

Internship title : **RSIP Career Basic ML 100**
Project ID : **SPS_PRO_289**
Project Title : **Diabetic Mellitus Prediction using IBM Auto AI**
Team : **SB**

Project Description:

Diabetes mellitus is a chronic disease characterized by hyperglycemia. It may cause many complications. According to the growing morbidity in recent years, in 2040, the world's diabetic patients will reach 642 million, which means that one of the ten adults in the future is suffering from diabetes. There is no doubt that this alarming figure needs great attention. With the rapid development of machine learning, machine learning has been applied to many aspects of medical health for accurate predictions.

Solution:

This project prevents the people from the avalanche by priory informing them there is a chance to the occurrence of avalanche or not. The model gets the data from the IOT based sensors. After that we want to process those data using a suitable algorithm, then our model displays whether the avalanche occurs or not and how strength it was. To analyze the data coming from different sensors we are applying various machine learning algorithms. If there is a chance of avalanche then the notification will be sent to people so that they can take decisions accordingly and the model is being built in Auto AI.

ACKNOWLEDGEMENT

It's been an enjoyable journey since the very first day of my remote summer internship offered by SMART BRIDGE team. During these unprecedented times, when there was no hope left of getting a summer internship extraneously then I got to know about this valuable opportunity which they were offering. The team and their members have burned the midnight oil to make this reality happen and to enlighten so many young minds with remarkably worthy experience.

I am thankful to the whole Smart Bridge team which made us learn things that are worth it and have huge scope in future in the field of Artificial Intelligence, Machine learning, IoT and integrated model making using the advanced algorithms.

Being from a non-technical branch, I not only learned the concepts but have gained interests towards the development of those in a real life.

The application plays an important role while learning because they make us think substantially for a solution to any problem and this is what the arena of the mentors who have taught us was, while teaching.

Special thanks to Mr Surya Tej Sir (mentor to my machine learning course).

DATA-COLLECTION

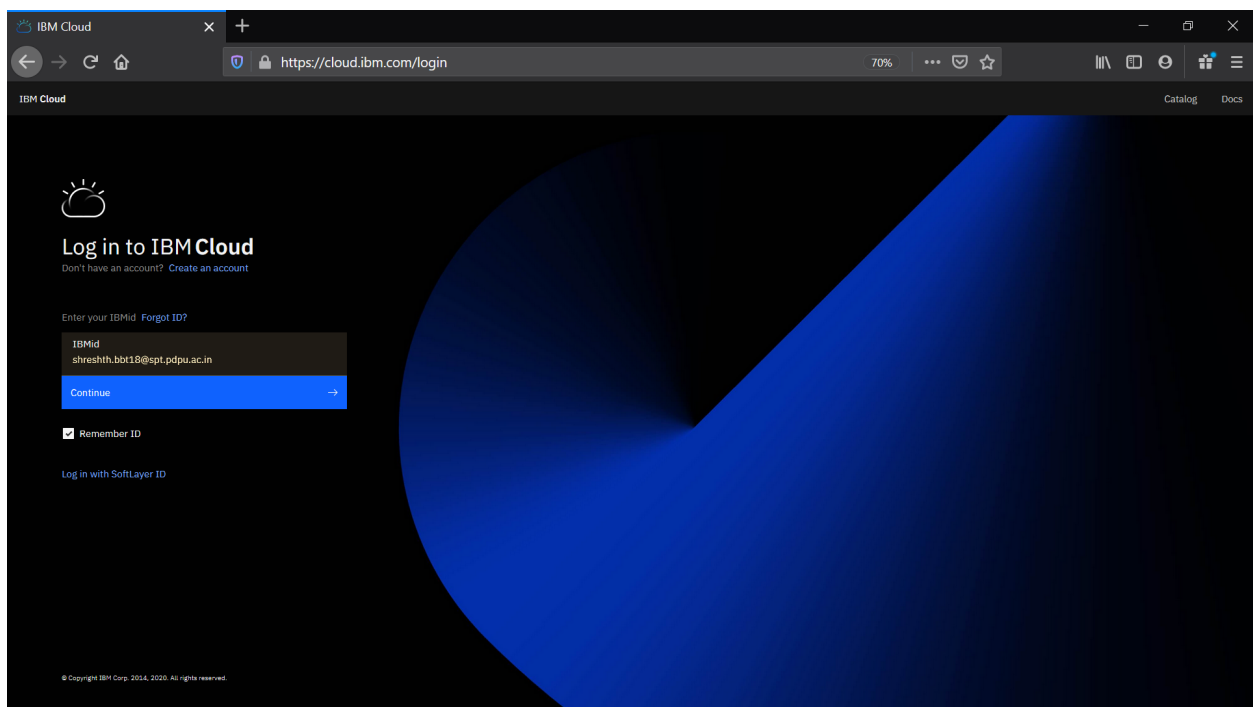
- the data was given as a list of features and I had to predict that whether or not person is a diabetic patient.
- it was given in the csv format in excel having 9 columns and 786 rows. rows were the values of the given 9 features wherein 9th feature was the one I had to predict about.
- 8 features which are the part of deployment are : preg, plas, pres, skin, test, mass, pedi, age
- 1 feature which tells that whether person is diabetic or not. I had to classify here with 1 and 0 as former signifies that the person is a patient of diabetic mellitus and the later shows that person is free from diabetic mellitus.
- [DATA SET LINK](#) : here I had to proceed with the dataset. link redirected me to the kaggle service from where data set file was present .

IBM CLOUD ACCOUNT

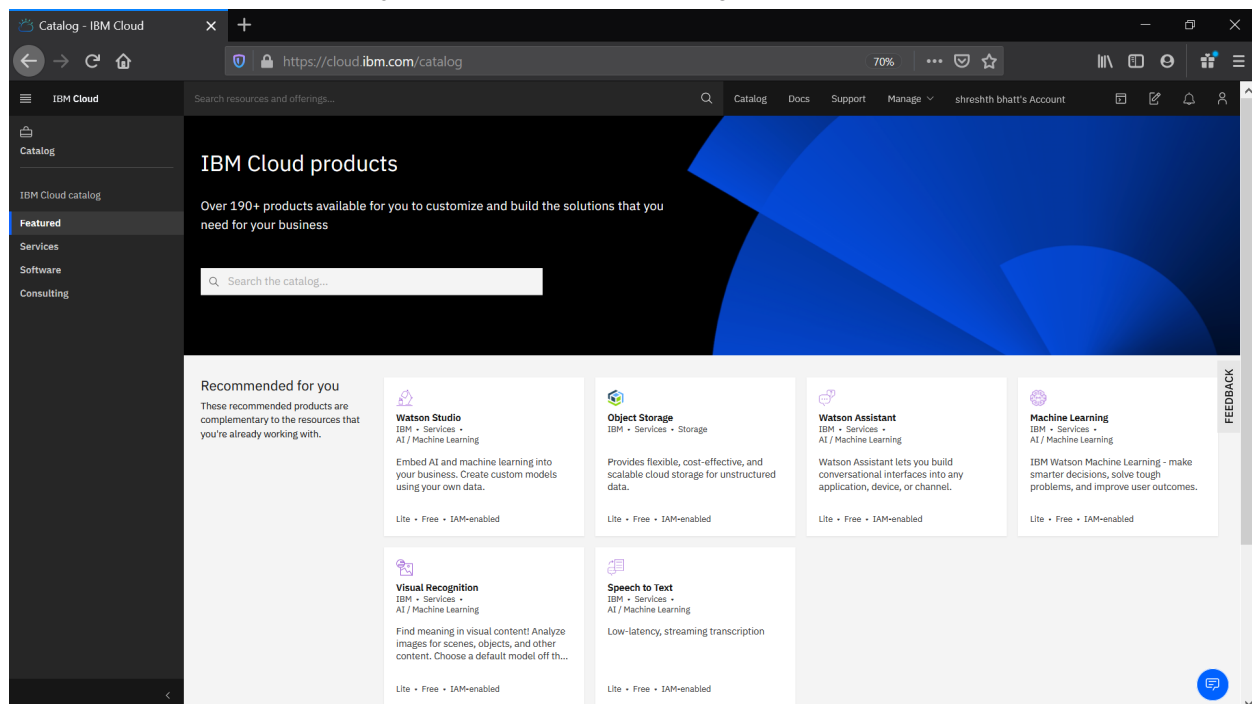
To get an access to IBM Cloud Services , one needs to create an account on it .

here, I will conclude the steps from creating IBM cloud account till Machine Learning Service with the help of images in a proceeding order.

