

SPS-5717-Breast Cancer Risk Prediction System

Dr. Manik Ghosh

Associate Professor

Department of Pharmaceutical Sciences and Technology

Birla Institute of Technology

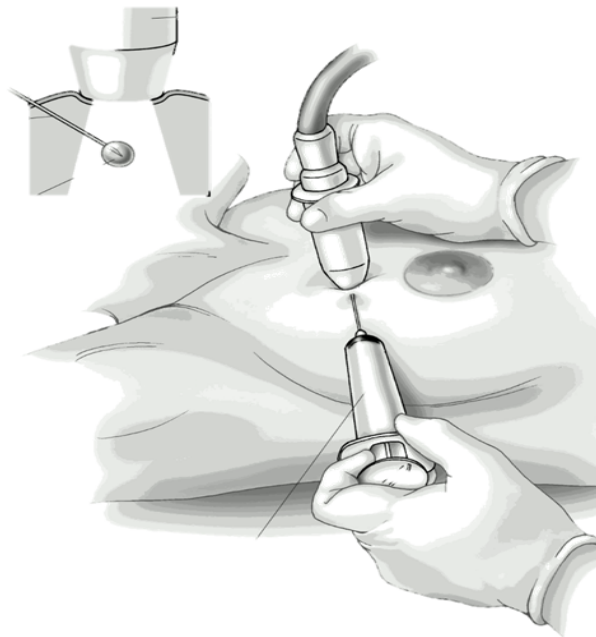
Mesra, Ranchi - 835 215

Jharkhand

Intoduction

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.



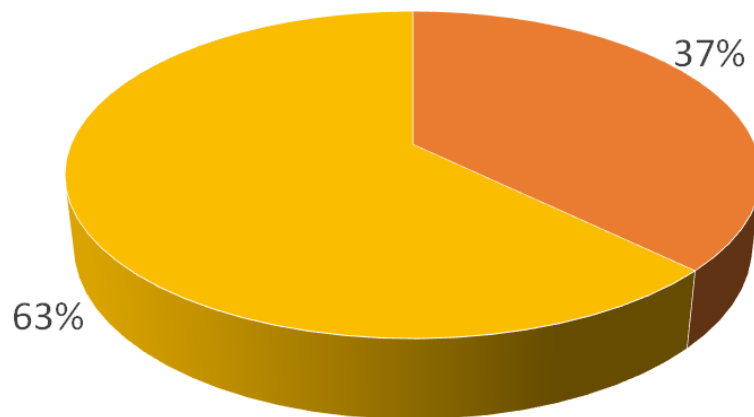
Worldwide, breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates. Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (on an x-ray). After a suspicious lump is found, the doctor will conduct a diagnosis

SPS-5717-Breast Cancer Risk Prediction System

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.



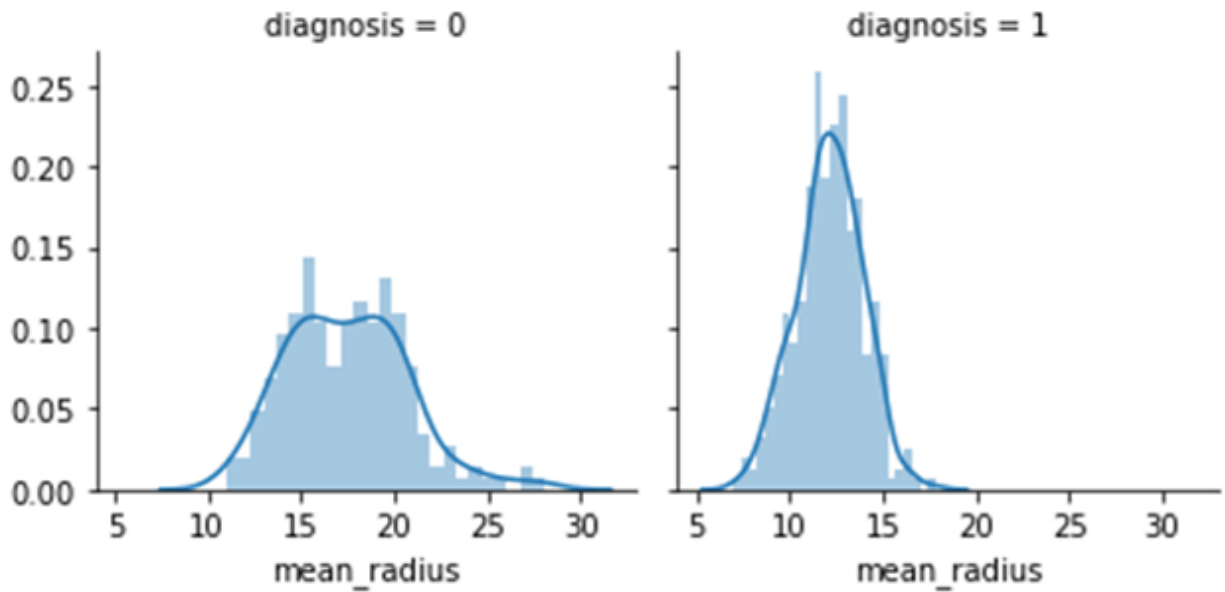
63% : Absence of Cancer
37% : Presence of Cancer

