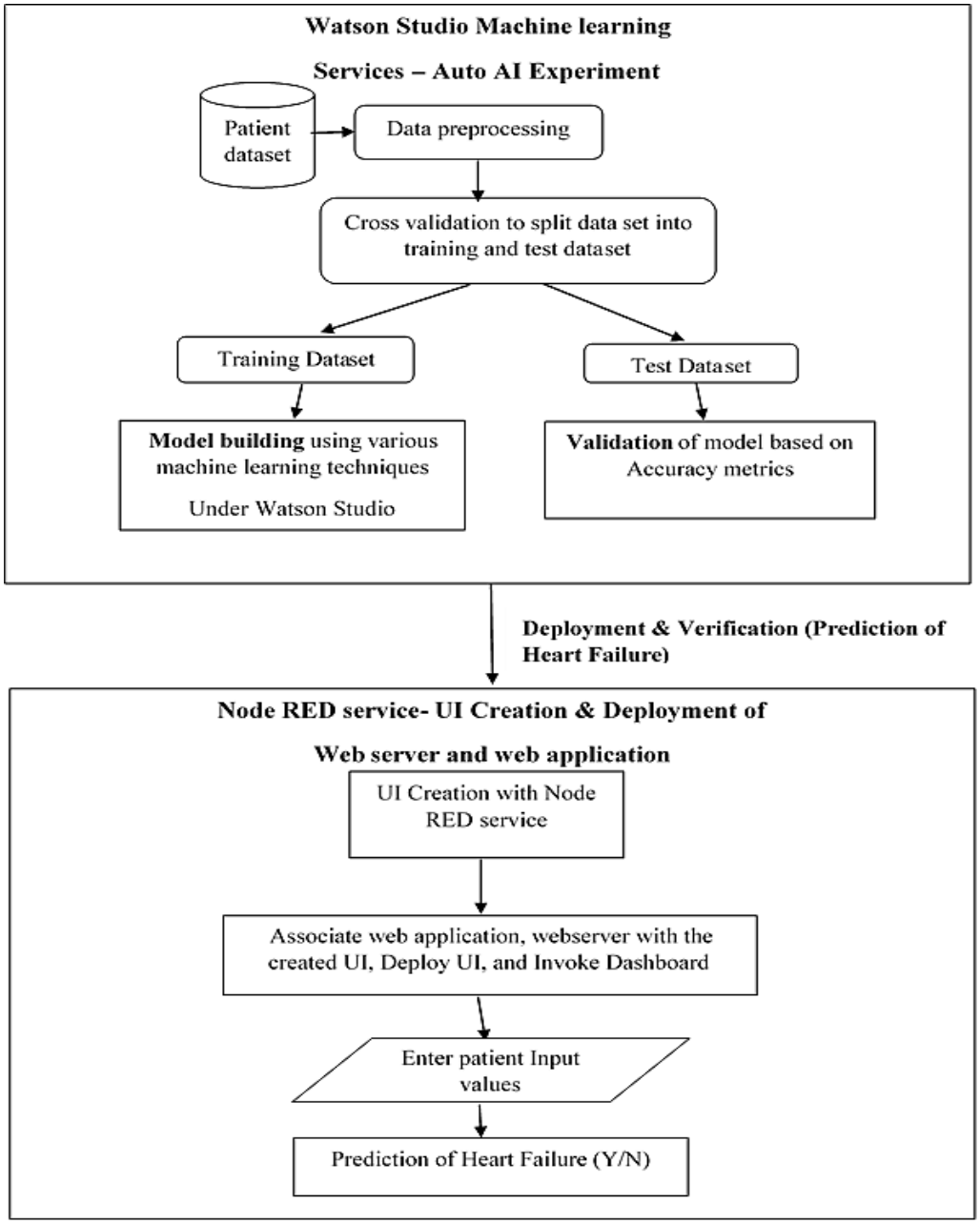
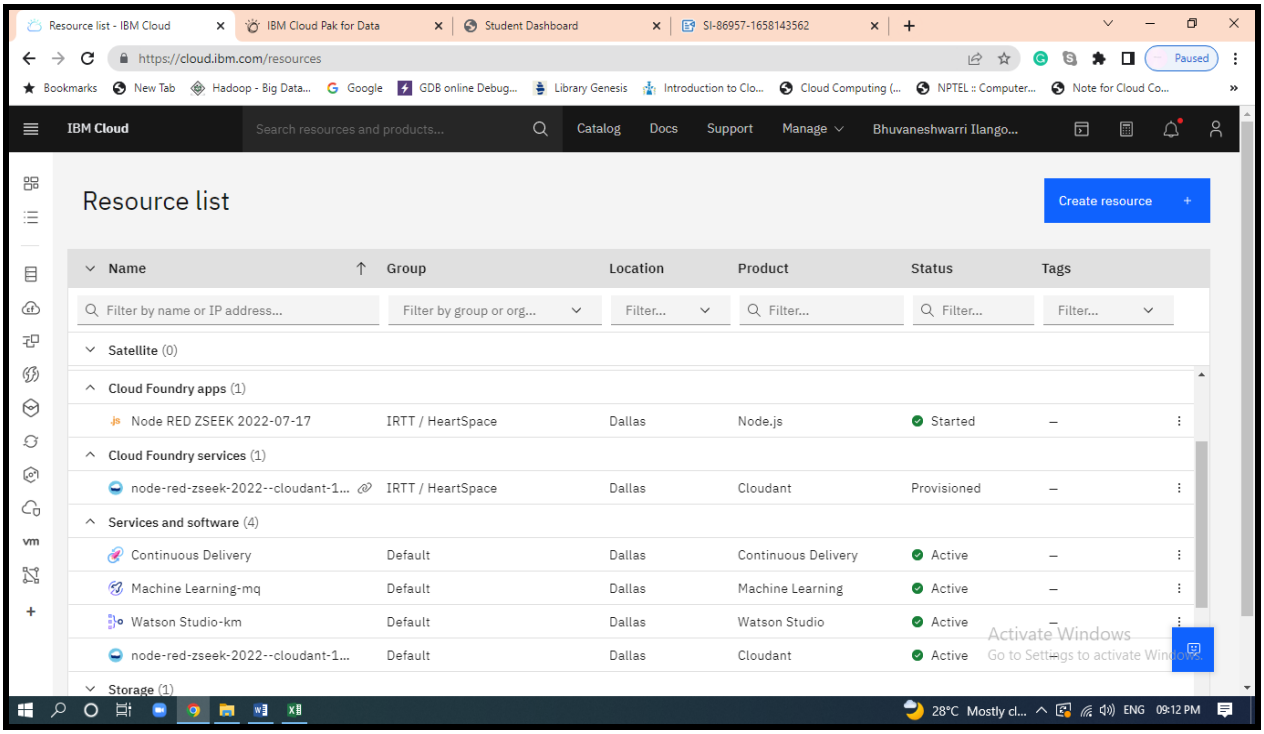
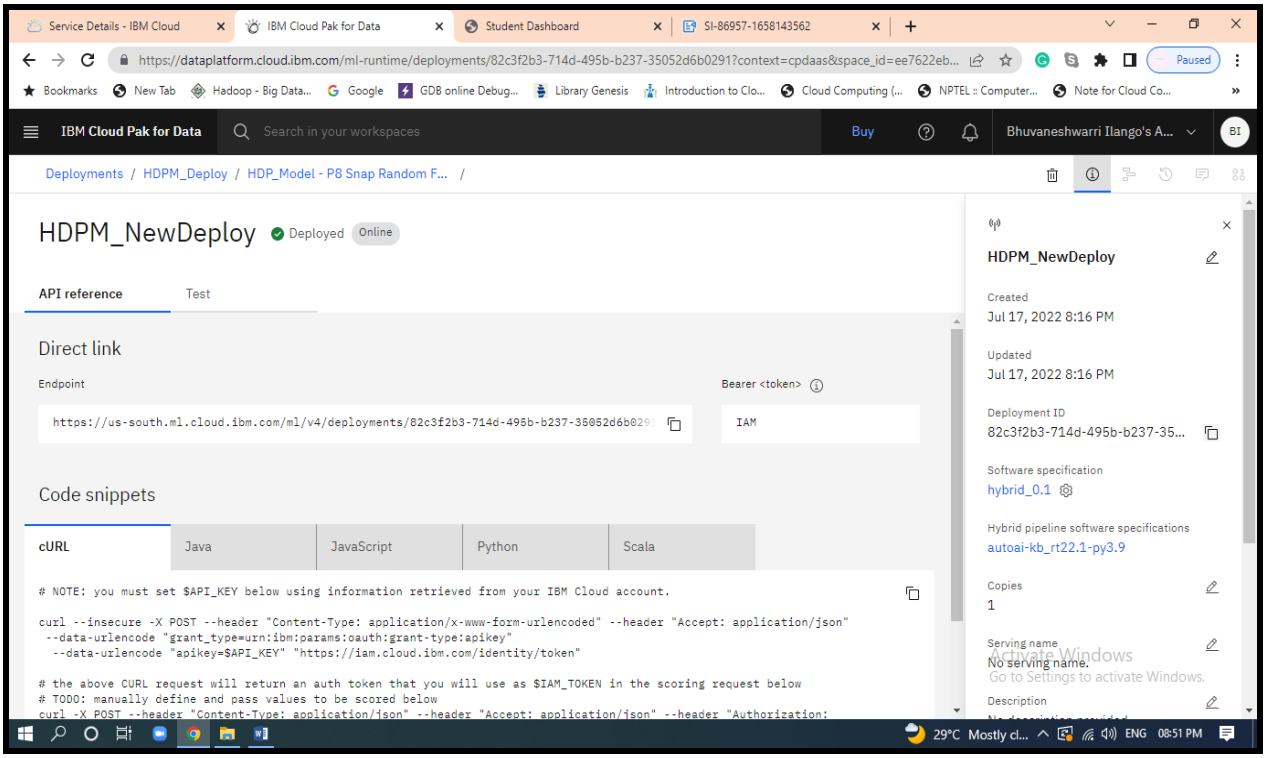
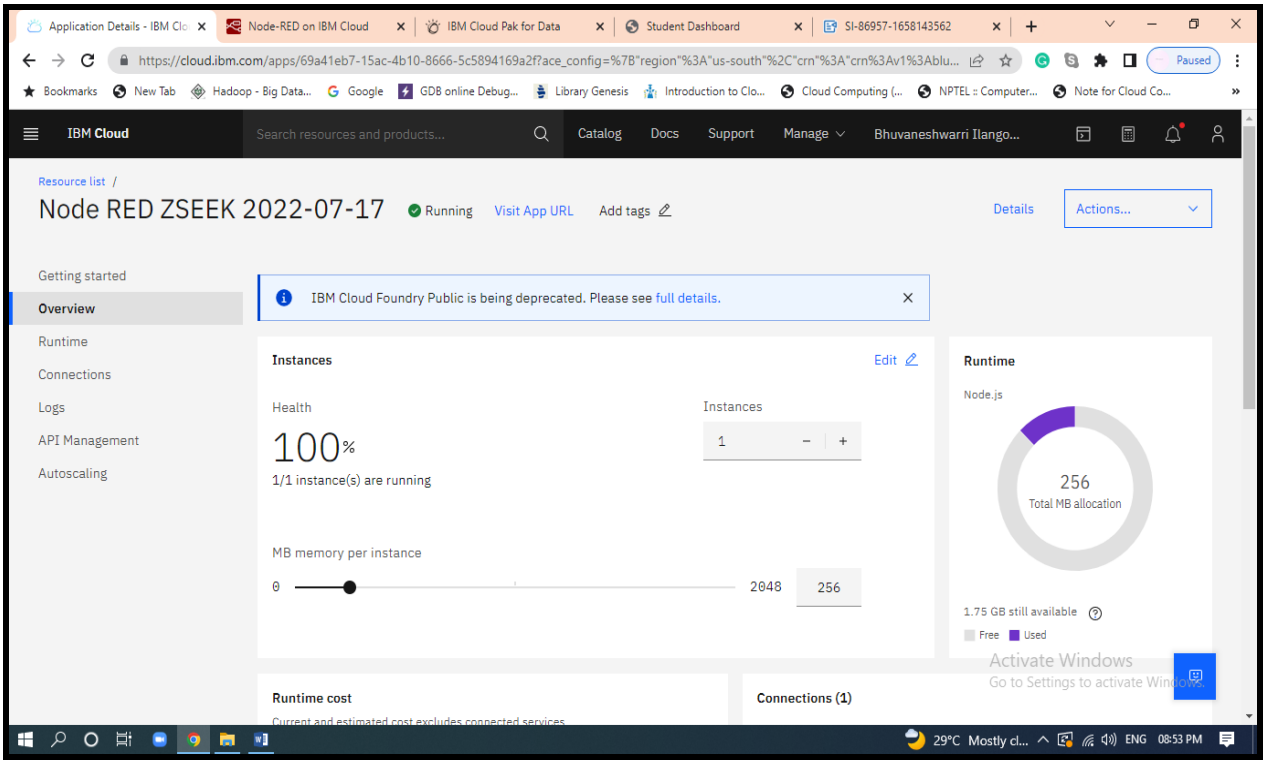