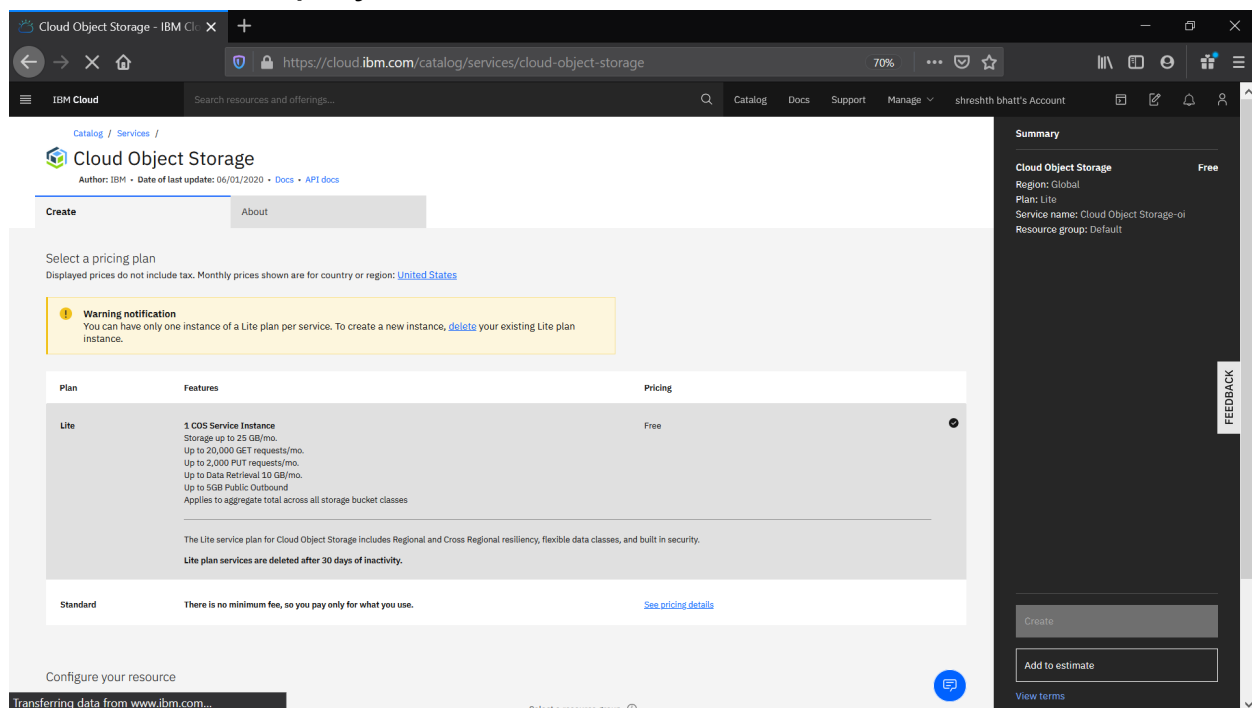
STEP-1: Create an IBM Cloud account and log in into it using certain valid credentials.



STEP-2: In the Header section, Catalog Tab is present to create resources necessary for making a project.



STEP-3: on the same page, create CLOUD STORAGE for storing all the data of the project.



STEP-4: Now, create Watson Studio to keep the projects in it.

The screenshot shows the IBM Cloud Watson Studio service page. The browser address bar displays <https://cloud.ibm.com/catalog/services/watson-studio>. The page header includes the IBM Cloud logo, a search bar, and navigation links for Catalog, Docs, Support, and Manage. The user's account name, 'shreshth bhatt's Account', is visible in the top right.

The main content area is titled 'Watson Studio' and includes the text 'Author: IBM • Date of last update: 06/12/2020 • Docs'. Below this, there are two tabs: 'Create' (selected) and 'About'. The 'Create' tab contains a 'Select a region' dropdown menu with 'London' selected, and a 'Select a pricing plan' section. A warning notification states: 'Warning notification: You can have only one instance of a Lite plan per service. To create a new instance, delete your existing Lite plan instance.'

Plan	Features	Pricing
Lite	<ul style="list-style-type: none">1 authorized user50 capacity unit-hours monthly limitEnvironment = a of capacity units required per hour<ul style="list-style-type: none">• 1 vCPU + 4 GB RAM = 0.5• 2 vCPU + 8 GB RAM = 1• 4 vCPU + 16 GB RAM = 2Decision Optimization = Environment + 5	Free

Below the table, a note states: 'The Lite plan for Watson Studio offers everything you need to become a better data scientist or domain expert in a collaborative environment. Lite plan services are deleted after 30 days of inactivity.'

The right sidebar shows a 'Summary' section with the following details: 'Watson Studio', 'Region: London', 'Plan: Lite', 'Service name: Watson Studio-45', and 'Resource group: Default'. At the bottom of the sidebar, there are buttons for 'Create', 'Add to estimate', and 'View terms'.

STEP-5: now, according to the services need, install the required one, here I chosen Machine Learning Services.

The screenshot shows the IBM Cloud Machine Learning service page. The browser address bar displays <https://cloud.ibm.com/catalog/services/machine-learning>. The page header includes the IBM Cloud logo, a search bar, and navigation links for Catalog, Docs, Support, and Manage. The user's account name, 'shreshth bhatt's Account', is visible in the top right.

The main content area is titled 'Machine Learning' and includes the text 'Author: IBM • Date of last update: 04/09/2020 • Docs'. Below this, there are two tabs: 'Create' (selected) and 'About'. The 'Create' tab contains a 'Select a region' dropdown menu with 'London' selected, and a 'Select a pricing plan' section. A warning notification states: 'Warning notification: You can have only one instance of a Lite plan per service. To create a new instance, delete your existing Lite plan instance.'

Plan	Features	Pricing
Lite	<ul style="list-style-type: none">Service Instance5 model deployments5,000 predictions50 capacity unit-hours (CUH) included:Capacity Type:<ul style="list-style-type: none">• 1 (one) NVIDIA K80 GPU = 3 capacity units required per hour• 1 (one) NVIDIA V100 GPU = 10 capacity units required per hour• 1 vCPU and 4 GB RAM = 0.5 capacity units required per hour• 2 vCPU and 8 GB RAM = 1 capacity units required per hour• 4 vCPU and 16 GB RAM = 2 capacity units required per hour• 8 vCPU and 32 GB RAM = 4 capacity units required per hour• 16 vCPU and 64 GB RAM = 8 capacity units required per hourAuto AT<ul style="list-style-type: none">• 8 vCPU and 32 GB RAM = 20 capacity units required per hour	Free

The right sidebar shows a 'Summary' section with the following details: 'Machine Learning', 'Region: London', 'Plan: Lite', 'Service name: Machine Learning-q4', and 'Resource group: Default'. At the bottom of the sidebar, there are buttons for 'Create', 'Add to estimate', and 'View terms'.

MODEL BUILDING:

Now comes the most important part of the project : to build the model using installed services in our watson studio having certain instance name.

I have chosen the most optimised model through accuracy score and then predicted about that which model could give the more appropriate values.

