Final Model Comparison Report

Dataset Overview

- 2500 synthetic candidate profiles were generated with realistic feature distributions.
- Key Features included technical test scores, coding scores, communication skills, ATS resume scores, etc.
- Success target variable was generated with weighted logic and randomness for realism.

Data Preprocessing

- Only numeric features were selected.
- StandardScaler was used for scaling the features.
- Data was split into 80% training and 20% testing sets.

Models Trained

- Logistic Regression
- Random Forest Classifier
- XGBoost Classifier

Hyperparameter Tuning

- Hyperparameters tuned using RandomizedSearchCV and GridSearchCV.
- Best parameters were selected for final model retraining.

Final Model Performances

Logistic Regression:

- Accuracy: 86.56%
- Precision (0/1): 0.88 / 0.85
- Recall (0/1): 0.87 / 0.86
- F1-Score (0/1): 0.87 / 0.86

Final Model Comparison Report

Random Forest:

- Accuracy: 85.44%
- Precision (0/1): 0.87 / 0.84
- Recall (0/1): 0.86 / 0.85
- F1-Score (0/1): 0.86 / 0.85

XGBoost:

- Accuracy: 85.28%
- Precision (0/1): 0.87 / 0.84
- Recall (0/1): 0.85 / 0.85
- F1-Score (0/1): 0.86 / 0.85

Conclusion

- Logistic Regression achieved the highest overall performance in accuracy and f1-score.
- Random Forest and XGBoost also performed strongly but slightly behind LR.
- Further improvements could involve feature engineering or balancing techniques like SMOTE if needed.