

Breast Cancer Risk Prediction using IBM Auto AI

➤ Introduction :-

➤ Overview:

This project is about predicting the risk of Breast Cancer using IBM Watson Auto AI Experiment. This comes under the category of "Machine Learning". The main objective of project is to predict if diagnosis is required for Breast Cancer Patient or not based on the previously present data.

➤ Purpose:

Worldwide, breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates. Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (on an x-ray). After a suspicious lump is found, the doctor will conduct a diagnosis to determine whether it is cancerous and, if so, whether it has spread to other parts of the body. . Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed potential for improving the diagnostic accuracy.

➤ Literature Survey :-

➤ Existing Problem:

Breast cancer is one of the main causes of cancer death worldwide. Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed potential for improving the diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.

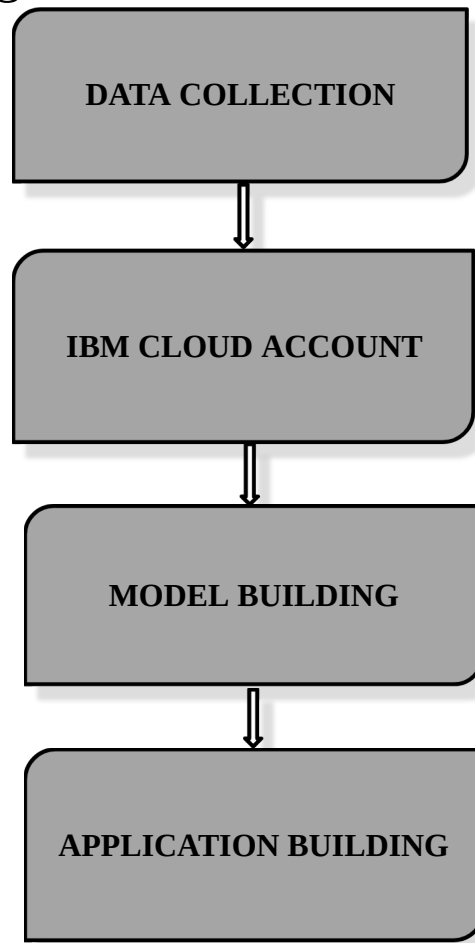
➤ Proposed Solution:

Here we are developing a machine learning model where in the model gets trained by considering the parameters such as: Radius ,Texture, Perimeter, Area, Smoothness, Concavity, Concaveness, Compactness here all these parameters are taken in mean, se and overall values are been taken. And the model is been trained using Auto AI service in IBM Watson cloud and that can be deployed in an

application such as web or mobile applications.

➤ Theoretical survey :-

➤ Block Diagram:



➤ Hardware/Software Designing:

