(venv) priyanshuparashar@Priyanshu-2 Fake_job_detection % python train_model.py

📥 Loading dataset... (0%)

✅ Dataset loaded! (10%)

✅ Data cleaned! (20%)

✅ Categorical features encoded! (30%)

⏳ Generating BERT embeddings... (40%)

100%|[redacted]| 15000/15000 [11:21<00:00, 22.00it/s]

✅ BERT embeddings generated! (50%)

✅ Features combined! (60%)

🔄 Applying SMOTE... (65%)

/Users/priyanshuparashar/Fake_job_detection/venv/lib/python3.9/site-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6 and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of the scikit-learn developer API.

 warnings.warn(

✅ SMOTE applied! (70%)

✅ Data scaled! (75%)

✅ Train-test split done! (80%)

🚀 Training Random Forest... (85%)

Fitting 3 folds for each of 5 candidates, totalling 15 fits

✅ Best RF Params: {'n_estimators': 400, 'min_samples_split': 3, 'max_depth': 25} (90%)

🚀 Training XGBoost... (92%)

Fitting 3 folds for each of 8 candidates, totalling 24 fits

[CV] END ..learning_rate=0.02, max_depth=6, n_estimators=100; total time= 26.3s

[CV] END ..learning_rate=0.02, max_depth=6, n_estimators=100; total time= 26.5s

[CV] END ..learning_rate=0.02, max_depth=6, n_estimators=100; total time= 26.7s

[CV] END ..learning_rate=0.02, max_depth=6, n_estimators=200; total time= 1.1min

[CV] END ..learning_rate=0.02, max_depth=6, n_estimators=200; total time= 1.1min

[CV] END ..learning_rate=0.02, max_depth=6, n_estimators=200; total time= 1.1min

[CV] END ..learning_rate=0.05, max_depth=6, n_estimators=100; total time= 44.2s

[CV] END ..learning_rate=0.05, max_depth=6, n_estimators=100; total time= 45.6s

[CV] END ..learning_rate=0.05, max_depth=6, n_estimators=100; total time= 46.0s

[CV] END ..learning_rate=0.05, max_depth=6, n_estimators=200; total time= 1.6min

[CV] END ..learning_rate=0.05, max_depth=6, n_estimators=200; total time= 1.5min

[CV] END ..learning_rate=0.05, max_depth=6, n_estimators=200; total time= 1.6min

[CV] END .learning_rate=0.02, max_depth=10, n_estimators=100; total time= 5.7min

[CV] END .learning_rate=0.02, max_depth=10, n_estimators=100; total time= 6.1min

[CV] END .learning_rate=0.02, max_depth=10, n_estimators=100; total time= 6.5min

[CV] END .learning_rate=0.05, max_depth=10, n_estimators=100; total time= 8.0min

[CV] END .learning_rate=0.05, max_depth=10, n_estimators=100; total time= 8.2min

[CV] END .learning_rate=0.05, max_depth=10, n_estimators=100; total time= 8.8min

[CV] END .learning_rate=0.02, max_depth=10, n_estimators=200; total time=14.0min

[CV] END .learning_rate=0.02, max_depth=10, n_estimators=200; total time=13.6min

[CV] END .learning_rate=0.02, max_depth=10, n_estimators=200; total time=14.2min

[CV] END .learning_rate=0.05, max_depth=10, n_estimators=200; total time= 9.1min

[CV] END .learning_rate=0.05, max_depth=10, n_estimators=200; total time= 8.5min

[CV] END .learning_rate=0.05, max_depth=10, n_estimators=200; total time= 3.5min

✅ Best XGBoost Params: {'learning_rate': 0.05, 'max_depth': 10, 'n_estimators': 200} (93%)

✅ XGBoost trained! (95%)

🚀 Training Hybrid Stacking Model... (97%)

✅ Stacking Model trained! (98%)

✅ Hybrid Model Accuracy: 0.8335 (99%)

🎯 Model & artifacts saved for Streamlit! (100%)

(venv) priyanshuparashar@Priyanshu-2 Fake_job_detection %