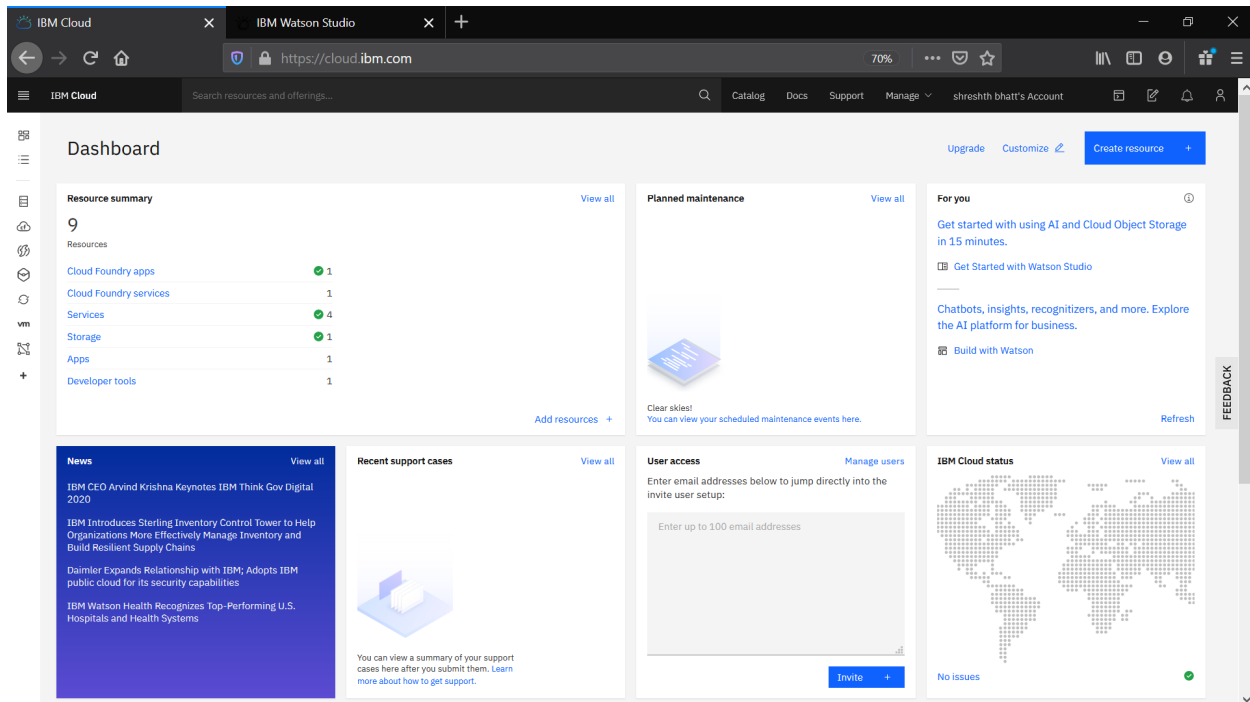
And the chosen model was deployed to give the outputs.
here , we can test, implement and see the result in our watson studio machine learning service.

The model was not build manually but with the help of Auto AI experiment.

The model with highest accuracy score was selected and deployed.

Given below were the steps I took to build a AUTO AI MODEL:

STEP-1: Click on the IBM CLOUD Tab given in the top left corner of the window next to navigation button.



STEP-2: On the Dashboard window, click on the services :

Resource list

Name	Group	Location	Offering	Status	Tags
Continuous Delivery	Default	Dallas	Continuous Delivery	Active	-
Watson Studio-ht	Default	London	Watson Studio	Active	-
node-red-qitcy-cloudant-1593697028353	Default	London	Cloudant	Active	-
pm-20-wf	Default	London	Machine Learning	Active	cpda...

STEP-3: click on Watson Studio to proceed further as we have to add Auto AI experiment to our project which is already present in Watson Studio.

Watson Studio-ht

Get Started

Documentation

Community

STEP-4: choose the project name by hovering the cursor on the instance name of the project(diabetic_mellitus)

The screenshot shows the IBM Watson Studio web interface. At the top, there's a navigation bar with the IBM Watson Studio logo and a user profile for 'shreshth bhatt'. Below the navigation bar, a large banner says 'Welcome shreshth!' with the subtitle 'Watson Studio • Watson Machine Learning'. A callout box titled 'Start by creating a project' explains that a project is used to organize resources. Below this, a 'Create a project' button is visible. The main content area is titled 'Recently updated projects' and features a table with one project listed: 'diabetic_mellitus'. The table has columns for Name, Role, Collaborators, Date created, and Last updated. The project 'diabetic_mellitus' is listed with the role 'Admin', one collaborator, and was created and last updated on 'Jul 02, 2020'. A 'New project +' link is to the right of the table. Below the table, there's a section for 'Watson services' with a 'View all (1)' link and an 'Add service +' link.

Service Details - IBM Cloud x IBM Watson Studio x IBM Watson Studio +

https://eu-gb.dataplatform.cloud.ibm.com/home?context=wdp&apps=data_science_experience 80%

IBM Watson Studio Upgrade shreshth bhatt's Account SB

Welcome shreshth!

Watson Studio • Watson Machine Learning

Start by creating a project
A project is how you organize your resources to work with data and collaborate with team members.

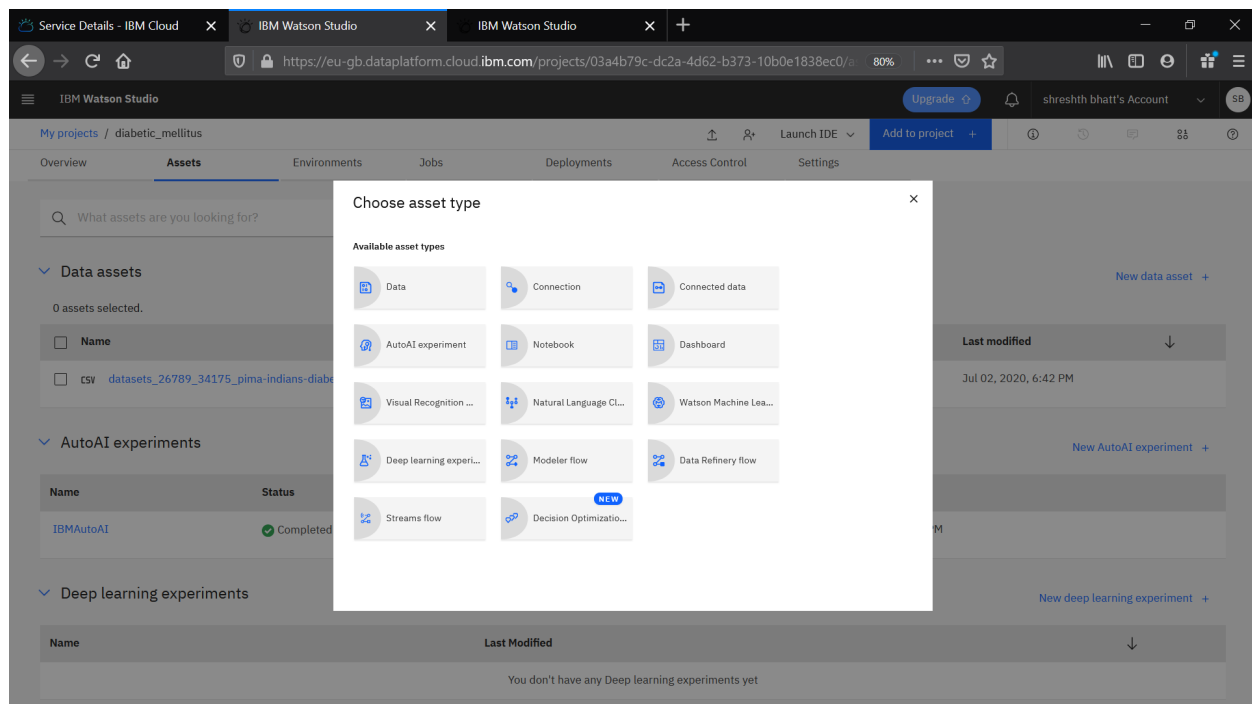
Create a project
Create a project, and then add the tools and assets you need.

Recently updated projects [View all \(1\)](#) [New project +](#)

Name	Role	Collaborators	Date created	Last updated
diabetic_mellitus	Admin	1	Jul 02, 2020	Jul 02, 2020

Watson services [View all \(1\)](#) [Add service +](#)

STEP-5: At the top right corner select the Add to Project option and choose AUTO AI Experiment.



STEP-6: Fill the following credentials and run the AI experiment.

