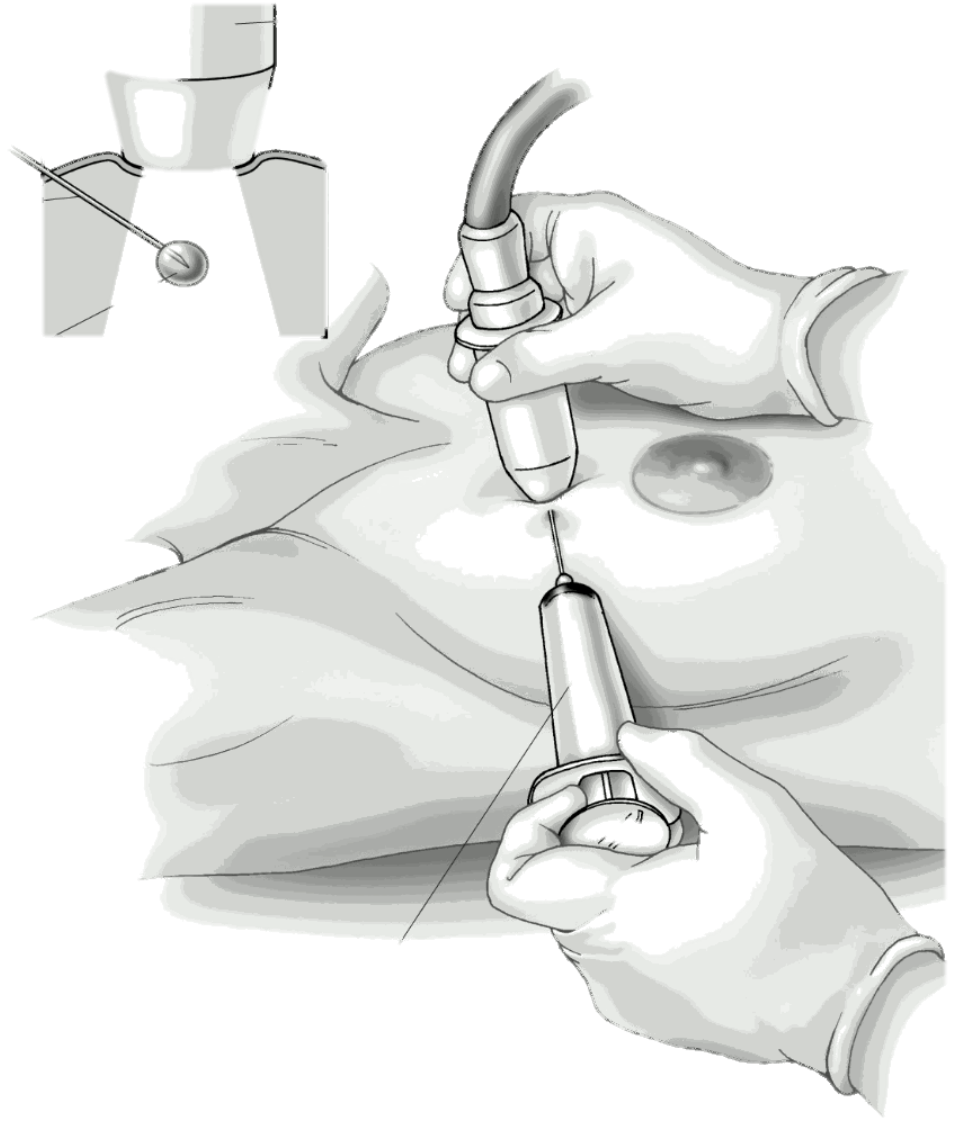
SPS-5717-Breast Cancer Risk Prediction System

Dr. Manik Ghosh
Department of Pharmaceutical Sciences and Technology
Birla Institute of Technology
Mesra, Ranchi - 835 215

1.Open watson studio
2.Create a project
3.add Auto AI experiment
4.create a machine learning instance
5.associate ML to project
6.Load the dataset to Cloud object storage
7.Select the prediction parameter in the dataset
8.Train the model
9.Deploy- we get an API
10.Buid web application using Node-red

