

# Day 42

## DIY

### Q1. Problem Statement: Support Vector Machine (SVM) Algorithm

You are given a dataset – “*breast\_cancer.csv*.” Load the dataset into a DataFrame. Now, Considering the “*diagnosis*” column as the target variable, perform the following tasks:

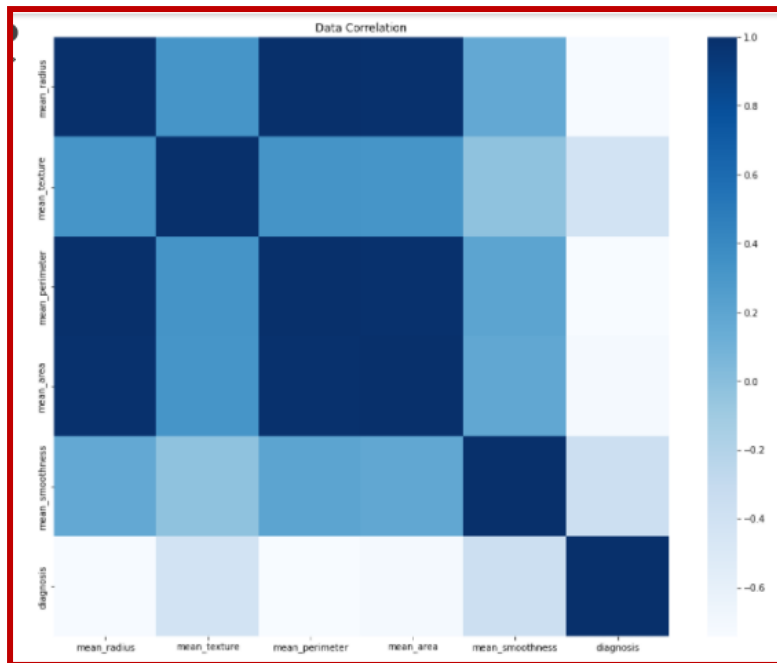
1. Declare feature vectors and the target variable and perform data analysis using correlation matrix and heatmap
2. Split the data into test and train fragments using `train_test_split()` function in 80:20 ratio (80% train, 20% test)
3. Build a Linear SVM and calculate the accuracy of the model
4. Build a Gaussian RBF SVM and calculate the accuracy of the model

### Dataset:

	mean_radius	mean_texture	mean_perimeter	mean_area	mean_smoothness	diagnosis
300	19.53	18.90	129.50	1217.0	0.11500	0
551	11.13	22.44	71.49	378.4	0.09566	1
9	12.46	24.04	83.97	475.9	0.11860	0
346	12.06	18.90	76.66	445.3	0.08386	1
121	18.66	17.12	121.40	1077.0	0.10540	0

### Sample Output:

1. Declare feature vectors and the target variable and perform data analysis using correlation matrix and heatmap.



- Split the data into test and train fragments using `train_test_split()` function in 80:20 ratio (80% train, 20% test)

```
After splitting the data-  
size of input train data is: 21872  
sizeof input test data is: 5504  
size of output train data is: 3744  
size of output test data is: 1016
```

- Build a Linear SVM and calculate the accuracy of the model.

```
Accuracy Score of Linear Model: 0.631578947368421
```

- Build a Gaussian RBF SVM and calculate the accuracy of the model.

```
Accuracy Score of Gaussian Model: 0.631578947368421
```