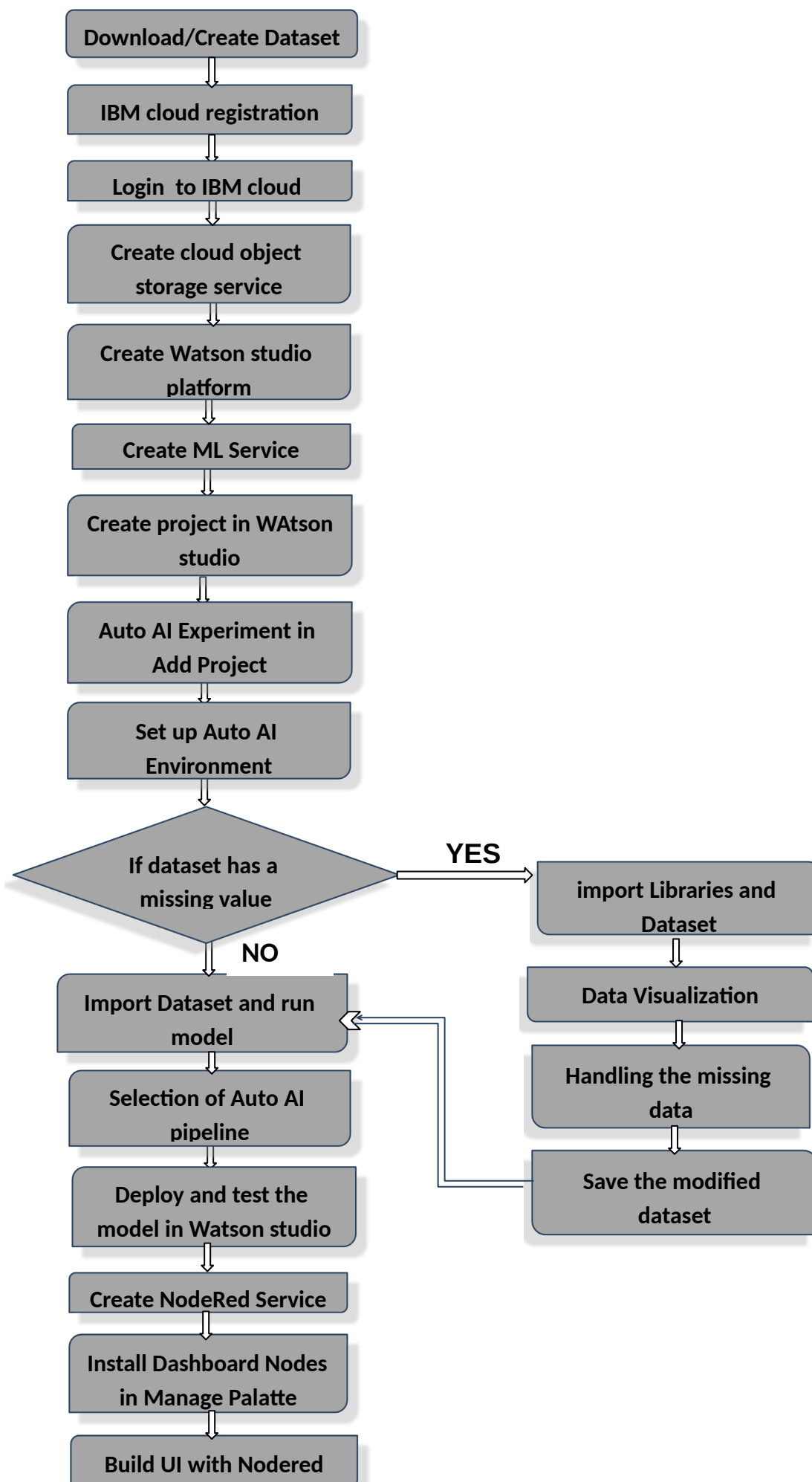
- Jupyter Notebook for Data Analysis
- IBM Cloud
- IBM Watson
- Auto AI Experiment
- NodeRed

➤ Experimental investigations :-

Upon analysing the dataset, we understand that -

- Since the dataset contains independent input parameters and an output is to be predicted, hence it comes under Supervised Learning.
- And since the output is categorical and not continuous, classification algorithm is to be used, but since we are using Auto AI here, any classification can not be applied externally.
- As the dataset contains all the important inputs to predict the output, there is no need to remove any input rows.

➤ Flow chart :-



➤ Result :-

The model is successfully trained and deployed using Auto AI Experiment and Node-red Service. Now "Breast Cancer Risk Prediction" Machine Learning model can predict the risk of breast cancer.

➤ Advantages and Disadvantages :-

➤ Advantages:

- This model helps in the prediction of diagnosis possibility for a patient with breast cancer by which we can easily stop the spread of cancer to different parts of the body.
- Improves the diagnostic accuracy.

➤ Disadvantages:

- Sometimes slow due to network glitch.
- Model cannot be edited yet in a more granular way.
- Limited in number of models it can choose from.
- Data must be good going in'
- If size of the dataset is large then the model takes a very large time for building.

