Effective Heart Disease Prediction Using IBM Auto AI Service

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.Among all fatal disease, heart attacks diseases are considered as the most prevalent. Medical practitioners conduct different surveys on heart diseases and gather information of heart patients, their symptoms and disease progression. Increasingly are reported about patients with common diseases who have typical symptoms. In this fast moving world people want to live a very luxurious life so they work like a machine in order to earn lot of money and live a comfortable life therefore in this race they forget to take care of themselves, because of this there food habits change their entire lifestyle change, in this type of lifestyle they are more tensed they have blood pressure, sugar at a very young age and they don't give enough rest for themselves and eat what they get and they even don't bother about the quality of the food if sick the go for their own medication as a result of all these small negligence it leads to a major threat that is the heart disease

The term 'heart disease' includes the diverse diseases that affect heart. The number of people suffering from heart disease is on the rise (health topics, 2010). The report from world health organization shows us a large number of people that die every year due to the heart disease all over the world. Heart disease is also stated as one of the greatest killers in Africa.

Data mining has been used in a variety of applications such as marketing, customer relationship management, engineering, and medicine analysis, expert prediction, web mining and mobile computing. Of late, data mining has been applied successfully in healthcare fraud and detecting abuse cases.

1.2 Purpose

Clinical decisions are often made based on doctor's insight and experience rather than on the knowledge rich data hidden in the dataset. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The proposed system will integrate clinical decision support with computer-based patient records (Data Sets). This will reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge rich environment which can help to significantly improve the quality of clinical decisions.

There are voluminous records in medical data domain and because of this, it has become necessary to use data mining techniques to help in decision support and prediction in the field of healthcare. Therefore, medical data mining contributes to business intelligence which is useful for diagnosing of diseaseHeart failure is a common event caused by CVDs and this dataset contains 9 features that can be used to predict mortality by heart failure.

2.Literature Survey

2.1 Existing Methods

Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods.

(Polaraju, Durga Prasad, & Tech Scholar, 2017) proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned

earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing.

(Deepika & Seema, 2017) focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine (SVM) and Artificial Neural Network (ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

(Beyene & Kamat, 2018) recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms. (Beyene & Kamat, 2018) suggested Heart Disease Prediction System using Data Mining Techniques. WEKA software used for automatic diagnosis of disease and to give qualities of services in healthcare centers. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms.

Chala Beyene recommended Prediction and Analysis the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in short time. The proposed methodology is also critical in healthcare organization with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex are some of the attributes are included to identify if the person has heart disease or not. Analyses of data set are computed using WEKA software.

(Soni, Ansari, & Sharma, 2011) proposed to use non- linear classification algorithm for heart disease prediction. It is proposed to use bigdata tools such as Hadoop Distributed File System (HDFS), Map reduce along with SVM for prediction of heart disease with optimized attribute set. This work made an investigation on the use of different data

mining techniques for predicting heart diseases. It suggests to use HDFS for storing large data in different nodes and executing the prediction algorithm using SVM in more than one node simultaneously using SVM. SVM is used in parallel fashion which yielded better computation time than sequential SVM.

(Science & Faculty, 2009) suggested heart disease prediction using data mining and machine learning algorithm. The goal of this study is to extract hidden patterns by applying data mining techniques. The best algorithm J48 based on UCI data has the highest accuracy rate compared to LMT. (Purushottam, Saxena, & Sharma, 2016) proposed an efficient heart disease prediction system using data mining. This system helps medical practitioner to make effective decision making based on the certain parameter. By testing and training phase a certain parameter, it provides 86.3% accuracy in testing phase and 87.3% in training phase.

(Kirmani, 2017) suggested multi disease prediction using data mining techniques. Nowadays, data mining plays vital role in predicting multiple disease. By using data mining techniques, the number of tests can be reduced. This paper mainly concentrates on predicting the heart disease, diabetes and breast cancer etc.,

(Sai & Reddy, 2017) proposed Heart disease prediction using ANN algorithm in data mining. Due to increasing expenses of heart disease diagnosis disease, there was a need to develop new system which can predict heart disease. Prediction model is used to predict the condition of the patient after evaluation on the basis of various parameters like heart beat rate, blood pressure, cholesterol etc. The

accuracy of the system is proved in java.

(A & Naik, 2016) recommended to develop the prediction system which will diagnosis the heart disease from patient's medical data set. 13 risk factors of input attributes have considered to build the system. After analysis of the data from the dataset, data cleaning and data integration was performed. He used k-means and naïve Bayes to predict heart disease. This paper is to build the system using historical heart database

that gives diagnosis. 13 attributes have considered for building the system. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used. 13 attributes with total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

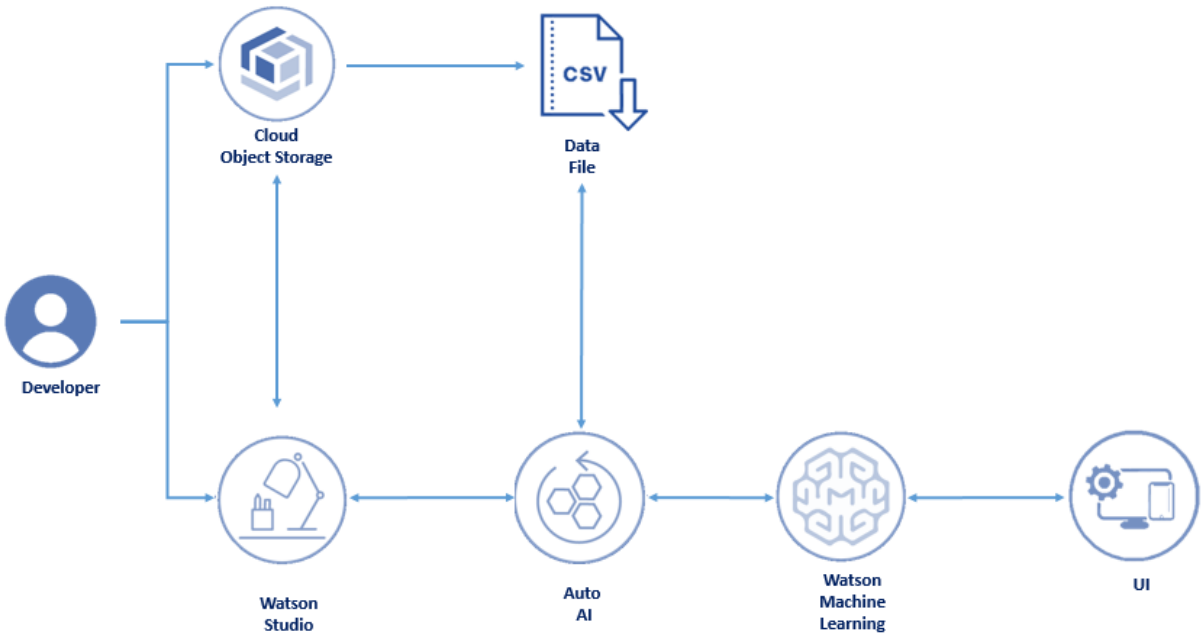
(Sultana, Haider, & Uddin, 2017) proposed an analysis of cardiovascular disease. This paper proposed data mining techniques to predict the disease. It is intended to provide the survey of current techniques to extract information from dataset and it will useful for healthcare practitioners. The performance can be obtained based on the time taken to build the decision tree for the system. The primary objective is to predict the disease with a smaller number of attributes.

2.2 Proposed solution

In this project, a model is build using Auto AI together with a web application where the prediction of heart failure can be view.