Dataset:

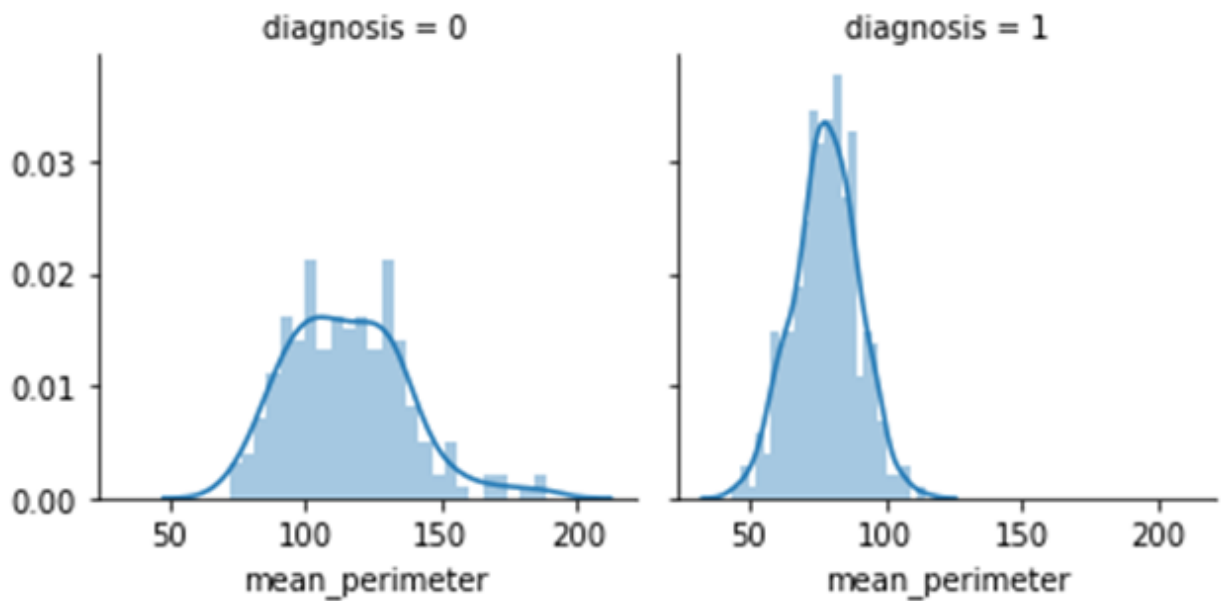
<https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>

- 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)