➤ Applications :-

Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed potential for improving the diagnostic accuracy. This can be used in various hospitals to predict the breast cancer risk as early as possible due to which the deaths may be reduced.

➤ Conclusion :-

Therefore the "Breast Cancer Risk Prediction" , Machine Learning model is created and the purpose of project is fulfilled.

➤ Future Scope :-

This model is mainly used for predicting breast cancer diagnostic accuracy, by upgrading the dataset the results may be more precisely accurate.

➤ Bibliography :-

- github.com
- [kaggle.com](https://www.kaggle.com)(for downloading dataset)
- [smartinternz.com](https://www.smartinternz.com)(for attaching links and pdfs in the platform)
- thesmartbridge.com

► Appendix :-

ScreenShots:

► Data Collection:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
32	18.63	25.11	124.8	1098	0.1064	0																	
33	11.84	18.7	77.93	440.6	0.1109	0																	
34	17.02	23.98	112.8	899.3	0.1197	0																	
35	19.27	26.47	127.9	1162	0.09401	0																	
36	16.13	17.88	107	807.2	0.104	0																	
37	16.74	21.59	110.1	869.5	0.0961	0																	
38	14.25	21.72	93.63	633	0.09823	0																	
39	13.03	18.42	82.61	523.8	0.08983	1																	
40	14.99	25.2	95.54	698.8	0.09387	0																	
41	13.48	20.82	88.4	559.2	0.1016	0																	
42	13.44	21.58	86.18	563	0.08162	0																	
43	10.95	21.35	71.9	371.1	0.1227	0																	
44	19.07	24.81	128.3	1104	0.09081	0																	
45	13.28	20.28	87.32	545.2	0.1041	0																	
46	13.17	21.81	85.42	531.5	0.09714	0																	
47	18.65	17.6	123.7	1076	0.1099	0																	
48	8.196	16.84	51.71	201.9	0.086	1																	
49	13.17	18.66	85.98	534.6	0.1158	0																	
50	12.05	14.63	78.04	449.3	0.1031	1																	
51	13.49	22.3	86.91	561	0.08752	1																	
52	11.76	21.6	74.72	427.9	0.08637	1																	
53	13.64	16.34	87.21	571.8	0.07685	1																	
54	11.94	18.24	75.71	437.6	0.08261	1																	
55	18.22	18.7	120.3	1033	0.1148	0																	
56	15.1	22.02	97.36	712.8	0.09056	0																	
57	11.52	18.75	73.34	409	0.09524	1																	
58	19.21	18.57	125.5	1152	0.1053	0																	
59	14.71	21.59	95.55	656.9	0.1137	0																	
60	13.05	19.31	82.61	527.2	0.0806	1																	
61	8.618	11.79	54.34	224.5	0.09752	1																	
62	10.17	14.88	64.55	311.9	0.1134	1																	
63	14.08	20.98	148.66	771.8	0.1243	1																	

► Importing Libraries and Dataset:

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [3]: dataset=pd.read_csv('datasets_56485_108594_Breast_cancer_data.csv')
dataset
```