Data from Microscopic Biopsy

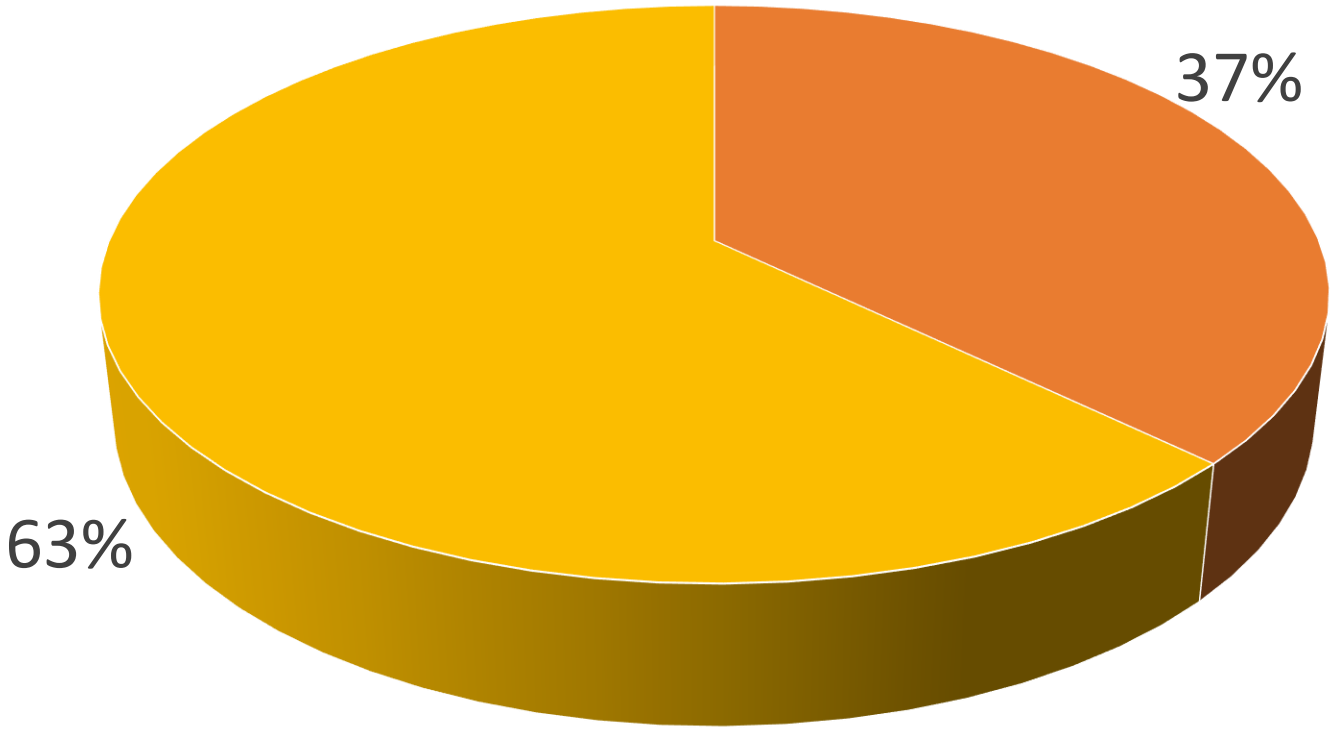
- As mentioned in UCI website, “Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image”.
- Moreover, FNA is a type of biopsy procedure where a very thin needle is inserted into an area of abnormal tissue or cells with a guide of CT scan or ultrasound monitors (figure). The collected sample is then transferred to a pathologist to study it under a microscope and examine whether cells in the biopsy are normal or not.



Breast Cancer

- 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.
- This breast cancer dataset was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg.
- 357 observations which account for 62.7% of all observations indicating the absence of cancer cells, 212 which account for 37.3% of all observations shows the presence of cancerous cell.