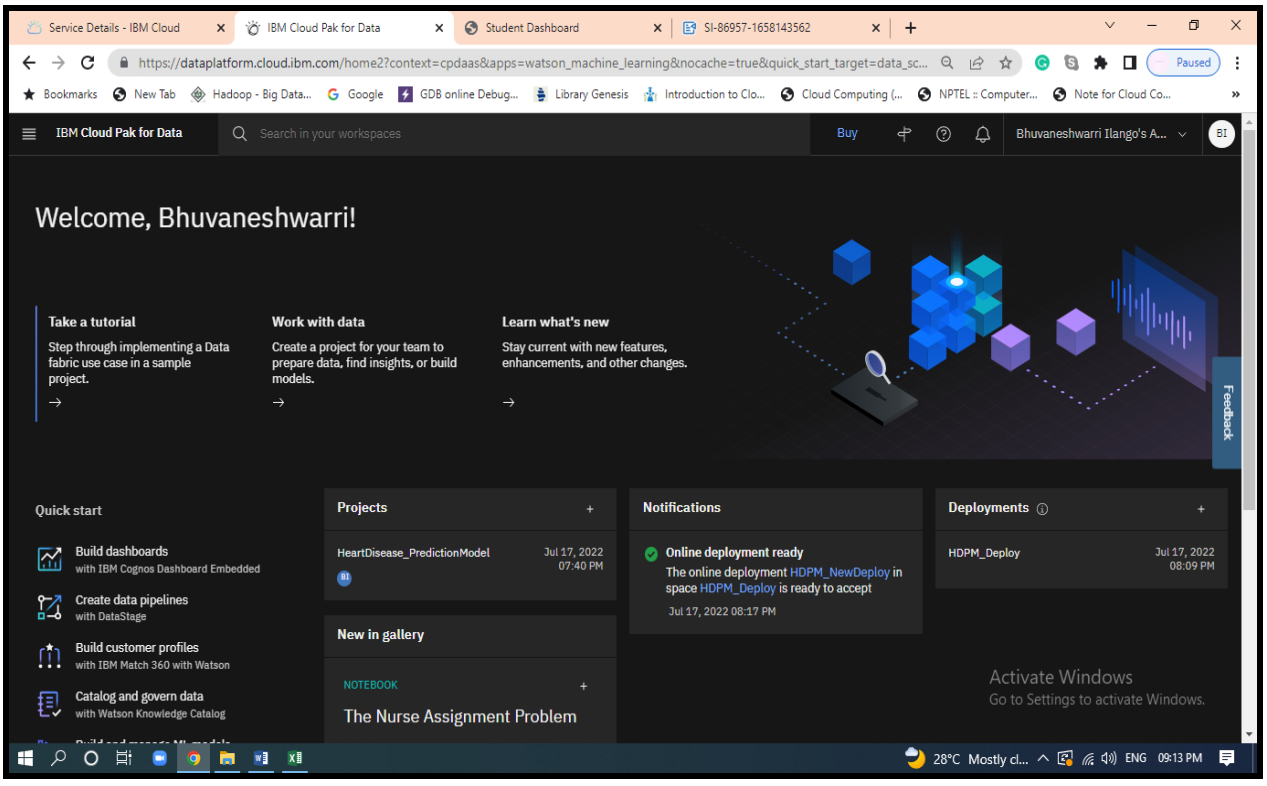
Figure.2 Flow work of Effective Heart Disease Prediction Model using IBM Auto AI

