Steps are:

1. [**Gathering Data**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#1)

Attribute Information:

1. ID number

* Diagnosis (M = malignant, B = benign)

Ten real-valued features are computed for each cell nucleus:

1. radius (mean of distances from center to points on the perimeter)

* texture (standard deviation of gray-scale values)
* perimeter
* area
* smoothness (local variation in radius lengths)
* compactness (perimeter^2 / area - 1.0)
* concavity (severity of concave portions of the contour)
* concave points (number of concave portions of the contour)
* symmetry
* fractal dimension ("coastline approximation" – 1)

we have data in zip file so first we unzip it .

After collecting data, we need to know what are the shape of this dataset

**(569, 33)**

* **2.** [**Exploratory Data Analysis**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#2)

**data.info()**

**it gives info about our data ,like (data type, null, no.of entry)**

**in our data we have no null values**

**To know how many unique values**

In [14]

data.diagnosis.value\_counts()

B 357

M 212

Name: diagnosis, dtype: int64

* **3.**[**Data Visualizations**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#3)

**creating a bar plot using the Python library, Matplotlib. Between b and m**

**creating a pairplot using the Seaborn library to visualize relationships between different pairs of variables in a dataset.**

**using the "hue" parameter to color the data points based on the "diagnosis" column**

**creates a scatter plot using Matplotlib and NumPy to visualize the relationship between the "texture\_mean" and "radius\_mean" columns in your dataset**

**Data Filtering**

Now, we have one categorical feature, so we need to convert it into numeric values using LabelEncoder from sklearn.preprocessing packages

#### **Find the correlation between other features, mean features only**

**Using heatmap we drown co-relation**

* **4.** [**Model Implementation.**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#4)

### Feature Selection

Select feature for predictions

* Take the dependent and independent feature for prediction

#### **Train Test Splitting**

* **Splite the dataset into TrainingSet and TestingSet by 30% and set the 15 fixed records**

**Perform Feature Standerd Scalling**

Standardize features by removing the mean and scaling to unit variance

The standard score of a sample x is calculated as:

* z = (x - u) / s
* **5.** [**ML Model Selecting and Model PredPrediction**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#5)

Now, we are ready to build our model for prediction, for the I made function for model building and preforming prediction and measure it's prediction and accuracy score.

#### **Arguments**

1. model => ML Model Object
2. Feature Training Set data
3. Feature Testing Set data
4. Targetd Training Set data
5. Targetd Testing Set data

**Model Implementing**

l

Now, Train the model one by one and show the classification report of particular models wise.

On the

**Precision, recall, f1-score ,support**

**For all this we have create and display confusion matrices**

**True Negative (TN) in (0,0)**

**False Positive (FP) in (0,1)**

**False Negative (FN) in (1,0)**

**True Positive (TP) in (1,1)**

### K-Fold Applying

### evaluating the performance of different machine learning models using cross-validation and displaying the accuracy scores for both the full dataset and the cross-validation scores

### Logistic Regression: Achieved a full-data accuracy of 0.9 and cross-validation scores ranging from 0.88 to 0.91.

### Random Forest Classifier: Achieved a perfect full-data accuracy of 1.0 and cross-validation scores ranging from 0.99 to 1.0.

### Decision Tree Classifier: Also achieved a perfect full-data accuracy of 1.0 and cross-validation scores of 1.0 for all folds.

### Support Vector Classifier (SVC): Achieved a full-data accuracy of 0.89 and cross-validation scores ranging from 0.88 to 0.90.

* **6.**[**HyperTunning the ML Model**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#6)

For HyperTunning we can use GridSearchCV to know the best performing parameters

* GridSearchCV implements a “fit” and a “score” method. It also implements “predict”, “predict\_proba”, “decision\_function”, “transform” and “inverse\_transform” if they are implemented in the estimator used.
* The parameters of the estimator used to apply these methods are optimized by cross-validated grid-search over a parameter grid.

**Observation**

Using this Algorithm, we can see that

* The best score is increases
* know the best estimator parametes for final model
* get the best parametes for it.
* [**Deploy Model**](https://www.kaggle.com/code/vikasukani/breast-cancer-prediction-using-machine-learning#7)