iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VIII **Month of publication:** August 2022

DOI: https://doi.org/10.22214/ijraset.2022.46087

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VIII August 2022- Available at www.ijraset.com

Clinical Implication of Machine Learning Based Cardiovascular Disease Prediction Using IBM Auto AI Service

M. Nirmala[1], V. Saravanan[2], A. R. Jayasudha[3], Princess Maria John[4], P. Privietha[5] , L. Mahalakshmi[6]
[1,3,4,5,6]*Department of Computer Applications,* [2]*Department of Information Technology*
[1]*Hindusthan College of Engineering and Technology,* [2]*Hindusthan College of Arts and Science*

Abstract*—Cardio vascular diseases 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 this dataset contains 11 features that can be used to predict mortality by heart failure. In this project, a model is built using the Random Forest Classifier Algorithm using Auto AI and a web application is created using Node Red Application and it showcases the prediction of heart failure in a web based format. The usage of IBM cloud environment for implementing the Machine learning Model using IBM Auto AI and Node Red Flows are created for the display of Web Application Structure. The complete paper explains the coordination among the Auto AI and Node red in the Cloud Platform.

Keywords— Cardio Vascular Disease, IBM Auto AI, Machine Learning, Node Red Application, Watson Studio.

I. INTRODUCTION

Machine Learning [1] commonly deals with big data where the size of the data is massive and the data can be both in structured and unstructured format It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Various techniques in data mining and neural networks have been employed to find out the severity of heart disease among humans. The nature of heart disease is complex and hence, the disease must be handled carefully. The proposed method which we use has 10 attributes for heart disease prediction and the problem is carried out using IBM AUTO AI service.

II. LITERATURE SURVEY / RELATED WORK

Lippi et al. [2] focused on the possibility of cardiovascular disease during the COVID-19 pandemic. ,e nationwide quarantine has compelled the government to implement various forms of lockdown to reduce the transmission of COVID-19. As a result of these restrictions, all citizens remain at home, resulting in physical inactivity. Although the WHO has established clear guidelines on the amount of physical activity required to maintain adequate health, strict quarantine, on the other hand, has increased the risk of cardiovascular mortality. After quarantine, negative health effects are observed. As a result, the authors proposed the fact that it is necessary to maintain physical exercise even during quarantine to avoid unfavourable cardiovascular consequences, has influenced the current research study's Computational Intelligence and Neuroscience design

The random forest algorithm was used in the study [3] to predict the occurrence of heart disease in patients. A total of 303 samples from the Kaggle dataset were considered and the metrics used to evaluate performance were accuracy, sensitivity, and specificity. In the classification of heart disease, the algorithm achieved a prediction rate of 93.3%.

S. Krishnan J. Geetha S [4] has made a system that predicts the developing potential results of Heart Disease. Their aftereffects of this system give the chances of happening heart disease to the extent rate. They have considered datasets used are organized similar to therapeutic parameters. Their structure evaluates those parameters using the information mining plan strategy. Their datasets were set up in python programming using two standard Machine Learning Algorithm to be explicit Decision Tree Algorithm and Naive Bayes Algorithm and have exhibited the best estimation among these two to the extent the precision level of heart illness.

K.G Dinesh, K.A.raj, K.D.Santhosh, V. M.eswari [5] has talked about heart illness expectation and performed information pre-preparing utilizes strategies like the removal of noisy data, removal of missing data, filling default values if applicable and classification of attributes for prediction and decision making at different levels. Their exhibition of the finding model is acquired by utilizing techniques like order, exactness, affectability and particularity examination.

This has proposed a forecast model to anticipate whether people have heart illness or not and to give mindfulness or finding on that. They have done examination by comparing the accuracies of applying rules with the individual consequences of Support Vector Machine, Gradient Boosting, Random backwoods, Naive Bayes classifier and calculated relapse on the dataset taken in a district to display an exact model of foreseeing cardiovascular ailment.

Senthil Kumar Mohan et al,[6] proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which strategy that objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady. The expectation model is created with various blends of highlights and a few known arrangement strategies. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM) they likewise educated about Diverse data mining approaches and expectation techniques, Such as, KNN, LR, SVM, NN, and Vote have been fairly famous of late to distinguish and predict heart disease.

Avinash Golande et al,[7] proposed Heart Disease Prediction Using Effective Machine Learning Techniques in which Specialists utilize a few data mining strategies that are available to support the authorities or doctors distinguish the heart disease. Usually utilized methodology utilized are decision tree, k- closest 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). The following area obviously gives subtleties of systems that were utilized in the examination.

Lakshmana Rao et al,[8] Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc..). So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. The idea of CHD ailment is bewildering, in addition, in this manner, the disease must be dealt with warily. Not doing early identification, may impact the heart or cause sudden passing. The perspective of therapeutic science furthermore, data burrowing is used for finding various sorts of metabolic machine learning a procedure that causes the framework to gain from past information tests, models without being expressly customized. Machine learning makes rationale dependent on chronicled information

Marimuthu M et al [9], proposed Heart disease is one of the prevalent disease that can lead to reduce the lifespan of human beings nowadays. Each year 17.5 million people are dying due to heart disease. Life is dependent on component functioning of heart, because heart is necessary part of our body.

Heart disease is a disease that effects on the function of heart. An estimate of a person's risk for coronary heart disease is important for many aspects of health promotion and clinical medicine. A risk prediction model may be obtained through multivariate regression analysis of a longitudinal study

Balakrishnan et al [10] proposed Machine learning is a technique converts the raw clinical data into an informational data that helps for decision making and prediction. Cardiovascular disease is one of the major causes of mortality around the world. It is considered in a large scale, so prediction of cardiovascular disease is more important in the clinical survey analysis as day by day it gets increased. The amount of data in the health club is huge. As cardiovascular is one of the major causes for death there are some data analytical techniques that predicts the occurrence of cardiovascular disease. It can be achieved through selecting a correct combination of prediction models and features.

Prediction models were developed using different classification techniques based on feature selection and there are certain algorithms which provide varied and improved accuracy.

Here prediction model is developed using Random Forest classification technique - Method for classification, regression by constructing a multitude of decision trees at training time. Developed by aggregating tree Avoids over fitting can deal with large number of features. Helps with feature selection based on importance where necessary features only classified. Pre-processing will be done first considering the clinical data. It will be spited into train and test data with which accuracy can be achieved.

III. RESEARCH METHODOLOGY

The various methods adopted during the research process have been portrayed. This is a Descriptive Research problem where the study of Medical data set is explored. It performs the prediction of Heart Disease of the patients from the medical data set by applying various methodologies with respect to Machine Learning using IBM Auto AI.

A. Research Data

The data collected from secondary data sources are tabulated in the Table 1.

Table 1 : Data Source Details

Dataset characteristics	Multivariate
Number of Instances	10800
Number of Attributes	10
Attribute Type	Categorical and Numerical
Link	https://github.com/IBM/predictive-model-on-watson-ml/blob/master/data/patientdataV6.csv

B. Proposed System Method Of Analysis

The project helps us to predict the heart failure of human being given various parameters like the details listed in

Table 2 : Patient Features

FEATURES	DATA TYPE
AVGHEARTBEATSPERMIN	NUMERIC
PALPITATIONSPERDAY	NUMERIC
CHOLESTEROL	NUMERIC
BMI	NUMERIC
AGE	NUMERIC
SEX	CATEGORICAL (F/M)
FAMILY HISTORY	CATEGORICAL (Y/N)
SMOKERLAST5YRS	CATEGORICAL (Y/N)
EXERCISEMINPERWEEK	NUMERIC
HEARTFAILURE	CATEGORICAL (Y/N) (TARGET VARIABLE)

C. Block diagram

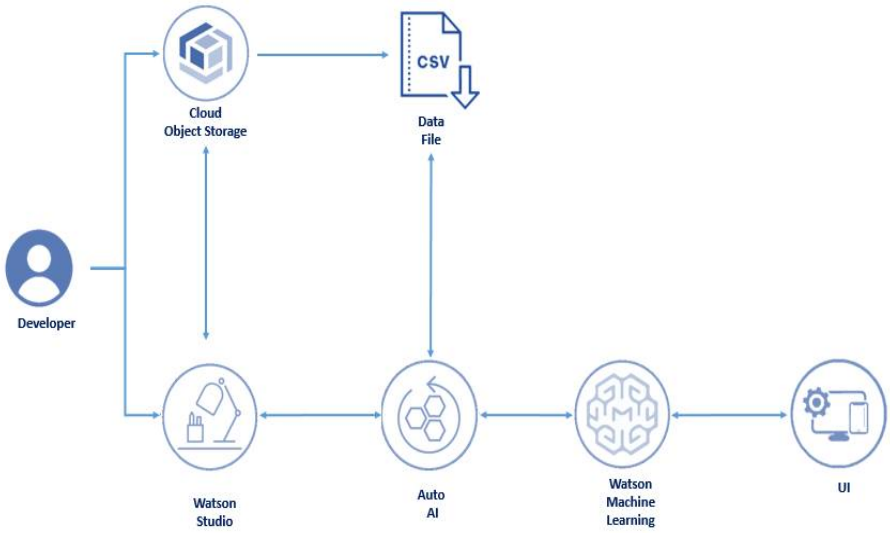


Figure 1 : Block Diagram

1) *Cloud Object Storage:* The developer creates the Cloud object storage in the ibm cloud. Cloud Object storage is a service offered by IBM for storing and accessing the unstructured data. Objects are pieces of data that is uploaded in the cloud storage.

2) *IBM Watson Studio:* IBM Watson studio is an integrated environment designed to develop, manage models and deploy AI powered Applications. It is Software as a Service. A project is created in IBM Watson Studio
3) *AUTO AI:* The Auto AI graphical tool in Watson Studio analyzes the data and discovers data transformations, algorithms, and parameter settings that work best for the predictive modeling problem. Auto AI displays the results as model candidate pipelines ranked on a leaderboard and the specific model can be appropriately chosen by the developer based on the result obtained.