3.Theoretical Analysis

3.1Block Diagram



3.2 Software designing

Services Used:

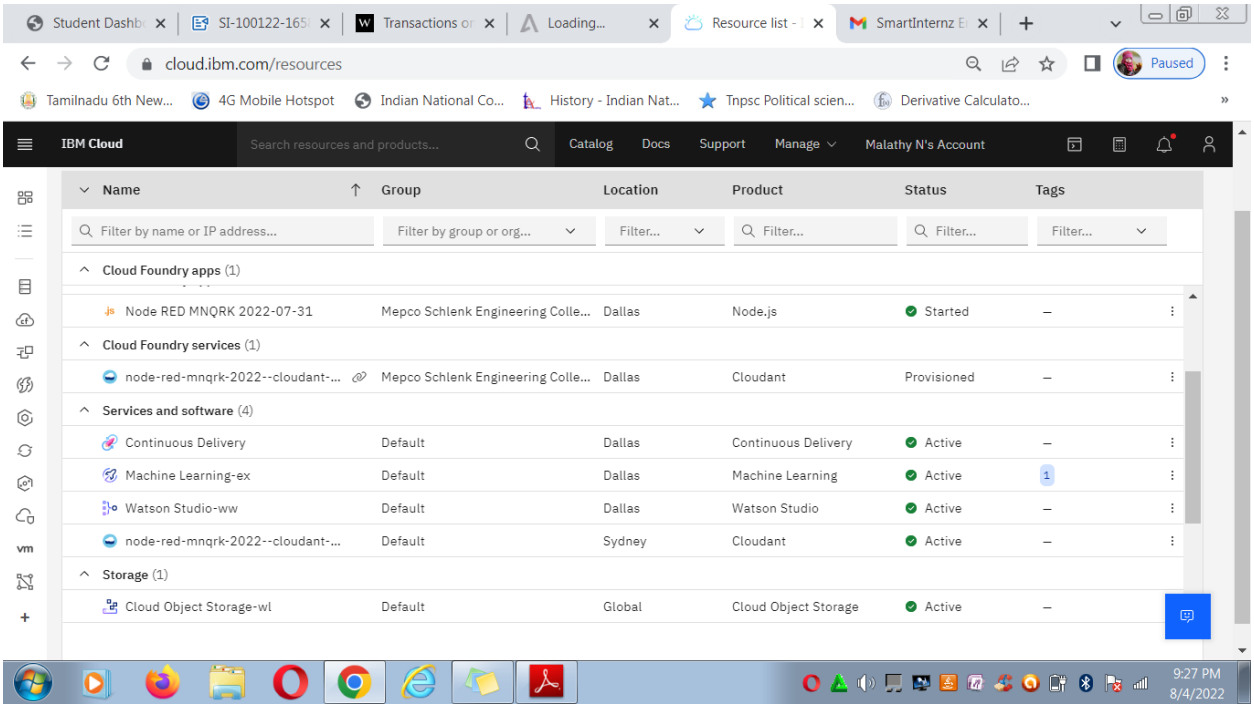
- IBM Watson Studio
- IBM Watson Machine Learning
- Node-RED
- IBM Cloud Object Storage

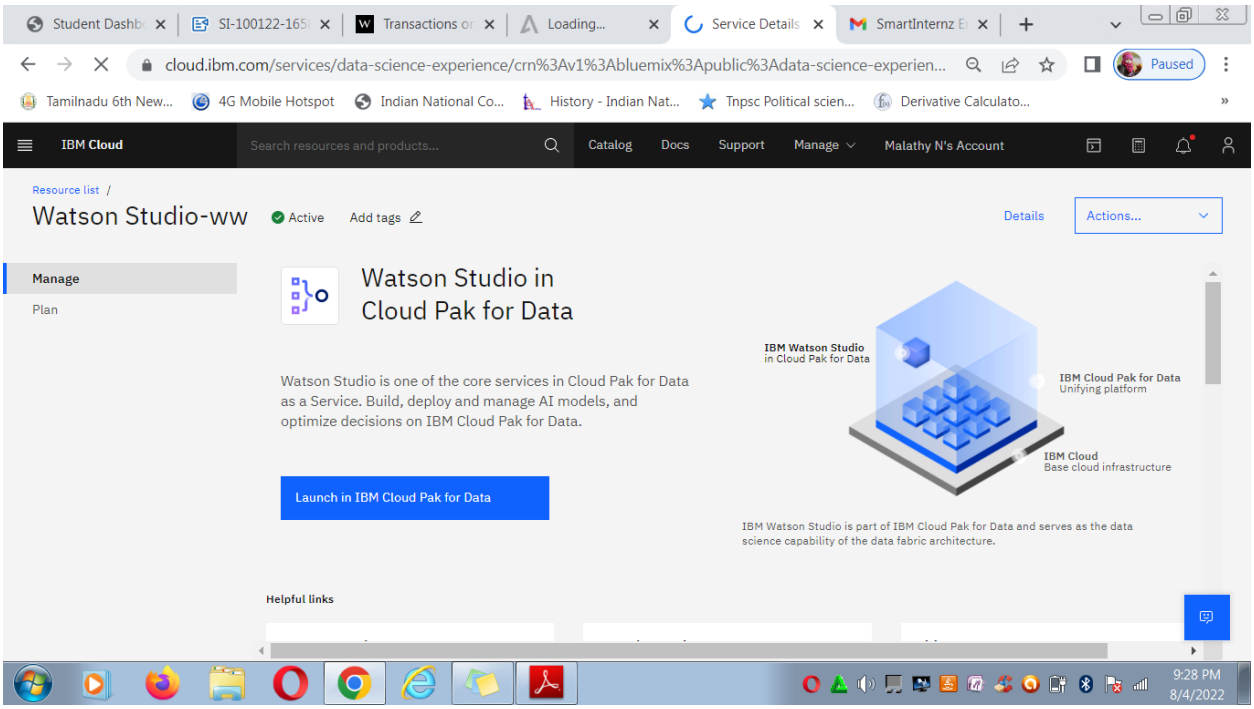
4.Experimental Investigations

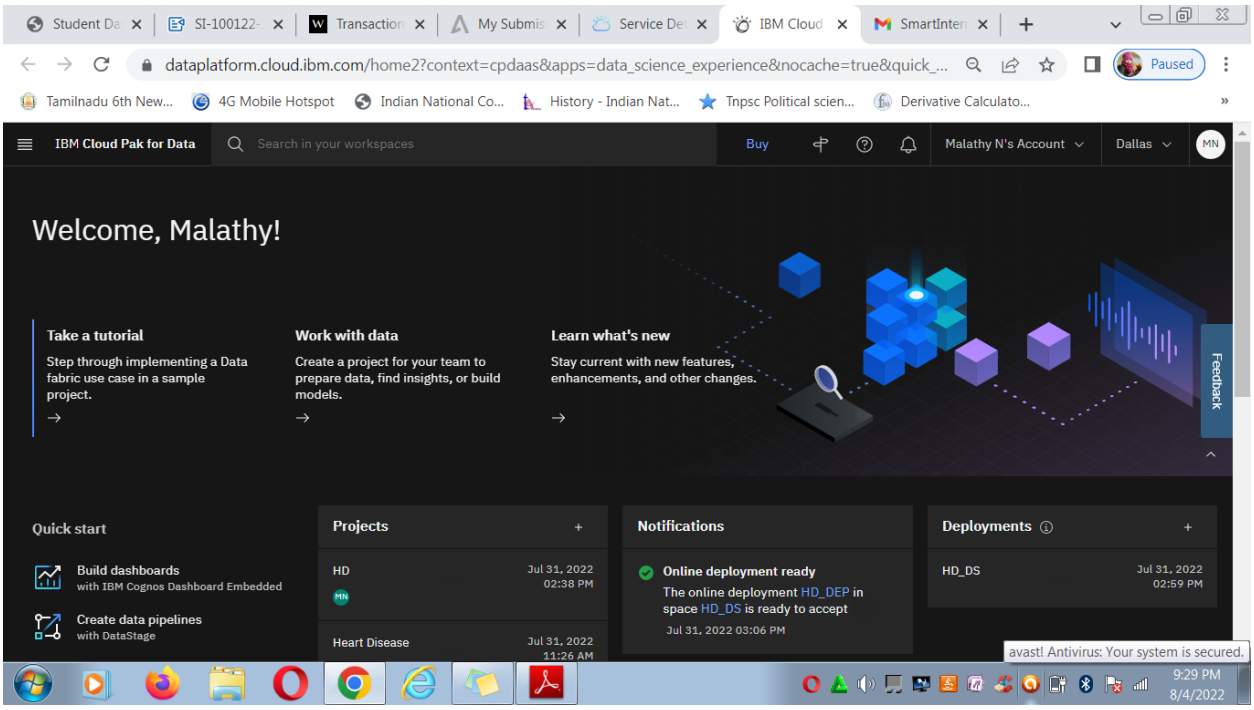
In this Activity, We gonna build a machine learning model that predicts heart failure based on the following parameters

