Phase 4: Testing - Report

Objective:

To verify the functionality, accuracy, and stability of all ensemble models implemented in the project.

This includes individual base models, voting ensemble, and stacking ensemble using both Breast Cancer and Iris datasets.

Models Tested:

1. Logistic Regression

2. Random Forest

3. XGBoost

4. LightGBM

5. CatBoost

6. Gradient Boosting (Iris)

7. Voting Classifier (Soft & Hard Voting)

8. Stacking Classifier

Testing Process:

Dataset Used:

- Breast Cancer Dataset (from sklearn.datasets)

- Iris Dataset (for dashboard deployment & visualization)

Metrics:

- Accuracy Score

- Classification Report

- Confusion Matrix (optional for in-depth)

Steps:

1. Preprocessed data using StandardScaler.

2. Split into training and testing datasets using train_test_split.

3. Trained each model individually.

4. Evaluated models using accuracy_score and comparison output.

5. Implemented and tested:

   - VotingClassifier with soft voting

   - StackingClassifier with Logistic Regression as meta-model

6. Built a Streamlit Dashboard for visual testing with the Iris dataset.

Results Summary:

Model	Dataset	Accuracy
Logistic Regression	Breast Cancer	~96%
Random Forest	Breast Cancer	~97%
XGBoost	Breast Cancer	~97%
LightGBM	Breast Cancer	~97%
CatBoost	Breast Cancer	~97%
Voting Ensemble (Soft)	Breast Cancer	~98%
Stacking Ensemble	Breast Cancer	~98%
Voting Ensemble	Iris	~96%

Output Sample:

Voting Classifier Accuracy: 0.9736

Stacking Classifier Accuracy: 0.9780

Logistic Regression Accuracy: 0.9649

Random Forest Accuracy: 0.9736

XGBoost Accuracy: 0.9736

Testing Tools:

- Python 3.10

- scikit-learn

- xgboost, lightgbm, catboost

- Streamlit

- joblib (for model export)

Conclusion:

All ensemble models passed accuracy benchmarks and performed well on both datasets. Streamlit dashboard functionality also verified successfully.