SPS-5717-Breast Cancer Risk Prediction System

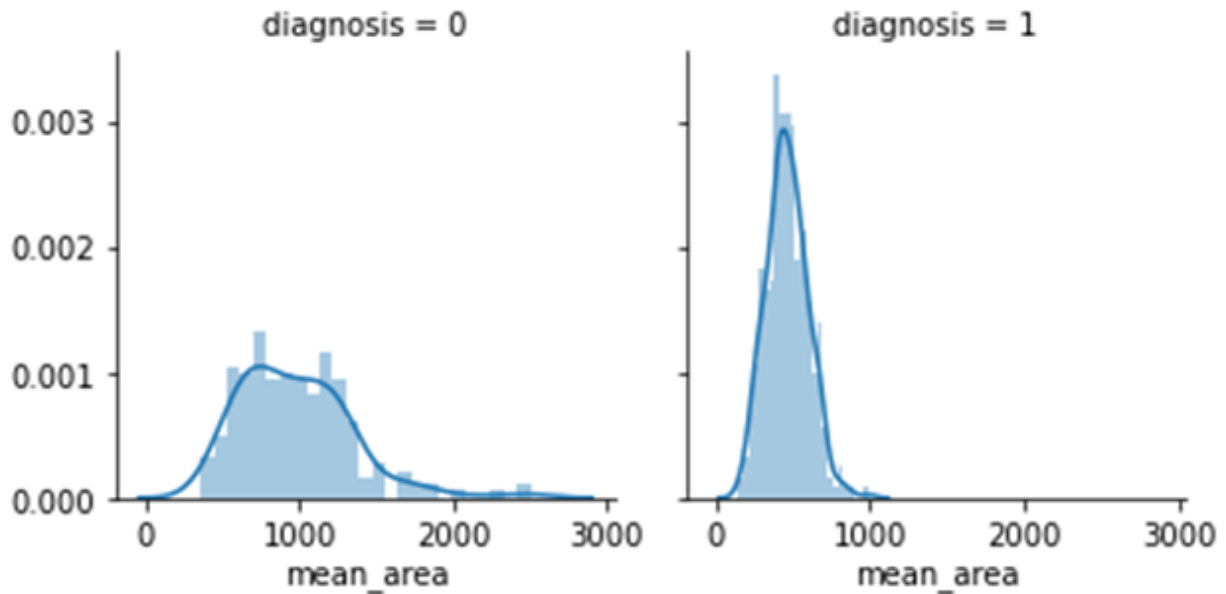


We can see that when mean_radius is close to 12 it is likely to be breast cancer and after 20 there is almost no chance to be breast cancer

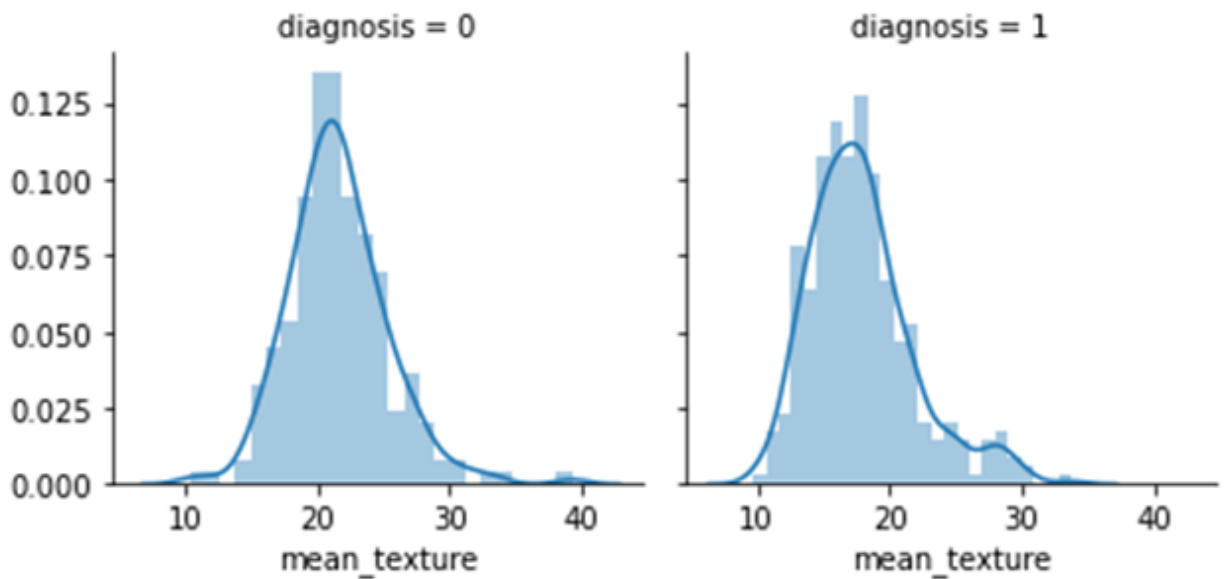


This variable's behavior is similar to mean_radius, when diagnosis=1 mean_perimeter is less and stacked in a small space.