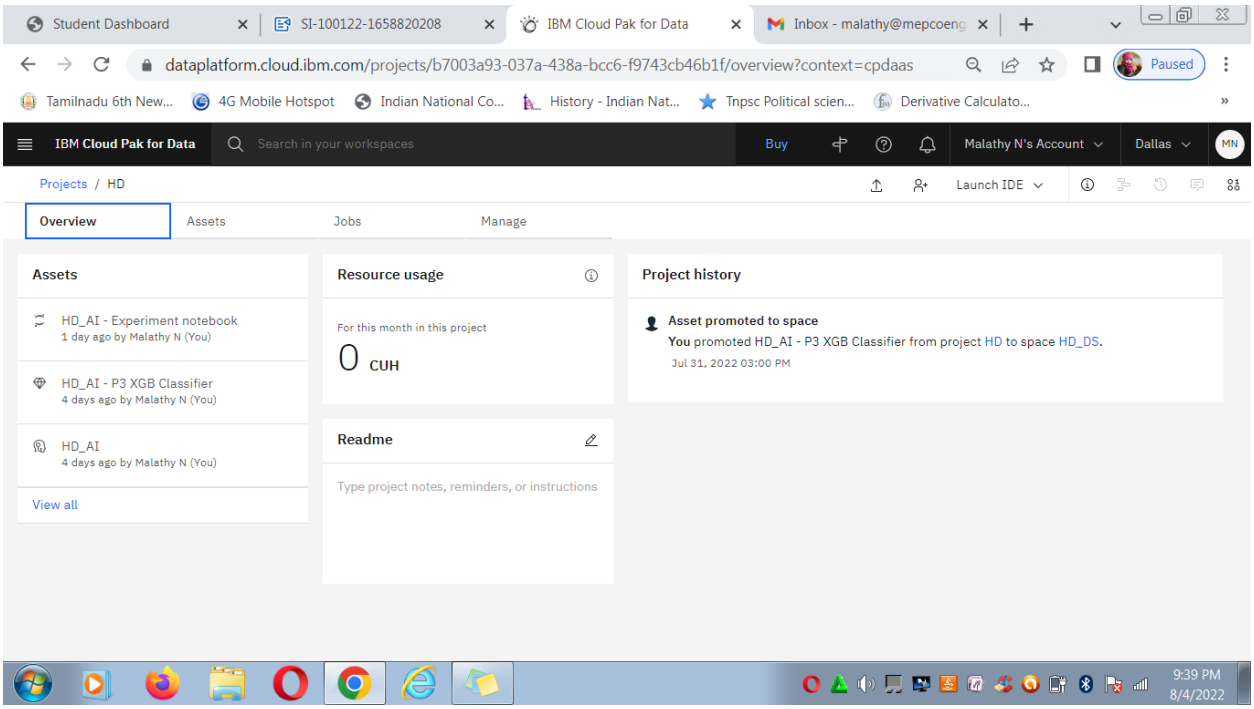
- AVGHEARTBEATSPERMIN
- PALPITATIONSPERDAY
- CHOLESTEROL
- BMI
- AGE
- SEX
- FAMILY HISTORY
- SMOKERLAST5YRS
- EXERCISEMINPERWEEK