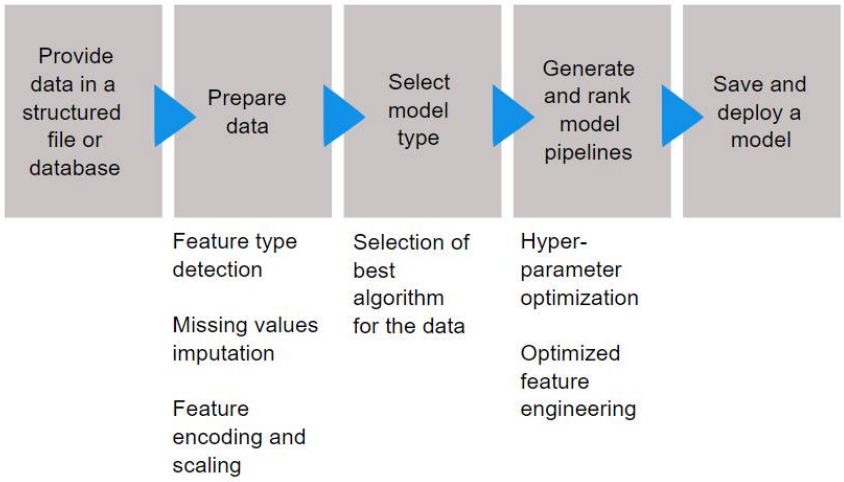


Figure 2 : AUTO AI Pipeline Format

4) *Node Red Service:* It is a programming tool for wiring together hardware devices, API's and Online services. It is a browser based editor that makes it easy to wire together flows using wide range of nodes in the palette that can be deployed in its run time in a single click. It helps in deploying the ML model as a web server

D. Hardware and Software Specification

Table 3 : Hardware Specifications

Processor	Intel(R) Core(TM) i3-3227U CPU @ 1.90GHz 1.90 GHz
Ram	4 GB.
HDD	100 GB.
Monitor type	15 Inch VGA.
Keyboard	110Keys Keyboard

Table 4 : Software Specifications

Operating system	Windows 10
Web Browser	Chrome, Mozilla firefox
Set up	Watson Studio
Storage	IBM Cloud Platform
Application Development	Node Red Service
Machine Learning Service	AUTO AI

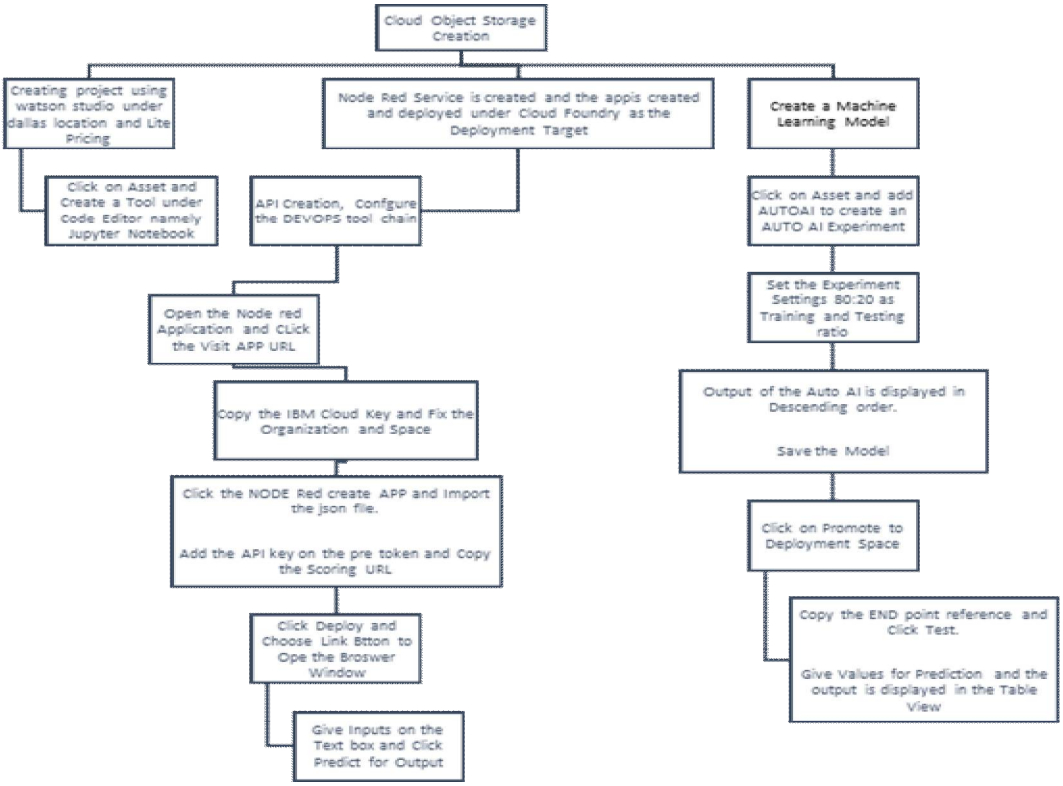


Figure 3 : Flow Chart of the Complete Experiment

IV. EXPERIMENTAL INVESTIGATIONS

Create the cloud Object storage in the IBM Cloud Platform. The cloud object is names as cloud object storage – 5m

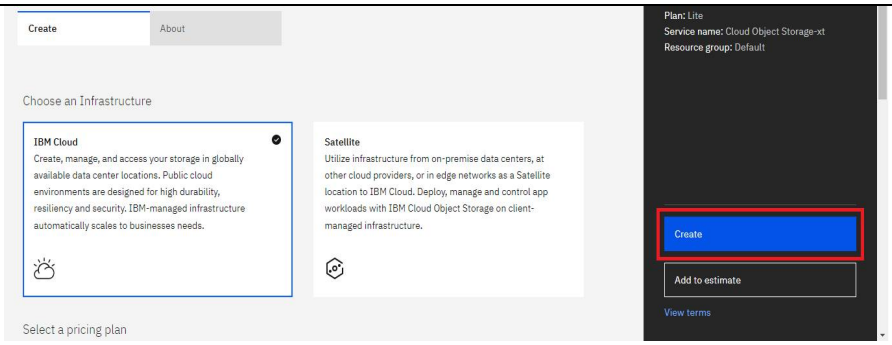


Figure 4 : Cloud Object Storage Creation

Create the Watson Studio called as WATSON STUDIO NIRMALA with the location as Dallas and the plan as Lite

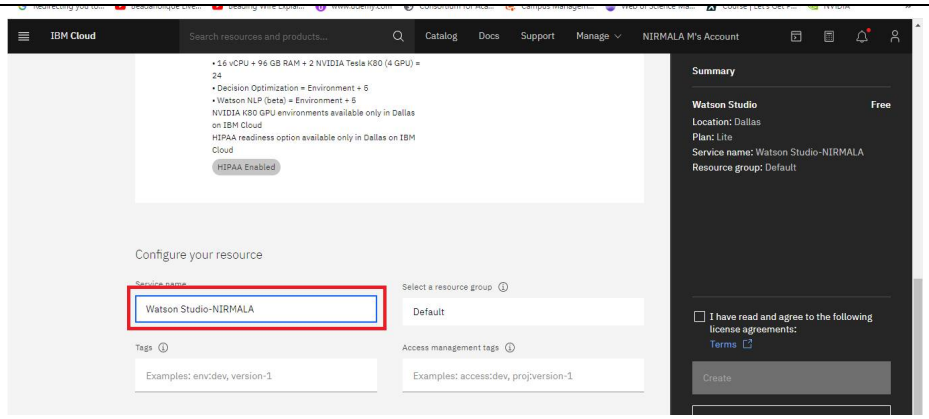


Figure 5 : Watson Studio Creation

After the Watson Studio creation Click on the Launch in IBM Cloud Pak for Data. Click on the new project for a New project as NIRMALA_HEART_DISEASE_PREDICTION_PROJ.

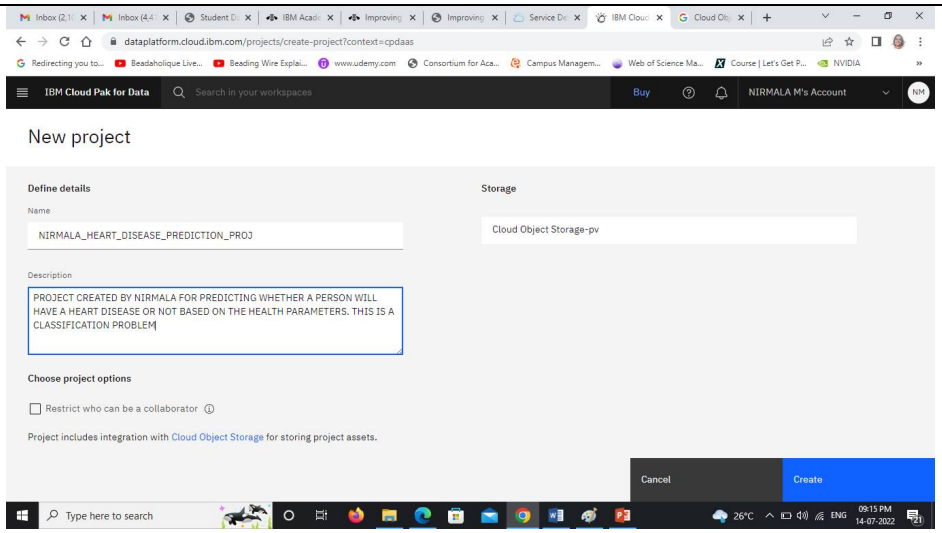


Figure 6 : Project Creation

Click on Asset and select the new tool to create an operational or configuration asset. The tool created is Code Editors and choose Jupyter Notebook

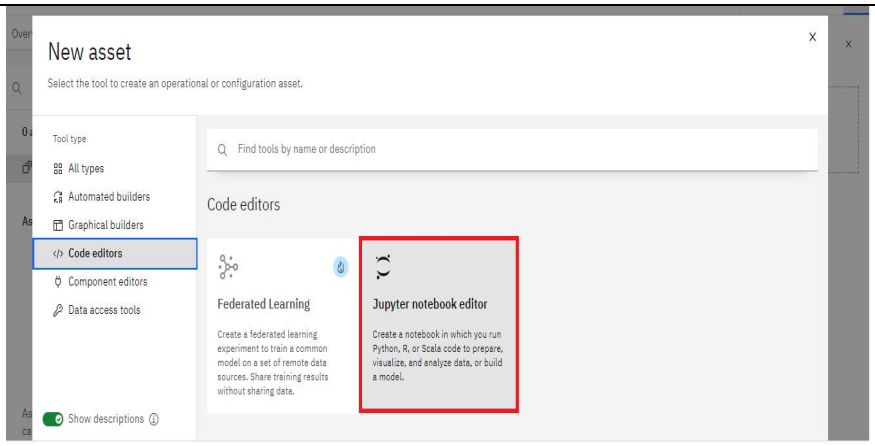


Figure 7 : Selection of Code Editor

The new notebook called as NIRMALA_HEART_NOTEBOOK is created

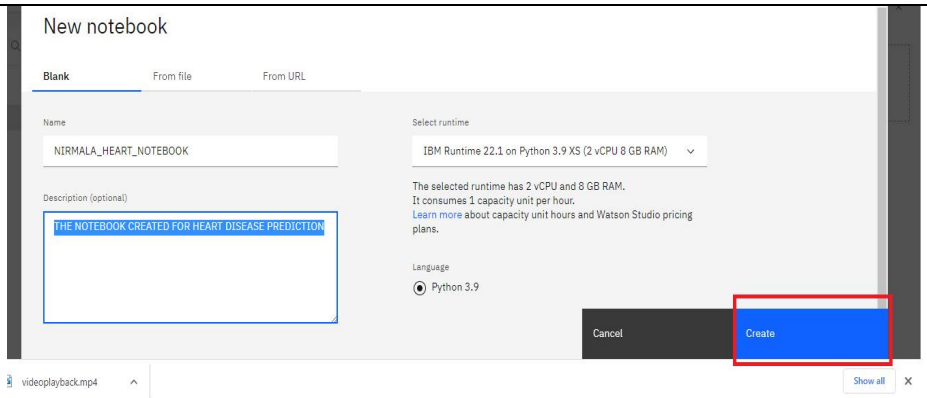


Figure 8 : Jupyter Notebook Creation

Node Red Service
It is a programming tool for wiring together hardware devices, API's and Online services. A node red service called as NODE RED NIRMALA is created with the location as dallas and pricing plan as Lite