Total Data: 569



63% : Absence of Cancer

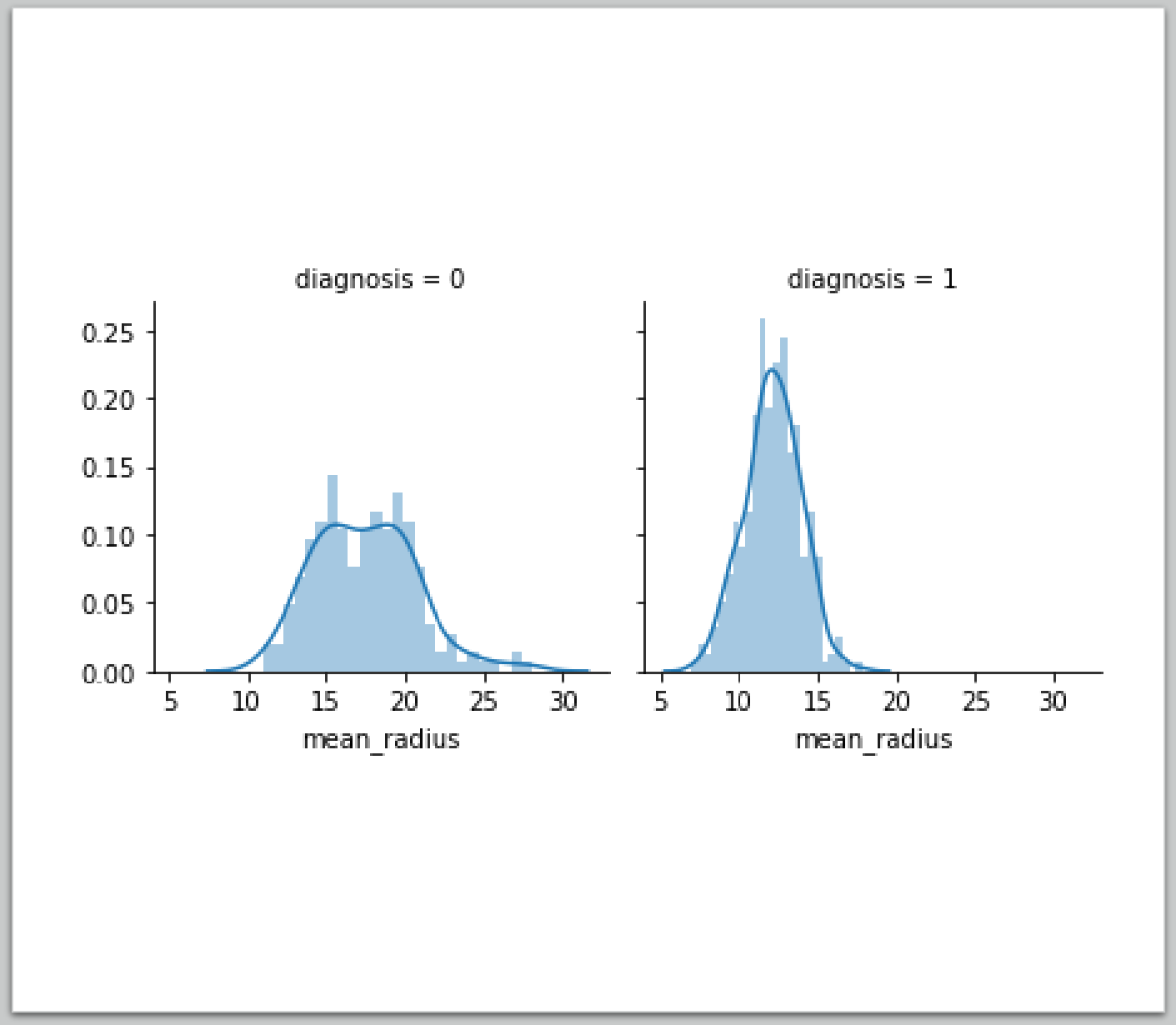
37% : Presence of Cancer

Total Data: 569

- **mean_radius: 14.1±3.52 Median: 13.4 (6.98 – 28.1)**
- **mean_texture: 19.3±4.3 Median: 18.8 (9.71 – 39.3)**
- **mean_perimeter: 92±24.3 Median: 86.2 (43.8 – 189)**
- **mean_area: 655±352 Median: 551 (144 – 2500)**
- **mean_smoothness: 0.1±0.01 Median: 0.1 (0.05 – 0.16)**

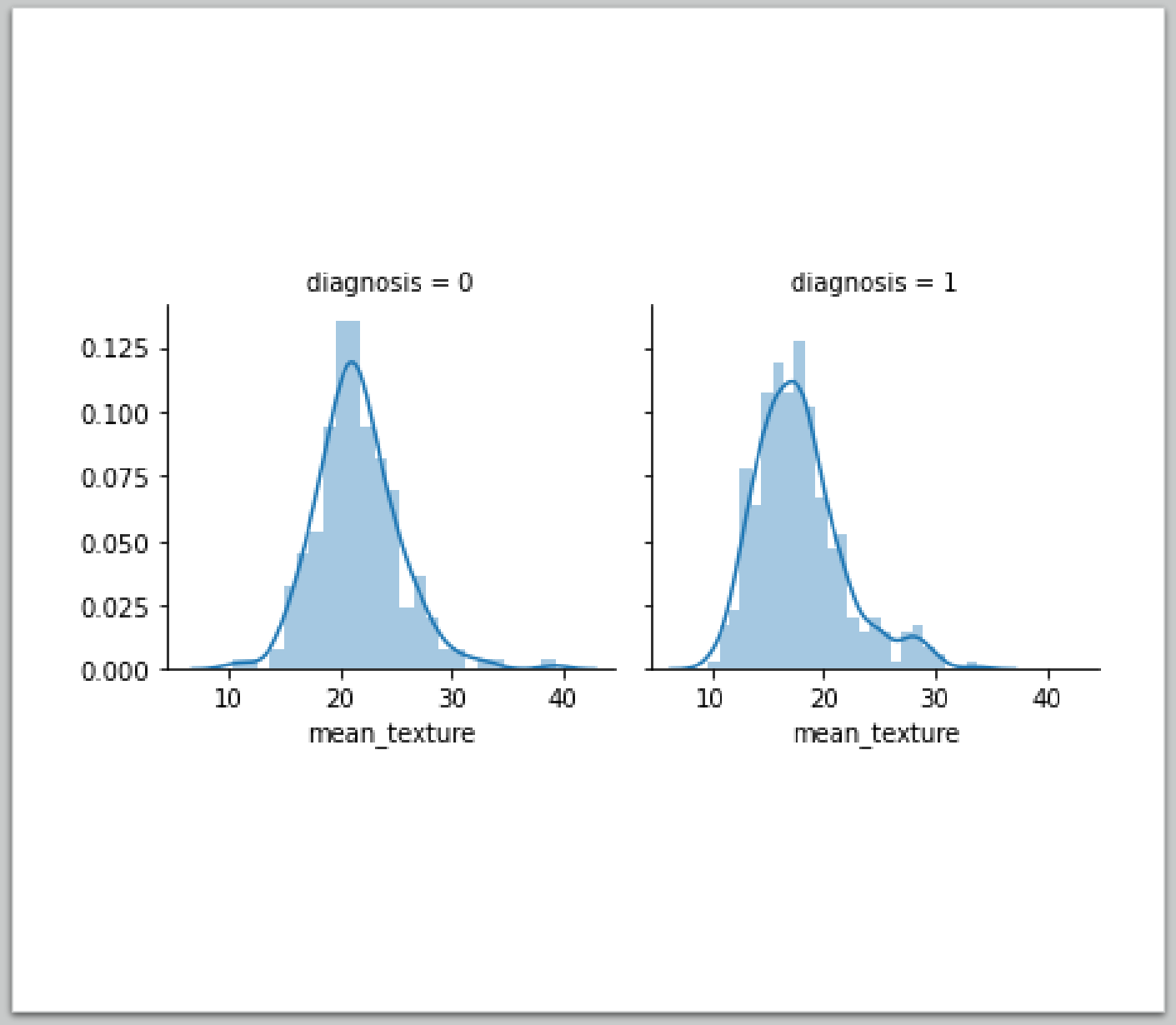
Radius

- We can see that when mean_radius is close to 12 it is likely to be breast cancer and after 20 there is almost no chance to be breast cancer