```
Out[3]:
```

	mean_radius	mean_texture	mean_perimeter	mean_area	mean_smoothness	diagnosis
0	17.99	10.38	122.80	1001.0	0.11840	0
1	20.57	17.77	132.90	1326.0	0.08474	0
2	19.69	21.25	130.00	1203.0	0.10960	0
3	11.42	20.38	77.58	386.1	0.14250	0
4	20.29	14.34	135.10	1297.0	0.10030	0
...
564	21.56	22.39	142.00	1479.0	0.11100	0
565	20.13	28.25	131.20	1261.0	0.09780	0
566	16.60	28.08	108.30	858.1	0.08455	0
567	20.60	29.33	140.10	1265.0	0.11780	0
568	7.76	24.54	47.92	181.0	0.05263	1

569 rows x 6 columns

► Creating IBM Account:

IBM Cloud

Search resources and offerings...

Dashboard

Upgrade Customize Create resource +

Resource summary View all

11 Resources

- Cloud Foundry apps 1
- Cloud Foundry services 1
- Services 4
- Storage 1
- Apps 2
- Developer tools 2

Add resources +

Planned maintenance View all

For you

Get started with using AI and Cloud Object Storage in 15 minutes.

Get Started with Watson Studio

Watson Assistant lets you build conversational interfaces into any application, device, or channel.

Use Watson Assistant

Clear skies! You can view your scheduled maintenance events here. Refresh

News View all

IBM CEO Arvind Krishna Keynotes IBM Think Gov Digital 2020

IBM Introduces Sterling Inventory Control Tower to Help Organizations More Effectively Manage Inventory and Build Resilient Supply Chains