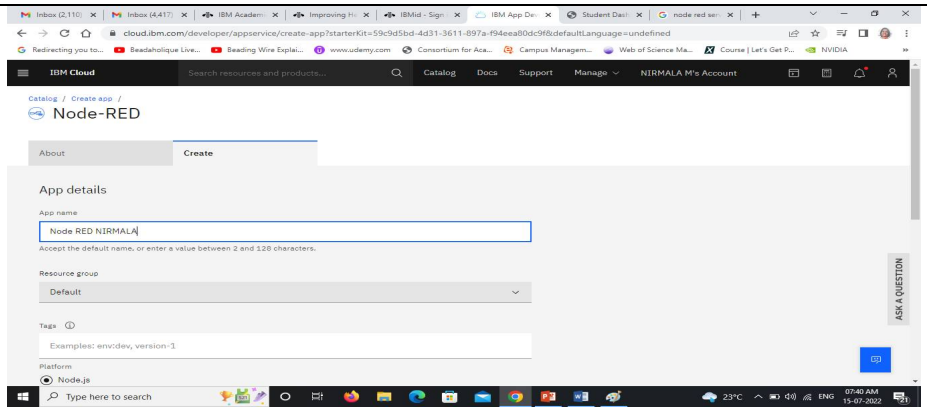


Figure 9 : Node Red Service Creation

After the App is created Deploy the App. Choose cloud foundry as the deployment target

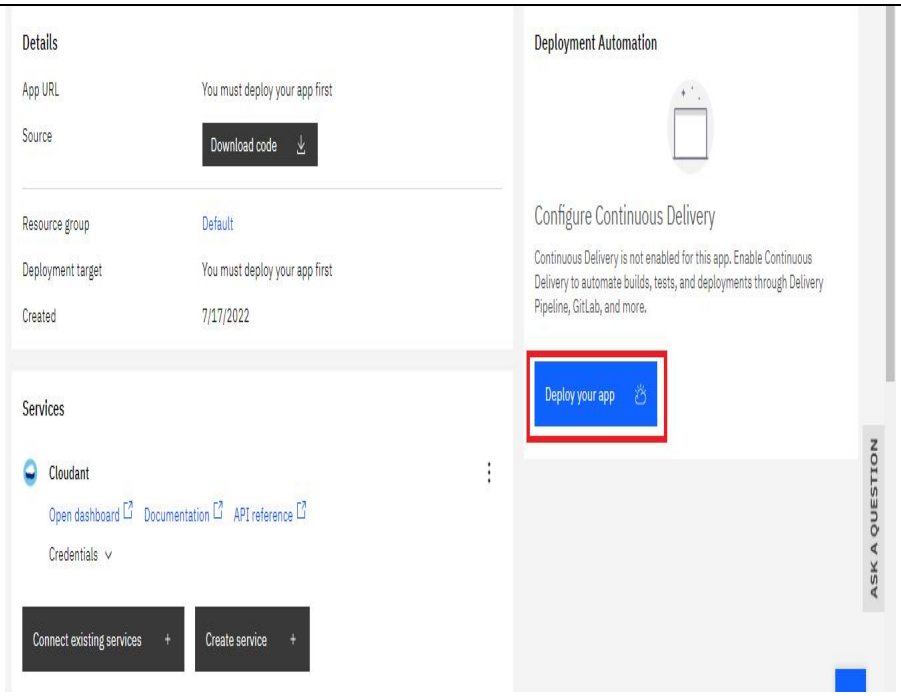


Figure 10 : App Deployment

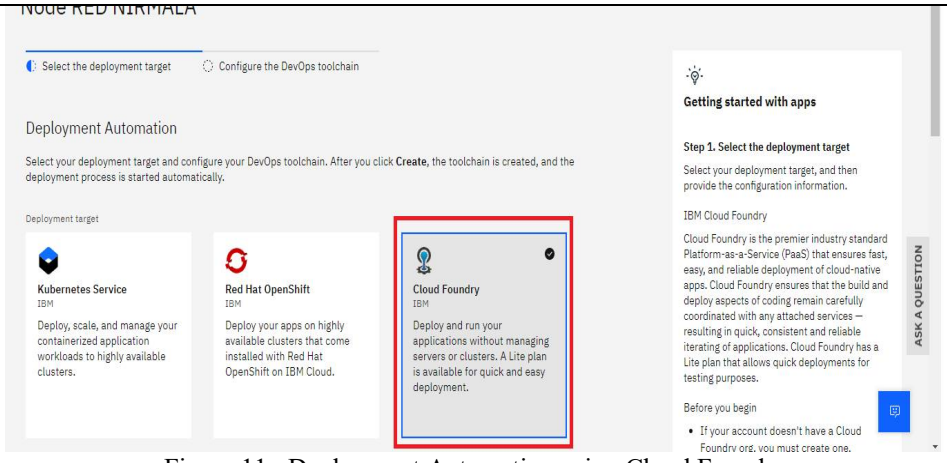
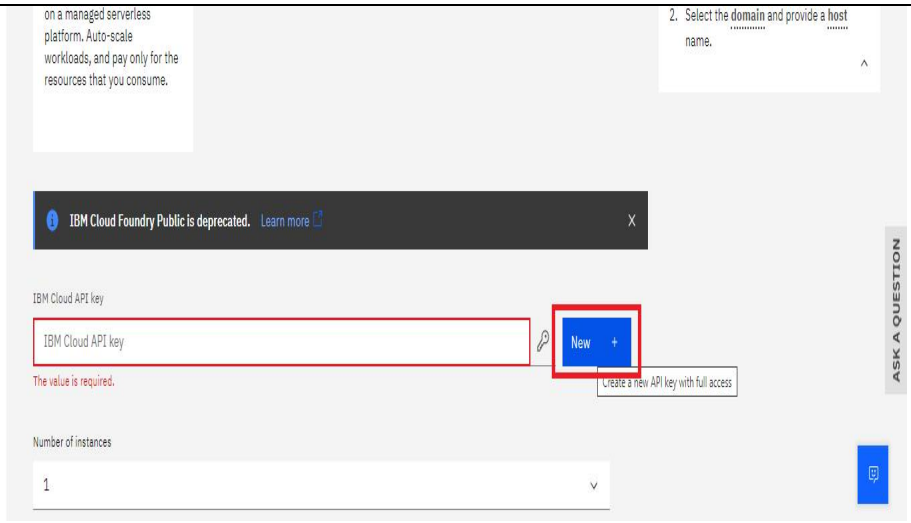
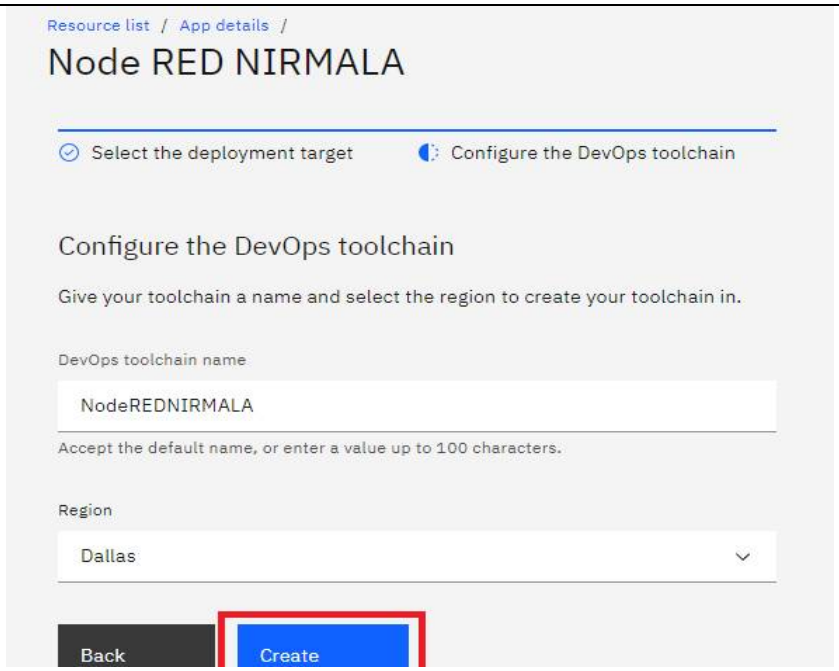
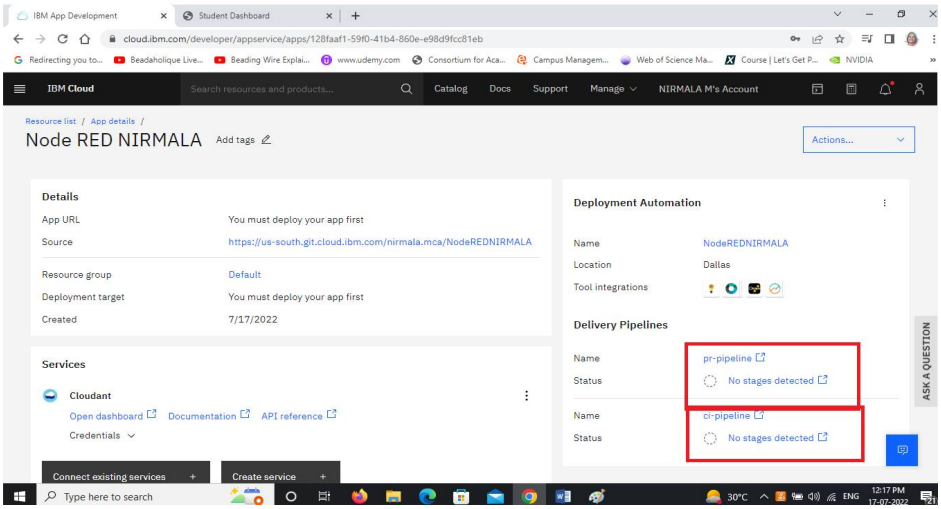
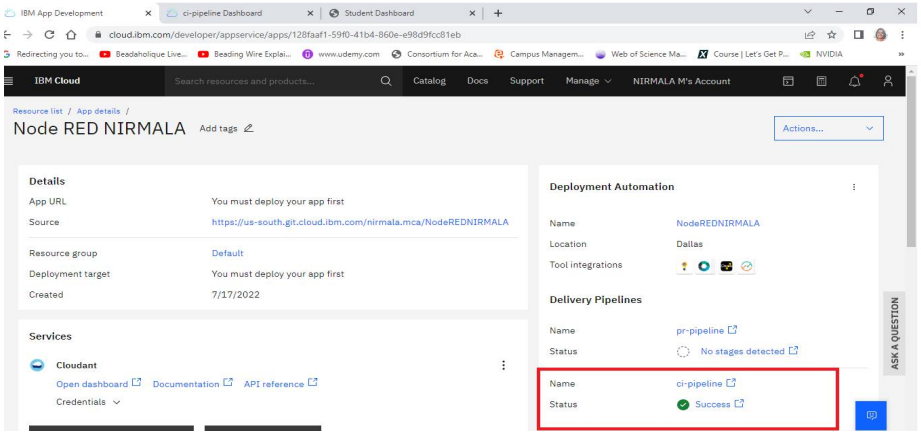
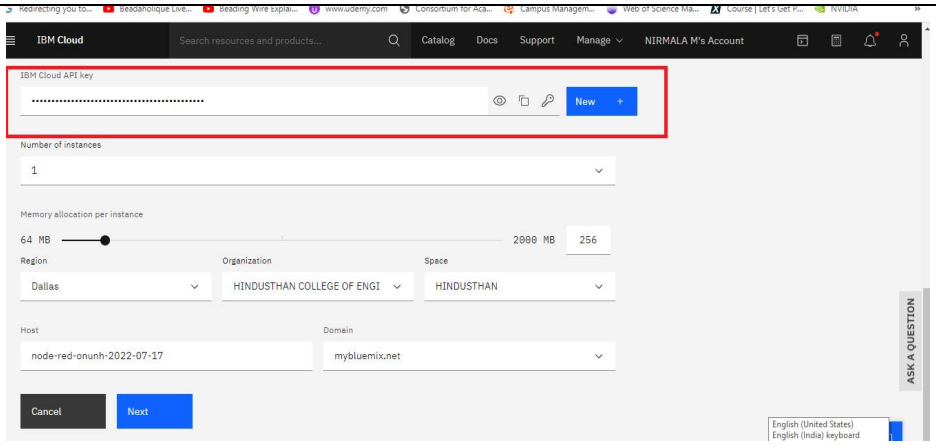