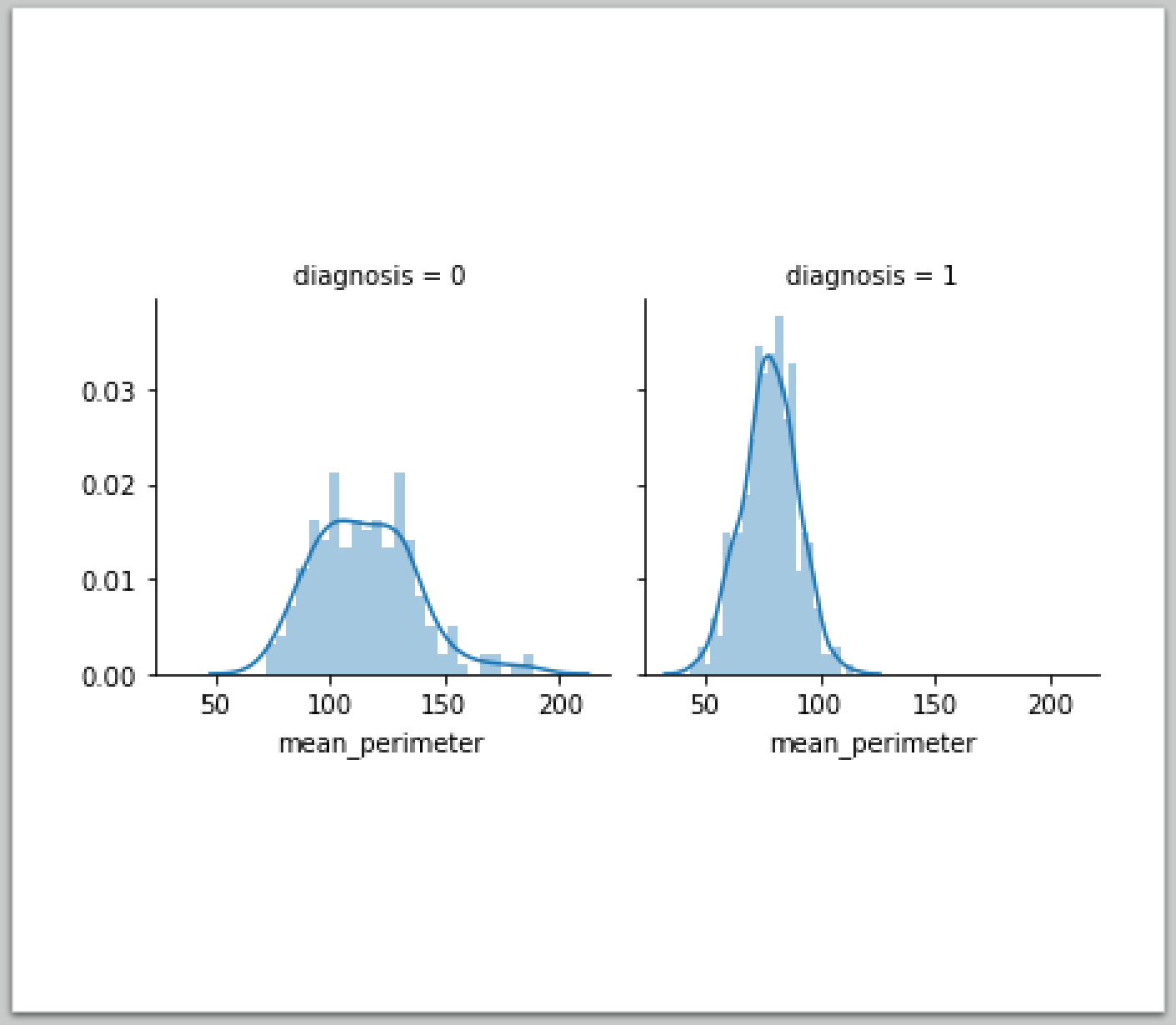
Texture

- We can say that when diagnosis is 1 mean_texture is likely to be closer to 18 while at diagnosis=0 mean is at near 21



