

Data Preparation

- A CSV file containing breast cancer diagnosis data is loaded.
- Missing values in the data are filled with the mean value of each feature using ``SimpleImputer``.

Data Splitting

- The data is split into training, validation, and test sets at a ratio of 60% / 20% / 20% using ``train_test_split``.
- The validation set is used to tune the hyperparameters of the model, while the test set remains independent for performance evaluation.

Classifier Selection and Hyperparameter Tuning

- The user is prompted to choose a classifier from three options: SVM, Logistic Regression, Random Forest.
- For the selected classifier, the corresponding model is initialized using ``Pipeline``, including feature scaling and the classifier itself.
- For each classifier, a grid of hyperparameters is defined to search for the best values using ``GridSearchCV`` and cross-validation on the training set.
- The best found parameters for the selected classifier are displayed.

Training and Performance Evaluation

- The selected classifier is trained using the best found parameters on the training set.
- The performance of the classifier is evaluated on the test set, computing metrics: Accuracy, Precision, Recall, F1 Score for the positive class 'M' (malignant).

```
Best parameters for Logistic Regression: {'classifier__C': 1}
Performance metrics for Logistic Regression:
Accuracy: 0.9824561403508771
Precision: 1.0
Recall: 0.9411764705882353
F1 Score: 0.9696969696969697
```

```
Best parameters for SVM: {'classifier__C': 1, 'classifier__kernel': 'rbf'}
Performance metrics for SVM:
Accuracy: 0.9824561403508771
Precision: 1.0
Recall: 0.9411764705882353
F1 Score: 0.9696969696969697
```

```
Best parameters for Random Forest: {'classifier__max_depth': 20, 'classifier__n_estimators': 100}
Performance metrics for Random Forest:
Accuracy: 0.9649122807017544
Precision: 0.9411764705882353
Recall: 0.9411764705882353
F1 Score: 0.9411764705882353
```