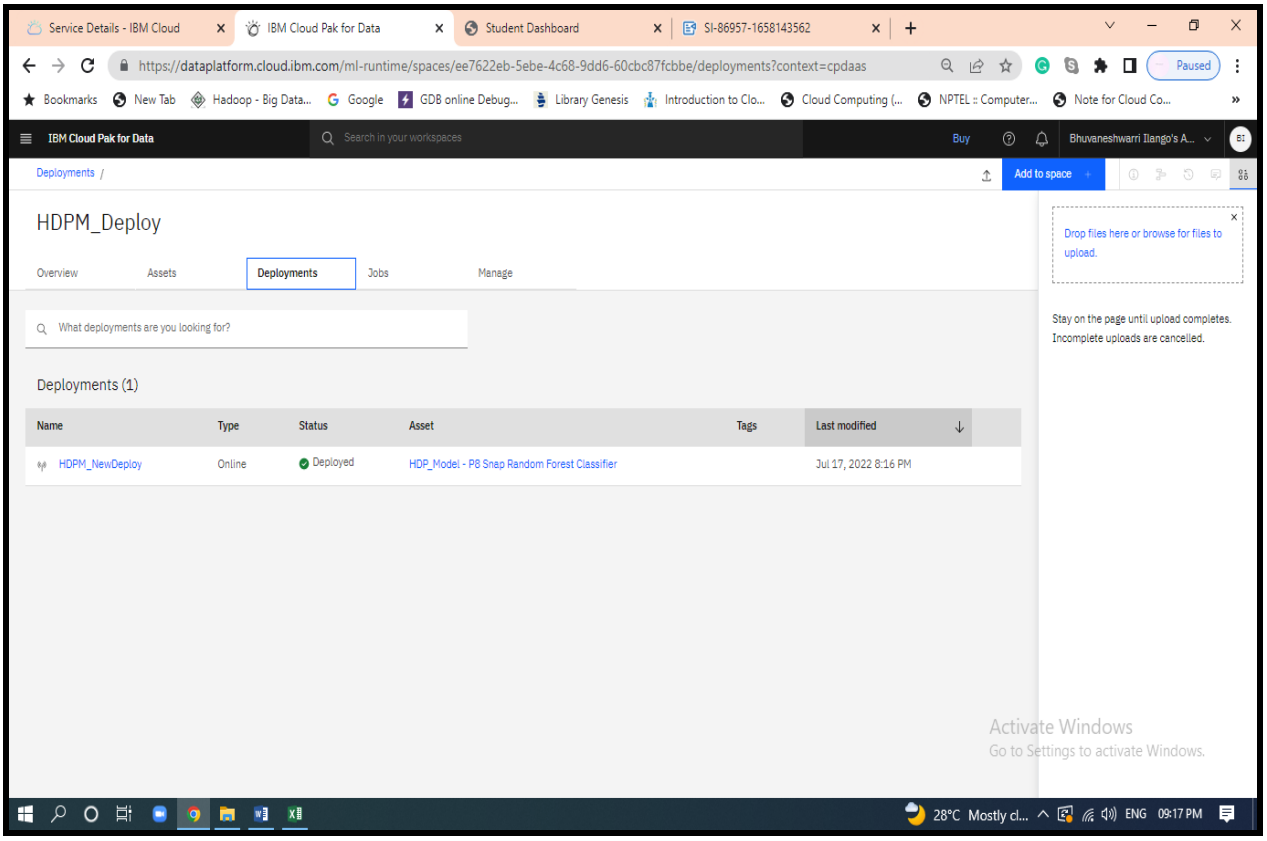
6 RESULT (Sample screen shots)