SPS-5717-Breast Cancer Risk Prediction System

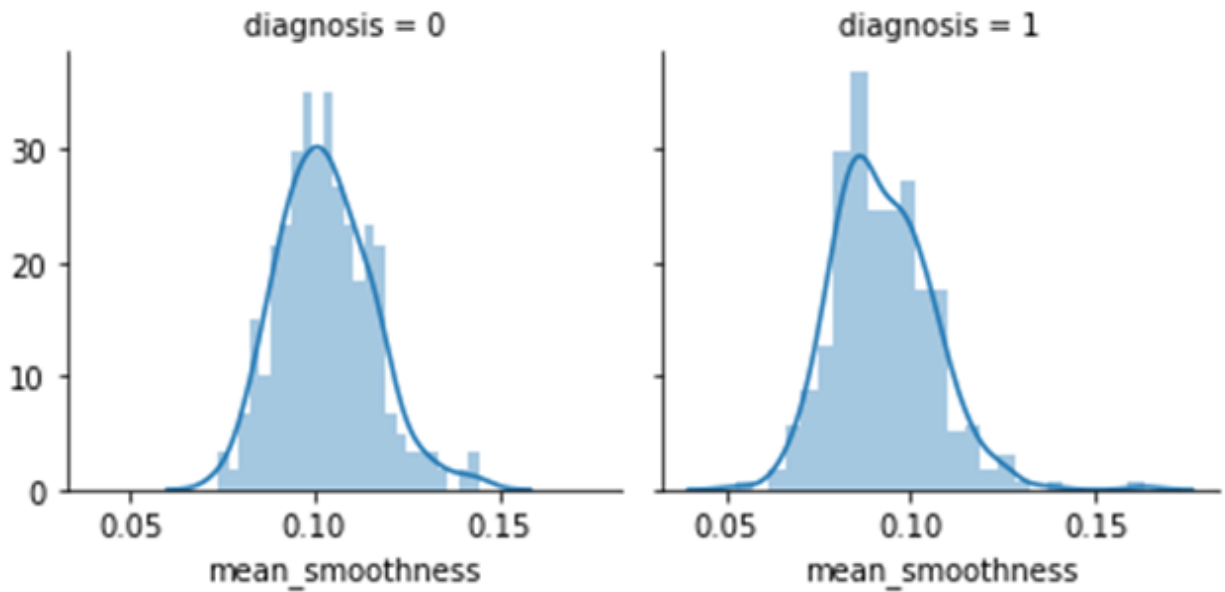


Again it is similar to perimeter and radius but different from other we can see diagnosis=0 for every area



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

SPS-5717-Breast Cancer Risk Prediction System



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.

Creation and Deployment of Machine Learning Project

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-mg0E3eoLwmPIJWCM2SkRL54DZU

Deployment ID:

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

Associate Asset:

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

SPS-5717-Breast Cancer Risk Prediction System

Creation of Project in Watson Studio

The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes the IBM Cloud Pak for Data logo, an 'Upgrade' button, and a user profile for 'Manik Ghosh's Account'. The main content area displays the 'BreastCancerProject' deployment page. The project is marked as 'Deployed' and 'Online'. The 'API reference' tab is selected, showing a 'Test' button. The 'Enter input data' section contains five input fields: 'mean_radius' (Double), 'mean_texture' (Double), 'mean_perimeter' (Double), 'mean_area' (Double), and 'mean_perimeter' (Double). The 'Result' section is empty. A right-hand sidebar provides details about the project, including its creation and update dates (Oct 19, 2020 5:21 PM), deployment ID (417bb628-03b2-4717-8773-86...), software specification (hybrid_0.1), and hybrid pipeline software specifications (autoai-kb_3.1-py3.7). The bottom of the screen shows a Windows taskbar with various application icons and a system clock indicating 7:12 PM on 10/19/2020.