Figure 11 : Deployment Automation using Cloud Foundry

An API key is created for this application and Downloaded	 Figure 12 : API Key Creation
Configure the DevOps toolchain. Give a toolchain name and select the region to create your toolchain	 Figure 13 : Configure of DevOps Tool Chain

From No stages detected it should change to success	 Figure 14 : Pipeline Progress Status
It is changed to Progress and then it is changed to Success	 Figure 15 : Pipeline Success Status
To open the node red application, Click on the dashboard, Choose cloud foundry apps and choose the Node red Application. Click the Visit App URL and Copy the IBM Cloud App Key, Choose the Region as Dallas, and choose the Organization as Hindusthan College of Engineering and Technology and the space as Hindusthan which has been already created.	 Figure 16 : Node Red Configuration

Get into the Node Red Instance on the IBM Cloud

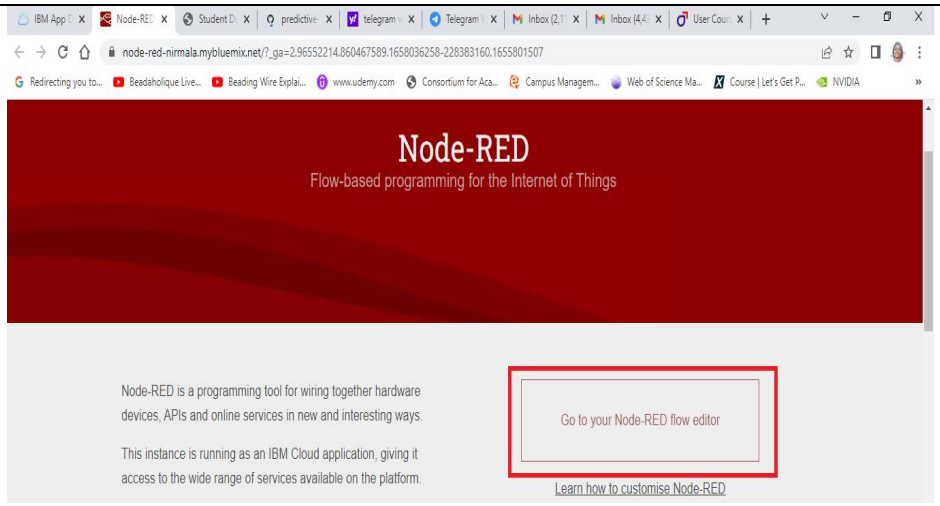


Figure 17 : Node Red Editor

A project is how you organize your resources to achieve a particular goal. In the Watson Studio NIRMALA an already created project called as NIRMALA_HEART_DISEASE_PREDICTION_PROJ is already present.

Assume the Jupyter Notebook is created using New asset option and also the data set is loaded

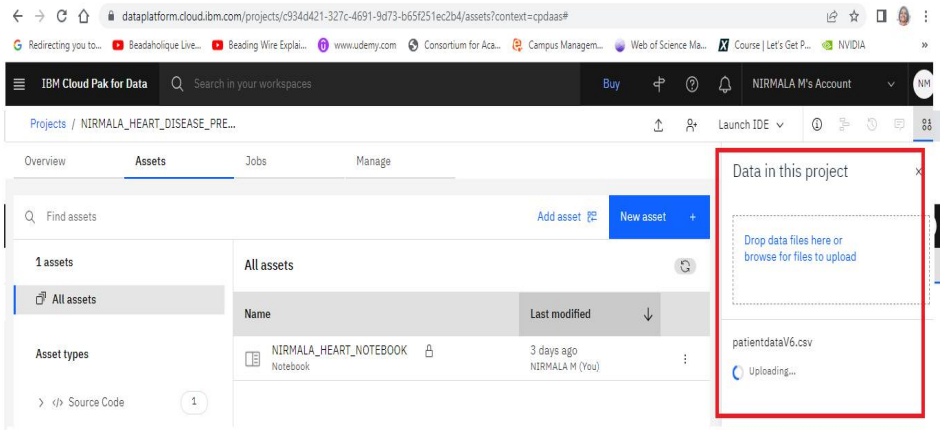


Figure 18 : Jupyter Notebook created using Asset Option

Assume there is a Machine Learning Model 8n already created

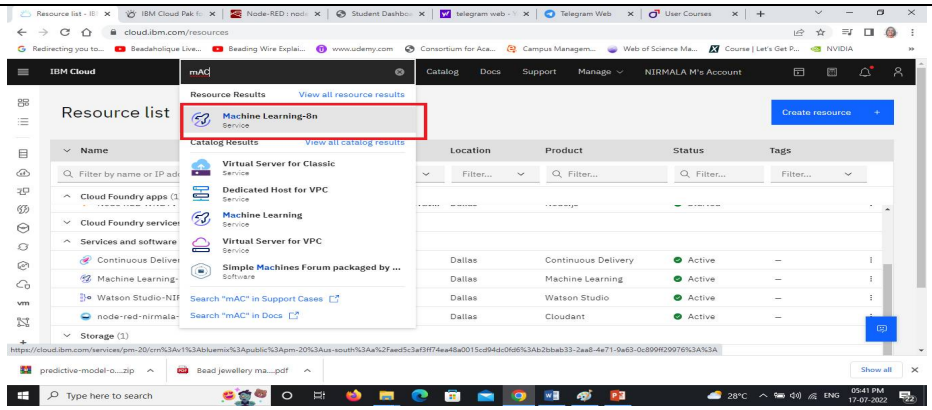


Figure 19 : Machine Learning Model Created

To add AUTO AI, Click on New Asset and Choose AUTO AI and create an AUTO AI experiment

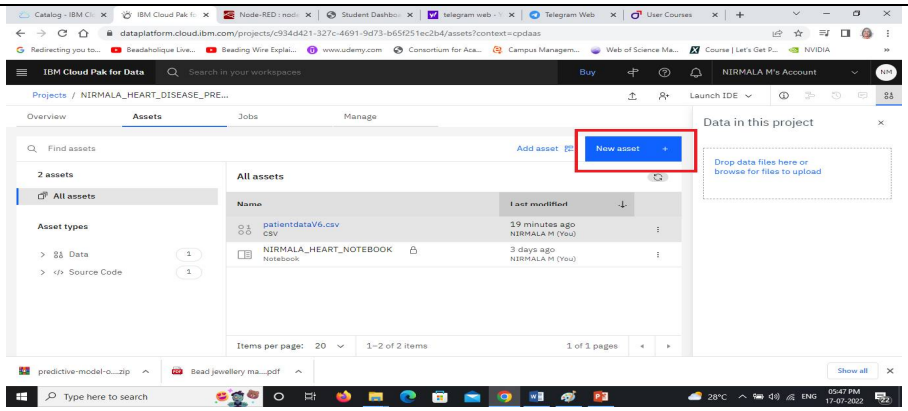


Figure 20 : AUTO AI Creation

An AUTO AI Experiment called as HEART_DISEASE_PREDICTION_AI is created and associated with the Machine learning-8n service with 8Vcpu and 32 GB RAM

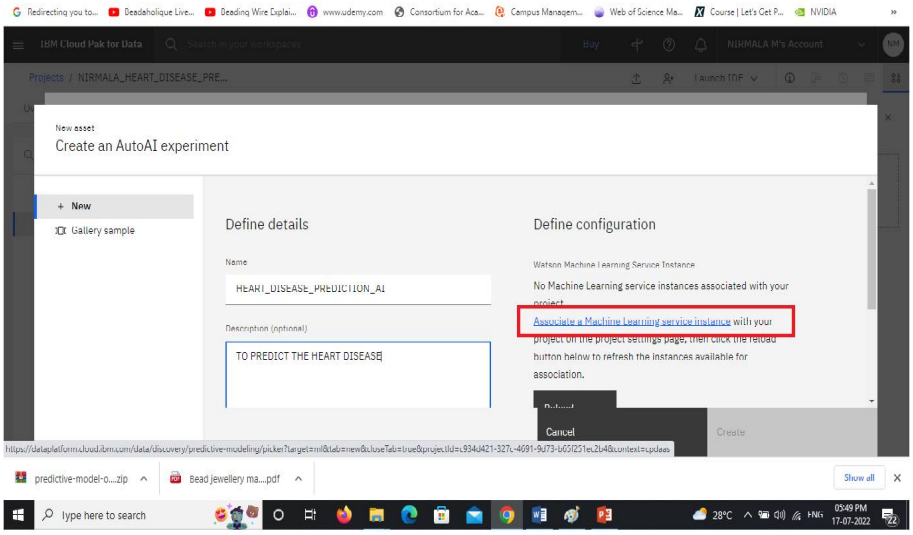


Figure 21 : AUTO AI Associated with Machine Learning

If required change the Experimental Settings and Run the Experiment with 80:20 as Training and Testing data set ratio.