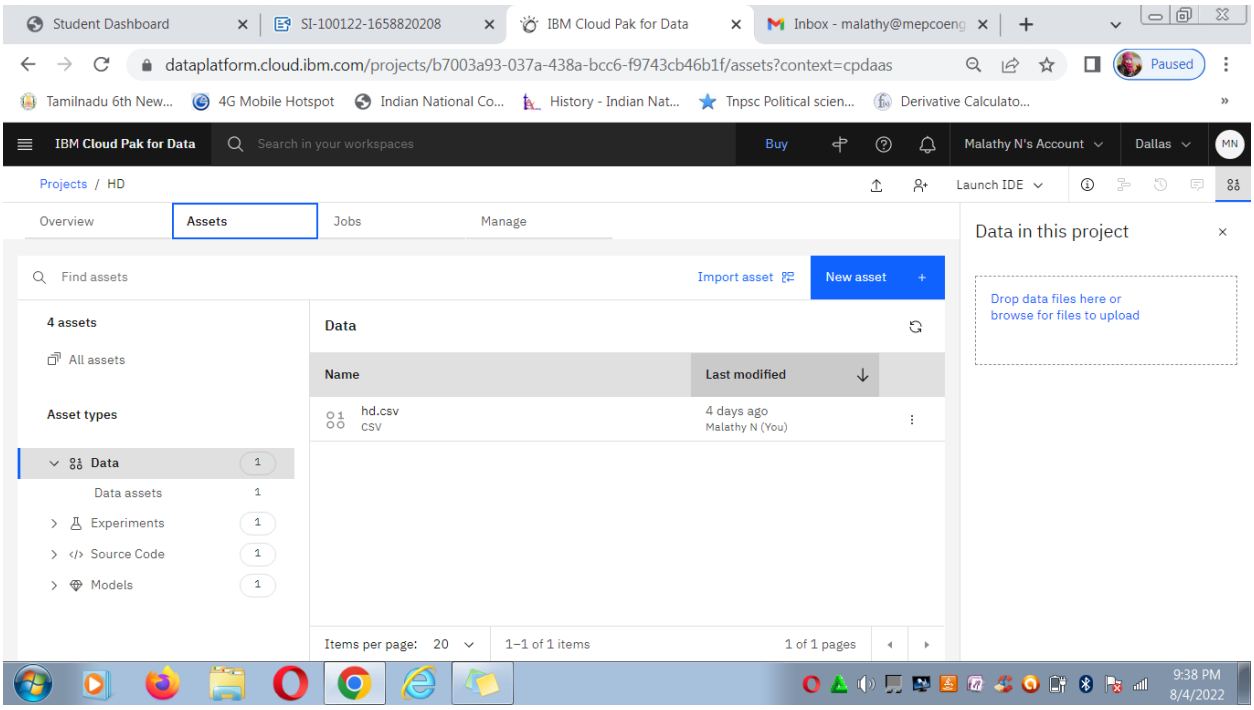
Created services are as follows

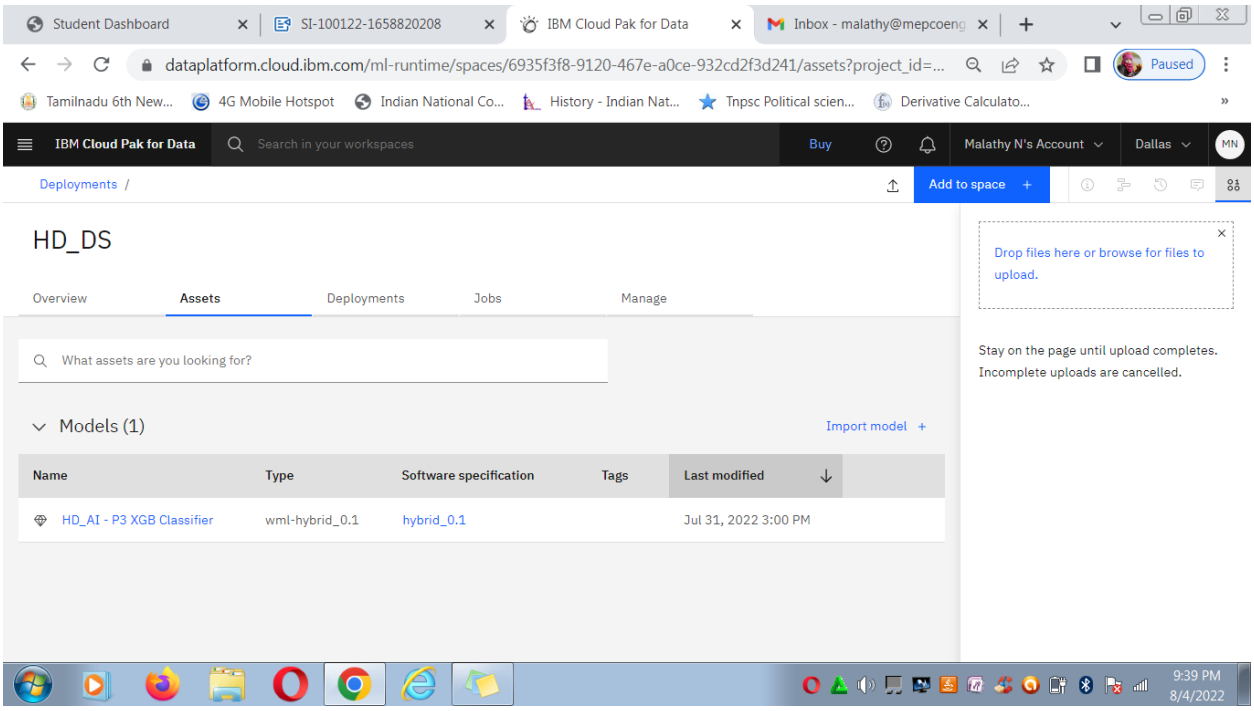


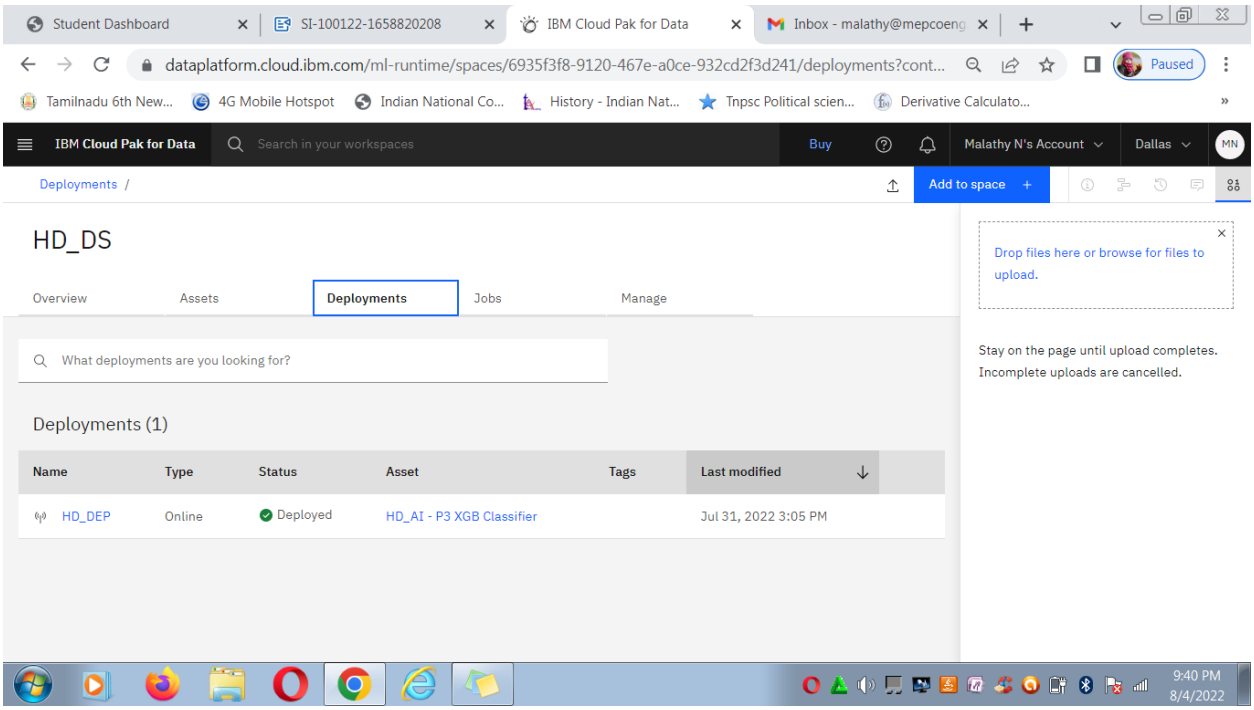

Student Dashb
SI-100122-165
Transactions or
Loading...
Service Details
SmartInternz E
cloud.ibm.com/services/data-science-experience/crn%3Av1%3Abluemix%3Apublic%3Adata-science-experien...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud
Search resources and products...
Catalog
Docs
Support
Manage
Malathy N's Account
Resource list /
Watson Studio-ww
Active
Add tags
Details
Actions...
Manage
Plan
Watson Studio in Cloud Pak for Data
Watson Studio is one of the core services in Cloud Pak for Data as a Service. Build, deploy and manage AI models, and optimize decisions on IBM Cloud Pak for Data.
Launch in IBM Cloud Pak for Data
IBM Watson Studio in Cloud Pak for Data
IBM Cloud Pak for Data Unifying platform
IBM Cloud Base cloud infrastructure
IBM Watson Studio is part of IBM Cloud Pak for Data and serves as the data science capability of the data fabric architecture.
Helpful links
9:28 PM
8/4/2022

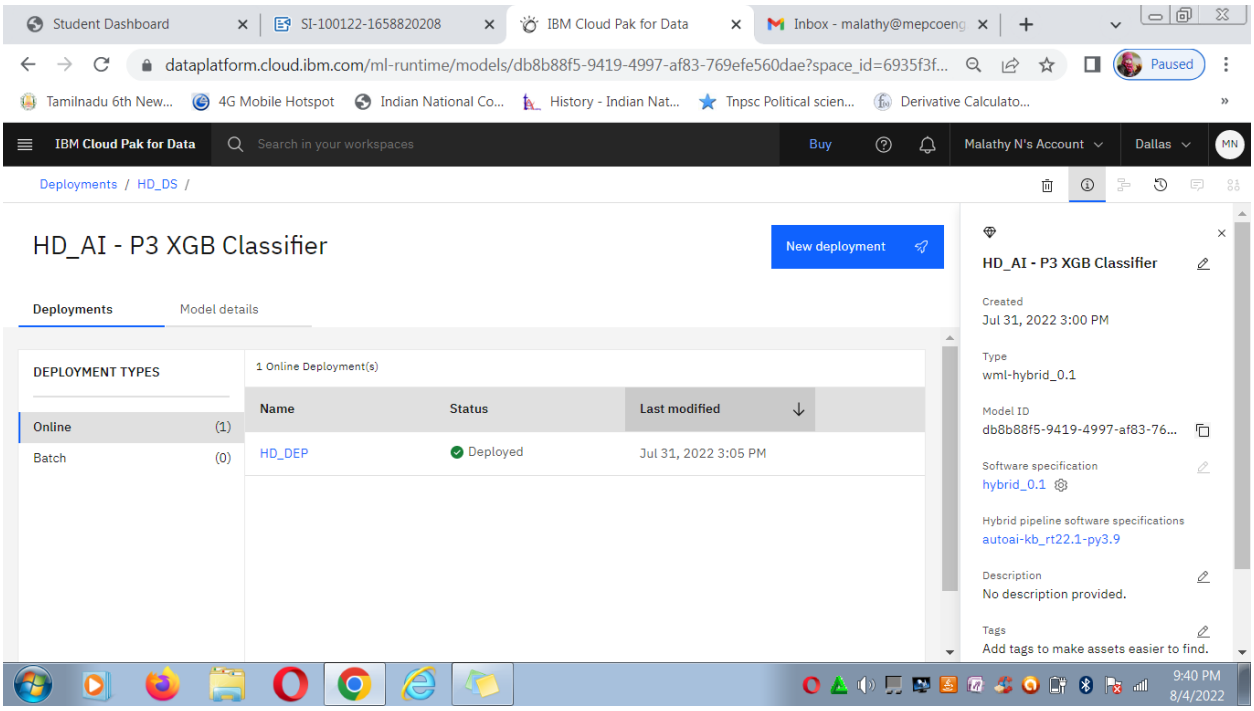

Student Da
SI-100122-
Transaction
My Submis
Service Del
IBM Cloud
SmartIntern
dataplatform.cloud.ibm.com/home2?context=cpdaas&apps=data_science_experience&nocache=true&quick_...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Welcome, Malathy!
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
Projects
HD
Jul 31, 2022 02:38 PM
Heart Disease
Jul 31, 2022 11:26 AM
Notifications
Online deployment ready
The online deployment HD_DEP in space HD_DS is ready to accept
Jul 31, 2022 03:06 PM
Deployments
HD_DS
Jul 31, 2022 02:59 PM
avast! Antivirus: Your system is secured.
9:29 PM
8/4/2022