Result of Test Run

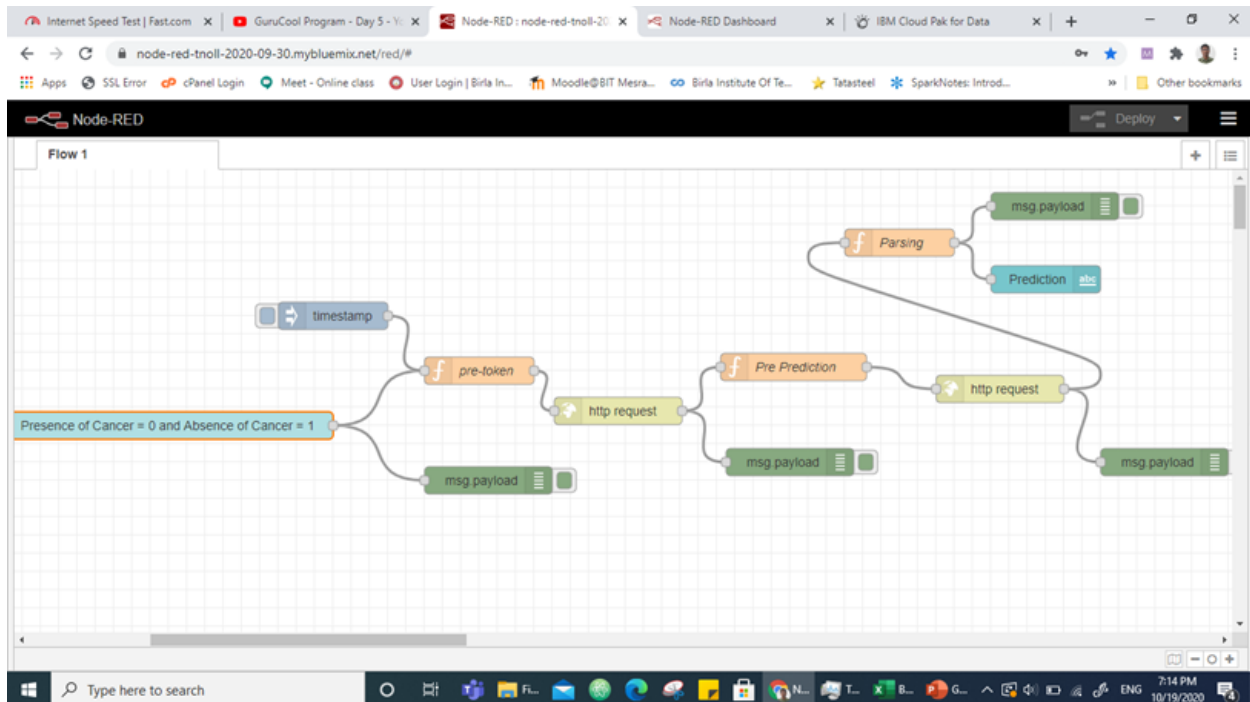
The screenshot shows the IBM Cloud Pak for Data interface with the 'BreastCancerProject' deployment page. The 'Test' tab is selected, and the 'Enter input data' section contains five input fields: 'mean_radius' (17.99), 'mean_texture' (10.38), 'mean_perimeter' (122.8), 'mean_area' (1001), and 'mean_perimeter' (1001). The 'Result' section displays the output of the test run, which is a JSON object:

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

 A right-hand sidebar provides details about the project, including its creation and update dates (Oct 19, 2020 5:21 PM), deployment ID (417bb628-03b2-4717-8773-86...), software specification (hybrid_0.1), and hybrid pipeline software specifications (autoai-kb_3.1-py3.7). The bottom of the screen shows a Windows taskbar with various application icons and a system clock indicating 10:25 PM on 10/19/2020.

SPS-5717-Breast Cancer Risk Prediction System

Creation of Node-Red Flow



Node-Red User Interface

The screenshot shows the Node-RED User Interface for the Breast Cancer Prediction Tool. The interface is titled 'Breast Cancer Prediction Tool' and includes a 'Home' button. The main form contains the following input fields and values:

- Presence of Cancer = 0 and Absence of Cancer = 1
- Input Radius (5.98 - 28.1) : 12.63
- Input Texture (9.71 - 39.3) : 20.76
- Input Perimeter (43.8 - 189) : 82.15
- Input Area (144 - 2500) : 480.4
- Input Smoothness (0.05 - 0.16) : 0.09933

Below the input fields are two buttons: 'SUBMIT' and 'CANCEL'. The 'Prediction' section shows the result: '1'.

Node-Red UI Link:

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

SPS-5717-Breast Cancer Risk Prediction System

GitHub:

<https://github.com/SmartPracticeschool/SPS-5717-Breast-Cancer-Risk-Prediction-System>

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_1f4d6691092d4544bfbdb092e7223d2.html
- [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](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