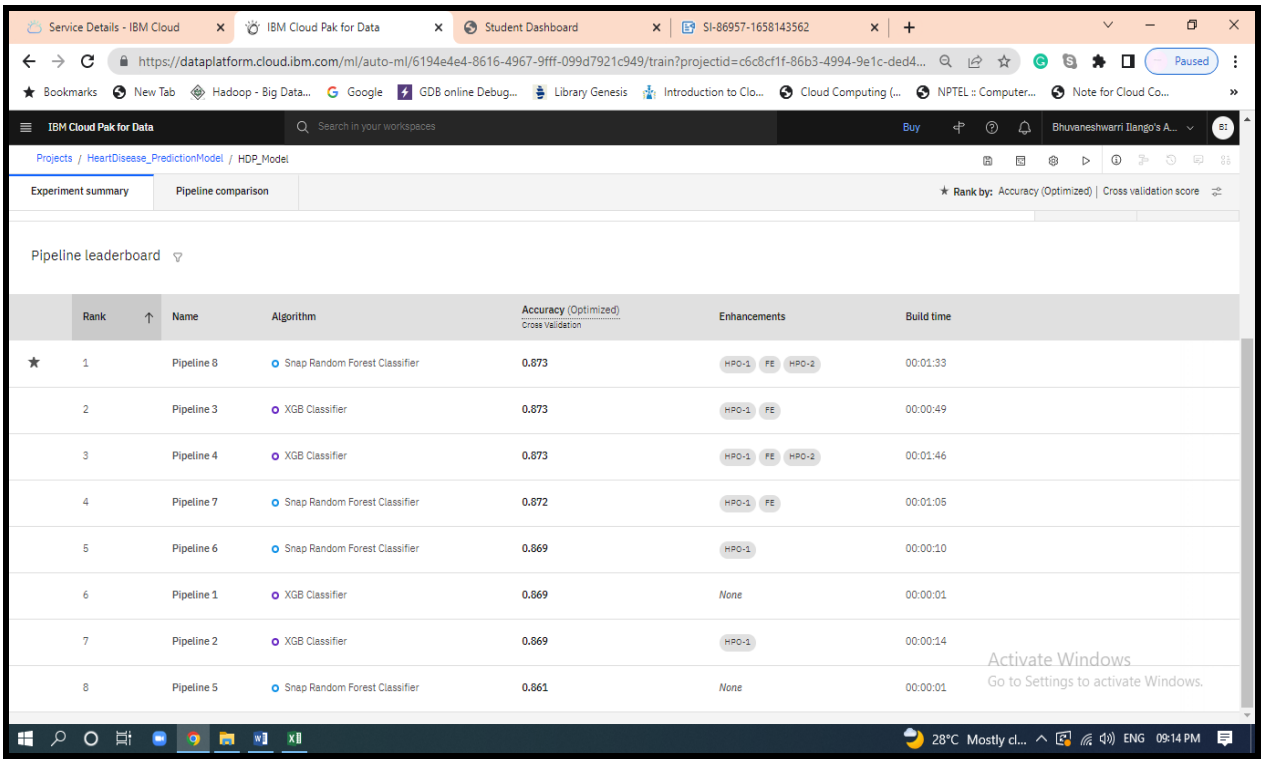


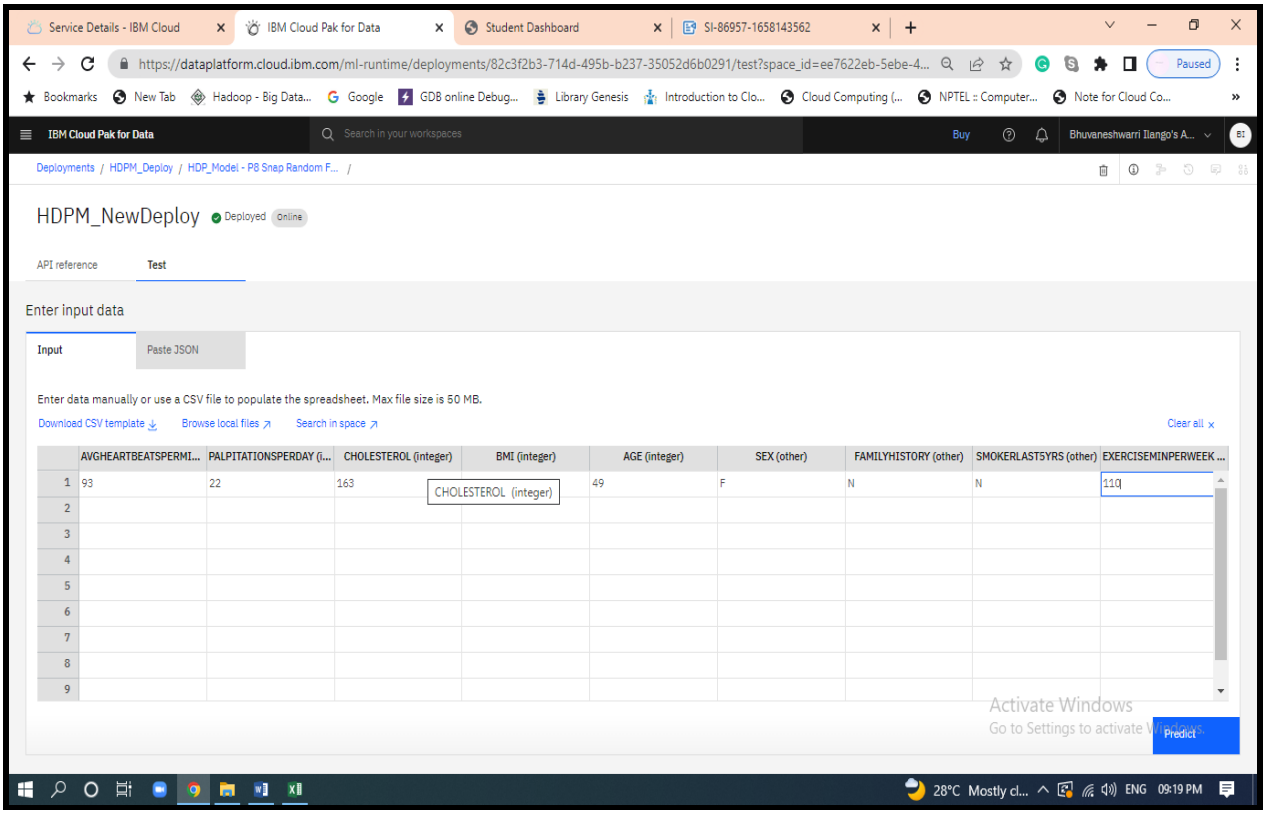


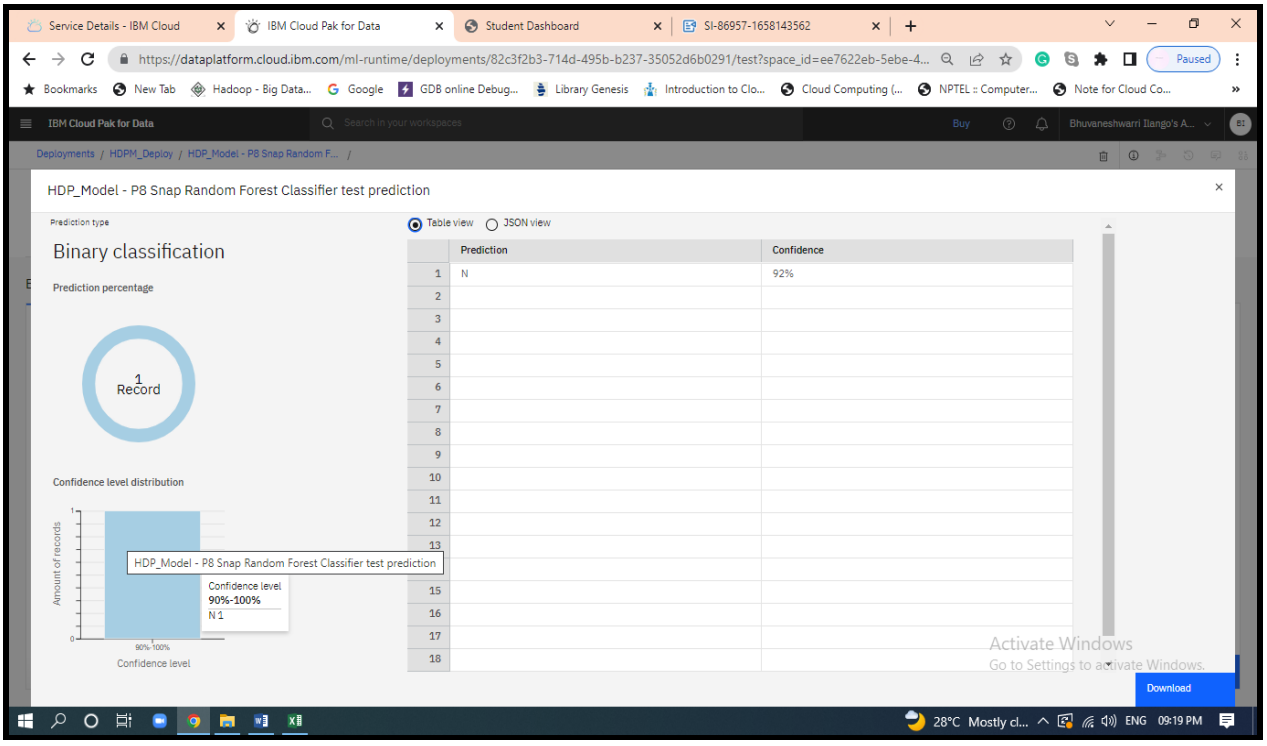

Application Details - IBM Cl
Node-RED on IBM Cloud
IBM Cloud Pak for Data
Student Dashboard
SI-86957-1658143562
https://cloud.ibm.com/apps/69a41eb7-15ac-4b10-8666-5c5894169a2f?ace_config=%7B"region"%3A"us-south"%2C"crn"%3A"crn%3Av1%3Ablu...
Paused
Bookmarks
New Tab
Hadoop - Big Data...
Google
GDB online Debug...
Library Genesis
Introduction to Clo...
Cloud Computing (...
NPTEL :: Computer...
Note for Cloud Co...
IBM Cloud
Search resources and products...
Catalog
Docs
Support
Manage
Bhuvaneshwarri Ilango...
Resource list /
Node RED ZSEEK 2022-07-17
Running
Visit App URL
Add tags
Details
Actions...
Getting started
Overview
Runtime
Connections
Logs
API Management
Autoscaling
IBM Cloud Foundry Public is being deprecated. Please see full details.
Instances
Edit
Health
100%
1/1 instance(s) are running
Instances
1
MB memory per instance
0
2048
256
Runtime
Node.js
256
Total MB allocation
1.75 GB still available
Free
Used
Activate Windows
Go to Settings to activate Windows.
Runtime cost
Connections (1)
29°C Mostly cl...
ENG
08:53 PM

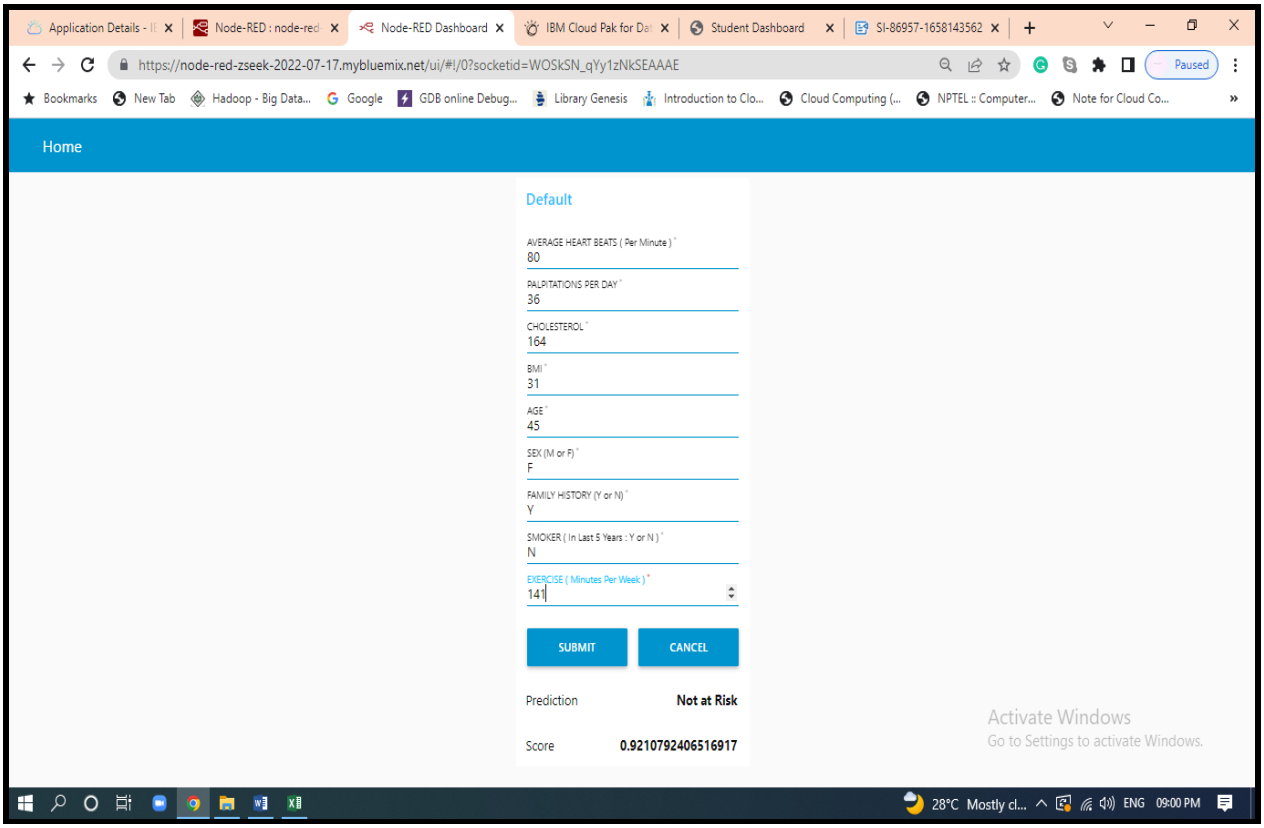

Service Details - IBM Cloud
IBM Cloud Pak for Data
Student Dashboard
