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.
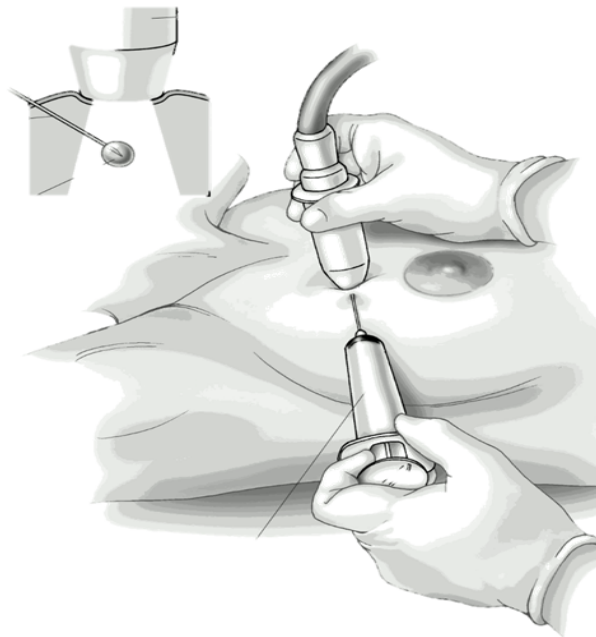

Worldwide, breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates. Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (on an x-ray). After a suspicious lump is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

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.
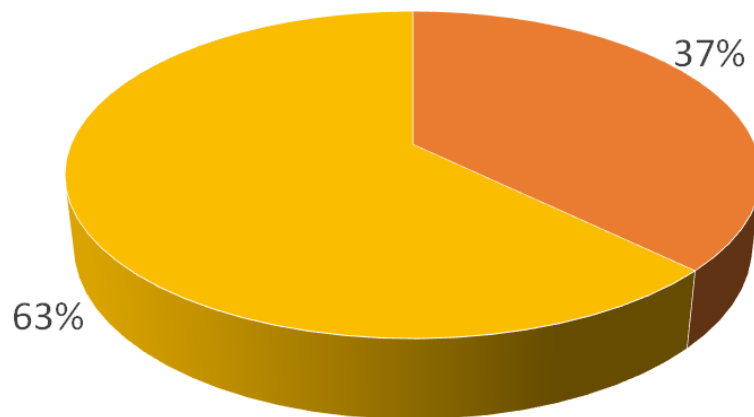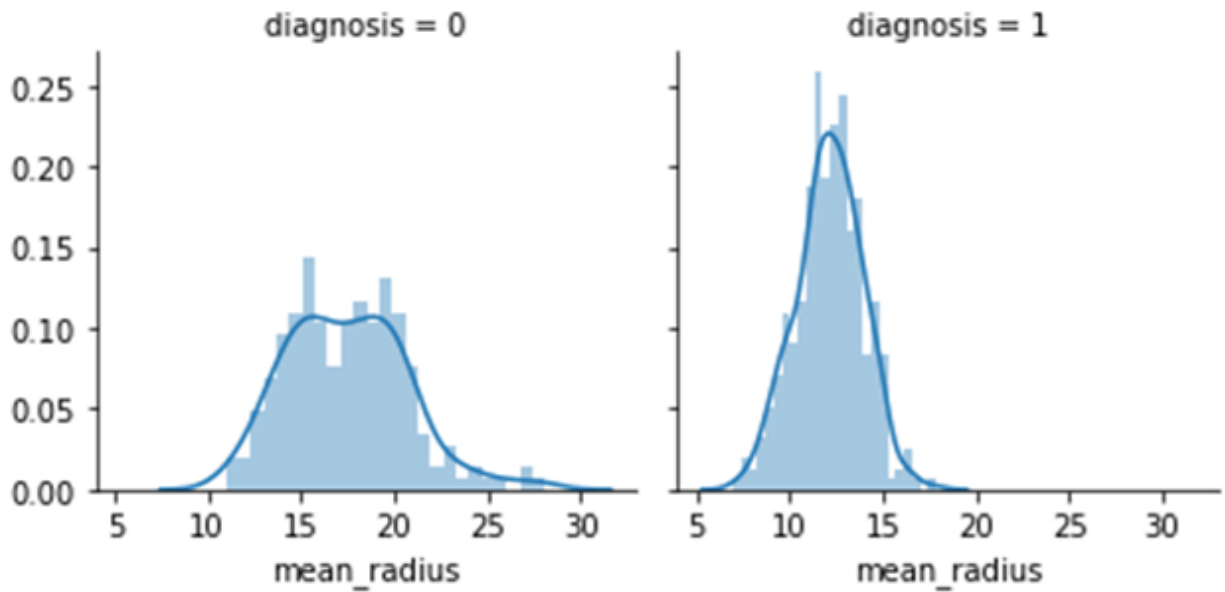