Daimler Expands Relationship with IBM; Adopts IBM public cloud for its security capabilities

Recent support cases View all

User access Manage users

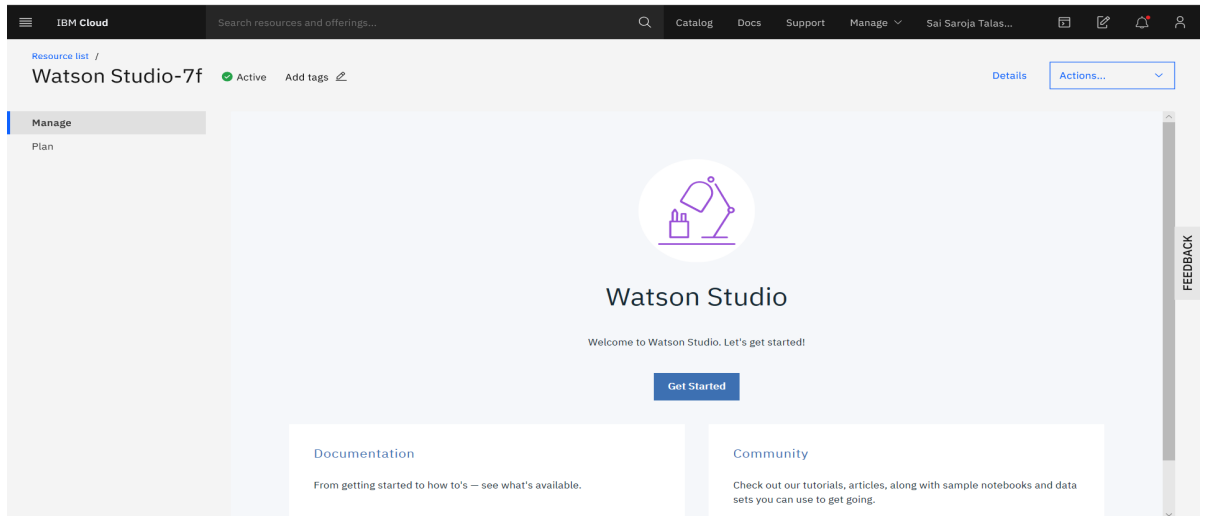
Enter email addresses below to jump directly into the invite user setup:

Enter up to 100 email addresses

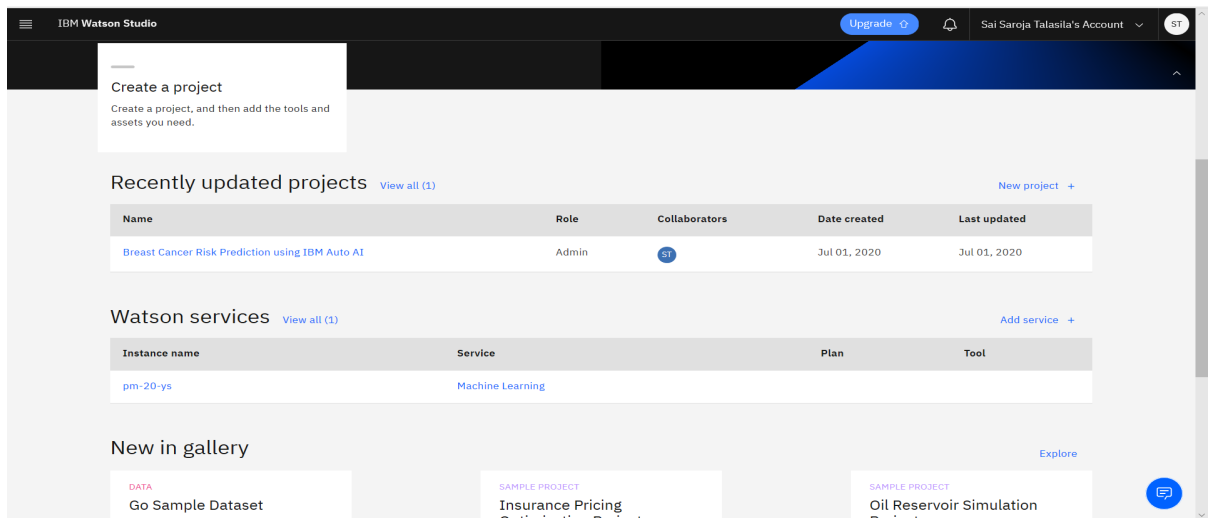
IBM Cloud status View all

FEEDBACK

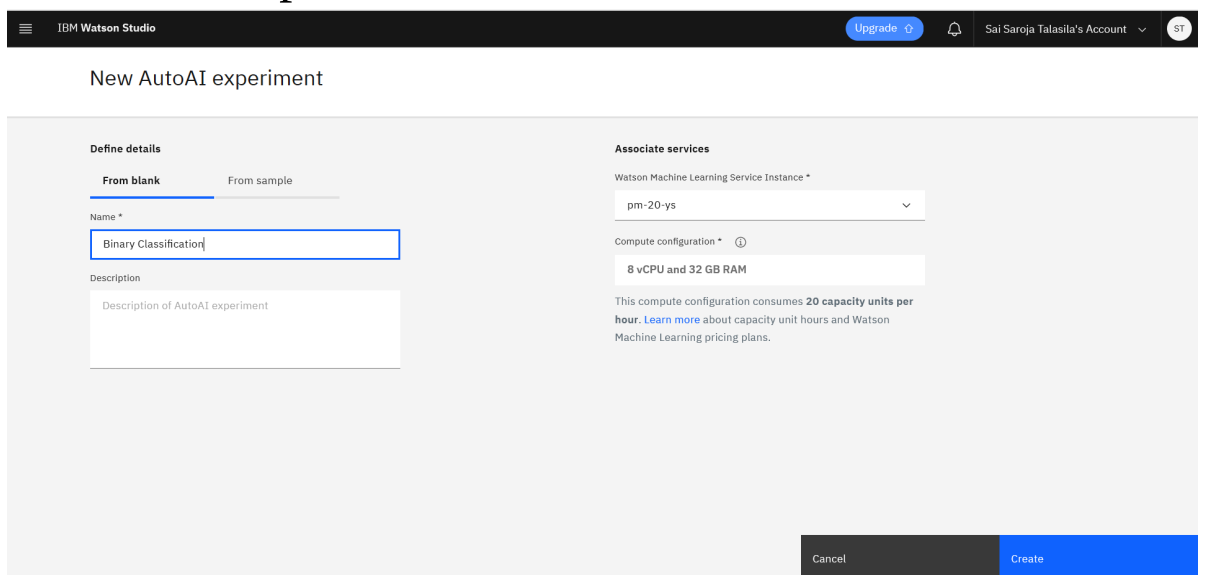
► Creating Watson Studio Platform:



► Creating Project and Watson Instance Service:



► Auto AI Setup:



► Importing Dataset:

IBM Watson Studio

My projects / Breast Cancer Risk Prediction usi...

Overview Assets Environments Jobs Deployments Access Control Settings

What assets are you looking for?

Data assets

0 assets selected.

Name	Type	Created by	Last modified
csv datasets_56485_108594_Breast_cancer_data.csv	Data Asset	Sai Saroja Talasila	Jul 01, 2020, 05:47 PM
csv datasets_56485_108594_Breast_cancer_data.csv	Data Asset	Sai Saroja Talasila	Jul 01, 2020, 05:42 PM

AutoAI experiments

New AutoAI experiment +

Name	Status	Model type	Last modified
Binary Classification	Completed	Binary Classification	Jul 01, 2020, 05:53 PM

Deep learning experiments

New deep learning experiment

► Running Experiment:

IBM Watson Studio

My projects / Breast Cancer Risk Prediction usi... / Binary Classification

Experiment summary Pipeline comparison Rank by: Accuracy (Optimized) Score: Cross validation Holdout