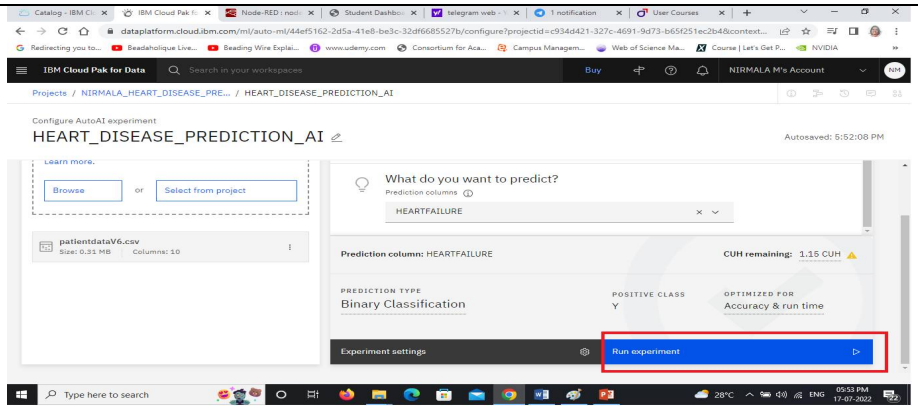


Figure 22 : Experimental Settings & Run Experiment

Output of AUTO AI Execution

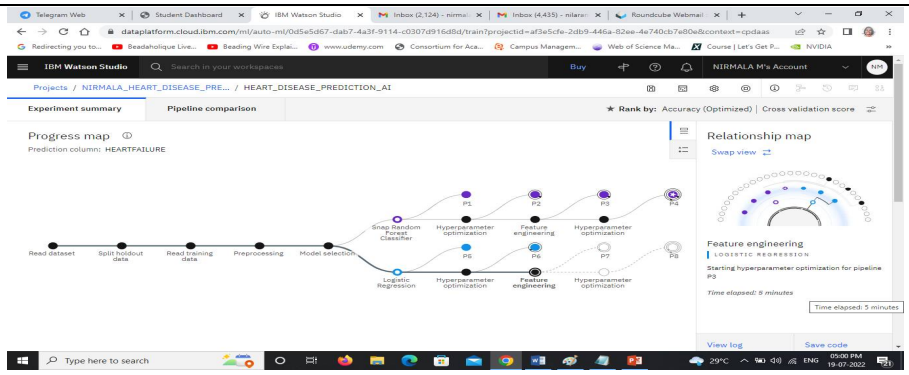


Figure 23 : AUTO AI Output

Output of AUTO AI after the Execution and the model Algorithms with its accuracy is displayed in the descending order. The Higher ranking algorithm is in the Top position with respect to accuracy. In this case the snap Random Forest Classifier comes with an accuracy of 87% and the Logistic Regression comes with 79%

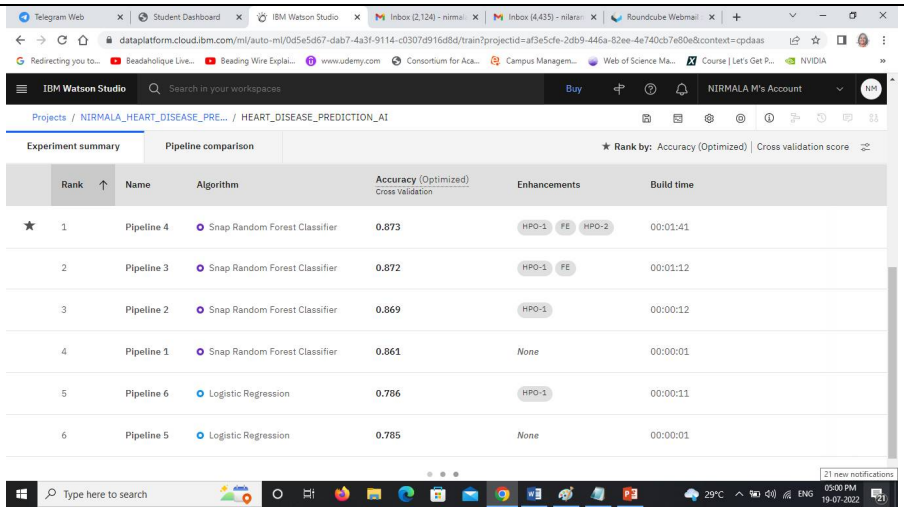


Figure 24 : Model Algorithms Ranking

Once the pipeline creation is complete, you can view and compare the ranked pipelines in a leaderboard. Choose Save model from the action menu for the pipeline with the highest accuracy or low error rate This saves the pipeline as a Machine Learning asset in your project. A notification gives you the link to view the saved model in your project.

