BREAST CANCER EXPLORATORY DATA ANALYSIS

THIS PROJECT ANALYSES THE DATA OF BREAST CANCER PATIENTS WITH UNIQUE IDs TO PRIDICT WEATHER THE CANCER IS BENIGN OR MALIGNANT

DATA COMPRISES OF "id","diagnosis","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","fractional\_dimensions\_se","radius\_worst","texture\_worst","perimeter\_worst","area\_worst","smoothness\_worst","compactness\_worst","concavity\_worst","concave points\_worst","symmetry\_worst","fractional\_dimension\_worst"

LANGUAGES USED :PYTHON

1. WE HAVE IMPORTED LIBRARIES LIKE PANDAS,NUMPY,SEABORN ETC
2. DATA SET IS IMPORTED AND READ

DATA ANALYSIS

1. INSPECTING AND CLEANING THE DATA IS DONE
2. FOR THIS WE HAVE USED THE FOLLOWING STEPS:

* Printing dimensions of data
* Viewing column heading
* Inspecting the target variable
* Identifying unique values
* Checking null values
* Checking missing values
* Viewing data statistics

DATA VISUALIZATION

finding corrolation between data

analyze target values

plotting corrolation,distribution of mean

plot bivarate relation

train test splitting

LOGISTIC REGRESSION

* 1. pridict train test data
  2. calculate accuracy

GAUSSION NAÏVE BAYES

* 1. pridict train test data
  2. calculate accuracy

DECISION TREE

* 1. import decision tree classifier
  2. hyperparameter optimization
  3. run the grid
  4. pridict train test data
  5. calculate the accuracy

RANDOM FOREST

* 1. import library of random forest classifier
  2. hyperparameter optimization
  3. run the grid
  4. pridict train test data
  5. calculate the accuracy

SUPPORT VECTOR MACHINE

* + 1. create support vector classifier

1. hyperparameter optimization
2. run the grid
3. pridict train test data
4. calculate the accuracy

K-NEAREST NEIGHBORS

* 1. import library of k-neighbors classifier model
  2. hyperparameter optimization
  3. run the grid
  4. pridict train test data
  5. calculate accuracy

EVALUATION AND COMPARISION OF ALL DATA MODELS

we evaluate and compare all the data analyzed in this step.