SI-86957-1658143562
https://dataplatform.cloud.ibm.com/home2?context=cpdaas&apps=watson_machine_learning&nocache=true&quick_start_target=data_sc...
Paused
Bookmarks
New Tab
Hadoop - Big Data...
Google
GDB online Debug...
Library Genesis
Introduction to Clo...
Cloud Computing (...
NPTEL :: Computer...
Note for Cloud Co...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Bhuvaneshwarri Ilango's A...
Welcome, Bhuvaneshwarri!
Take a tutorial
Step through implementing a Data fabric use case in a sample project.
Work with data
Create a project for your team to prepare data, find insights, or build models.
Learn what's new
Stay current with new features, enhancements, and other changes.
Feedback
Quick start
Build dashboards
with IBM Cognos Dashboard Embedded
Create data pipelines
with DataStage
Build customer profiles
with IBM Match 360 with Watson
Catalog and govern data
with Watson Knowledge Catalog
Projects
HeartDisease_PredictionModel
Jul 17, 2022
07:40 PM
New in gallery
NOTEBOOK
The Nurse Assignment Problem
Notifications
Online deployment ready
The online deployment HDPM_NewDeploy in space HDPM_Deploy is ready to accept
Jul 17, 2022 08:17 PM
Deployments
HDPM_Deploy
Jul 17, 2022
08:09 PM
Activate Windows
Go to Settings to activate Windows.
28°C Mostly cl...
ENG
09:13 PM