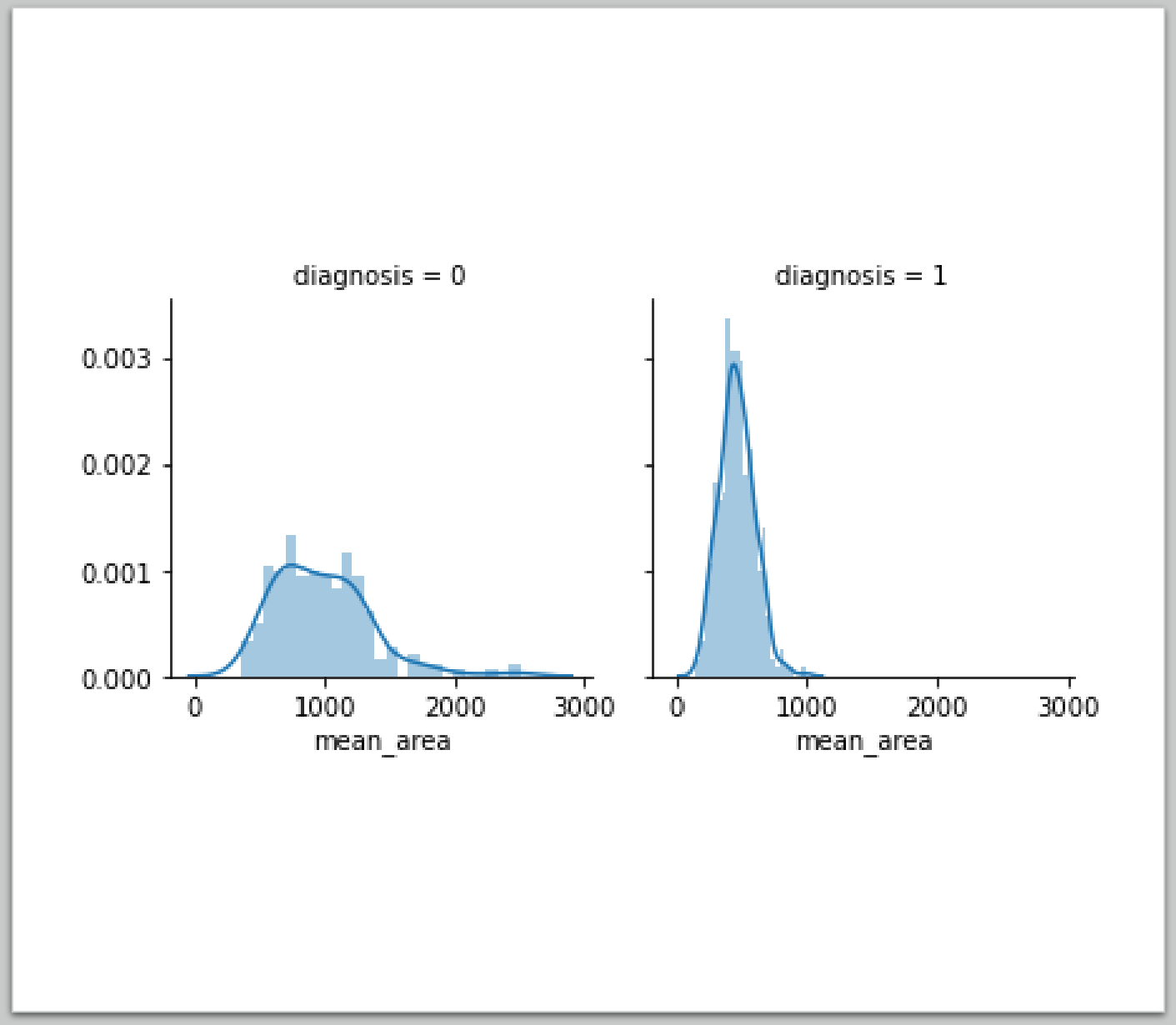
Perimeter

- This variable's behavior is similar to mean_radius, when diagnosis=1 mean_perimeter is less and stacked in a small space.



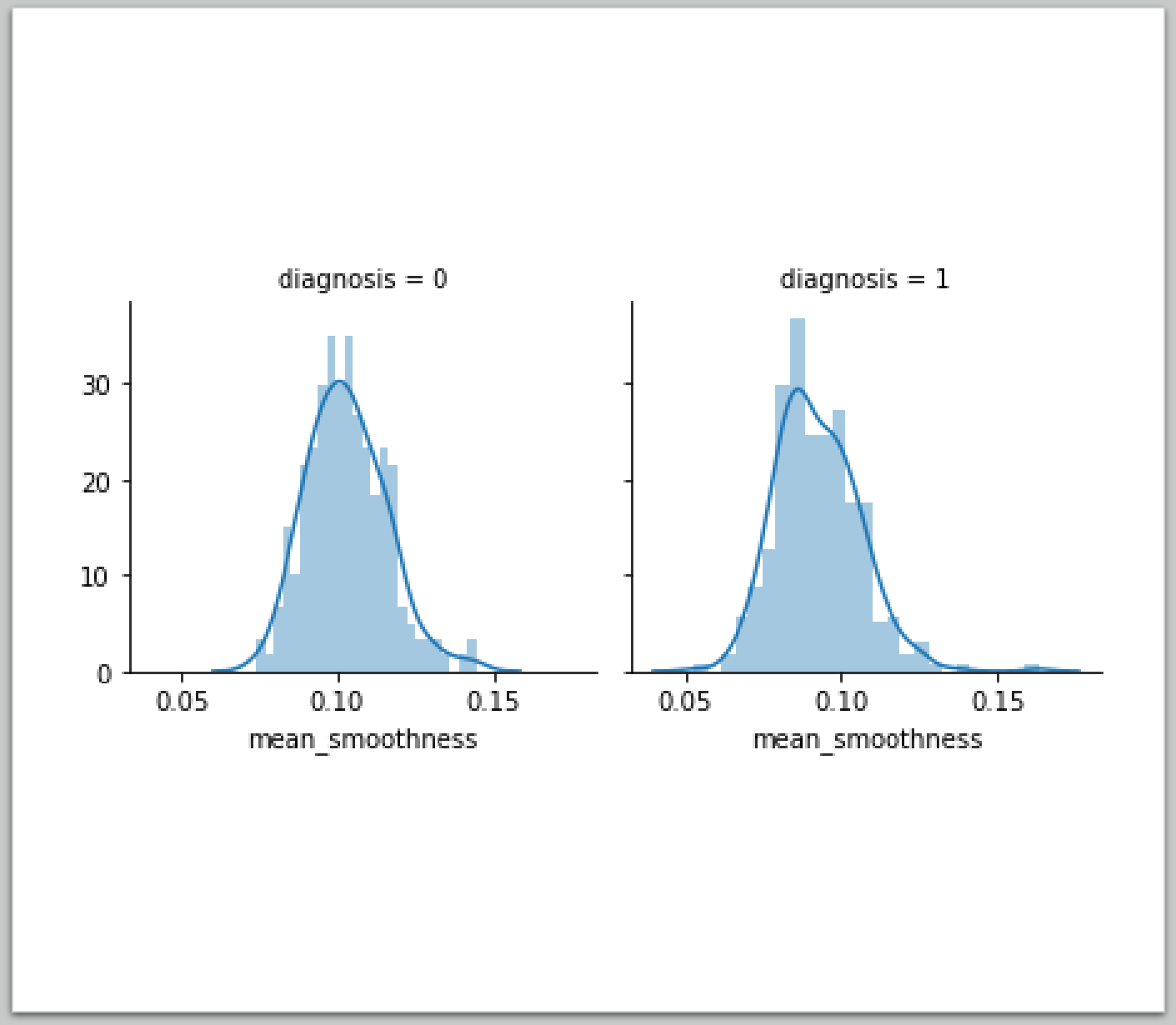
Area

- Again it is similar to perimeter and radius but different from other we can see diagnosis=0 for every area



Smoothness

- It is similar to texture variable. For every case of diagnosis, distribution is similar but when diagnosis is 1 it is less than other case.