Relationship map

Prediction column: diagnosis

FEATURE TRANSFORMERS

PIPELINES

TOP ALGORITHMS

datasets_56485_10...

Progress map

Swap view

Experiment completed

8 PIPELINES GENERATED

8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

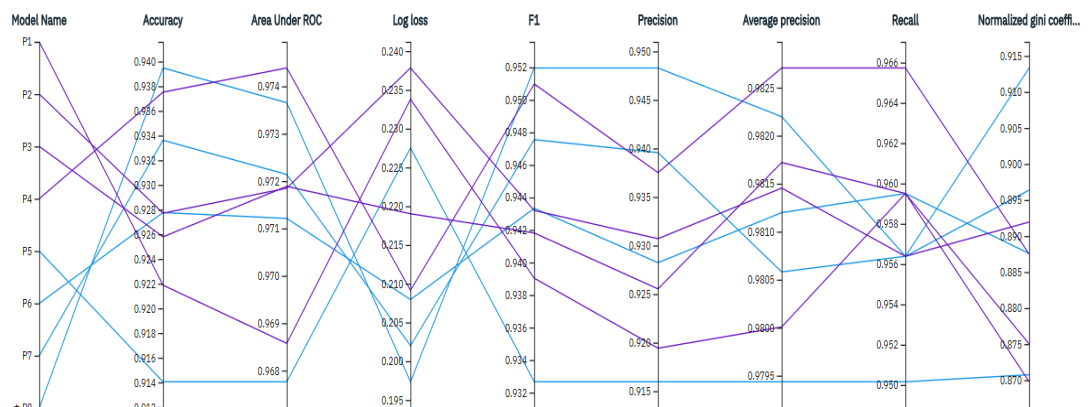
Time elapsed: 5 minutes

View full log

► Metric Analysis:

Metric chart

Prediction column: diagnosis



► Selection of Auto AI Pipeline:

Pipeline leaderboard

	Rank	↑	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
>	★ 1		Pipeline 8	XGB Classifier	0.939	HPO-1 FE HPO-2	00:00:51
>	2		Pipeline 4	Gradient Boosting Classifier	0.938	HPO-1 FE HPO-2	00:00:09
>	3		Pipeline 7	XGB Classifier	0.934	HPO-1 FE	00:00:56
>	4		Pipeline 6	XGB Classifier	0.928	HPO-1	00:00:13
>	5		Pipeline 2	Gradient Boosting Classifier	0.928	HPO-1	00:00:04
>	6		Pipeline 3	Gradient Boosting Classifier	0.926	HPO-1 FE	00:00:32
>	7		Pipeline 1	Gradient Boosting Classifier	0.922	None	00:00:01
>	8		Pipeline 5	XGB Classifier	0.914	None	00:00:01

► Model Deployment and Testing:

IBM Watson Studio

Upgrade

Sai Saroja Talasila's Account

ST

My projects / Breast Cancer Risk Prediction usi... / Binary Classification - P8 XGBCla... / Classifier

Classifier

OverviewImplementationTest

Enter input data

24

mean_perimeter

47

mean_area

180

mean_smoothness

0.08

Predict