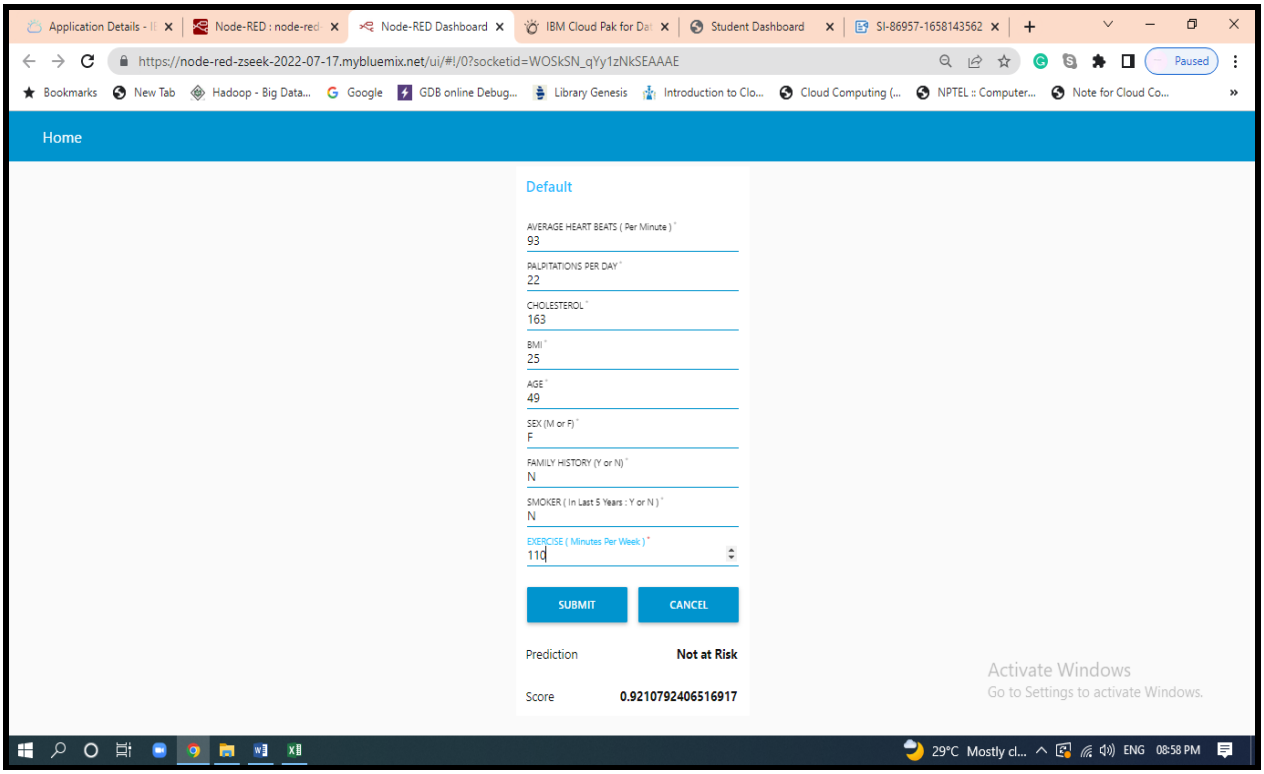

Service Details - IBM Cloud
IBM Cloud Pak for Data
Student Dashboard
SI-86957-1658143562
https://dataplatform.cloud.ibm.com/ml-runtime/spaces/ee7622eb-5ebe-4c68-9dd6-60cbc87fcbbe/deployments?context=cpdaas
IBM Cloud Pak for Data
Search in your workspaces
Buy
Deployments /
Add to space
HDPM_Deploy
Overview
Assets
Deployments
Jobs
Manage
What deployments are you looking for?
Deployments (1)
Name
Type
Status
Asset
Tags
Last modified
HDPM_NewDeploy
Online
Deployed
HDP_Model - P8 Snap Random Forest Classifier
Jul 17, 2022 8:16 PM
Drop files here or browse for files to upload.
Stay on the page until upload completes.
Incomplete uploads are cancelled.
Activate Windows
Go to Settings to activate Windows.
28°C Mostly cl...
ENG
09:17 PM