The screenshot shows the 'New AutoAI experiment' form in the IBM Watson Studio interface. The form is divided into two main sections: 'Define details' and 'Associate services'.

Define details

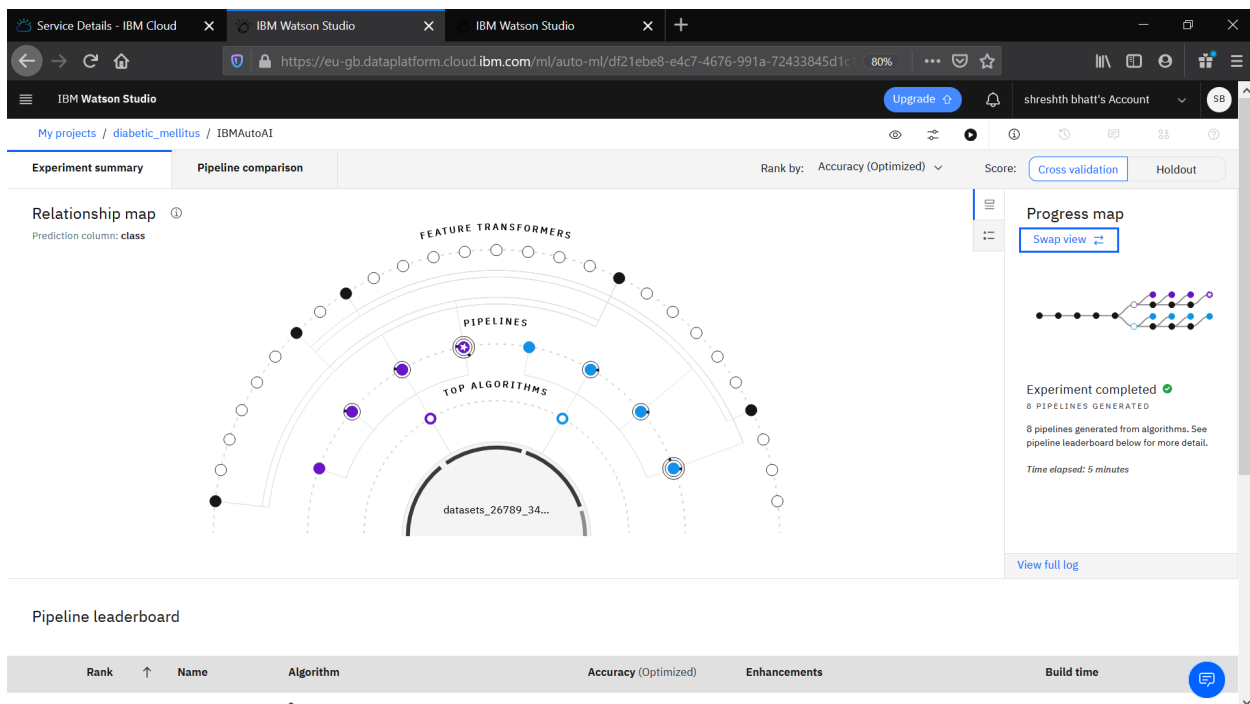
- From blank** (selected) / From sample
- Name ***: Name of AutoAI experiment
- Description**: Description of AutoAI experiment

Associate services

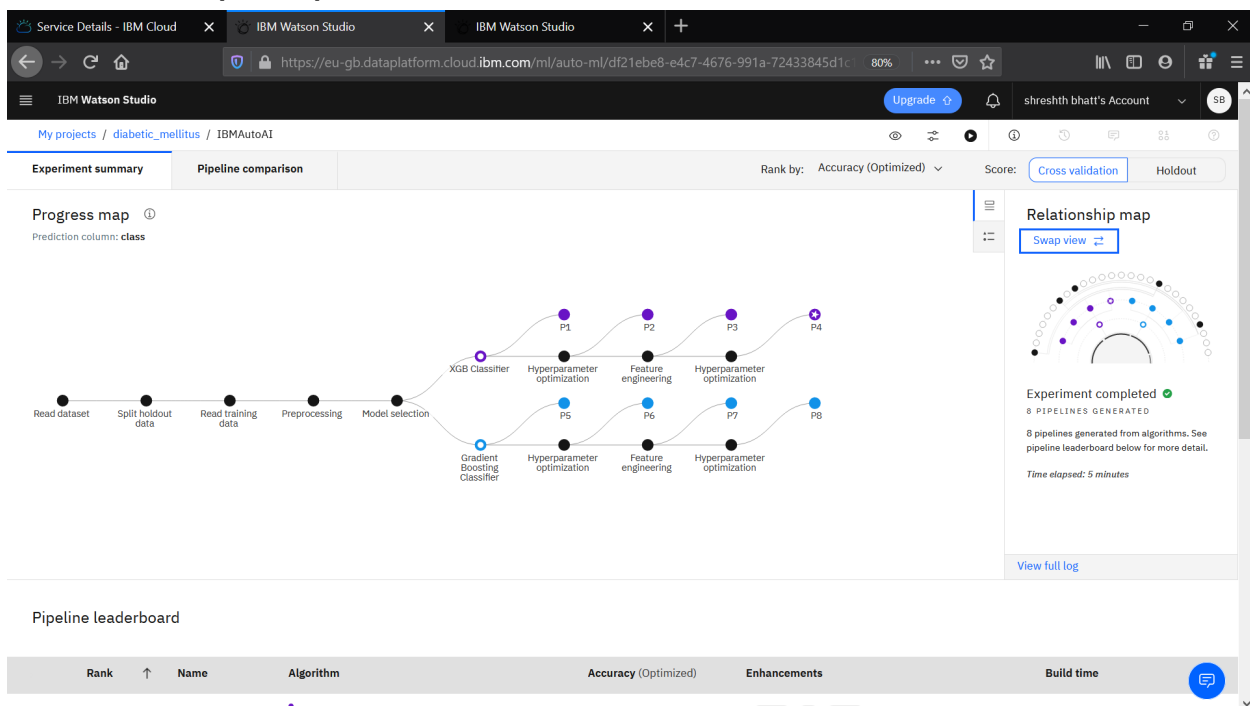
- Watson Machine Learning Service Instance ***: pm-20-wf
- Compute configuration ***: 8 vCPU and 32 GB RAM
- Notes**: This compute configuration consumes 20 capacity units per hour. [Learn more](#) about capacity unit hours and Watson Machine Learning pricing plans.

At the bottom right, there are two buttons: 'Cancel' and 'Create'.

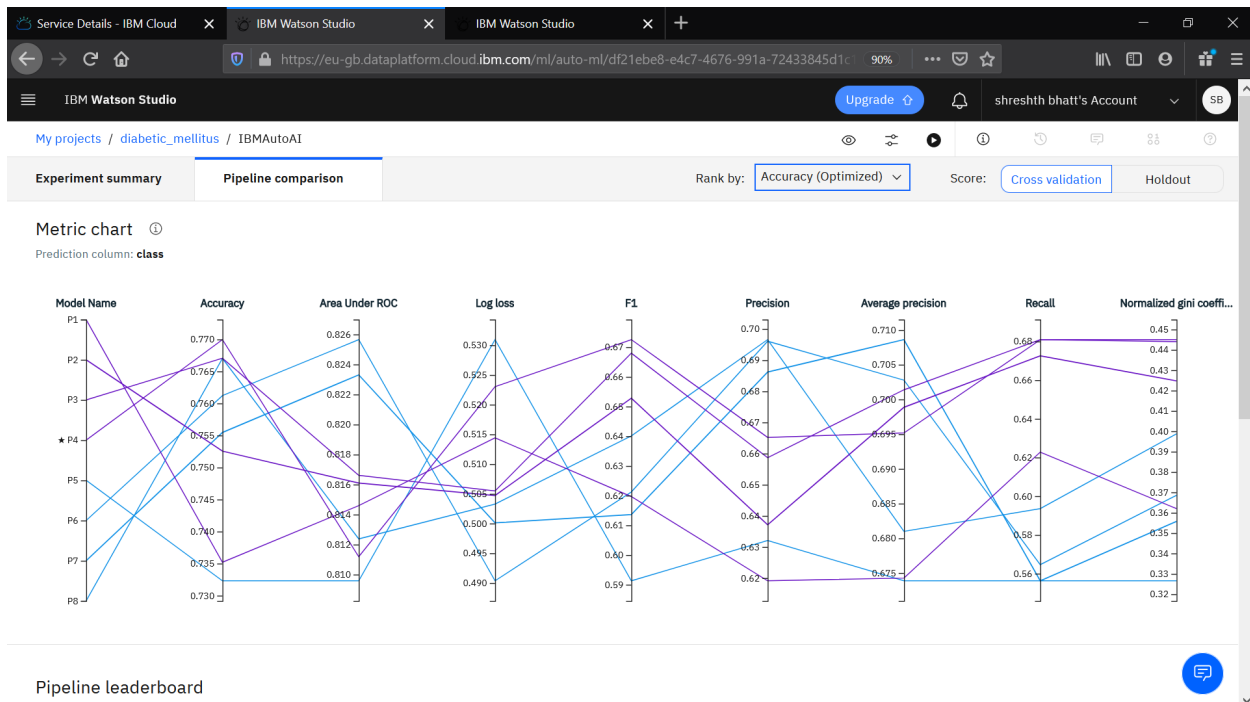
Now, lets look at Experiment Summary: Progress Map:



Relationship Map:



Pipeline comparison(Metric Chart):



Pipeline Leaderboard:

Rank by: **Accuracy (Optimized)** | Score: **Cross validation** | Holdout

Pipeline leaderboard

Rank	Name	Algorithm	Accuracy (Optimized)	Average pre...	F1	Log loss	Normalized ...	Precision	Recall	ROCAU
★ 1	Pipeline 4	XGB Classifier	0.770	0.695	0.673	0.523	0.444	0.665	0.681	0.811
2	Pipeline 3	XGB Classifier	0.767	0.701	0.668	0.505	0.445	0.659	0.681	0.817
3	Pipeline 8	Gradient Boosting Classifier	0.767	0.681	0.640	0.503	0.399	0.697	0.594	0.812
4	Pipeline 6	Gradient Boosting Classifier	0.761	0.703	0.621	0.490	0.369	0.696	0.565	0.826
5	Pipeline 7	Gradient Boosting Classifier	0.755	0.709	0.614	0.500	0.356	0.686	0.556	0.823
6	Pipeline 2	XGB Classifier	0.753	0.699	0.653	0.505	0.425	0.637	0.673	0.816
7	Pipeline 1	XGB Classifier	0.735	0.674	0.620	0.514	0.362	0.619	0.623	0.815
8	Pipeline 5	Gradient Boosting Classifier	0.732	0.674	0.591	0.531	0.327	0.632	0.556	0.810

How many pipelines are generated in auto AI experiment in IBM?

By default, **AutoAI** will choose the top two performing algorithms of the ones it considers, and use that algorithm **to generate 8 pipelines** that you **can** view and compare, but you **can** change the number from 1 **to** 4.

What is scoring in machine learning?

Scoring is also called prediction, and is the process of generating values based on a trained **machine learning** model, given some new input data. The values or **scores** that are created can represent predictions of future values, but they might also represent a likely category or outcome.

Best Accuracy Score(optimised):	Pipeline 4
Algorithm used	: XGB Classifier
Accuracy score	: 0.770

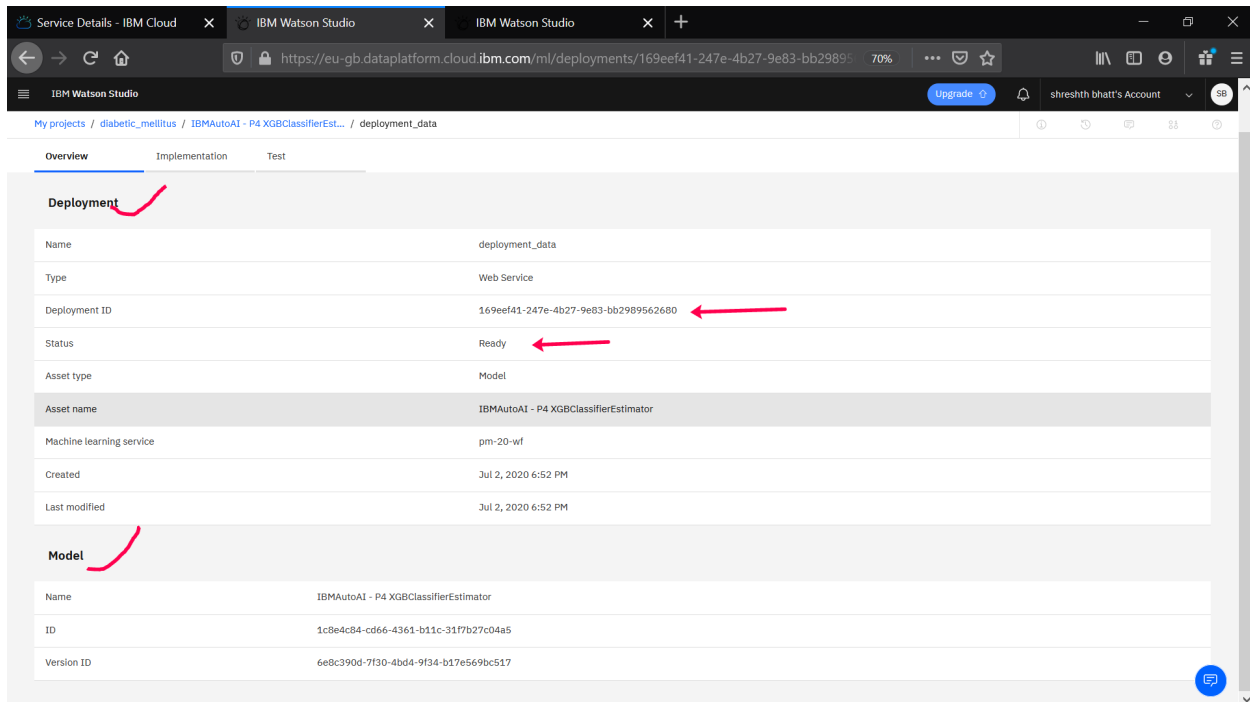
Now comes the Deployment Part :

The screenshot shows the IBM Watson Studio web interface. The browser tabs at the top include 'Service Details - IBM Cloud', 'IBM Watson Studio', and another 'IBM Watson Studio' tab. The address bar shows the URL: <https://eu-gb.dataplatform.cloud.ibm.com/projects/03a4b79c-dc2a-4d62-b373-10b0e1838ec0/>. The page header includes 'IBM Watson Studio', an 'Upgrade' button, and the user's account 'shreshth bhatt's Account'. The main navigation bar has tabs for 'Overview', 'Assets', 'Environments', 'Jobs', 'Deployments' (which is selected and highlighted with an orange arrow), 'Access Control', and 'Settings'. Below the navigation bar, the 'Deployments' section is displayed. It contains a table with the following data:

Name	Type	Status	Actions
deployment_data (highlighted with a red arrow)	Web Service	Ready	⋮

A blue chat icon is visible in the bottom right corner of the interface.

Here, as one can see that we ready to test our deployed model and all the credentials about the deployment model are given.



The screenshot shows the IBM Watson Studio interface. The browser address bar displays the URL: <https://eu-gb.dataplatform.cloud.ibm.com/ml/deployments/169eef41-247e-4b27-9e83-bb2989562680>. The page title is "Service Details - IBM Cloud". The breadcrumb navigation shows: "My projects / diabetic_mellitus / IBMAutoAI - P4 XGBClassifierEst... / deployment_data". The "Overview" tab is selected, showing the "Deployment" section. A red checkmark is next to the "Deployment" heading. The deployment details are as follows:

Property	Value
Name	deployment_data
Type	Web Service
Deployment ID	169eef41-247e-4b27-9e83-bb2989562680
Status	Ready
Asset type	Model
Asset name	IBMAutoAI - P4 XGBClassifierEstimator
Machine learning service	pm-20-wf
Created	Jul 2, 2020 6:52 PM
Last modified	Jul 2, 2020 6:52 PM

Below the deployment details is the "Model" section, also marked with a red checkmark. The model details are as follows:

Property	Value
Name	IBMAutoAI - P4 XGBClassifierEstimator
ID	1c8e4c84-cd66-4361-b11c-31f7b27c04a5
Version ID	6e8c390d-7f30-4bd4-9f34-b17e569bc517

Red arrows point to the "Deployment ID" and "Status" fields in the deployment table.