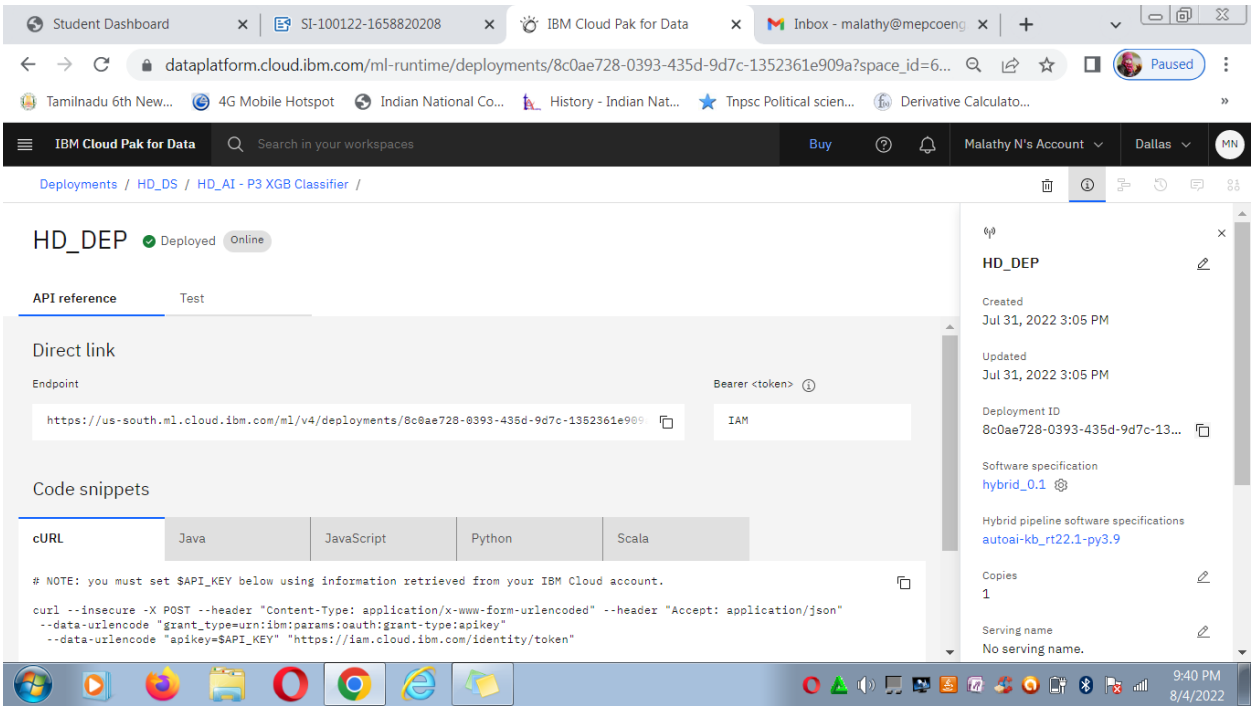

Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/projects/b7003a93-037a-438a-bcc6-f9743cb46b1f/overview?context=cpdaas
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Projects / HD
Launch IDE
Overview
Assets
Jobs
Manage
Assets
HD_AI - Experiment notebook
1 day ago by Malathy N (You)
HD_AI - P3 XGB Classifier
4 days ago by Malathy N (You)
HD_AI
4 days ago by Malathy N (You)
View all
Resource usage
For this month in this project
0 CUH
Readme
Type project notes, reminders, or instructions
Project history
Asset promoted to space
You promoted HD_AI - P3 XGB Classifier from project HD to space HD_DS.
Jul 31, 2022 03:00 PM
9:39 PM
8/4/2022

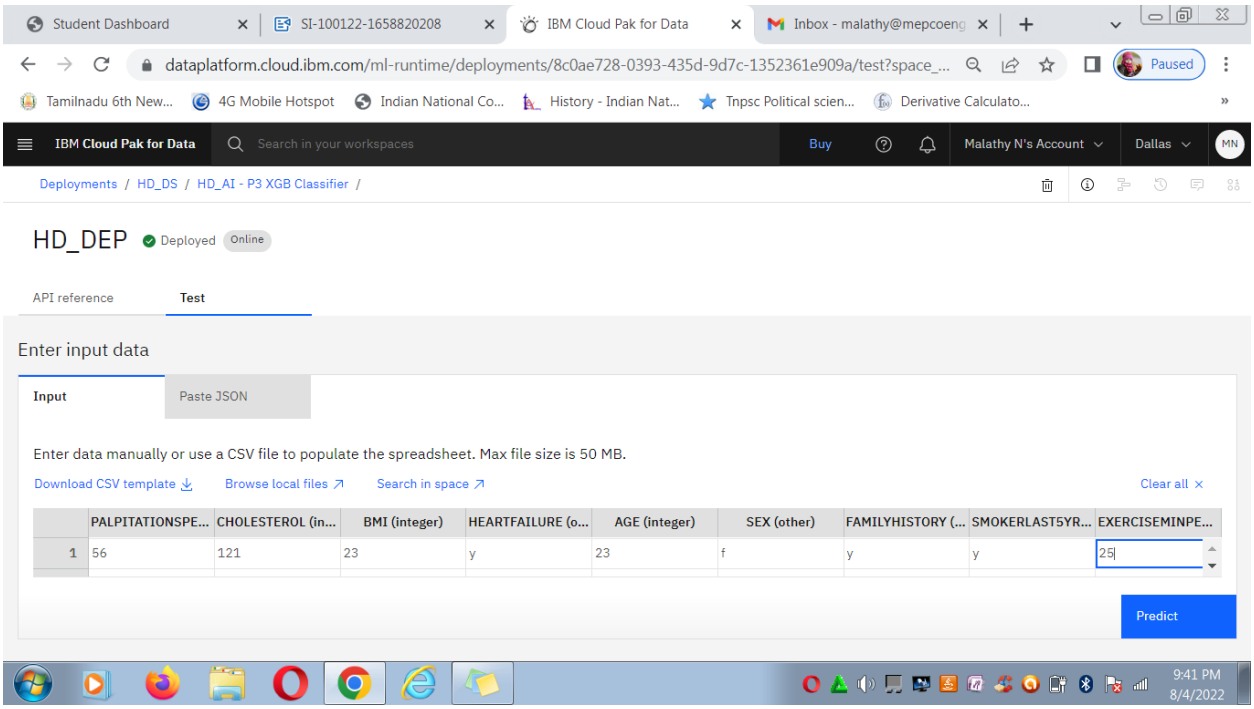

Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/projects/b7003a93-037a-438a-bcc6-f9743cb46b1f/assets?context=cpdaas
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Projects / HD
Launch IDE
Overview
Assets
Jobs
Manage
Find assets
Import asset
New asset
4 assets
All assets
Asset types
Data 1
Data assets 1
Experiments 1
Source Code 1
Models 1
Data
Name
Last modified
hd.csv
CSV
4 days ago
Malathy N (You)
Items per page: 20
1–1 of 1 items
1 of 1 pages
Data in this project
Drop data files here or
browse for files to upload
9:38 PM
8/4/2022

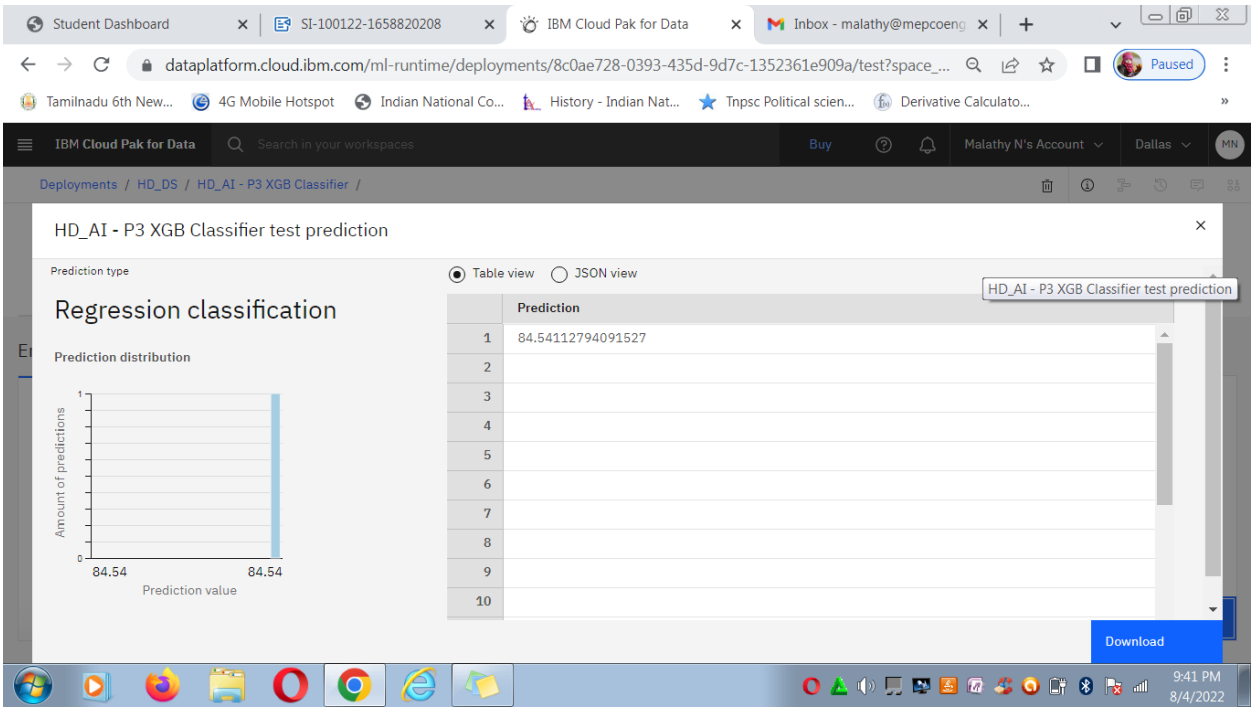

Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/ml-runtime/spaces/6935f3f8-9120-467e-a0ce-932cd2f3d241/assets?project_id=...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Deployments /
Add to space
HD_DS
Overview
Assets
Deployments
Jobs
Manage
What assets are you looking for?
Models (1)
Import model
Name
Type
Software specification
Tags
Last modified
HD_AI - P3 XGB Classifier
wml-hybrid_0.1
hybrid_0.1
Jul 31, 2022 3:00 PM
Drop files here or browse for files to upload.
Stay on the page until upload completes.
Incomplete uploads are cancelled.
9:39 PM
8/4/2022