Breast Cancer Project

Deployed | Online

API reference | Test

Endpoint

https://us-south.ml.cloud.ibm.com/ml/v4/deployments/45a3ffd1-a4af-4af5-abdf-b0fb78580b95/or

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-url

# the above CURL request will return an auth token that you will use as $IAM_TOKEN in the scoring request below
# TODO: manually define and pass values to be scored below
curl -X POST --header "Content-Type: application/json" --header "Accept: application/json" --header "Authorization: Bearer $IAM_T
```

Breast Cancer Project

Created
Oct 17, 2020 11:30 PM

Updated
Oct 17, 2020 11:30 PM

Deployment ID
45a3ffd1-a4af-4af5-abdf-b0fb7...

Software specification
hybrid_0.1

Hybrid pipeline software specifications
autoai-kb_3.1-py3.7

Copies
1

Description
No description provided.

Associated asset

Internet Speed Test | Fast.com | GuruCool Program - Day 5 - Y | Node-RED : node-red-tnoll-20 | Node-RED Dashboard | IBM Cloud Pak for Data

dataplatform.cloud.ibm.com/ml-runtime/deployments/417bb628-03b2-4717-8773-8652335eed76/test?space_id=30043465-834c-42e5-a2c8-dec1eacc6449&context=c...

Apps | SSL Error | cPanel Login | Meet - Online class | User Login | Birla In... | Moodle@BIT Mesra... | Birla Institute Of Te... | Tatasteel | SparkNotes: Introd... | Other bookmarks

BreastCancerProject

Deployed | Online

API reference | **Test**

Enter input data

mean_radius
Double

mean_texture
Double

mean_perimeter
Double

mean_area
Double

Predict

Result

BreastCancerProject

Created
Oct 19, 2020 5:21 PM

Updated
Oct 19, 2020 5:21 PM

Deployment ID
417bb628-03b2-4717-8773-86...

Software specification
hybrid_0.1

Hybrid pipeline software specifications
autoai-kb_3.1-py3.7

Copies
1

Description
No description provided.

Associated asset

Student Dashboard | SPS-5717-Breast Cance | SmartPracticeschool inv | Upload files · SmartPrac | GitHub | IBM Cloud Pak for Data

dataplatform.cloud.ibm.com/ml-runtime/deployments/417bb628-03b2-4717-8773-8652335eed76/test?space_id=30043465-834c-42e5-a2c8-dec1eacc6449&context=c...

Apps | SSL Error | cPanel Login | Meet - Online class | User Login | Birla In... | Moodle@BIT Mesra... | Birla Institute Of Te... | Tatasteel | SparkNotes: Introd... | Other bookmarks

BreastCancerProject

Deployed | Online

API reference | Test

Enter input data

mean_radius
17.99

mean_texture
10.38

mean_perimeter
122.8

mean_area
1001

Predict

Result