Here, under Implementation, useful information like Scoring End point URL, token, ML instance ID, payload scoring are given which are needed while creating Nodered UI application:

The screenshot shows the IBM Watson Studio interface for a deployment named 'deployment_data'. The 'Implementation' tab is active, displaying the following details:

Property	Value
Scoring End-point	https://eu-gb.ml.cloud.ibm.com/v4/deployments/169eef41-247e-4b27-9e83-bb2989562680/predictions
Authorization: Bearer <token>	Review the WML authentication documentation for details about generating IAM tokens.
ML-Instance-ID	The "ML-Instance-ID" HTTP header must be populated with the WML Instance Id, which can be obtained as described here
Content-type: application/json	Required if the request body is sent in JSON format.

Below the implementation details, there is a 'Code Snippets' section with tabs for cURL, Java, JavaScript, Python, and Scala. The Python tab is selected, showing the following code:

```
import urllib3, requests, json

# NOTE: generate iam_token and retrieve ml_instance_id based on provided documentation
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + iam_token, 'ML-Instance-ID': ml_instance_id}

# NOTE: manually define and pass the array(s) of values to be scored in the next line
payload_scoring = {'input_data': [{'fields': ['preg', 'plac', 'pres', 'skin', 'test', 'mass', 'ped1', 'age'], 'values': [array_of_values_to_be_scored, another_array_of_values_to_be_scored]}]}

response_scoring = requests.post('https://eu-gb.ml.cloud.ibm.com/v4/deployments/169eef41-247e-4b27-9e83-bb2989562680/predictions', json=payload_scoring, headers=header)
print("Scoring response")
print(json.loads(response_scoring.text))
```

Zoom the image to get the look at how predicted model gave the correct output when input was given as same as from the original dataset

The screenshot shows the IBM Watson Studio interface for a deployment named 'deployment_data'. The 'Test' tab is selected, indicated by a red arrow. On the left, the 'Enter input data' section contains a 'Predict' button and input fields for 'mass' (33.6), 'pedi' (0.627), and 'age' (50). The 'mass' field is highlighted. On the right, the output is displayed as a JSON object. A red arrow points to the 'prediction' value, which is '1', labeled 'output'. The JSON structure is as follows:

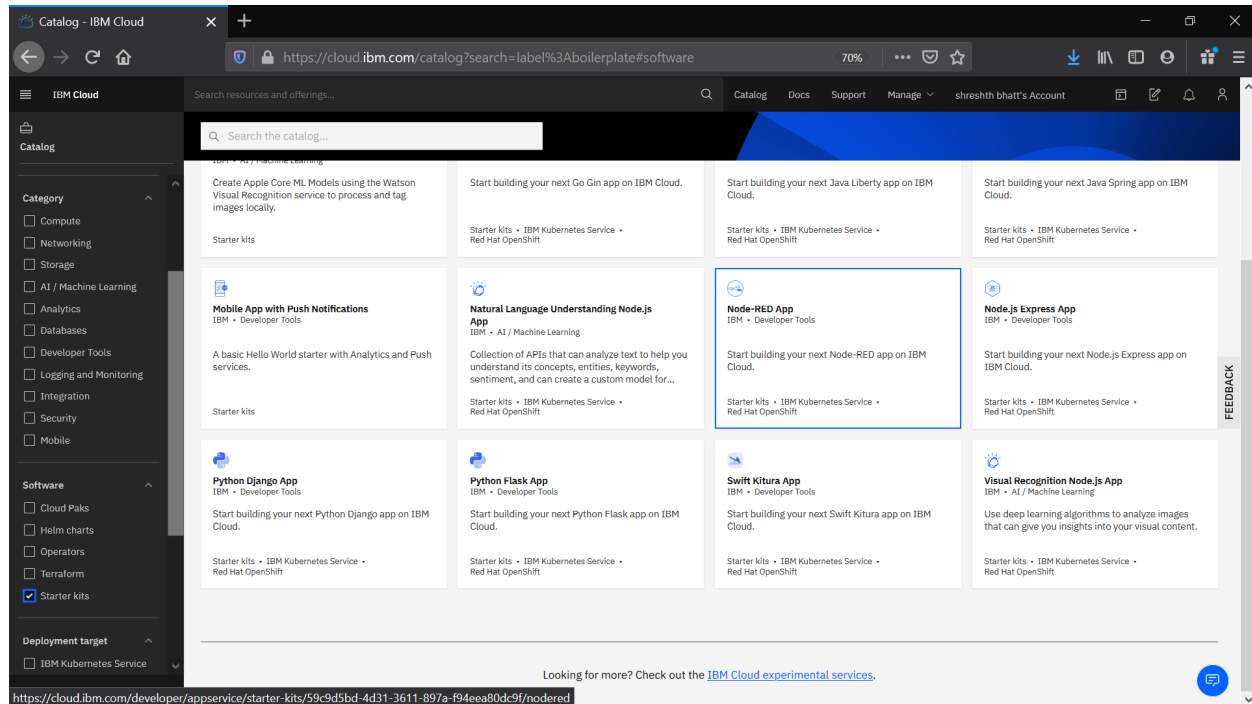
```
{
  "predictions": [
    {
      "fields": [
        "prediction",
        "probability"
      ],
      "values": [
        1,
        [
          0.1366117000579834,
          0.8633882999420166
        ]
      ]
    }
  ]
}
```

values from original dataset were provided in "enter input data" section and correct output was displayed according to the data set along with two other values: prediction & probability.

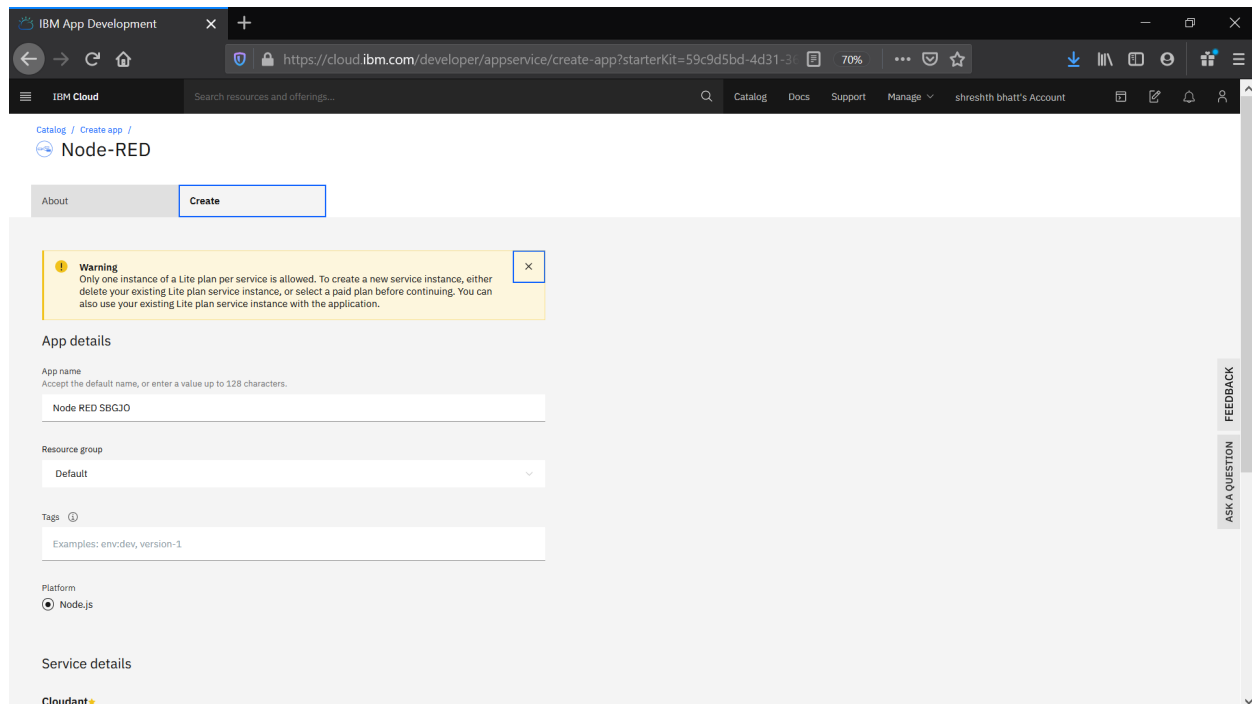
And finally this is how I have created model using Auto AI experiment and output can be easily tested in the watson studio itself.

APPLICATION BUILDING

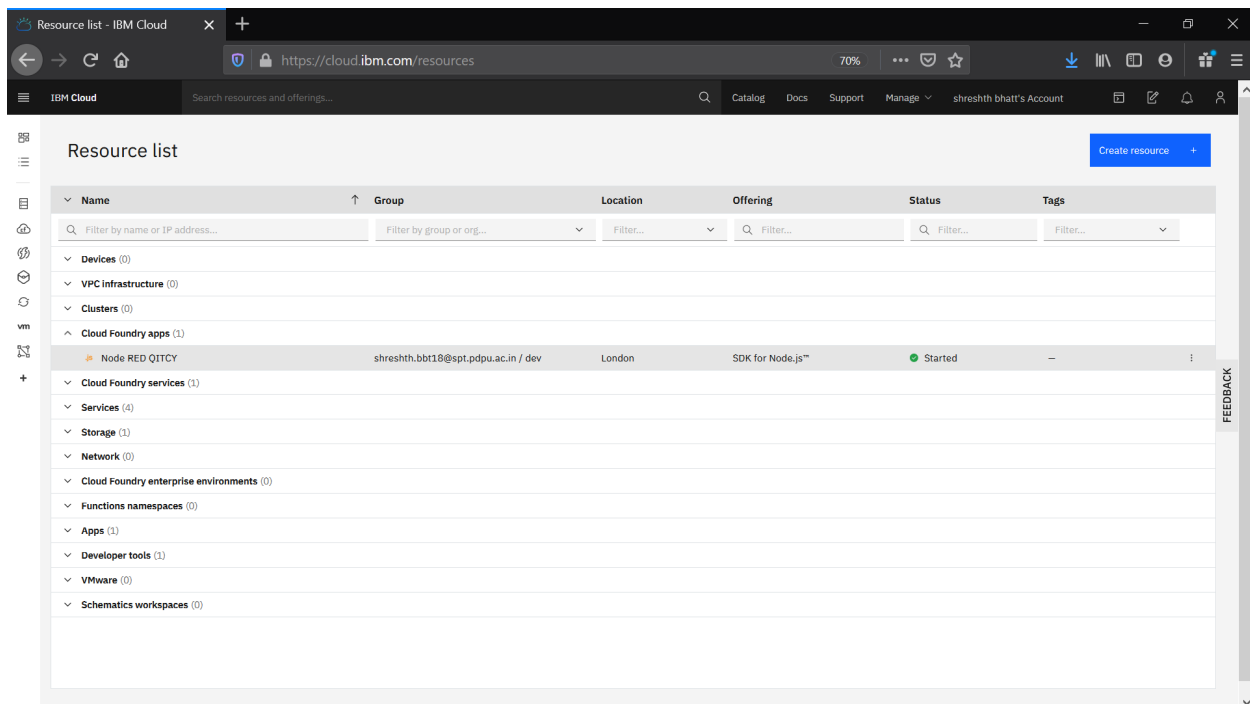
Step 1: Go to the catalog tab once again and install NODERED App



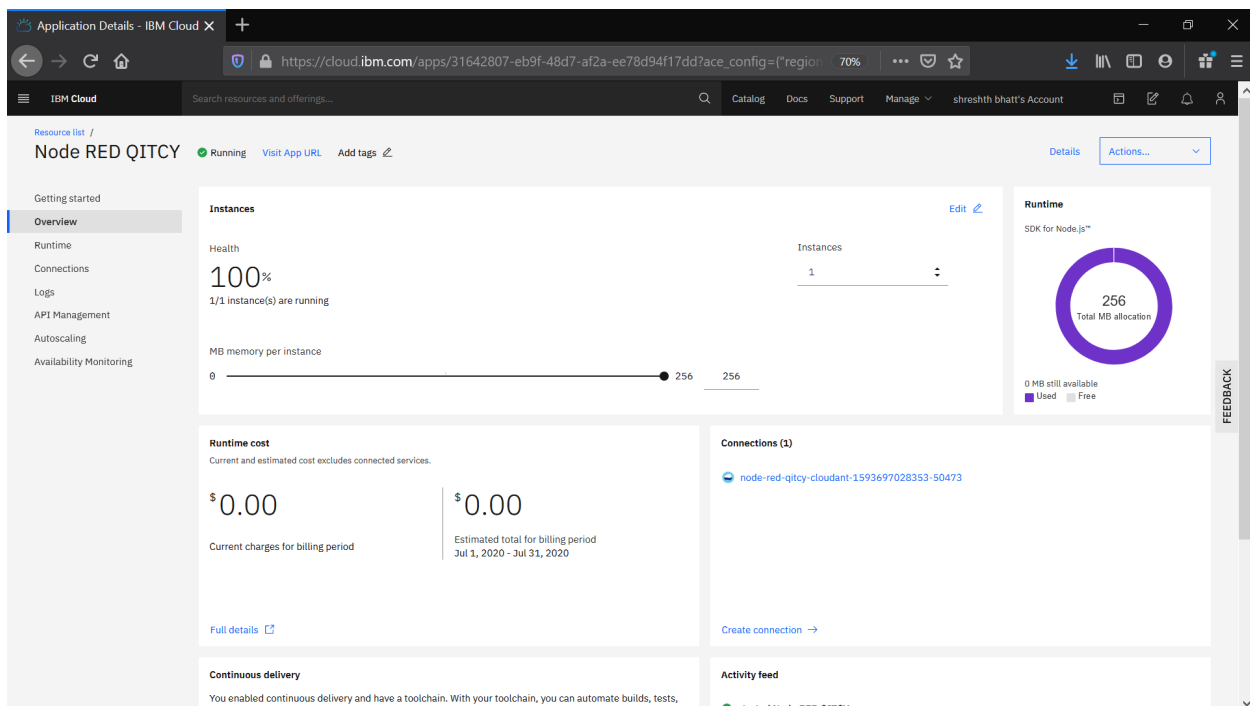
STEP-2: Create the NODERED instance & this is How Node Red Service is started.



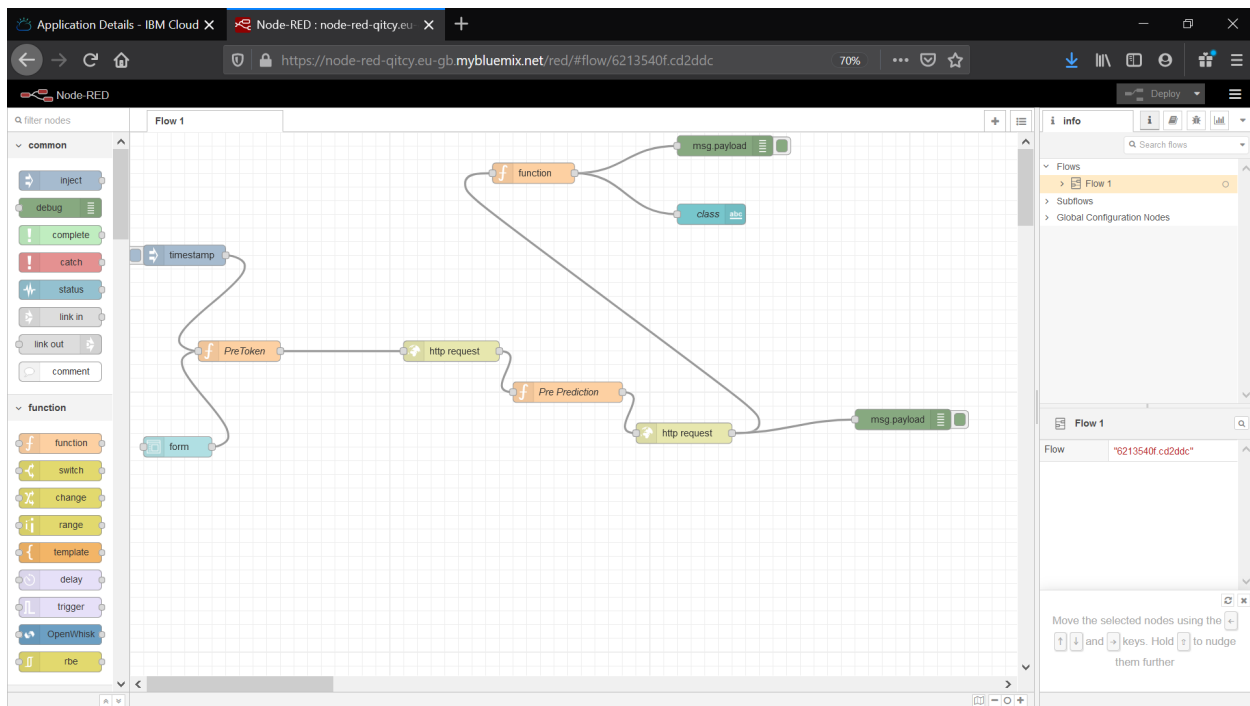
STEP-3: Get Access to the NODERED app in IBM CLOUD Dashboard window under Cloud Foundry Apps Section.



STEP-4: Once App window opens, click on Visit App URL.



STEP-5: After the Step 4, NodeRed Editor window get open which has collection of predefined nodes installed and consists of debug,output and link to the User Interface.



after creating the nodered service and creating the app, now comes, installation of dashboard nodes which are needed to build an UI Web application.
So, here we go to the navigation button and then manage palette from where we install the dashboard nodes.

Modification of Nodes according to our given model, I have concluded it through some steps using images.

FORM NODE

The screenshot shows the Node-RED web interface. On the left, a flow diagram includes a 'timestamp' node followed by a 'PreToken' node, which is connected to a 'form' node. The 'form' node is highlighted. The right sidebar shows the 'info' panel with details for the 'form' node (Node ID: 'af0652c7ab008', Type: 'ui_form').

Edit form node

Properties:

- Group: [Home] Default
- Size: auto
- Label: optional label

Form elements:

Label	Name	Type	Required	Rows	Remove
preg	pg	Number	<input checked="" type="checkbox"/>		
plac	pl	Number	<input checked="" type="checkbox"/>		
pres	pr	Number	<input checked="" type="checkbox"/>		
Skin	sk	Number	<input checked="" type="checkbox"/>		
test	tt	Number	<input checked="" type="checkbox"/>		

Buttons: submit, cancel

Topic: optional msg topic

Name:

Enabled: ☐

PRE-TOKEN NODE

The screenshot shows the Node-RED web interface. On the left, a flow diagram includes a 'timestamp' node followed by a 'PreToken' node, which is connected to a 'form' node. The 'PreToken' node is highlighted. The right sidebar shows the 'info' panel with details for the 'PreToken' node (Node ID: '7b1cc748f4feb', Type: 'function').

Edit function node

Properties:

- Name: PreToken

Setup: Function Close

```
1 global.set("pg",msg.payload.pg)
2 global.set("pl",msg.payload.pl)
3 global.set("pr",msg.payload.pr)
4 global.set("sk",msg.payload.sk)
5 global.set("tt",msg.payload.tt)
6 global.set("ms",msg.payload.ms)
7 global.set("pd",msg.payload.pd)
8 global.set("sg",msg.payload.sg)
9 var apiKey="2duhP12Wp1d4uouR5qIsK3UW6IAjfc_xZB1566Ef7Y";
10 msg.headers={"content-type":"application/x-www-form-urlencoded"}
11 msg.payload={"grant_type":"urn:ibm:params:oauth:grant-type:apikey","apikey":apiKey}
12 return msg;
```

Outputs: 1

Enabled: ☐

TOKEN URL NODE

The screenshot shows the Node-RED web interface in a browser. The main workspace displays a flow with nodes: 'inject', 'debug', 'complete', 'catch', 'status', 'link in', 'link out', 'comment', 'form', 'PreToken', 'http request', and 'Pre Prediction'. The 'http request' node is selected, and its configuration panel is open on the right. The configuration shows:

- Method: POST
- URL: https://iam.cloud.ibm.com/identity/token
- Enable secure (SSL/TLS) connection: ☐
- Use authentication: ☐
- Enable connection keep-alive: ☐
- Use proxy: ☐
- Return: a parsed JSON object
- Name: Name

A tip at the bottom of the configuration panel states: "Tip: If the JSON parse fails the fetched string is returned as-is."

PRE-PREDICTION NODE

The screenshot shows the Node-RED web interface with the 'Pre Prediction' function node selected. The configuration panel on the right shows the function code:

```
1 var pg = global.get('pg')
2 var pl = global.get('pl')
3 var pr = global.get('pr')
4 var sk = global.get('sk')
5 var tt = global.get('tt')
6 var ms = global.get('ms')
7 var pd = global.get('pd')
8 var ag = global.get('ag')
9 var token=msg.payload.access_token
10 var instance_id="748924ca-e407-4984-8cbc-f72edb732ac7"
11 msg.headers={"Content-Type": "application/json","Authorization":"Bearer "+token,"ML-Instance-ID":instance_id}
12 msg.payload={"input_data": [{"fields": {"preg", "plas", "pres", "skin", "test", "mass", "pedi", "age"}, "values": [[pg,pl,pr,sk,tt,ms,pd,ag]]}]}
13 return msg;
```

The 'Outputs' field is set to 1. The 'Enabled' checkbox is checked.

FUNCTION NODE

Application Details - IBM Cloud x Node-RED : node-red-qitcy.eu x +

https://node-red-qitcy.eu-gb.mybluemix.net/red/#flow/6213540f.cd2ddc 70%

Node-RED

Filter nodes

common

- inject
- debug
- complete
- catch
- status
- link in
- link out
- comment

function

- function
- switch
- change
- range
- template
- delay
- trigger
- Open/Close
- rbe

Flow 1

timestamp

PreToken

http

form

Edit function node

Delete Cancel Done

Properties

Name

Setup Function Close

```
1 msg.payload=msg.payload.predictions[0].values[0][0];
2 return msg;
```

Outputs 1

Enabled

info

Search flows

Flows

- Flow 1

Subflows

Global Configuration Nodes

function: 293ace90.ed3342

Node "293ace90.ed3342"

Type function

show more

Hold down **ctrl** when you **click** on a node port to enable quick-wiring

UI NODERED WEB APPLICATION WITH OUR PREDICTION DATA INPUT

Application Details - IBM Cloud x Node-RED : node-red-qitcy.eu x Node-RED Dashboard x +

https://node-red-qitcy.eu-gb.mybluemix.net/ui/#/0?socketid=om22YRi77zsfPl_AAAG 70%

Home

Default

text	1
preq	10
plus	198
pres	121
Skin	98
test	4
mass	67
ped	1.2
age	20

SUBMIT CANCEL

CONCLUSION:

- **Predictive** analytics and **machine learning** go hand-in-hand, as **predictive** models typically include a **machine learning** algorithm. These models can be trained over time to respond to new data or values, delivering the results the business needs.
- Logistic Regression which comes under Supervised Learning is the algorithm used here as output was in binary and we know that binary classification can only predict such models.
- AI lifecycle management automates data preparation, model development and hyper parameter optimization as an end-to-end data science and AI development.
- Node-RED can speed up development time and make app development more accessible to coders and non-coders alike.

THANKS & CHEERS,
SHRESHTH BHATT
(RSIP- CAREER BASIC ML100)