63% : Absence of Cancer
37% : Presence of Cancer

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)

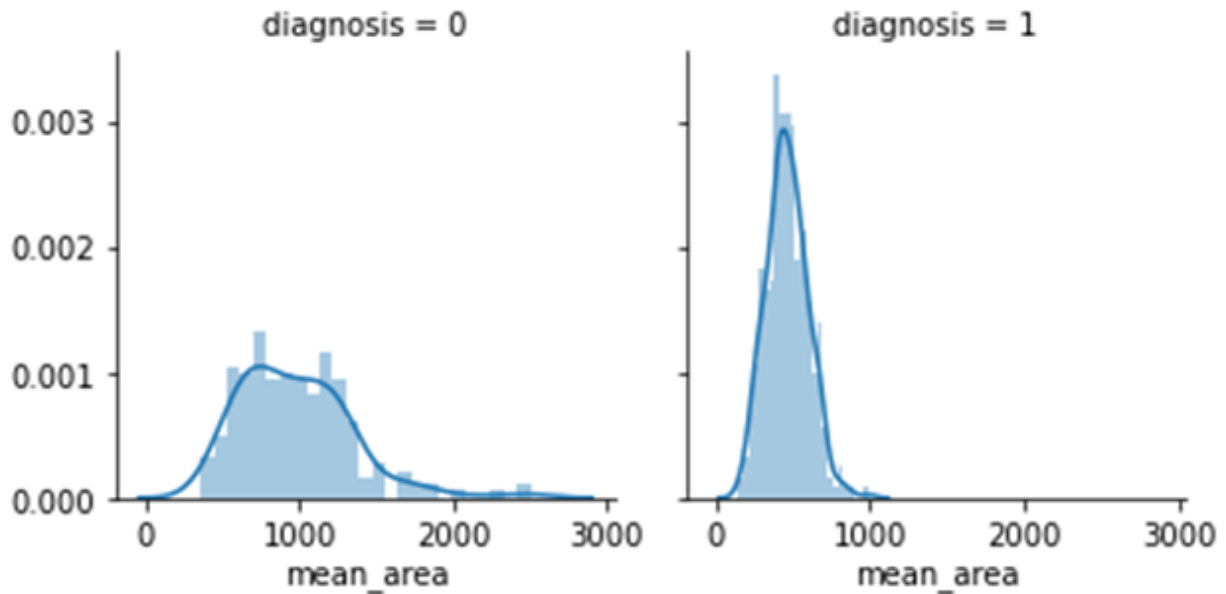# SPS-5717-Breast Cancer Risk Prediction System