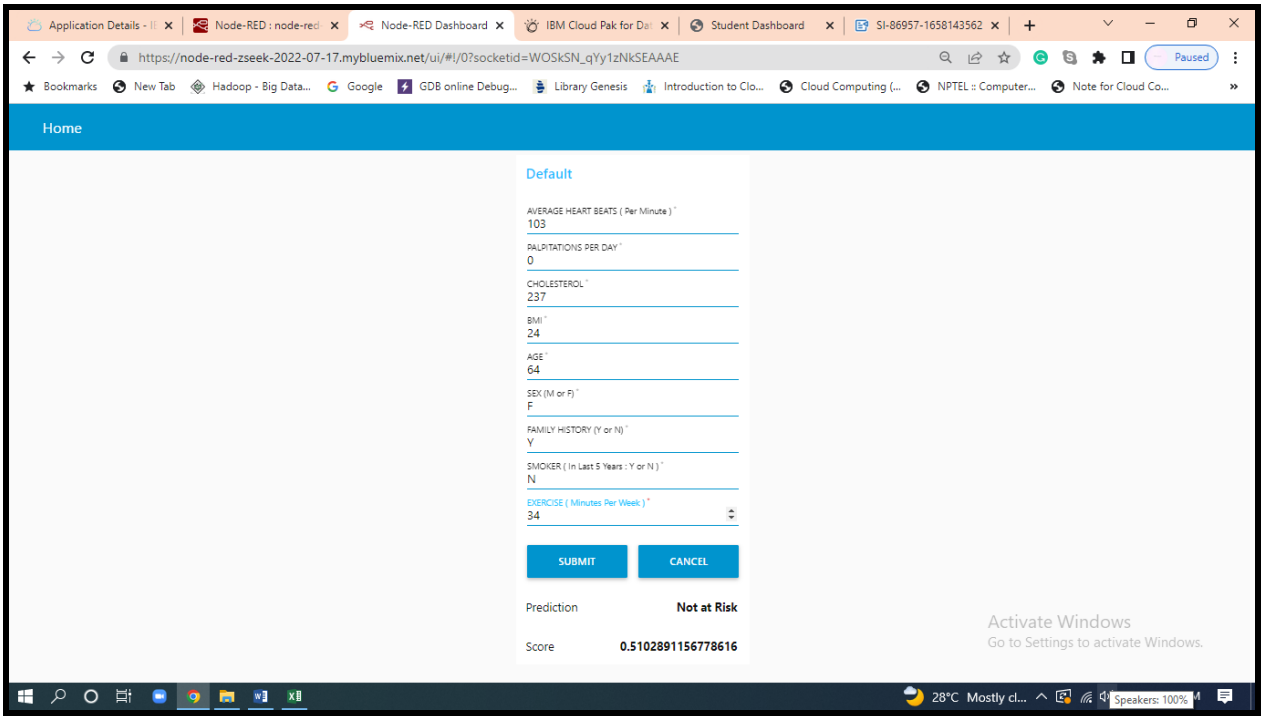

Service Details - IBM Cloud
IBM Cloud Pak for Data
Student Dashboard
SI-86957-1658143562
https://dataplatform.cloud.ibm.com/ml/auto-ml/6194e4e4-8616-4967-9fff-099d7921c949/train?projectid=c6c8cf1f-86b3-4994-9e1c-ded4...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Projects / HeartDisease_PredictionModel / HDP_Model
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 8 Snap Random Forest Classifier 0.873 HPO-1 FE HPO-2 00:01:33
2 Pipeline 3 XGB Classifier 0.873 HPO-1 FE 00:00:49
3 Pipeline 4 XGB Classifier 0.873 HPO-1 FE HPO-2 00:01:46
4 Pipeline 7 Snap Random Forest Classifier 0.872 HPO-1 FE 00:01:05
5 Pipeline 6 Snap Random Forest Classifier 0.869 HPO-1 00:00:10
6 Pipeline 1 XGB Classifier 0.869 None 00:00:01
7 Pipeline 2 XGB Classifier 0.869 HPO-1 00:00:14
8 Pipeline 5 Snap Random Forest Classifier 0.861 None 00:00:01
Activate Windows
Go to Settings to activate Windows.
28°C Mostly cl...
ENG
09:14 PM

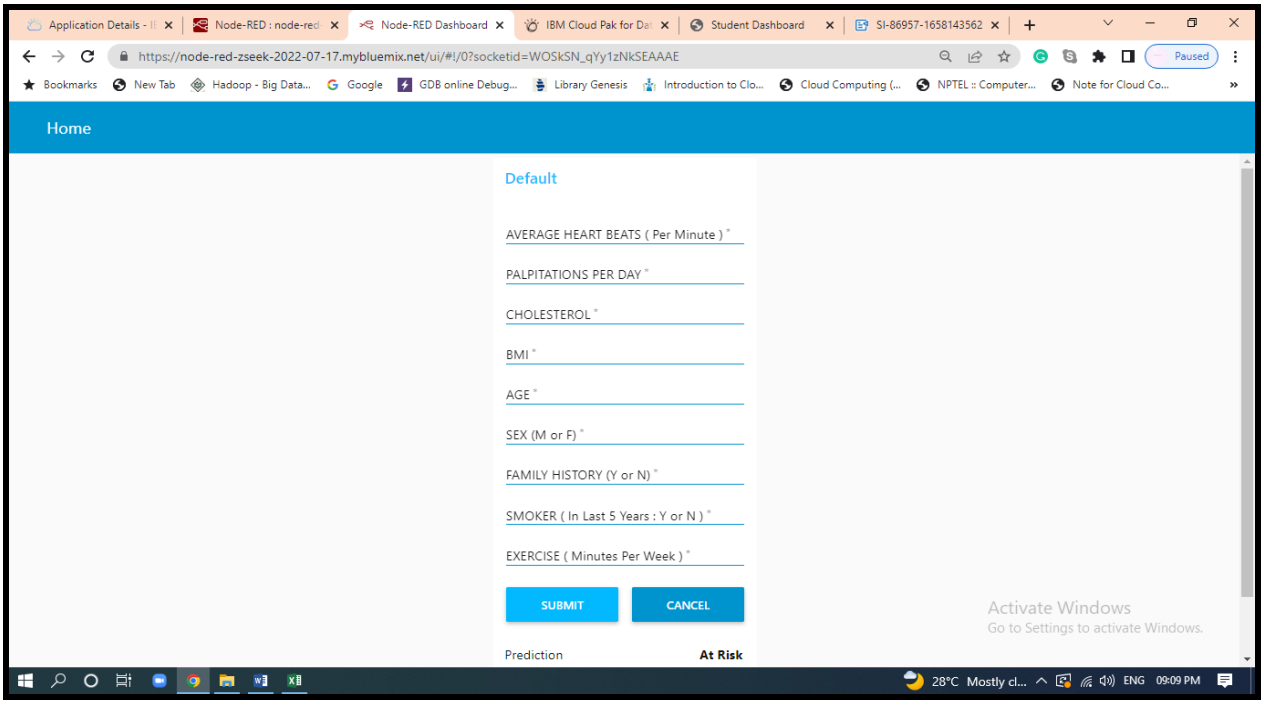

Service Details - IBM Cloud
IBM Cloud Pak for Data
Student Dashboard
SI-86957-1658143562
https://dataplatform.cloud.ibm.com/ml-runtime/deployments/82c3f2b3-714d-495b-b237-35052d6b0291/test?space_id=ee7622eb-5ebe-4...
IBM Cloud Pak for Data
Deployments / HDPM_Deploy / HDP_Model - P8 Snap Random F...
HDPM_NewDeploy
Deployed
Online
API reference
Test
Enter input data
Input
Paste JSON
Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.
Download CSV template
Browse local files
Search in space
Clear all
AVGHEARTBEATSPERMI...
PALPITATIONSPERDAY (i...
CHOLESTEROL (integer)
BMI (integer)
AGE (integer)
SEX (other)
FAMILYHISTORY (other)
SMOKERLAST5YRS (other)
EXERCISEMINPERWEEK ...
1 93 22 163 49 F N N 110
CHOLESTEROL (integer)
Activate Windows
Predict
28°C Mostly cl...
ENG
09:19 PM


Service Details - IBM Cloud
IBM Cloud Pak for Data
Student Dashboard
SI-86957-1658143562
https://dataplatform.cloud.ibm.com/ml-runtime/deployments/82c3f2b3-714d-495b-b237-35052d6b0291/test?space_id=ee7622eb-5ebe-4...
HDP_Model - P8 Snap Random Forest Classifier test prediction
Prediction type
Binary classification
Prediction percentage
1 Record
Table view
JSON view
Prediction
Confidence
1 N 92%
Confidence level distribution
Amount of records
HDP_Model - P8 Snap Random Forest Classifier test prediction
Confidence level
90%-100%
N 1
90%-100%
Confidence level
Activate Windows
Download
28°C Mostly cl...
ENG
09:19 PM