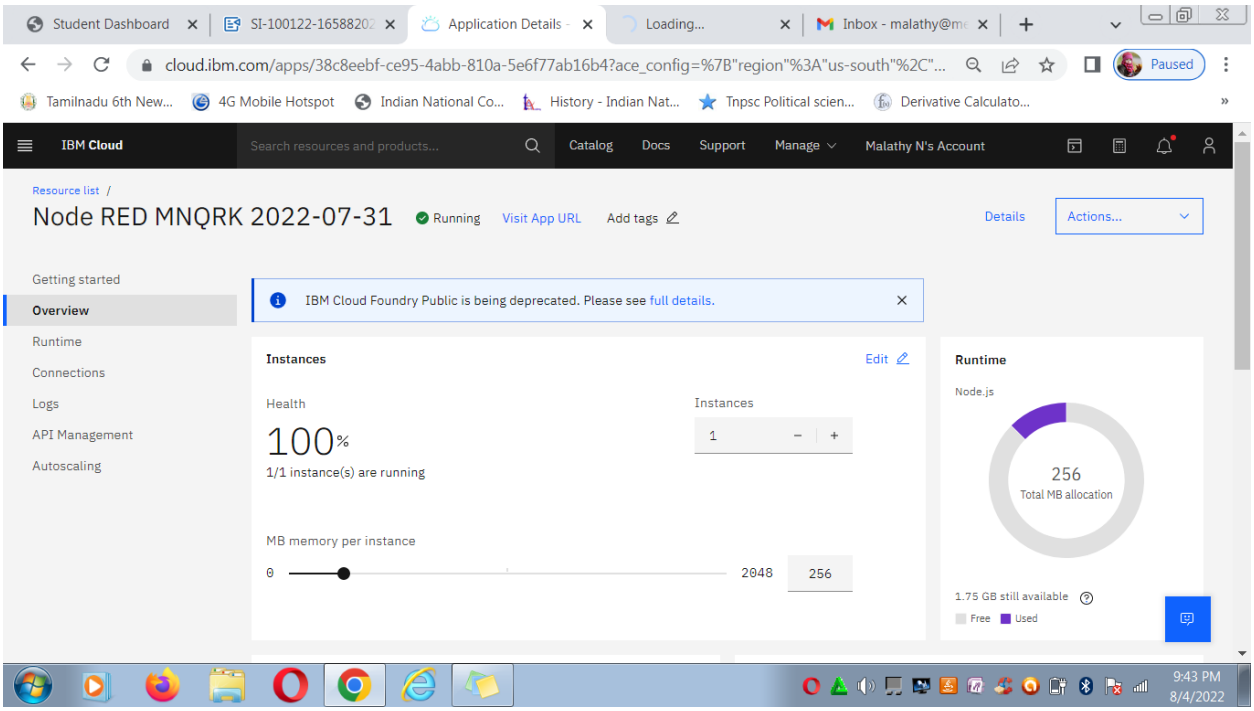

Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/ml-runtime/spaces/6935f3f8-9120-467e-a0ce-932cd2f3d241/deployments?cont...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Deployments /
Add to space
HD_DS
Overview
Assets
Deployments
Jobs
Manage
What deployments are you looking for?
Deployments (1)
Name
Type
Status
Asset
Tags
Last modified
HD_DEP
Online
Deployed
HD_AI - P3 XGB Classifier
Jul 31, 2022 3:05 PM
Drop files here or browse for files to upload.
Stay on the page until upload completes.
Incomplete uploads are cancelled.
9:40 PM
8/4/2022


Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/ml-runtime/models/db8b88f5-9419-4997-af83-769efe560dae?space_id=6935f3f...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Deployments / HD_DS /
HD_AI - P3 XGB Classifier
New deployment
Deployments
Model details
DEPLOYMENT TYPES
Online (1)
Batch (0)
1 Online Deployment(s)
Name
Status
Last modified
HD_DEP
Deployed
Jul 31, 2022 3:05 PM
HD_AI - P3 XGB Classifier
Created
Jul 31, 2022 3:00 PM
Type
wml-hybrid_0.1
Model ID
db8b88f5-9419-4997-af83-76...
Software specification
hybrid_0.1
Hybrid pipeline software specifications
autoai-kb_rt22.1-py3.9
Description
No description provided.
Tags
Add tags to make assets easier to find.
9:40 PM
8/4/2022

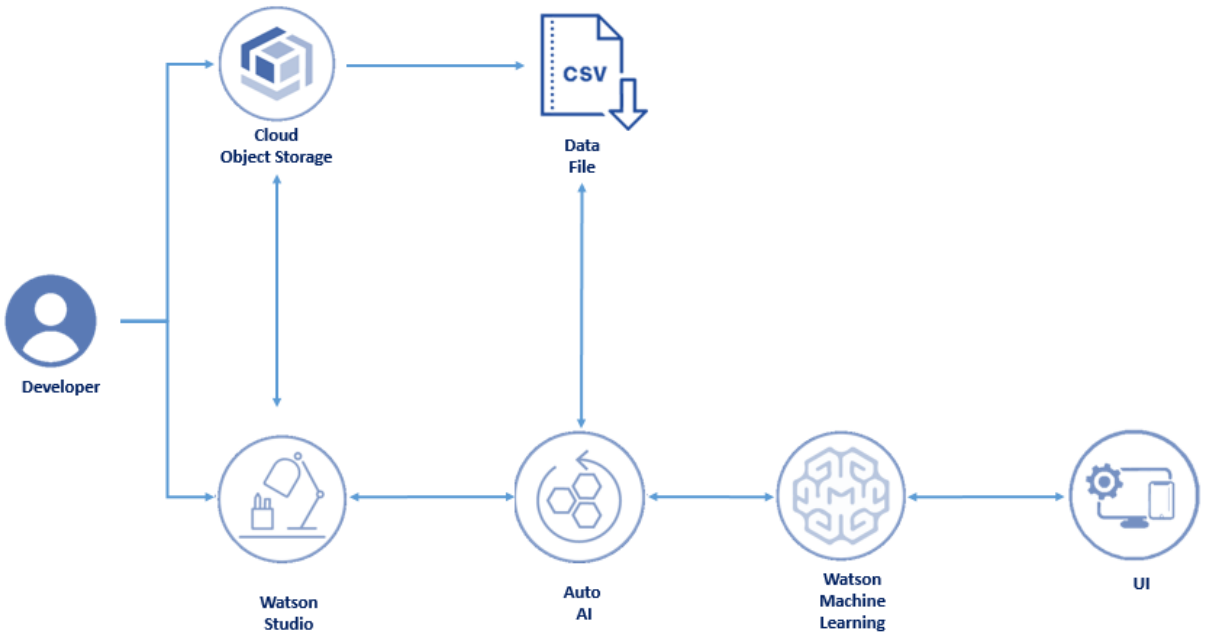

Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/ml-runtime/deployments/8c0ae728-0393-435d-9d7c-1352361e909a?space_id=6...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Deployments / HD_DS / HD_AI - P3 XGB Classifier /
HD_DEP
Deployed
Online
API reference
Test
Direct link
Endpoint
https://us-south.ml.cloud.ibm.com/ml/v4/deployments/8c0ae728-0393-435d-9d7c-1352361e909
Bearer <token>
IAM
Code snippets
cURL
Java
JavaScript
Python
Scala
NOTE: you must set $API_KEY below using information retrieved from your IBM Cloud account.
curl --insecure -X POST --header "Content-Type: application/x-www-form-urlencoded" --header "Accept: application/json"
--data-urlencode "grant_type=urn:ibm:params:oauth:grant-type:apikey"
--data-urlencode "apikey=$API_KEY" "https://iam.cloud.ibm.com/identity/token"
HD_DEP
Created
Jul 31, 2022 3:05 PM
Updated
Jul 31, 2022 3:05 PM
Deployment ID
8c0ae728-0393-435d-9d7c-13...
Software specification
hybrid_0.1
Hybrid pipeline software specifications
autoai-kb_rt22.1-py3.9
Copies
1
Serving name
No serving name.
9:40 PM
8/4/2022


Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/ml-runtime/deployments/8c0ae728-0393-435d-9d7c-1352361e909a/test?space_...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Search in your workspaces
Buy
Malathy N's Account
Dallas
MN
Deployments / HD_DS / HD_AI - P3 XGB Classifier /
HD_DEP
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
PALPITATIONSPE...
CHOLESTEROL (in...
BMI (integer)
HEARTFAILURE (o...
AGE (integer)
SEX (other)
FAMILYHISTORY (...
SMOKERLAST5YR...
EXERCISEMINPE...
1
56
121
23
y
23
f
y
y
25
Predict
9:41 PM
8/4/2022