```
{
  "predictions": [
    {
      "fields": [
        "prediction",
        "probability"
      ],
      "values": [
        1,
        [
          0.013977766036987304,
          0.9860222339630127
        ]
      ]
    }
  ]
}
```

► Node-Red Service creation:

Node-RED on IBM Cloud

Node-RED

Flow-based programming for the Internet of Things

Node-RED is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways.

This instance is running as an IBM Cloud application, giving it access to the wide range of services available on the platform.

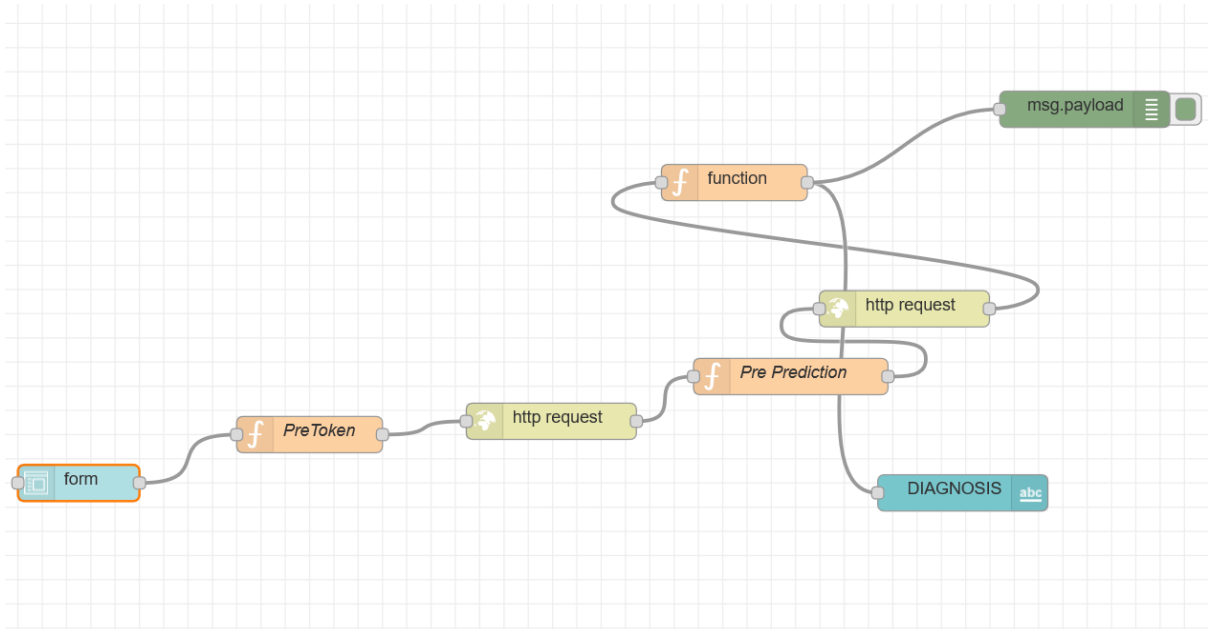
More information about Node-RED, including documentation, can be found at nodered.org.

Go to your Node-RED flow editor

[Learn how to customise Node-RED](#)

Customising your instance of Node-RED

► Building UI(User Interface) with Node-Red:



► Output in Node-Red App:

BREAST CANCER RISK PREDICTION

PARAMETERS

MEAN RADIUS *
13

MEAN TEXTURE *
14

MEAN PERIMETER *
87

MEAN AREA *
566

MEAN SMOOTHNESS *
0.09

SUBMIT **CANCEL**

DIAGNOSIS **1**