Application Details -
Node-RED : node-red
Node-RED Dashboard
IBM Cloud Pak for Da
Student Dashboard
SI-86957-1658143562
https://node-red-zseek-2022-07-17.mybluemix.net/ui/#!/0?socketid=WOSkSN_qYy1zNkSEAAAE
Paused
Bookmarks
New Tab
Hadoop - Big Data...
Google
GDB online Debug...
Library Genesis
Introduction to Clo...
Cloud Computing (...
NPTEL :: Computer...
Note for Cloud Co...
Home
Default
AVERAGE HEART BEATS (Per Minute)
80
PALPITATIONS PER DAY
36
CHOLESTEROL
164
BMI
31
AGE
45
SEX (M or F)
F
FAMILY HISTORY (Y or N)
Y
SMOKER (In Last 5 Years : Y or N)
N
EXERCISE (Minutes Per Week)
141
SUBMIT
CANCEL
Prediction
Not at Risk
Score
0.9210792406516917
Activate Windows
Go to Settings to activate Windows.
28°C Mostly cl...
ENG
09:00 PM


Application Details -
Node-RED : node-red
Node-RED Dashboard
IBM Cloud Pak for Da
Student Dashboard
SI-86957-1658143562
https://node-red-zseek-2022-07-17.mybluemix.net/ui/#!/0?socketid=WOSkSN_qYy1zNkSEAAAE
Paused
Bookmarks
New Tab
Hadoop - Big Data...
Google
GDB online Debug...
Library Genesis
Introduction to Clo...
Cloud Computing (...
NPTEL :: Computer...
Note for Cloud Co...
Home
Default
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
Activate Windows
Go to Settings to activate Windows.
29°C Mostly cl...
ENG
08:58 PM


Application Details -
Node-RED : node-red
Node-RED Dashboard
IBM Cloud Pak for Da
Student Dashboard
SI-86957-1658143562
https://node-red-zseek-2022-07-17.mybluemix.net/ui/#!/0?socketid=WOSkSN_qYy1zNkSEAAAE
Paused
Bookmarks
New Tab
Hadoop - Big Data...
Google
GDB online Debug...
Library Genesis
Introduction to Clo...
Cloud Computing (...
NPTEL :: Computer...
Note for Cloud Co...
Home
Default
AVERAGE HEART BEATS (Per Minute) *
103
PALPITATIONS PER DAY *
0
CHOLESTEROL *
237
BMI *
24
AGE *
64
SEX (M or F) *
F
FAMILY HISTORY (Y or N) *
Y
SMOKER (In Last 5 Years : Y or N) *
N
EXERCISE (Minutes Per Week) *
34
SUBMIT
CANCEL
Prediction
Not at Risk
Score
0.5102891156778616
Activate Windows
Go to Settings to activate Windows.
28°C Mostly cl...
Speakers: 100%


Application Details -
Node-RED : node-red
Node-RED Dashboard
IBM Cloud Pak for Da
Student Dashboard
SI-86957-1658143562
https://node-red-zseek-2022-07-17.mybluemix.net/ui/#!/0?socketid=WOSkSN_qYy1zNkSEAAAE
Paused
Bookmarks
New Tab
Hadoop - Big Data...
Google
GDB online Debug...
Library Genesis
Introduction to Clo...
Cloud Computing (...
NPTEL :: Computer...
Note for Cloud Co...
Home
Default
AVERAGE HEART BEATS (Per Minute) *
PALPITATIONS PER DAY *
CHOLESTEROL *
BMI *
AGE *
SEX (M or F) *
FAMILY HISTORY (Y or N) *
SMOKER (In Last 5 Years : Y or N) *
EXERCISE (Minutes Per Week) *
SUBMIT
CANCEL
Prediction
At Risk
Activate Windows
Go to Settings to activate Windows.
28°C Mostly cl...
ENG
09:09 PM

7 ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- No need to worry about infrastructure,platform and software.
- Initial setup and maintenance cost is reduced
- All under single umbrella, i.e., everything comes under single framework
- Improves health care services
- Reduces medical supervision procedures
- Ease of maintaining EHR (Electronic Health Care) records
- Proactively preparing for upcoming population health trends
- Acquiring new patients through personalized campaigns
- Reducing costs on appointment no show and readmission penalties
- Speeding up administrative tasks such as discharge procedures and insurance claims submission
- Preventing ransom and other cyberattacks by analysing ongoing transactions and assigning risk scores

DISADVANTAGES

- Internet Connectivity is mandatory
- More steps to remember during creation of different services
- User interface creation is little bit tougher.

8 APPLICATIONS

- Chronic disease prediction model
- Autism disorder prediction model
- Prediction modelling in healthcare system
- Preventing readmissions
- Managing population health
- Enhancing cybersecurity
- Increasing patient engagement and outreach
- Speeding up insurance claims submission
- Predicting suicide attempts
- Forecasting appointment no-shows
- Preventing readmissions
- Managing population health

- Enhancing cybersecurity
- Increasing patient engagement and outreach
- Speeding up insurance claims submission
- Predicting suicide attempts
- Forecasting appointment no-shows

9 CONCLUSION

The proposed effective heart disease prediction model is used to predict the HEART FAILURE target attribute of the patient using the following conditional attributes such as AVGHEARTBEATSPERMIN,FAMILYHISTORY,PALPITATIONSPERDAY,CHOLESTEROL, BMI, AGE,SEX,SMOKERLAST5YRS,EXERCISEMINPERWEEK of the Patient.The proposed model is built with IBM Watson Studio,Node-RED service,Auto AI service, Cloud Object Storage service(COS) and Machine Learning Service. The prediction model is developed with Snap Random Forest Classifier which has accuracy of 87.3%.

10 FUTURE SCOPE

The model creation,validation and deployment have taken lots of procedures and steps. The aim of the future work is to predict the target attribute by reducing the number of procedures and steps. The accuracy of the model is also somewhat less compared to already existing prediction models discussed under literature review. In order to improve the accuracy, pipeline structure and algorithm selection procedure will need to be optimized.

11 BIBILIOGRAPHY

1. Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology. 2018 Oct;7(2.8):684-7.
2. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE access. 2019 Jun 19;7:81542-54.
3. Rajdhan A, Agarwal A, Sai M, Ravi D, Ghuli P. Heart disease prediction using machine learning.International Journal of Research and Technology. 2020 Apr;9(04):659-62.
4. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Computer Science. 2020 Nov;1(6):1-6.
5. Patel J, TejalUpadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Disease. 2015 Sep;7(1):129-37.

6. Singh A, Kumar R. Heart disease prediction using machine learning algorithms. In2020 international conference on electrical and electronics engineering (ICE3) 2020 Feb 14 (pp. 452-457). IEEE.
7. Jagtap A, Malewadkar P, Baswat O, Rambade H. Heart disease prediction using machine learning. International Journal of Research in Engineering, Science and Management. 2019 Feb;2(2):352-5.
8. Khourdifi Y, Bahaj M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International Journal of Intelligent Engineering and Systems. 2019 Feb;12(1):242-52.
9. Gavhane A, Kokkula G, Pandya I, Devadkar K. Prediction of heart disease using machine learning. In2018 second international conference on electronics, communication and aerospace technology (ICECA) 2018 Mar 29 (pp. 1275-1278). IEEE.
10. Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using machine learning algorithms. In IOP conference series: materials science and engineering 2021 (Vol. 1022, No. 1, p. 012072). IOP Publishing.

APPENDIX