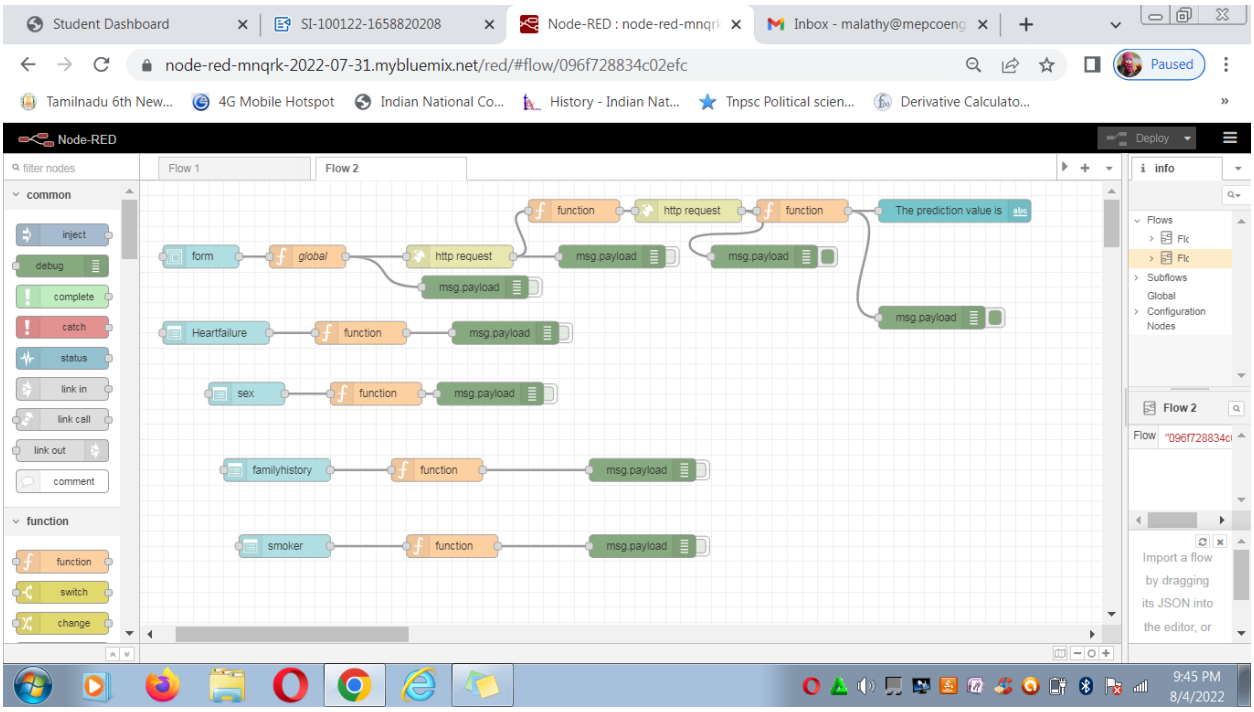

Student Dashboard
SI-100122-1658820208
IBM Cloud Pak for Data
Inbox - malathy@mepcoeng
dataplatform.cloud.ibm.com/ml-runtime/deployments/8c0ae728-0393-435d-9d7c-1352361e909a/test?space_...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud Pak for Data
Buy
Malathy N's Account
Dallas
MN
Deployments / HD_DS / HD_AI - P3 XGB Classifier /
HD_AI - P3 XGB Classifier test prediction
Prediction type
Regression classification
Prediction distribution
Amount of predictions
Prediction value
84.54
84.54
Table view
JSON view
HD_AI - P3 XGB Classifier test prediction
Prediction
1
84.54112794091527
2
3
4
5
6
7
8
9
10
Download
9:41 PM
8/4/2022

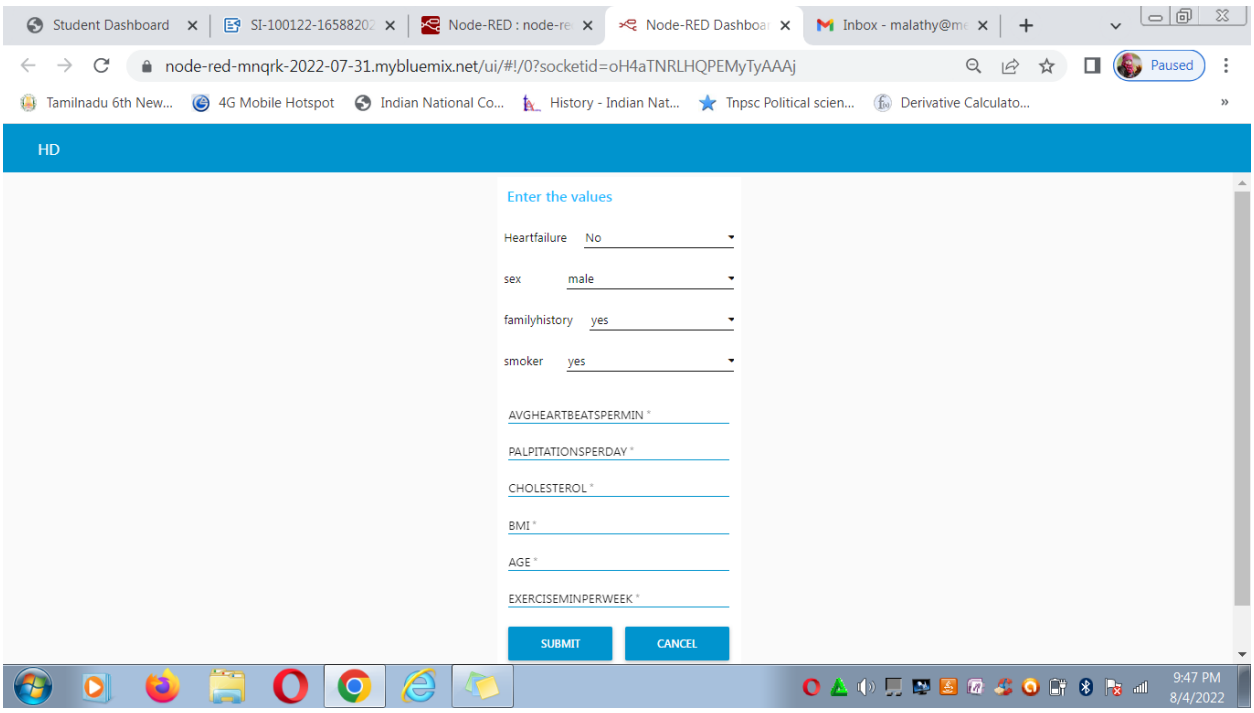

Student Dashboard
Application Details
Loading...
Inbox - malathy@m
cloud.ibm.com/apps/38c8eebf-ce95-4abb-810a-5e6f77ab16b4?ace_config=%7B"region"%3A"us-south"%2C"...
Paused
Tamilnadu 6th New...
4G Mobile Hotspot
Indian National Co...
History - Indian Nat...
Tnpsc Political scien...
Derivative Calculato...
IBM Cloud
Search resources and products...
Catalog
Docs
Support
Manage
Malathy N's Account
Resource list /
Node RED MNQRK 2022-07-31
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
9:43 PM
8/4/2022

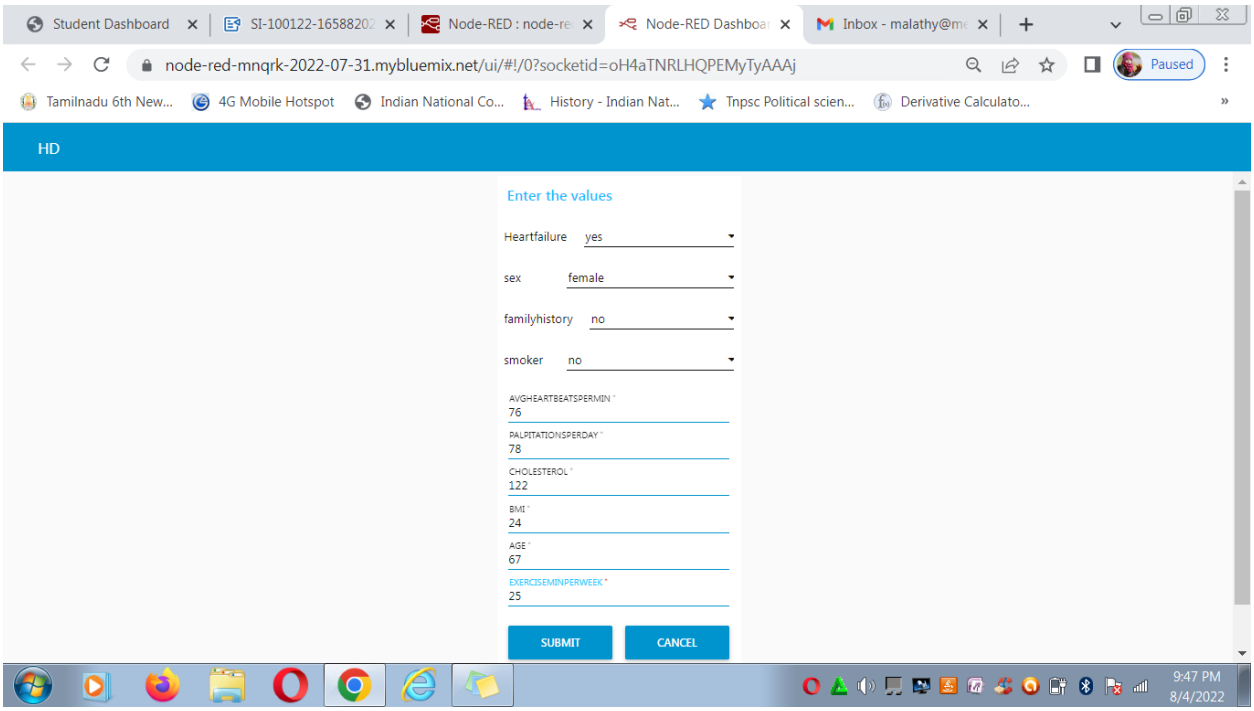
5.Flow Chart

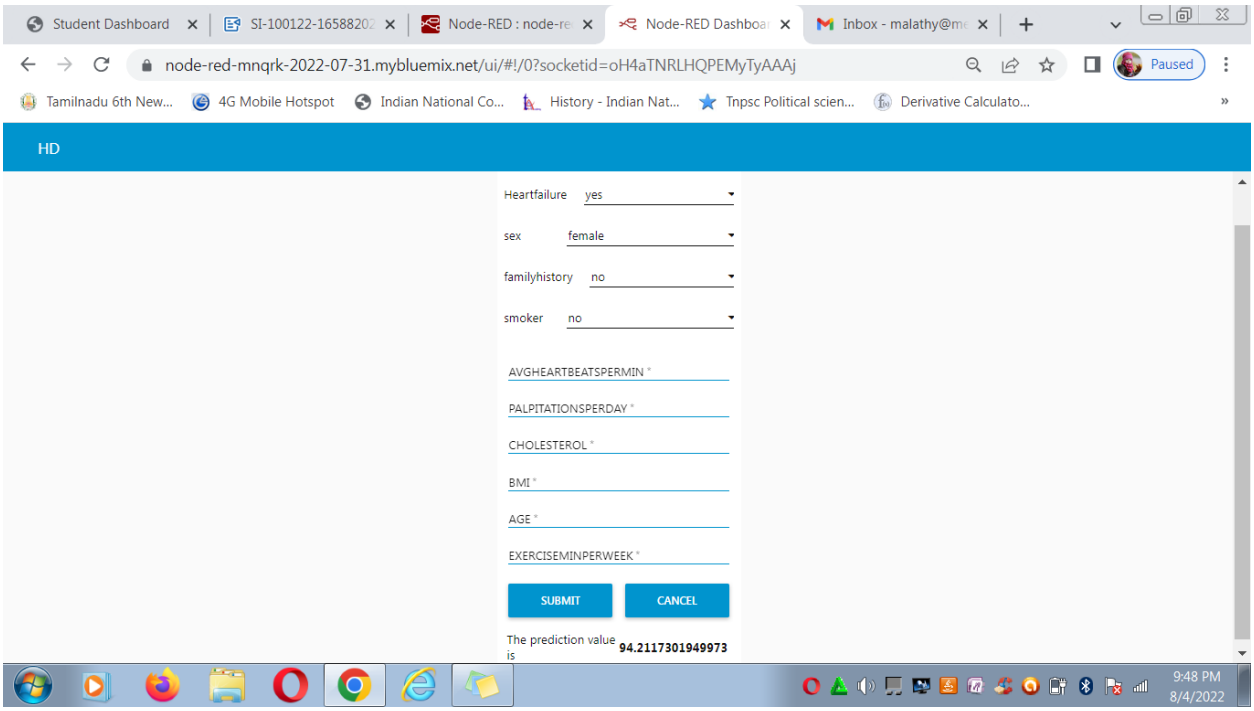

Cloud
Object Storage
CSV
Data
File
Developer
Watson
Studio
Auto
AI
Watson
Machine
Learning
UI

6.Result


Student Dashboard
SI-100122-1658820208
Node-RED : node-red-mnqr
Inbox - malathy@mepcoeng
node-red-mnqrk-2022-07-31.mybluemix.net/red/#flow/096f728834c02efc
Paused
Node-RED
Deploy
Flow 1
Flow 2
common
inject
debug
complete
catch
status
link in
link call
link out
comment
function
function
switch
change
form
global
http request
function
http request
function
The prediction value is
msg.payload
msg.payload
msg.payload
msg.payload
Heartfailure
function
msg.payload
sex
function
msg.payload
familyhistory
function
msg.payload
smoker
function
msg.payload
info
Flows
Subflows
Global
Configuration Nodes
Flow 2
Import a flow by dragging its JSON into the editor, or
9:45 PM
8/4/2022


Student Dashboard
Node-RED : node-re
Node-RED Dashboa
Inbox - malathy@mc
node-red-mnqrk-2022-07-31.mybluemix.net/ui/#!/0?socketid=oH4aTNRLHQPEMyTyAAAj
Paused
HD
Enter the values
Heartfailure No
sex male
familyhistory yes
smoker yes
AVGHEARTBEATSPERMIN *
PALPITATIONSPERDAY *
CHOLESTEROL *
BMI *
AGE *
EXERCISEMINPERWEEK *
SUBMIT
CANCEL
9:47 PM
8/4/2022





7.Advantages

It can be used in hospitals to predict the disorder at an earlier stage

8.Applications

It can be used in hospitals to predict the disorder at an earlier stage

9.Conclusion and Future work

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system. The proposed working model can also help in reducing treatment costs by providing Initial diagnostics in time. The model can also serve the purpose of training tool for medical students and will be a soft diagnostic tool available for physician and cardiologist. General physicians can utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research. In DM warehouse, generally, the dimensionality of the heart database is high, so identification and selection of significant attributes for better diagnosis of heart disease are very challenging tasks for future research.

References.

1. A, A. S., & Naik, C. (2016). Different Data Mining Approaches for Predicting Heart Disease, 277–281. https://doi.org/10.15680/IJIRSET.2016.0505545
2. Beyene, C., & Kamat, P. (2018). Survey on prediction and analysis the occurrence of heart disease using data mining techniques. International Journal of Pure and Applied Mathematics, 118(Special Issue 8), 165–173. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041895038&partnerID=40&md5=2f0b0c5191a82bc0c3f0daf67d73bc81
3. Brownlee, J. (2016). Naive Bayes for Machine Learning. Retrieved March 4, 2019, from https://machinelearningmastery.com/naive-bayes-for-machine-learning/

4. Kirmani, M. (2017). Cardiovascular Disease Prediction using Data Mining Techniques. Oriental Journal of Computer Science and Technology, 10(2), 520–528. https://doi.org/10.13005/ojcst/10.02.38
5. Polaraju, K., Durga Prasad, D., & Tech Scholar, M. (2017). Prediction of Heart Disease using Multiple Linear Regression Model. International Journal of Engineering Development and Research, 5(4), 2321–9939. Retrieved from www.ijedr.org
6. Purushottam, Saxena, K., & Sharma, R. (2016). Efficient Heart Disease Prediction System. In Procedia Computer Science (Vol. 85, pp. 962–969). https://doi.org/10.1016/j.procs.2016.05.288
7. Sai, P. P., & Reddy, C. (2017). International Journal of Computer Science and Mobile Computing HEART DISEASE PREDICTION USING ANN ALGORITHM IN DATA MINING. International Journal of Computer Science & Mobile Computing, 6(4), 168–172. Retrieved from www.ijcsmc.com
8. Science, C., & Faculty, G. M. (2009). Heart Disease Prediction Using Machine learning and Data Mining Technique. Ijcsc 0973-7391, 7, 1–9.
9. Soni, J., Ansari, U., & Sharma, D. (2011). Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers. Heart Disease, 3(6), 2385–2392.