Breast cancer prediction

*Model name:*

Breast cancer prediction

*Aim:*

The aim of this project is to predict the breast cancer.

*Content/summary:*

Here before making the predictions we need to analyse the data first, Analysing the data is very impotant. Then to handle process, and Analysis of a big data set we require some complex and efficient algorithm and technology and that is data science.

After analyzing the dataset I did the cleaning process. This is very important because, if we didn’t clean the data and make predictions then we cannot get the accurate results because the data may contain null values, missing values, wrong values or they can be in some wrong format. So, I clean the data by removing null values which are not necessary.

After cleaning the data I visualize the data, because by seeing the big dataset we didn’t actually or clearly can no about data, so I used visualization techniques to make it much more effective for understanding purpose. Here I used the techniques like Heatmap, Count plot, Scatter Plot, Distribution Plot, Line Plot ,Reg plot, Lm Plot. So that I can easily identify which column is to predict.

After cleaning the data it can be ready for data manipulation. Here I take the necessary column for prediction by manipulating the data. After manipulating the data, I build the model. Training and testing takes major part in bulding the model, so based on training values I predict the output by testing it with some other random values.

Here I used DecisionTree for predicting the Breast-cancer by training it with the previous data. Here I tested the accuracy of the model.

*Problem Statement:*

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.

*Input to the Model:*

It contains id, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, compactness\_mean, concavity\_mean, concave points\_mean, symmetry\_mean, fractal\_dimension\_mean, radius\_se,texture\_se, perimeter\_se, area\_se, smoothness\_se, compactness\_se, concavity\_se, concave points\_se, symmetry\_se, fractal\_dimension\_se, radius\_worst,texture\_worst, perimeter\_worst, area\_worst, smoothness\_worst, compactness\_worst, concavity\_worst, concave points\_worst, symmetry\_worst, fractal\_dimension\_worst

*Output to the model:*

To predict the Diognosis.

*How it can be helpful:*

It can be useful to predict breast-cancer at early stages so it can be cured very quickly and easily at low cost.

Steps:

1. Analysing the dataset
2. Cleaning the dataset
3. Data visualization
4. Data manipulation
5. Model building
6. Prediction

Libraries:

1.Numpy

2.Pandas

3.Seaborn

4.Matplotlib

5. RandomForestClassifier

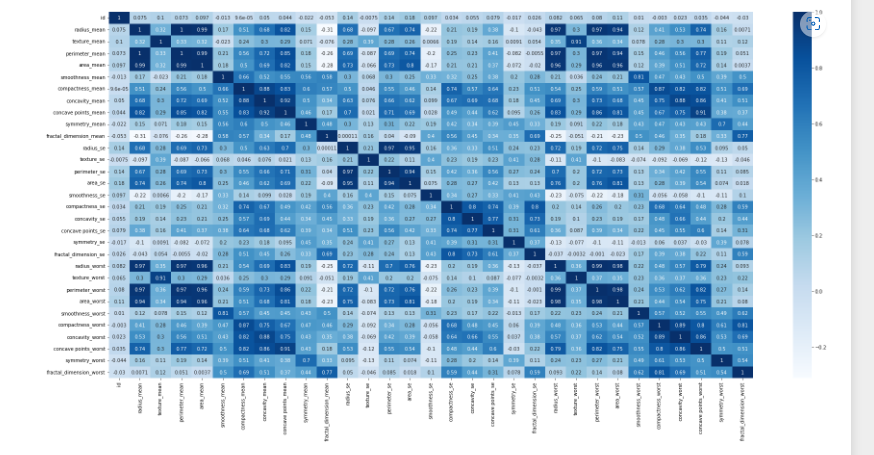
6. Scikit-learn

Data Visualization:

1.Heat plot:

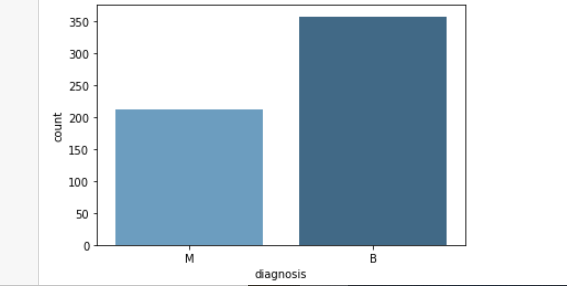
plt.figure(figsize=(30,14))

sns.heatmap(df.corr(),annot=True,cmap='Blues')



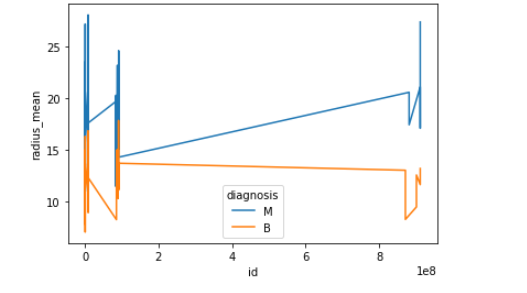
2.Count plot:

sns.countplot(x='diagnosis',data=df,palette='Blues\_d')



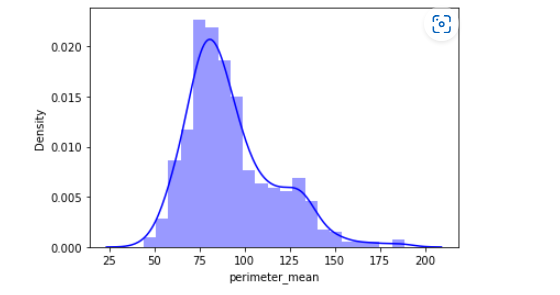
3.Lineplot:

sns.lineplot(x='id',y='radius\_mean',data=df,hue='diagnosis')



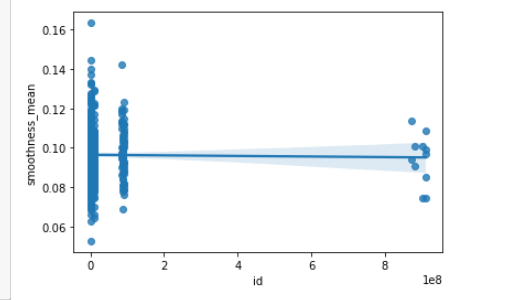
4.Distribution plot:

sns.distplot(df.perimeter\_mean,kde=True,color='b')



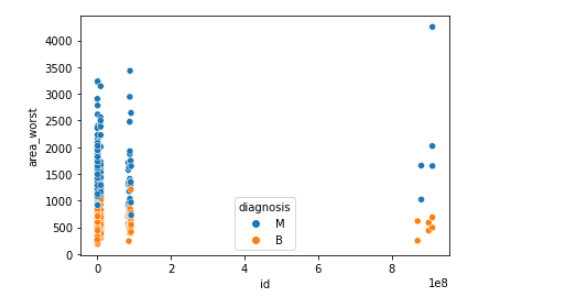
5.Regression plot:

sns.regplot(x='id',y='smoothness\_mean',data=df)



6.Scatter Plot:

sns.scatterplot(x='id',y='area\_worst',data=df,hue='diagnosis')



*Conclusion:*

We have multiple algorithms for training purposes like Decision tree, KNN,Support Vector Machine,Linear Regression etc…

Here I have used Decision tree for predicting the Breast-cancer.