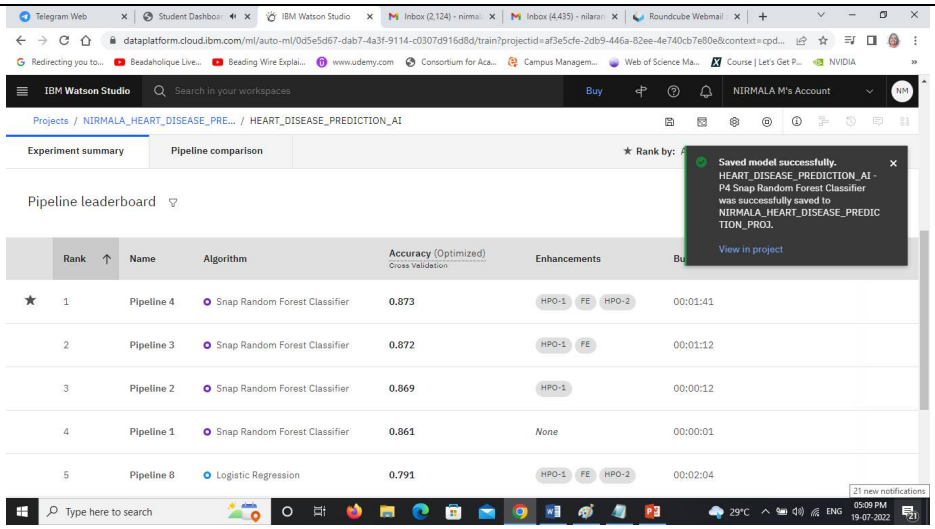
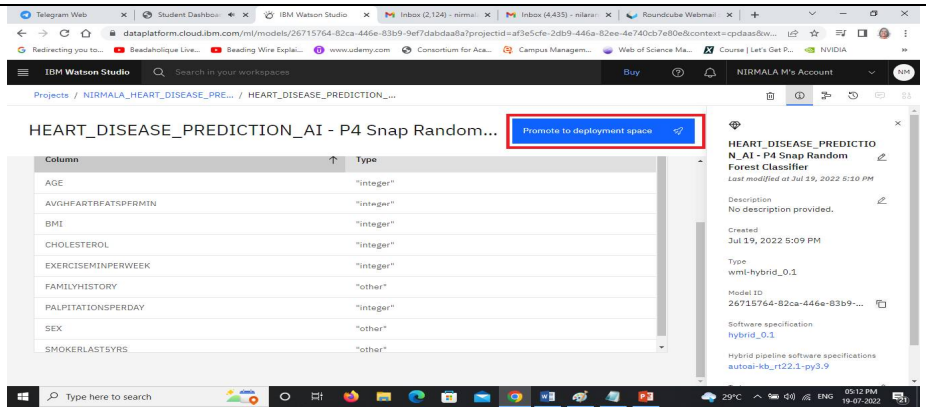
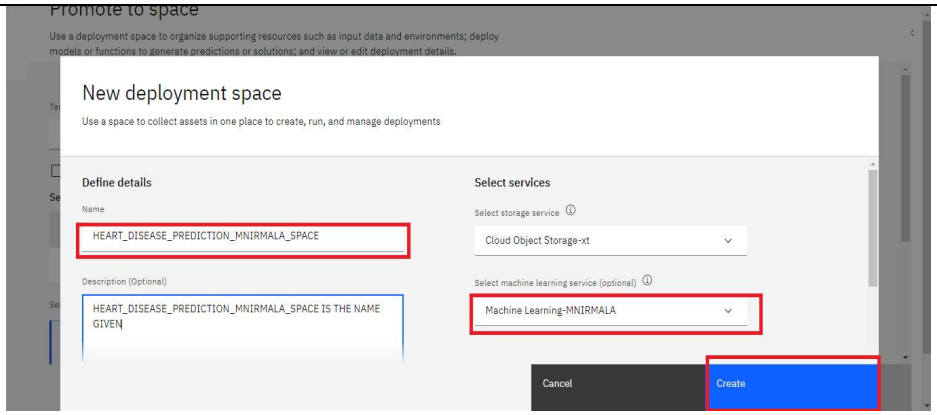
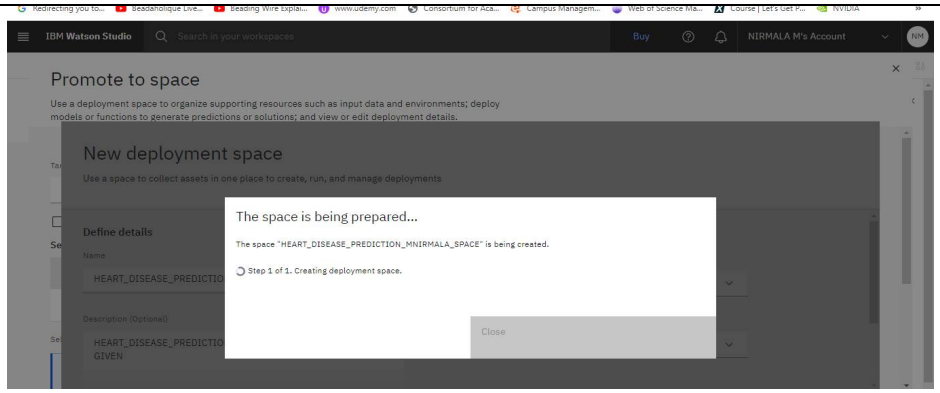
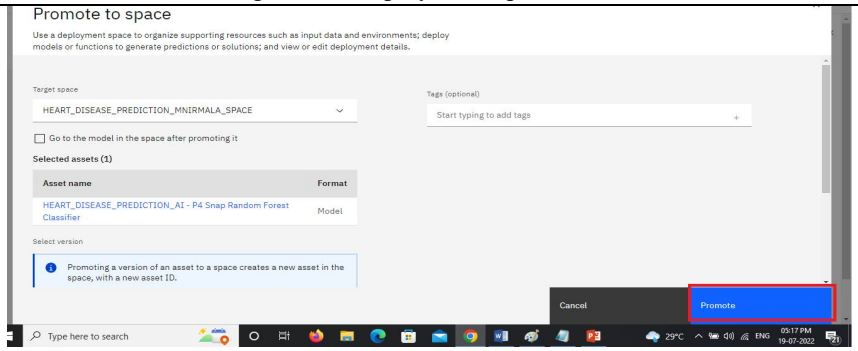
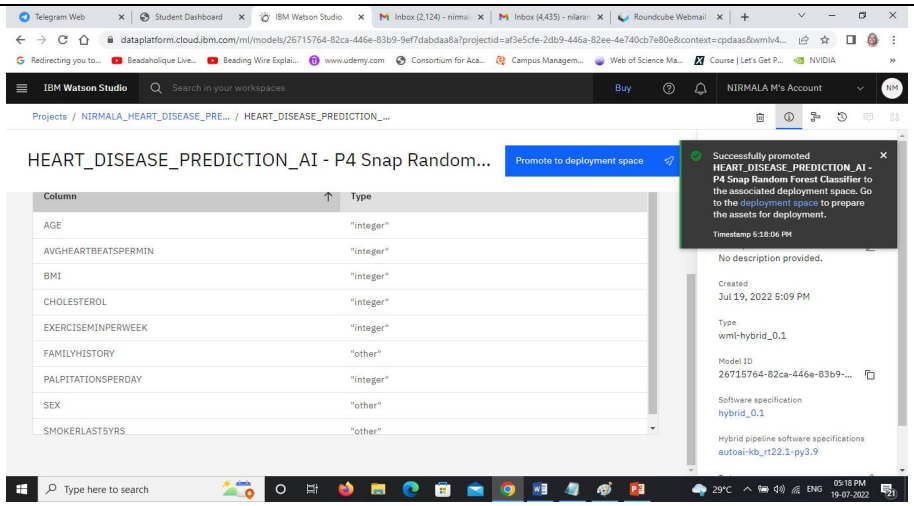
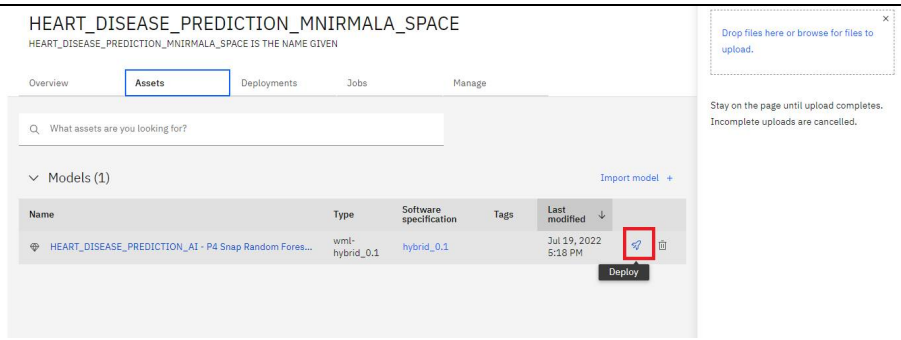
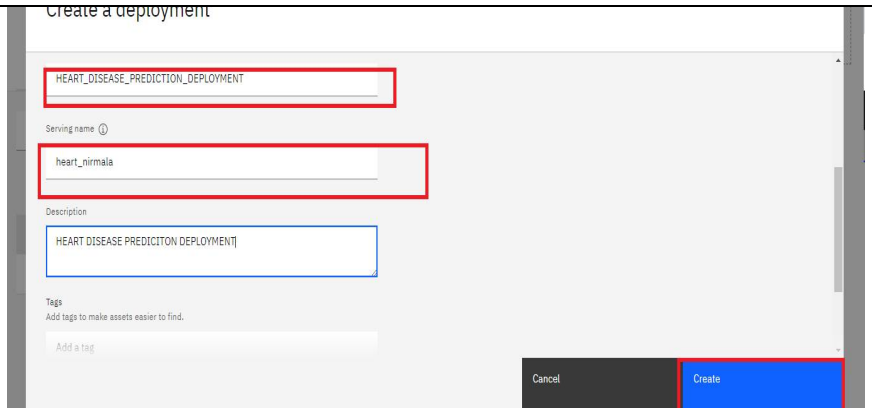
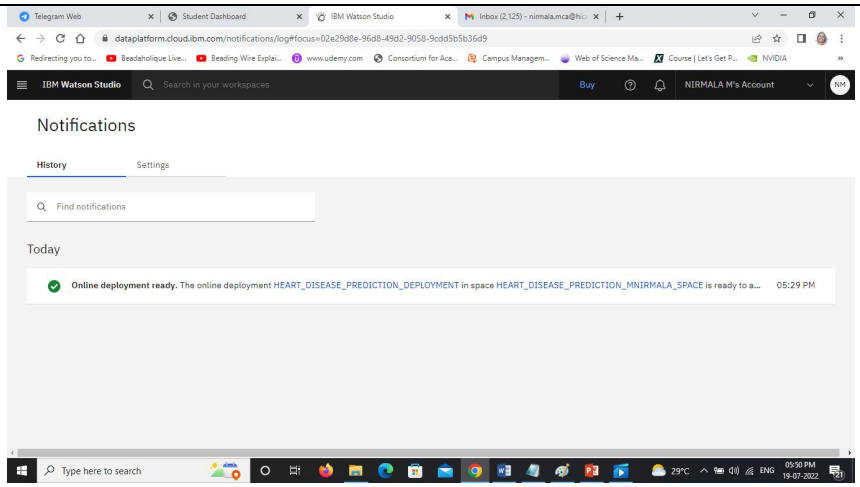
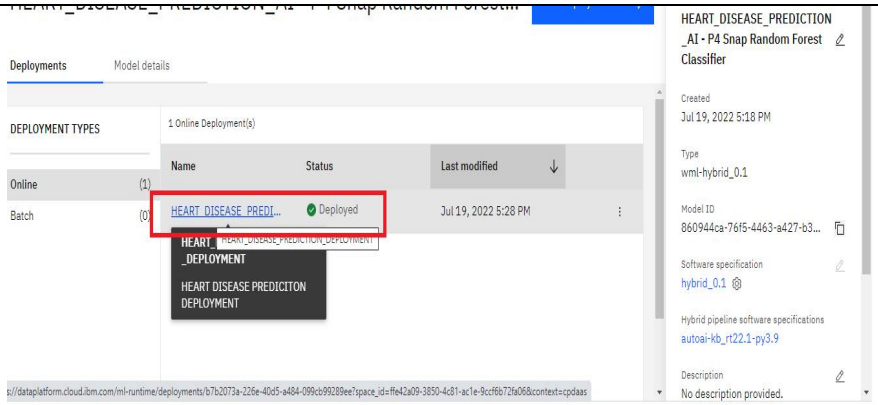
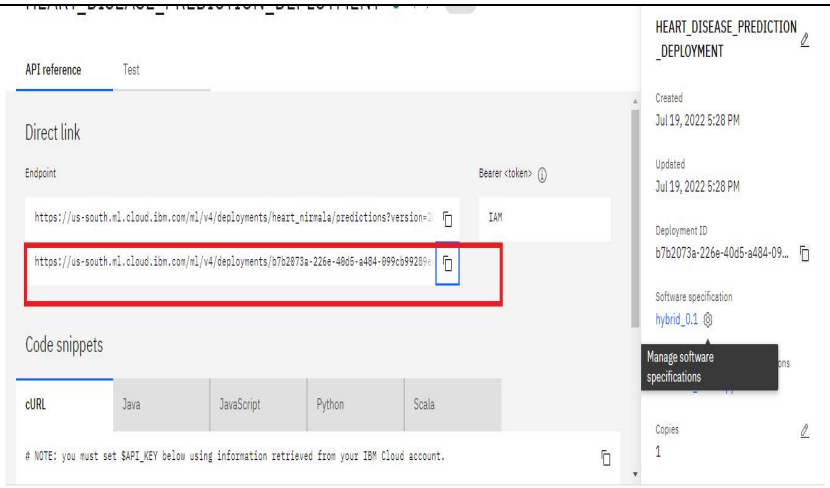
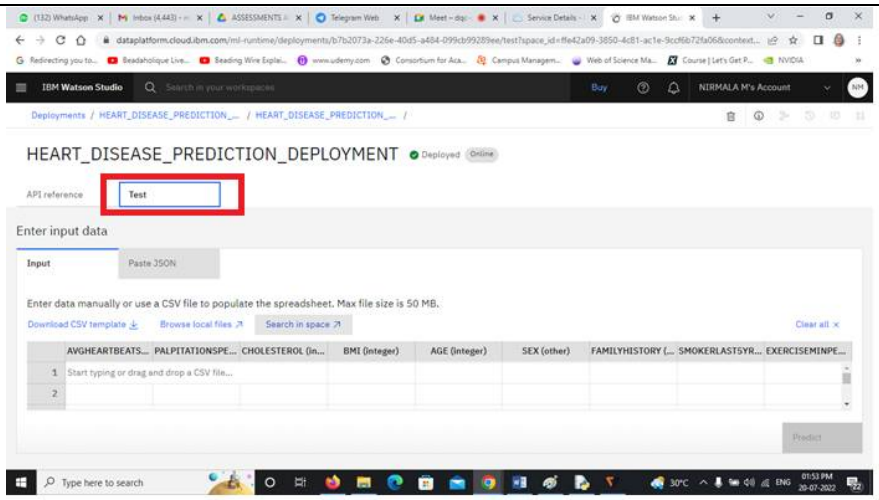
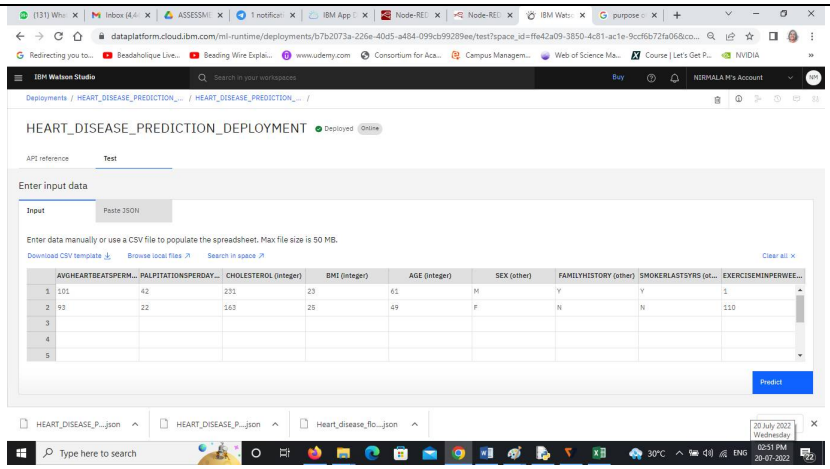
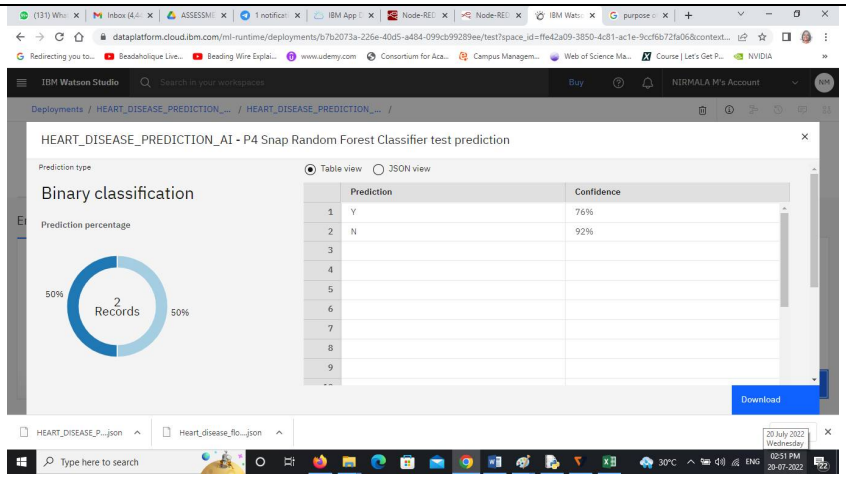


