EasyVisa – Visa Approval Prediction

Machine learning classification | Author: Sayan Nandi

Executive Summary

Built and evaluated supervised and ensemble machine learning models to predict whether a visa application is likely to be certified or denied. The solution emphasizes strong evaluation under class imbalance and interpretable outputs for real-world decision support.

Problem	Binary classification (imbalanced)
Best Performance	F1 = 0.83 | ROC-AUC = 0.79
Primary Metric	F1-score
Explainability	SHAP

Method Snapshot

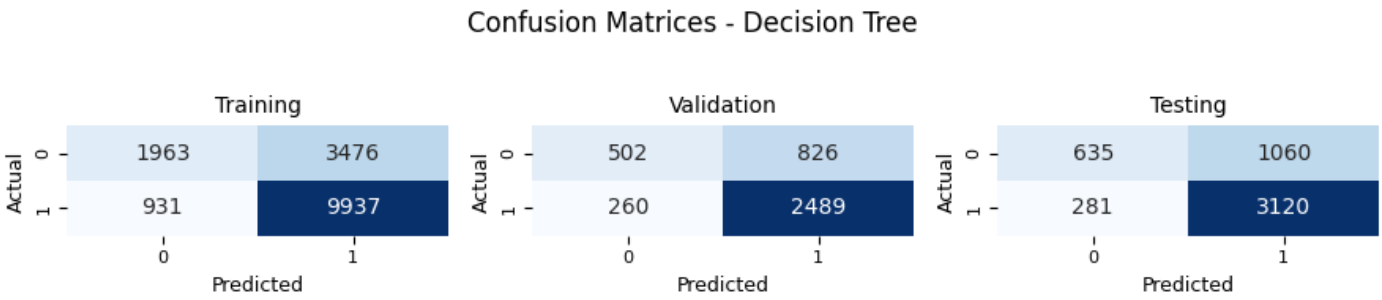
- Exploratory data analysis and preprocessing
- Feature engineering and categorical encoding
- Imbalanced classification using ensemble models
- Cross-validation and threshold tuning
- Explainability using SHAP

Data source: Dataset provided by Great Learning (PG Program in Data Science and Business Analytics) for academic use.