```
{
    "predictions": [
        {
            "fields": [
                "prediction",
                "probability"
            ],
            "values": [
                [
                    0,
                    [
                        0.9641520136987483,
                        0.03584798630125169
                    ]
                ]
            ]
        }
```


Show less

BreastCancerProject

Created
Oct 19, 2020 5:21 PM

Updated
Oct 19, 2020 5:21 PM

Deployment ID
417bb628-03b2-4717-8773-86...

Software specification
hybrid_0.1

Hybrid pipeline software specifications
autoai-kb_3.1-py3.7

Copies
1

Description
No description provided.

Associated asset
Breast Cancer Prediction - P8 LGB...

Type here to search | Day 5 | IBM Clo... | GURUC... | ENG | 10:25 PM 10/19/2020

BreastCancerProject - Deployment

Deployment ID

- 417bb628-03b2-4717-8773-8652335eed76

Associate Asset

- 797bc8f3-9c48-4c42-b803-9e67b45be257

Breast Cancer Project API

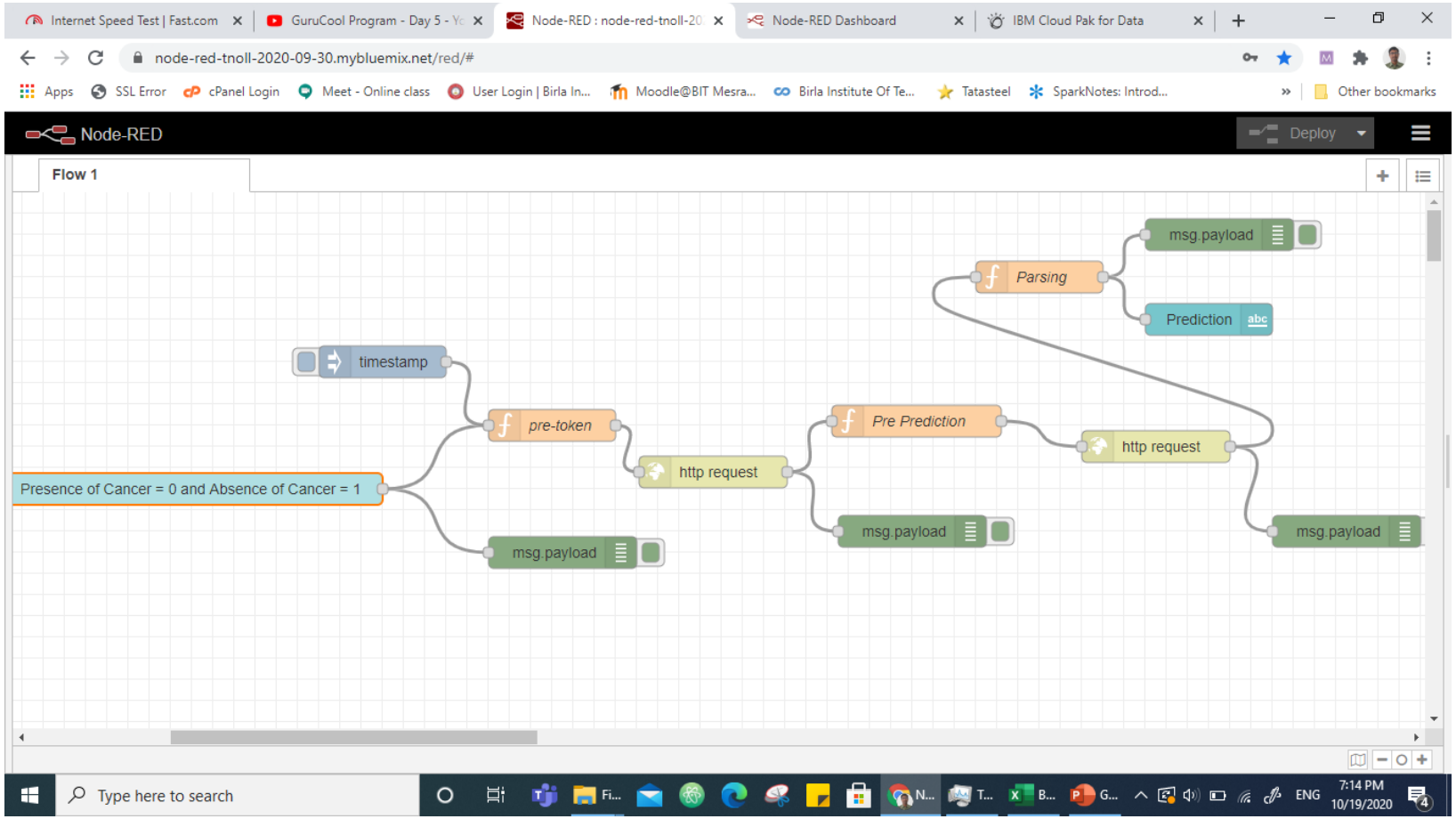
Endpoint:

https://us-south.ml.cloud.ibm.com/ml/v4/deployments/417bb628-03b2-4717-8773-8652335eed76/predictions?version=2020-10-19?version=2020-09-01

API Key:

uxYXB9RuHJw7VG-vn-mg0E3eoLwmPlJWCM2SkRL54DZU

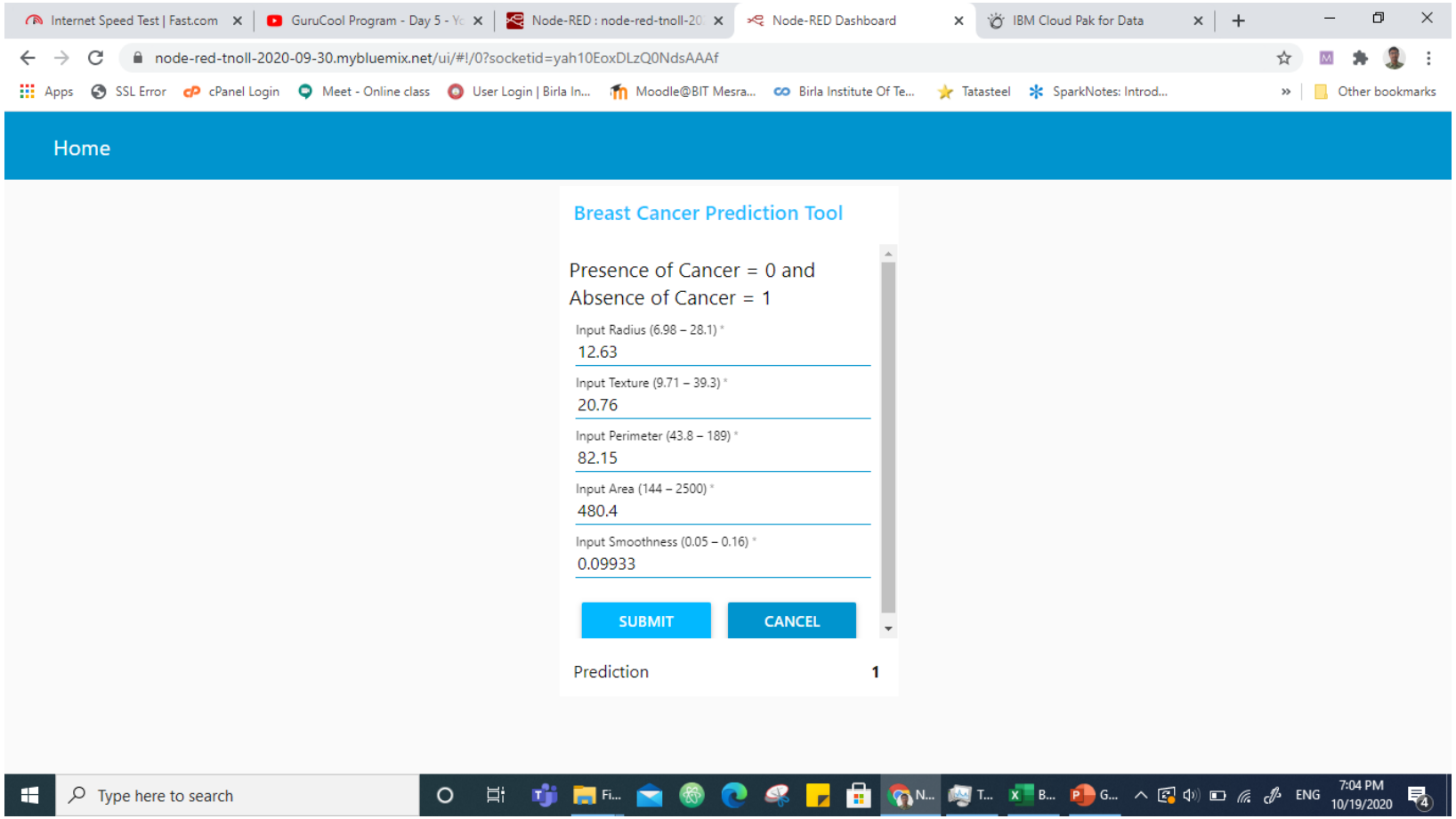
Node-Red Flow



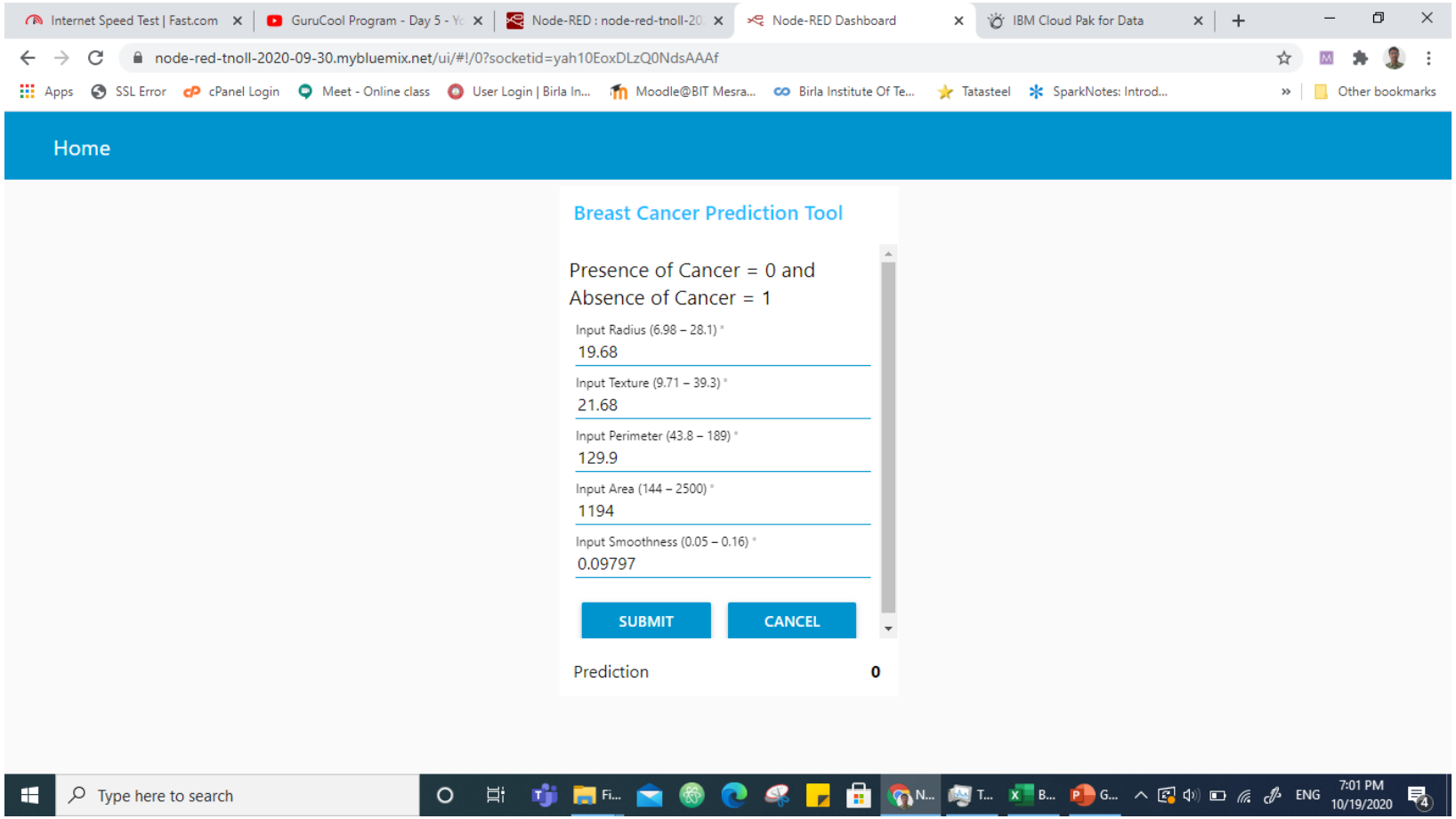
Node-Red UI Link

https://node-red-tnoll-2020-09-30.mybluemix.net/ui/#!/0?socketid=yah10EoxDLzQ0NdsAAA

Cancer Prediction : Absence of Cancer



Cancer Prediction : Presence of Cancer



Bibliography

- https://www.kaggle.com/onuralpsisman/breast-cancer-prediction
- https://www.kaggle.com/merishnasuwal/breast-cancer-prediction-dataset
- https://www.sciencedirect.com/science/article/pii/S2001037014000464
- https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3
- https://towardsdatascience.com/breast-cancer-cell-type-classifier-ace4e82f9a79
- https://rstudio-pubs-static.s3.amazonaws.com/344010_1f4d6691092d4544bfbddb092e7223d2.html
- https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
- https://smartinternz.com/Student/badge_workspace/5717
- https://www.youtube.com/results?search_query=gurucool+day+5