Figure 25 : Snap Random Forest Model Saved

Click on the Promote to Deployment Space	 Figure 26 : Promote to Deployment Space
New Deployment space is used to use a space to collect assets in one place to create ,run and manage deployments	 Figure 27 : New Deployment Space Creation
The new Deployment space called as HEART_DISEASE_PREDICTION_MNIRMALA_SPACE is created	 Figure 28 : Deployment Space Created
Promoting an version of an asset to a space creates a new asset in the space with a new Asset ID	 Figure 29 : Promote to Space

The HEART_DISEASE_PREDICTION_AI-P4 Snap Random Forest classifier is associated to the created deployment space called as HEART_DISEASE_PREDICTION_MNIRMALA_SPACE	 Figure 30 : Model Associated with Space
GOTO THE DEPLOYMENT_SPACE OR FROM THE NAVIGATION BAR, CHOOSE DEPLOYMENT SPACE	 Figure 31 : Deploy the Model
Click Create	 Figure 32 : Deployment Creation

The Online deployment HEART_DISEASE_PREDICTION_DEPLOYMENT in space HEART_DISEASE_PREDICTION_MNIRMALA_SPACE is ready to access	 Figure 33 : Online Deployment Space ready
Click on the Assets to goto the Created Model called as HEART_DISEASE_PREDICTION_AI-P4 Snap Random Forest classifier	 Figure 34 : Move to Created Model
Copy the end point	https://us-south.ml.cloud.ibm.com/ml/v4/deployments/heart_nirmala/predictions?version=2022-07-19
Endpoints specify where resources can be accessed by APIs and play a key role in guaranteeing the correct functioning of the software that interacts with it. In short, API performance relies on its ability to communicate effectively with API Endpoints.	 Figure 35 : Copy of End Points

Click on Test	 Figure 36 : Model Testing
Give the values for prediction and click predict	 Figure 37 : Entering Values in the Model
Output displayed is	 Figure 38 : Output of Model Testing

To create the node red service flow, click the Node red flow editor
Click on the menu bar on the right end and choose import and upload the provided json file

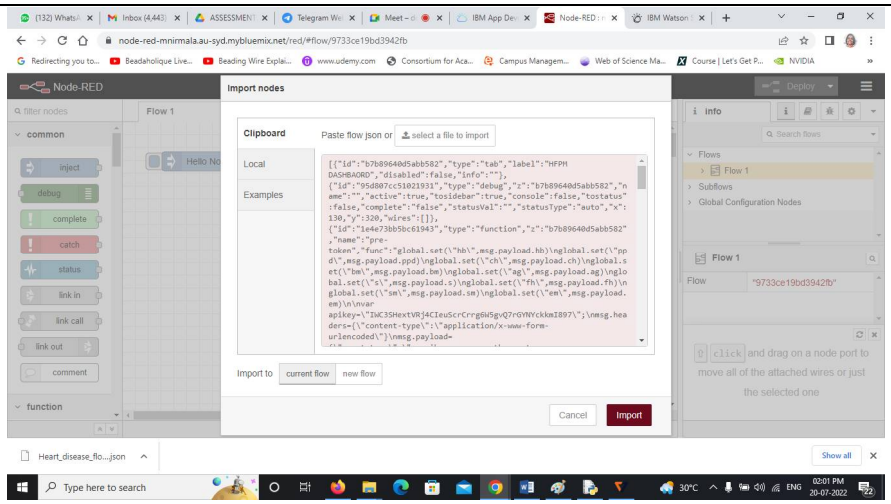


Figure 39 : Node Red Service Flow Screen

After the json file is imported to enable the hfpm dashboard flow install the dashboard-evi and press install for it to get enabled.

To do it choose palette and type dashboard-evi and install it

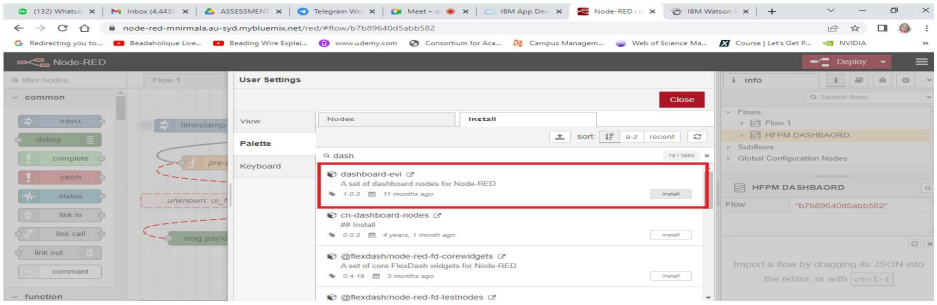


Figure 40 : Installation of dashboard

Output of dashboard

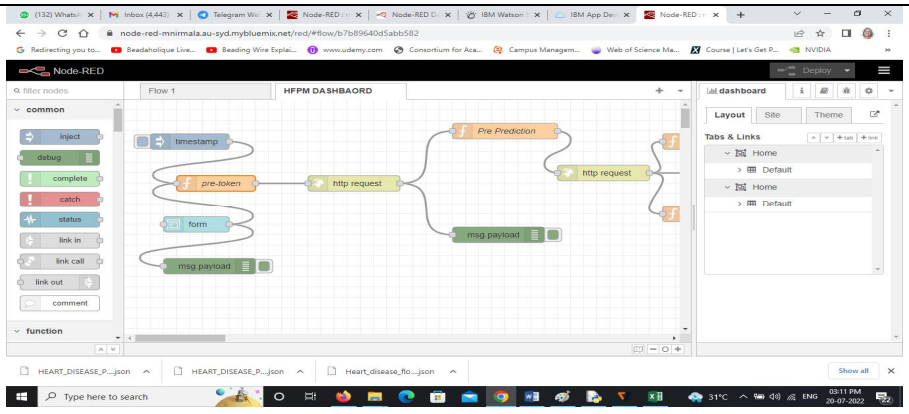
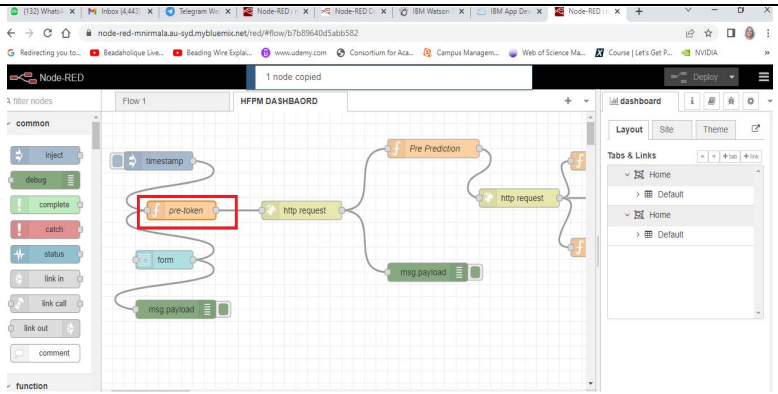
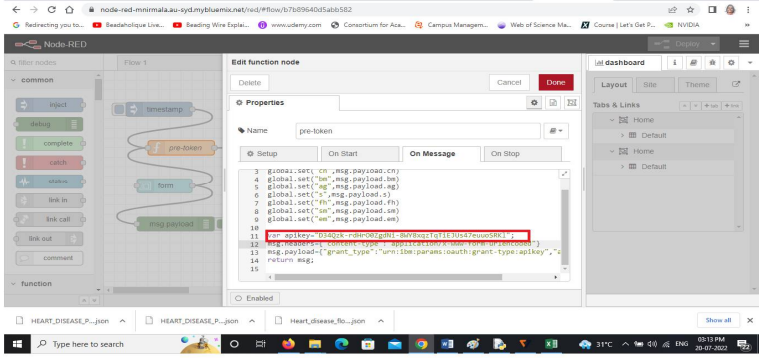
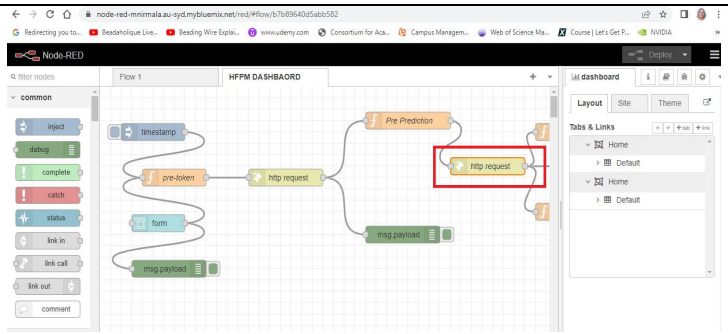
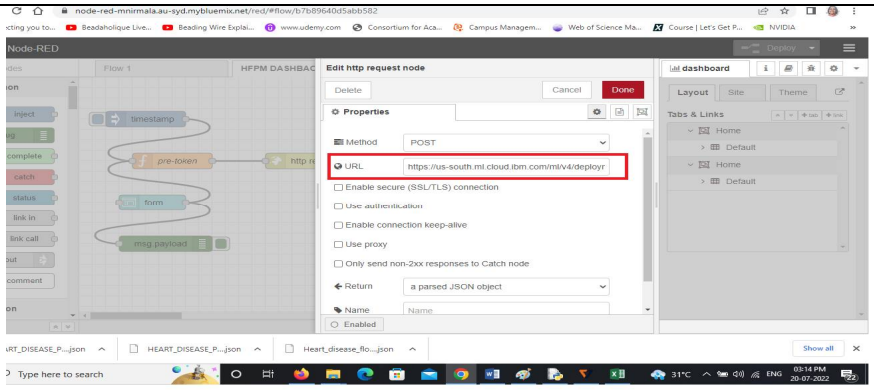
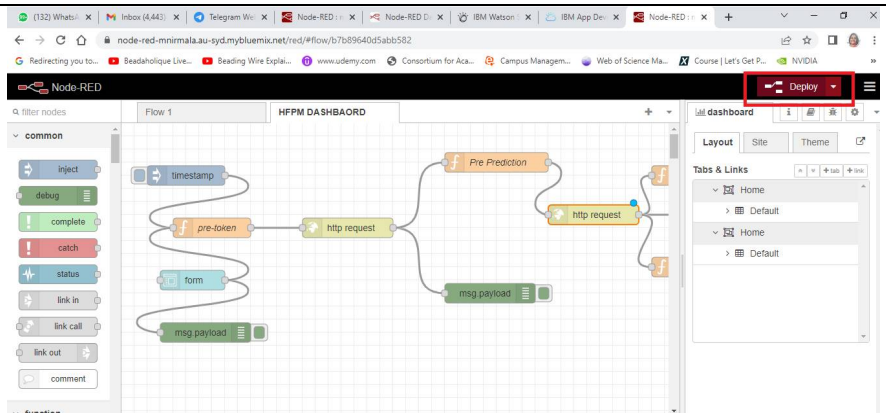
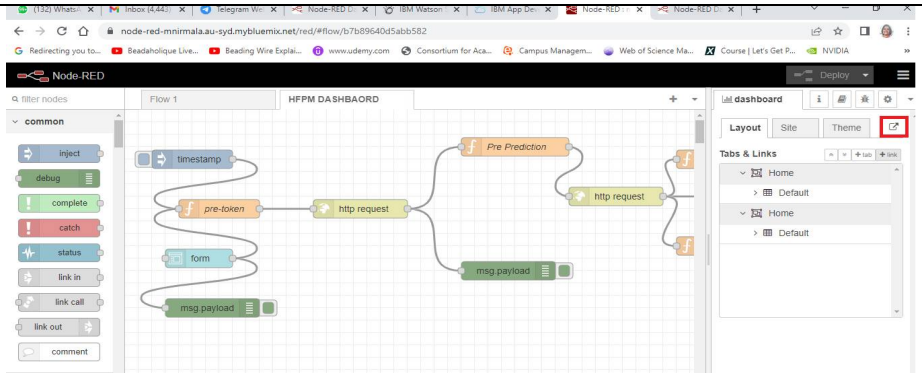
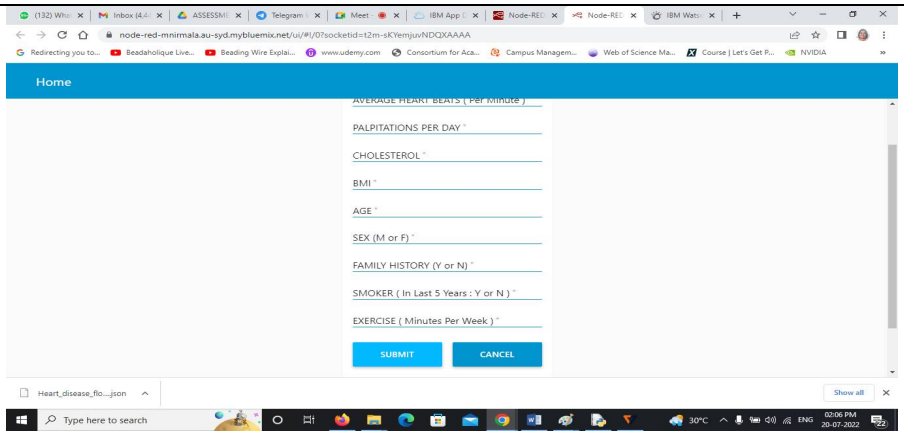
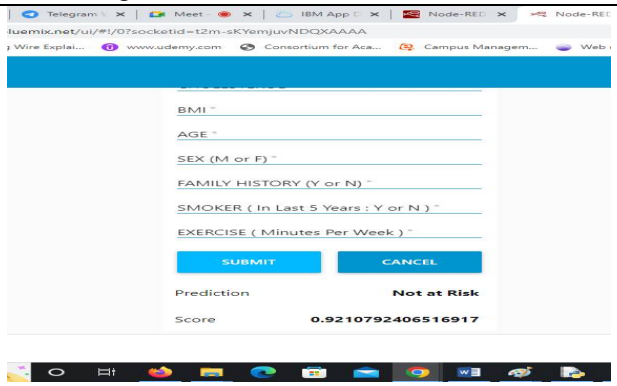