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
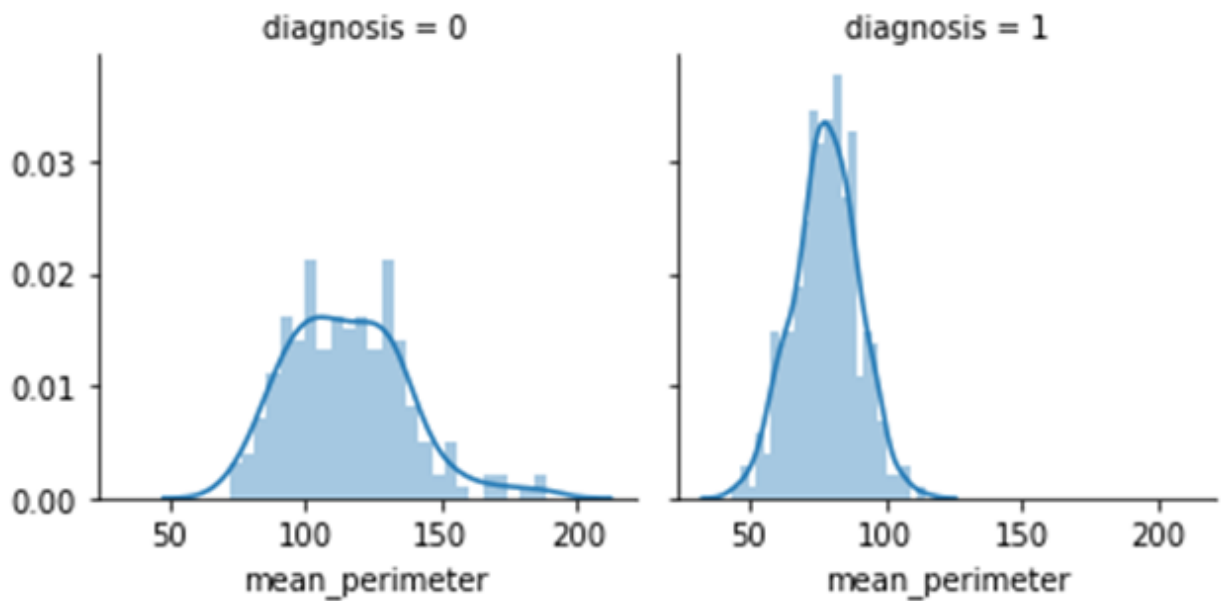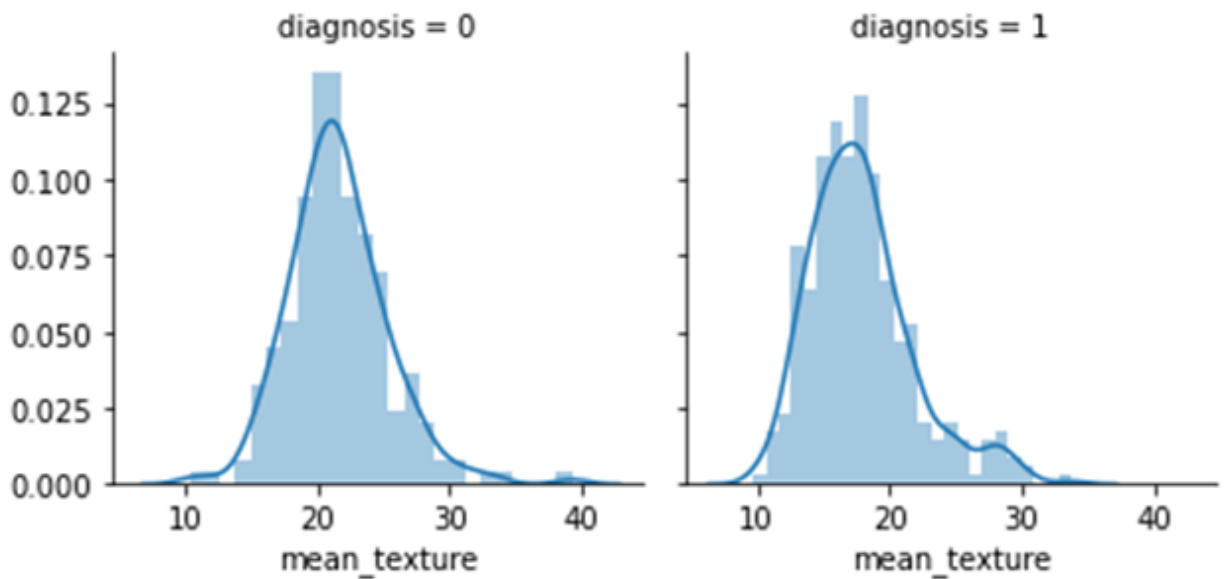

This variable's behavior is similar tomean_radius, when diagnosis=1 mean_perimeter is less and stacked in a small space.
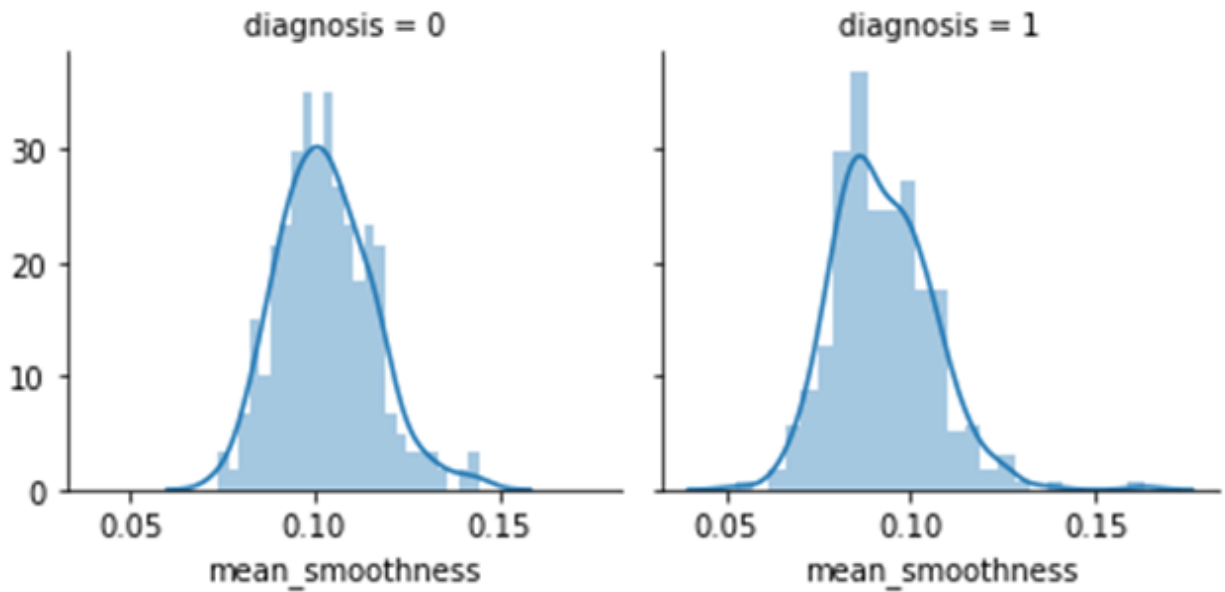
Again it is similar to perimeter and radius but different from other we can see diagnosis=0 for every area