EasyVisa – Visa Approval Prediction

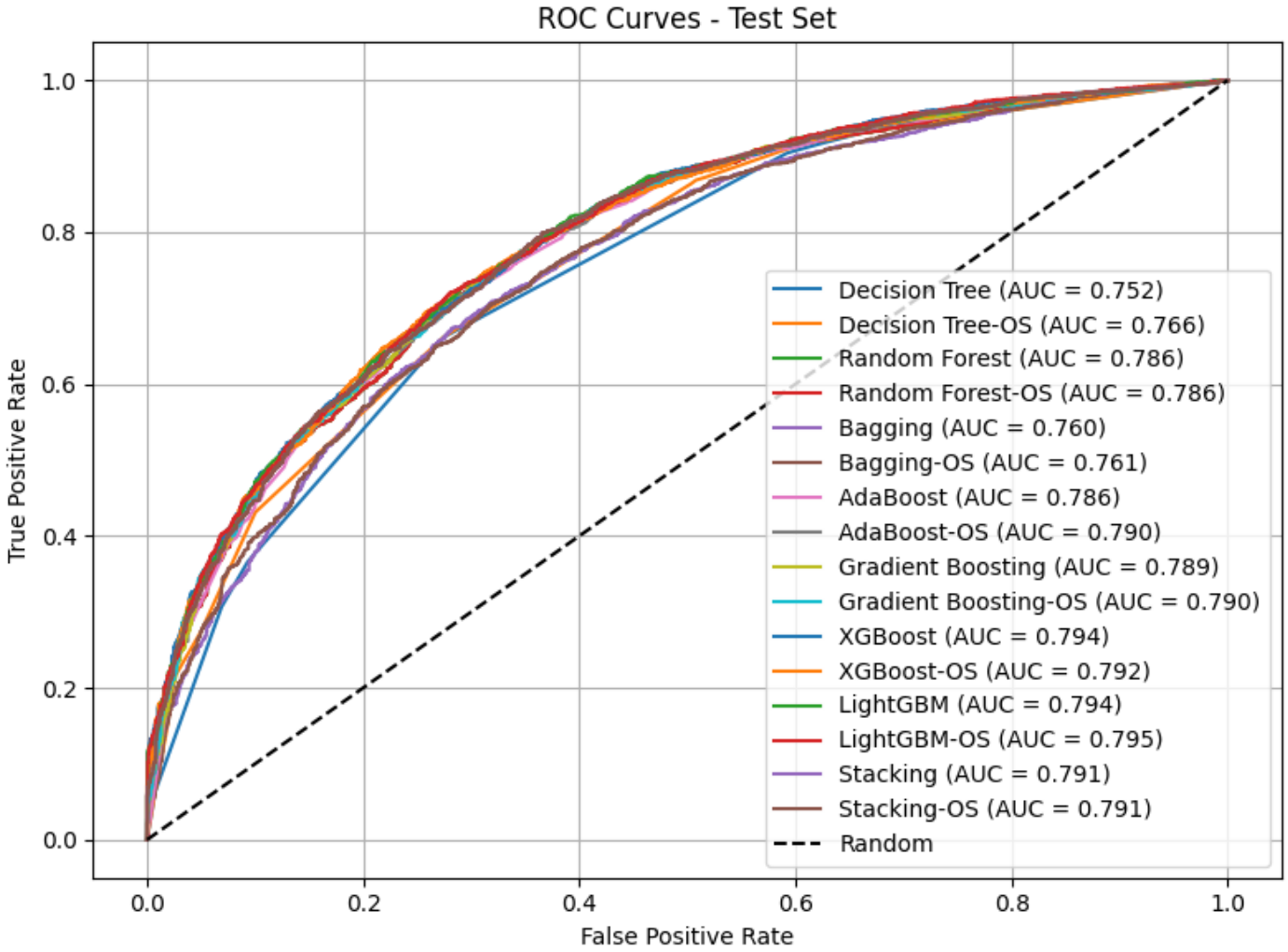
Modeling & Evaluation

Confusion Matrix Analysis



Confusion matrices (Train / Validation / Test)

ROC Curve Evaluation



ROC curves across evaluated models

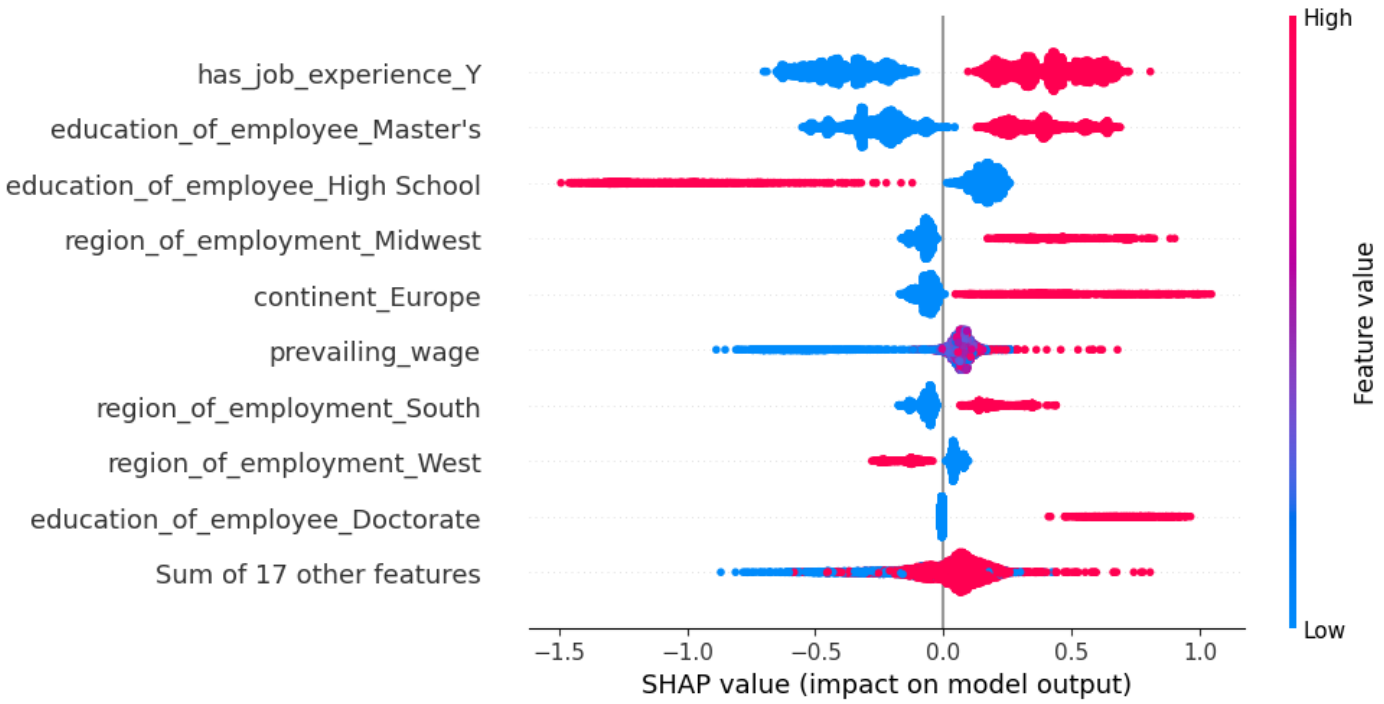
Key Takeaways

- Ensemble models outperform single estimators
- F1-score balances precision and recall under imbalance
- ROC-AUC confirms strong ranking performance