Figure 41 : Dashboard Output

Double click on the pretoken and paste the APIkey created and click done	
	Figure 42 : Pasting of API Key in Pretoken
Copy the API Key	
	Figure 43 : Pasted API Key in Pretoken
Click on the second http request and copy the Scoring URL in the URL Column	
	Figure 44 : Pasting of Scoring URL in HTTP
Copy this in the URL Column https://us-south.ml.cloud.ibm.com/ml/v4/deployments/b7b2073a-226e-40d5-a484-099cb99289ee/predictions?version=2022-07-19	
	Figure 45 : Scoring URL Pasted

Click on Deploy and after successfully deployed	 Figure 46 : Deployment Stage
Click the Link button on the left	 Figure 47 : Executing the Application
A window of the node red will be opened	 Figure 48 : Node Red Browser Window
Output of the Predicted result	 Figure 49 : Predicted Result Output1

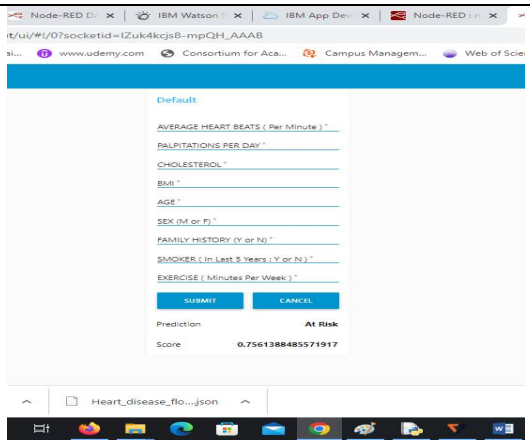
Output of the Predicted result The Application is Build using using Node-RED which takes inputs from the user and showcases the prediction on UI in the Browser window	

Figure 50 : Predicted Result Output2

V. EXPERIMENTAL RESULTS

The transformed data set is partitioned into training data set and the test data set where the training data is 80% of the whole data set and the remaining unused 20% is used as Test data set. The random state is set as 0. The parameters applied for various algorithms are depicted in

Table 5 : Experimented Results

Model Type	Accuracy
Random Forest	87%
Logistic Regression	79%

VI. CONCLUSION

The Machine learning methodology is rapidly increasing and the impact of the machine able to predict the result of a system by itself and also it is able to train a data over a period of time and also test the trained model with a different set of data to prove that the model is working efficiently and effectively. The problem is able to predict whether a person is at risk of heart failure. The data is clean and does not require any exploratory data analysis. The problem can be applied in a hospital environment to identify with the given parameters whether the person will have a heart disease or not. In this research study it has been apparently proved that Random Forest got the accuracy of 87% and Logistic Regression has got a accuracy of 79% and the model is working efficiently when connected with the node red application to display the output in the browser window.

VII. FUTURE SCOPE

The present study predicted whether patient is at risk of heart failure. The problem has to support with more parameters to be suitable for a generalized application. The web page creation can be further enhanced to make it a user friendly format.

REFERENCES

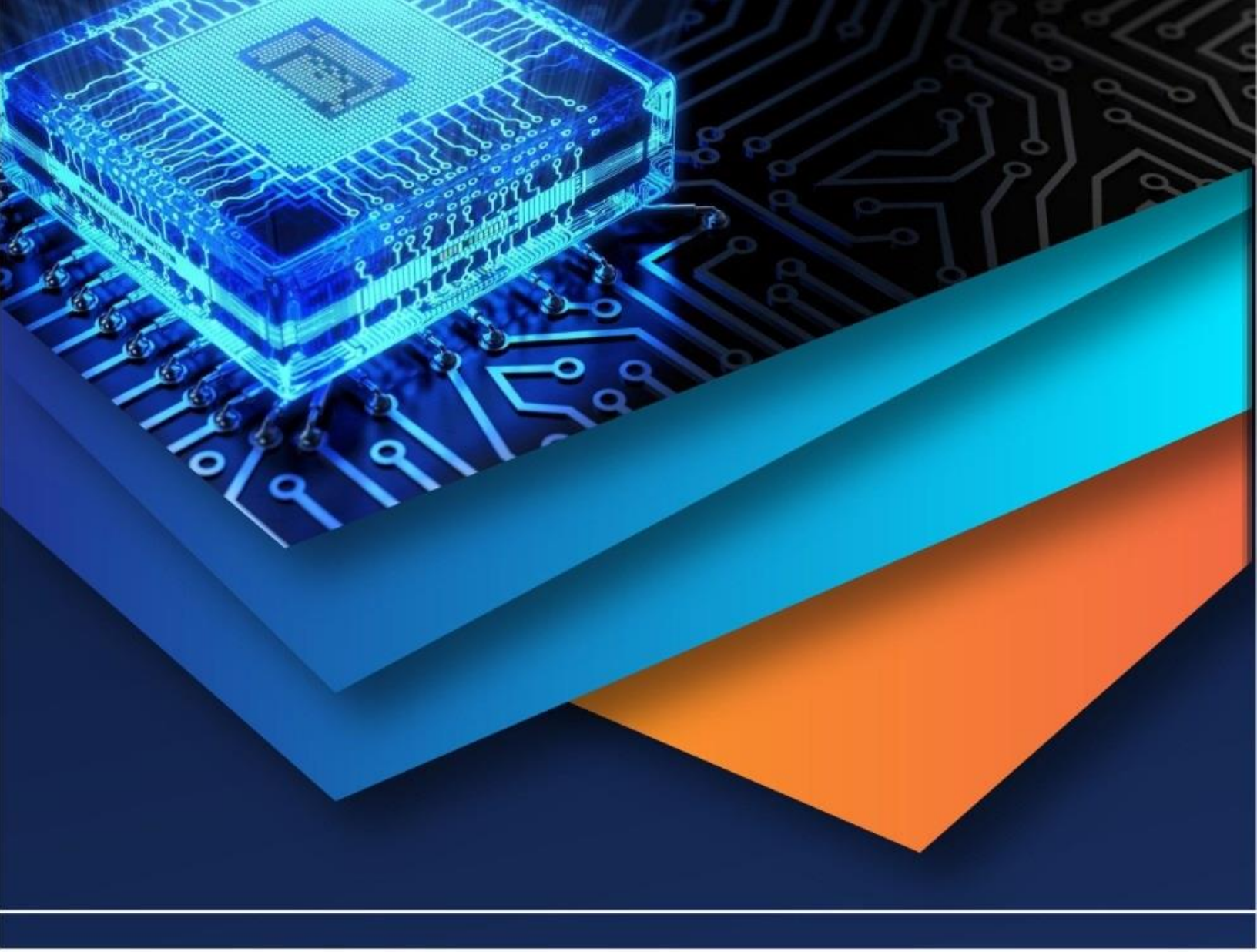
[1] Smola, Alex, and S.V.N. Vishwanathan. Introduction to Machine Learning. Cambridge University Press, 2008. N.p., 2008. Web.
[2] G. Lippi, B. M. Henry, and F. Sanchis-Gomar, “Physical inactivity and cardiovascular disease at the time of coronavirus disease 2019 (COVID-19),” European journal of preventive cardiology, vol. 27, no. 9, pp. 906–908, 2020.
[3] M. Pal and S. Parija, “Prediction of heart diseases using random forest,” in Journal of Physics: Conference Seriesvol. 1817, no. 1, IOP Publishing, Article ID 012009, 2021
[4] S.Krishnan J. Geetha S, "Forecast of Heart Disease Using Machine Learning Algorithms",First International Conference on Innovations in Information and Communication Technology ,2019.
[5] K.G.Dinesh, K A.garaj, K.D.Santhosh, V M.eswari. "Forecast of Cardiovascular Disease Utilizing Machine Learning Algorithms", 2018 International Conference on Current Trends towards Converging Technologies ,2018
[6] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava —Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019

[7] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019

[8] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar," Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019

[9] Marimuthu M, Abinaya A, 2018, A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach, International Journal of Computer Applications, Volume 181 – No. 18, pp20-25 .

[10] M. Balakrishnan, A.B. Arockia Christopher, P. Ramprakash and A. Logeswari, Prediction of Cardiovascular Disease using Machine Learning, Journal of Physics: Conference Series, Volume 1767, International E- Conference on Data Analytics, Intelligent Systems and Information Security & ICDIIS 2020 11-12 December 2020, Pollachi, India



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY