Breast Cancer Diagnosis

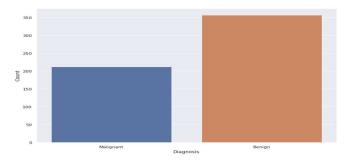| Team members	Contribution
Goutham Thota	Data analysis, Decision making, SVM
Tejaswi Chintapalli	Data analysis, Decision making, Random Forest
Shalaka Gadhani	Data analysis, Decision making, Logistic regression

Github link: https://github.com/Goutham-19/Breast-Cancer-Diagnosis.git
Cancer Dataset:

    Number of Instances: 569
    Number of Attributes: 30
    Target variable: 1
    Target classes: 'malignant', 'benign'
    Distribution of Target class:

Steps:
1. Since our target classes are imbalanced, we used stratified k-fold cross validation.
2. To predict the target class, we used SVM, Random Forest, Logistic regression
3. Hyper parameter tuning is performed to tune our parameters for the better model performance
4. Using grid search cv, we choose the optimum parameters
5. Next, we fitted our model using the optimal parameters
6. Finally, we predict our test data
7. All three ML models we used performed well
8. However, Random Forest outperforms the other models with the highest accuracy (0.988), precision (0.981), and recall (1.00).

	Model	Accuracy	Precision	recall
0	SVM	0.964912	0.981132	0.962963
1	RandomForest	0.988304	0.981818	1.000000
2	LogisticRegression	0.953216	0.962963	0.962963