

Assignment 3: Breast Cancer Diagnosis Classification Report

The objective of this report is to classify breast cancer diagnosis as malignant or benign using machine learning classifiers. The dataset used for this task contains features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass.

Data Loading and Preprocessing:

- The dataset "Assignment3-Breast-Cancer-Diagnose.csv" was loaded using pandas.
- Features and target variable were separated.
- Label encoding was applied to the target variable to encode 'M' (malignant) and 'B' (benign) as numerical values.

Classifier Implementation:

Three classifiers were implemented for the binary classification task:

1. **Decision Tree:** A single decision tree classifier.
2. **XGBoost:** An ensemble learning method known for its performance and scalability.
3. **Gradient Boosting (HistGradientBoostingClassifier):** A gradient boosting classifier that handles missing values.

Model Training and Evaluation:

- The dataset was split into training and testing sets using a 80-20 split.
- Each classifier was trained on the training data and evaluated on the testing data.
- Accuracy scores were computed for each classifier to evaluate their performance.

Encountered Errors:

Initially, the dataset contained missing values (NaNs), which caused errors when training the classifiers. Therefore, I replaced K-Nearest Neighbors (KNN) with Random Forest. However, it was later replaced with HistGradientBoostingClassifier to ensure compatibility with missing values and avoid errors.

Conclusion

In conclusion, the implemented classifiers achieved satisfactory accuracy in diagnosing breast cancer as malignant or benign. The use of HistGradientBoostingClassifier ensured compatibility with missing values in the dataset, resolving errors encountered during implementation. Further

optimizations and hyperparameter tuning could potentially improve the performance of the classifiers.

```
1 import pandas as pd

Decision Tree Accuracy: 0.9298245614835088
XGBoost Accuracy: 0.956140350877193
Gradient Boosting Accuracy: 0.9736842105263158

Decision Tree Accuracy: 0.9385964912280702

Decision Tree Classification Report:
precision    recall  f1-score   support
Benign       0.94      0.96      0.95        71
Malignant    0.93      0.91      0.92        43
accuracy      0.94      0.93      0.94       114
macro avg     0.94      0.93      0.93       114
weighted avg  0.94      0.94      0.94       114

XGBoost Accuracy: 0.956140350877193

XGBoost Classification Report:
precision    recall  f1-score   support
Benign       0.96      0.97      0.97        71
Malignant    0.95      0.93      0.94        43
accuracy      0.96      0.95      0.96       114
macro avg     0.96      0.95      0.95       114
weighted avg  0.96      0.96      0.96       114

Gradient Boosting Accuracy: 0.9736842105263158

Gradient Boosting Classification Report:
precision    recall  f1-score   support
Benign       0.97      0.99      0.98        71
Malignant    0.98      0.95      0.96        43
accuracy      0.97      0.97      0.97       114
macro avg     0.97      0.97      0.97       114
```