EasyVisa – Visa Approval Prediction

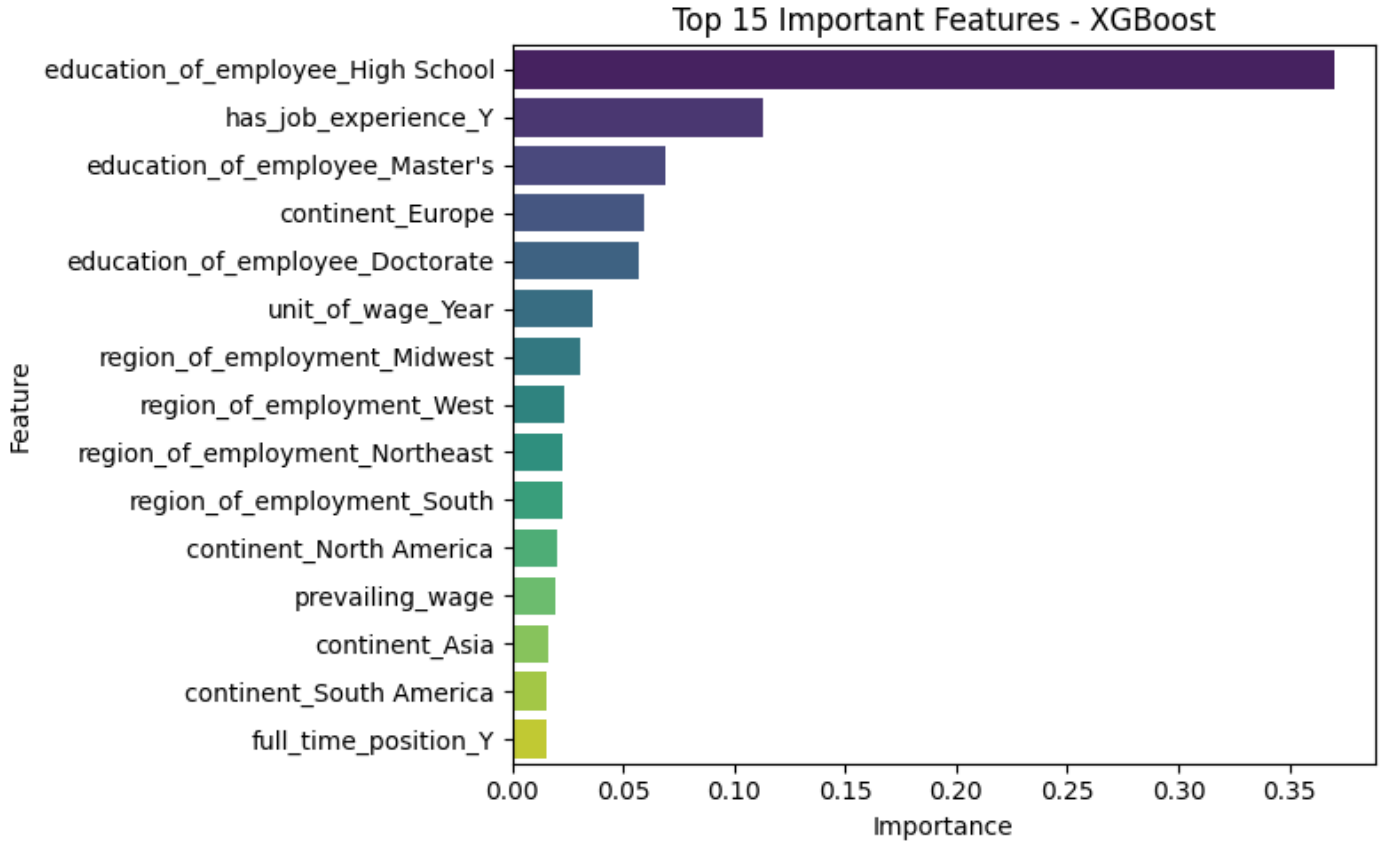
Explainability & Recommendations

SHAP Summary



SHAP summary of feature impact

Feature Importance



XGBoost feature importance

Business Recommendations

- Use model scores to prioritize review queues
- Monitor drift in key drivers and retrain periodically
- Deploy as batch scoring pipeline with explanations
- Log predictions and SHAP values for auditability