Data Preprocessing Report

Label Encoding: Transforms categorical variables into numerical form for model compatibility.

Data Splitting: Split data into 80% training and 20% testing using `train_test_split`, ensuring that models are trained on one subset and tested on another.

Pipelines:

Standard Scaler: Normalizes data for each model to improve performance and convergence.

Classifiers: Different classifiers are used in separate pipelines.

Hyperparameters:

Hyperparameters for Logistic Regression:

`{'classifier__C': [0.01, 0.1, 1, 10, 100]}`

Hyperparameters for Decision Tree:

`{'classifier__max_depth': [None, 10, 20, 30, 40, 50]}`

Hyperparameters for Random Forest:

`{'classifier__n_estimators': [10, 50, 100, 200], 'classifier__max_depth': [None, 10, 20, 30]}`

Hyperparameters for AdaBoost:

`{'classifier__n_estimators': [10, 50, 100, 200], 'classifier__learning_rate': [0.01, 0.1, 1, 10]}`

Hyperparameters for Gradient Boosting:

`{'classifier__n_estimators': [10, 50, 100, 200], 'classifier__learning_rate': [0.01, 0.1, 1, 10], 'classifier__max_depth': [3, 5, 7]}`

Model Training and Evaluation Report

GridSearchCV

Purpose: Finds the best hyperparameters for each model using cross-validation (`cv=5`).

Evaluation:

Best Parameters: Identified for each model.

Classification Report: Generated for each model using test data, providing precision, recall, f1-score, macro avg, weighted avg, and accuracy.

Detailed Analysis of Each Model

1. Logistic Regression

Logistic Regression Best Parameters: {'classifier__C': 0.1}

Logistic Regression Classification Report:

	precision	recall	f1-score	support
class 0	0.93	0.98	0.96	58
class 1	0.94	0.79	0.86	19
macro avg	0.94	0.89	0.91	77
weighted avg	0.94	0.94	0.93	77
accuracy	0.94			77

Precision:

- Class 0: 0.93 - Indicates that 93% of the predicted class 0 instances are correct.

- Class 1: 0.94 - Indicates that 94% of the predicted class 1 instances are correct.

Recall:

- Class 0: 0.98 - Shows that 98% of actual class 0 instances are correctly identified.

- Class 1: 0.79 - Shows that 79% of actual class 1 instances are correctly identified.

F1-Score:

- Class 0: 0.96 - Balanced measure of precision and recall for class 0, indicating strong performance.

- Class 1: 0.86 - Balanced measure of precision and recall for class 1, indicating decent performance.

Macro Average:

- Precision: 0.94 - Average precision across both classes, treating each class equally.

- Recall: 0.89 - Average recall across both classes, treating each class equally.

- F1-Score: 0.91 - Average F1-Score across both classes, treating each class equally.

Weighted Average:

- Precision: 0.94 - Average precision weighted by support, reflecting overall accuracy.

- Recall: 0.94 - Average recall weighted by support, reflecting overall identification rate.

- F1-Score: 0.93 - Average F1-Score weighted by support, reflecting overall balanced performance.

Accuracy:

- 0.94 - Indicates that 94% of the total instances are correctly classified.

Decision Tree Analysis

Decision Tree Best Parameters: {'classifier__max_depth': 10}
Decision Tree Classification Report:

	precision	recall	f1-score	support
class 0	0.96	0.90	0.93	58
class 1	0.74	0.89	0.81	19
macro avg	0.85	0.90	0.87	77
weighted avg	0.91	0.90	0.90	77
accuracy	0.90			77

Precision:

- Class 0: 0.96 - Indicates that 96% of the predicted class 0 instances are correct.

- Class 1: 0.74 - Indicates that 74% of the predicted class 1 instances are correct.

Recall:

- Class 0: 0.90 - Shows that 90% of actual class 0 instances are correctly identified.

- Class 1: 0.89 - Shows that 89% of actual class 1 instances are correctly identified.

F1-Score:

- Class 0: 0.93 - Balanced measure of precision and recall for class 0, indicating strong performance.

- Class 1: 0.81 - Balanced measure of precision and recall for class 1, indicating moderate performance.

Macro Average:

- **Precision:** 0.85 - Average precision across both classes, treating each class equally.

- **Recall:** 0.90 - Average recall across both classes, treating each class equally.

- **F1-Score:** 0.87 - Average F1-Score across both classes, treating each class equally.

Weighted Average:

- **Precision:** 0.91 - Average precision weighted by support, reflecting overall accuracy.

- **Recall:** 0.90 - Average recall weighted by support, reflecting overall identification rate.

- **F1-Score:** 0.90 - Average F1-Score weighted by support, reflecting overall balanced performance.

Accuracy:

- 0.90 - Indicates that 90% of the total instances are correctly classified.

Random Forest Analysis

Random Forest Best Parameters: {'classifier__max_depth': None,
'classifier__n_estimators': 200}

Random Forest Classification Report:

	precision	recall	f1-score	support
class 0	0.98	1.00	0.99	58
class 1	1.00	0.95	0.97	19
macro avg	0.99	0.97	0.98	77
weighted avg	0.99	0.99	0.99	77
accuracy	0.99			77

Precision:

- Class 0: 0.98 - Indicates that 98% of the predicted class 0 instances are correct.

- Class 1: 1.00 - Indicates that 100% of the predicted class 1 instances are correct.

Recall:

- Class 0: 1.00 - Shows that 100% of actual class 0 instances are correctly identified.

- Class 1: 0.95 - Shows that 95% of actual class 1 instances are correctly identified.

F1-Score:

- Class 0: 0.99 - Balanced measure of precision and recall for class 0, indicating near-perfect performance.

- Class 1: 0.97 - Balanced measure of precision and recall for class 1, indicating excellent performance.

Macro Average:

- **Precision:** 0.99 - Average precision across both classes, treating each class equally.

- **Recall:** 0.97 - Average recall across both classes, treating each class equally.

- **F1-Score:** 0.98 - Average F1-Score across both classes, treating each class equally.

Weighted Average:

- **Precision:** 0.99 - Average precision weighted by support, reflecting overall accuracy.

- **Recall:** 0.99 - Average recall weighted by support, reflecting overall identification rate.

- **F1-Score:** 0.99 - Average F1-Score weighted by support, reflecting overall balanced performance.

Accuracy:

- 0.99 - Indicates that 99% of the total instances are correctly classified.

AdaBoost Analysis

AdaBoost Best Parameters: {'classifier__learning_rate': 0.1,
'classifier__n_estimators': 50}

AdaBoost Classification Report:

	precision	recall	f1-score	support
class 0	0.95	1.00	0.97	58
class 1	1.00	0.84	0.91	19
macro avg	0.98	0.92	0.94	77
weighted avg	0.96	0.96	0.96	77
accuracy	0.96			77

Precision:

- Class 0: 0.95 - Indicates that 95% of the predicted class 0 instances are correct.

- Class 1: 1.00 - Indicates that 100% of the predicted class 1 instances are correct.

Recall:

- Class 0: 1.00 - Shows that 100% of actual class 0 instances are correctly identified.

- Class 1: 0.84 - Shows that 84% of actual class 1 instances are correctly identified.

F1-Score:

- Class 0: 0.97 - Balanced measure of precision and recall for class 0, indicating strong performance.

- Class 1: 0.91 - Balanced measure of precision and recall for class 1, indicating good performance.

Macro Average:

- **Precision:** 0.98 - Average precision across both classes, treating each class equally.

- **Recall:** 0.92 - Average recall across both classes, treating each class equally.

- **F1-Score:** 0.94 - Average F1-Score across both classes, treating each class equally.

Weighted Average:

- **Precision:** 0.96 - Average precision weighted by support, reflecting overall accuracy.

- **Recall:** 0.96 - Average recall weighted by support, reflecting overall identification rate.

- **F1-Score:** 0.96 - Average F1-Score weighted by support, reflecting overall balanced performance.

Accuracy:

- 0.96 - Indicates that 96% of the total instances are correctly classified.

Gradient Boosting Analysis

Gradient Boosting Best Parameters: {'classifier__learning_rate': 0.1, 'classifier__max_depth': 3, 'classifier__n_estimators': 100}

Gradient Boosting Classification Report:

	precision	recall	f1-score	support
class 0	0.98	0.98	0.98	58
class 1	0.95	0.95	0.95	19
macro avg	0.97	0.97	0.97	77
weighted avg	0.97	0.97	0.97	77
accuracy	0.97			77

Precision:

- Class 0: 0.98 - Indicates that 98% of the predicted class 0 instances are correct.

- Class 1: 0.95 - Indicates that 95% of the predicted class 1 instances are correct.

Recall:

- Class 0: 0.98 - Shows that 98% of actual class 0 instances are correctly identified.

- Class 1: 0.95 - Shows that 95% of actual class 1 instances are correctly identified.

F1-Score:

- Class 0: 0.98 - Balanced measure of precision and recall for class 0, indicating strong performance.

- Class 1: 0.95 - Balanced measure of precision and recall for class 1, indicating excellent performance.

Macro Average:

- **Precision:** 0.97 - Average precision across both classes, treating each class equally.

- **Recall:** 0.97 - Average recall across both classes, treating each class equally.

- **F1-Score:** 0.97 - Average F1-Score across both classes, treating each class equally.

Weighted Average:

- **Precision:** 0.97 - Average precision weighted by support, reflecting overall accuracy.

- **Recall:** 0.97 - Average recall weighted by support, reflecting overall identification rate.

- **F1-Score:** 0.97 - Average F1-Score weighted by support, reflecting overall balanced performance.

Accuracy:

- 0.97 - Indicates that 97% of the total instances are correctly classified.

Conclusion

Logistic Regression: Balanced performance with high precision and recall, suitable for general use. (4)

Decision Tree: Good but slightly less balanced, with some trade-offs in precision and recall for class 1. (5)

Random Forest: Best overall performance, with near-perfect metrics, highly reliable. (1)

AdaBoost: High precision, especially for class 1, with strong overall performance. (3)

Gradient Boosting: Excellent and balanced performance, making it a very reliable model. (2)