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-mg0E3eoLwmPlJWCM2SkRL54DZU

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

Result of Test Run

SPS-5717-Breast Cancer Risk Prediction System

Creation of Node-Red Flow
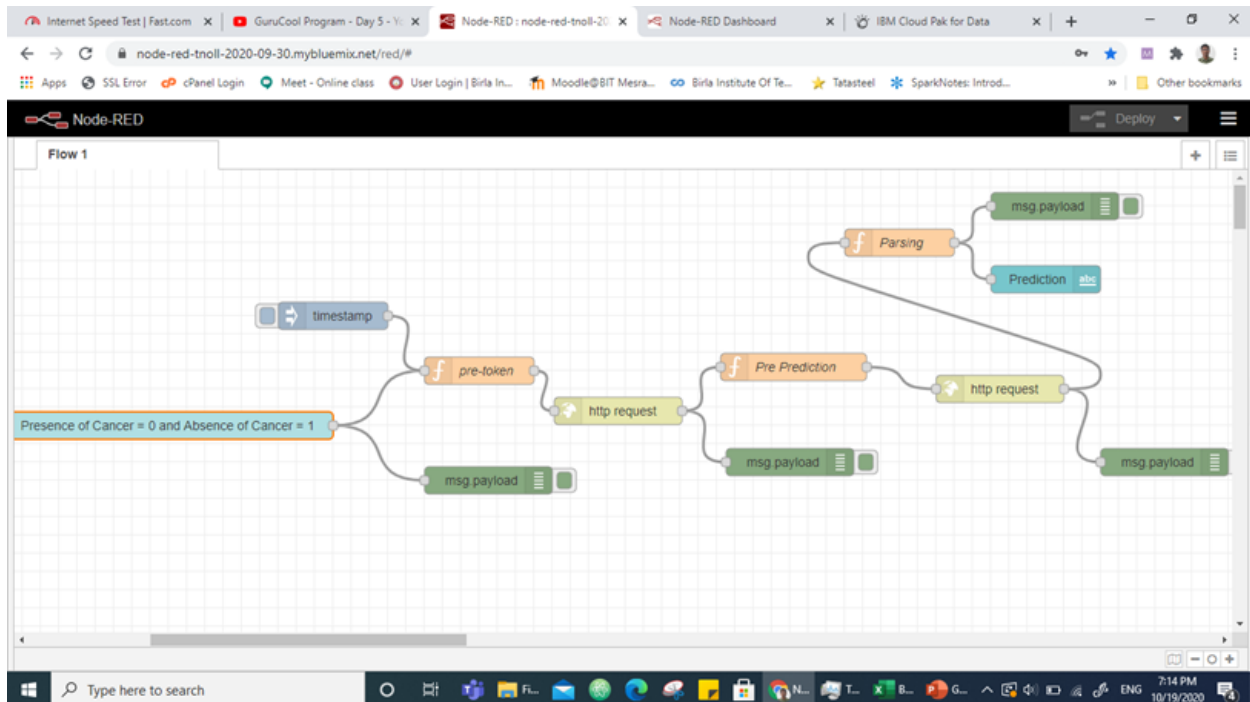
Node-Red User Interface

Node-Red UI Link:

https://node-red-tnoll-2020-09-30.mybluemix.net/ui/#!/0?socketid=yah10EoxDLzQ0Nds
AAA

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_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