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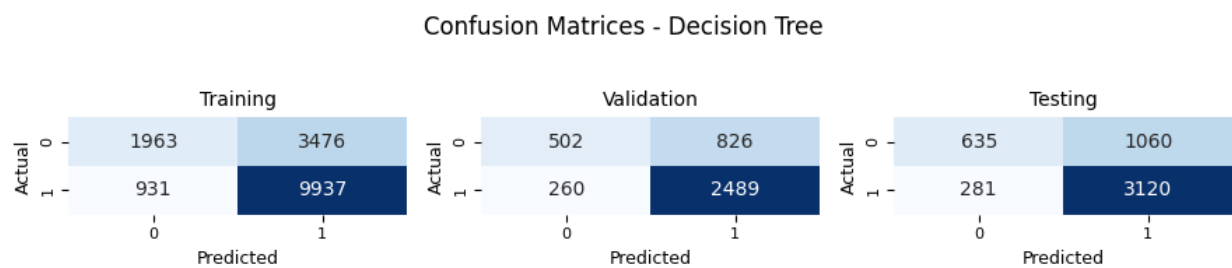
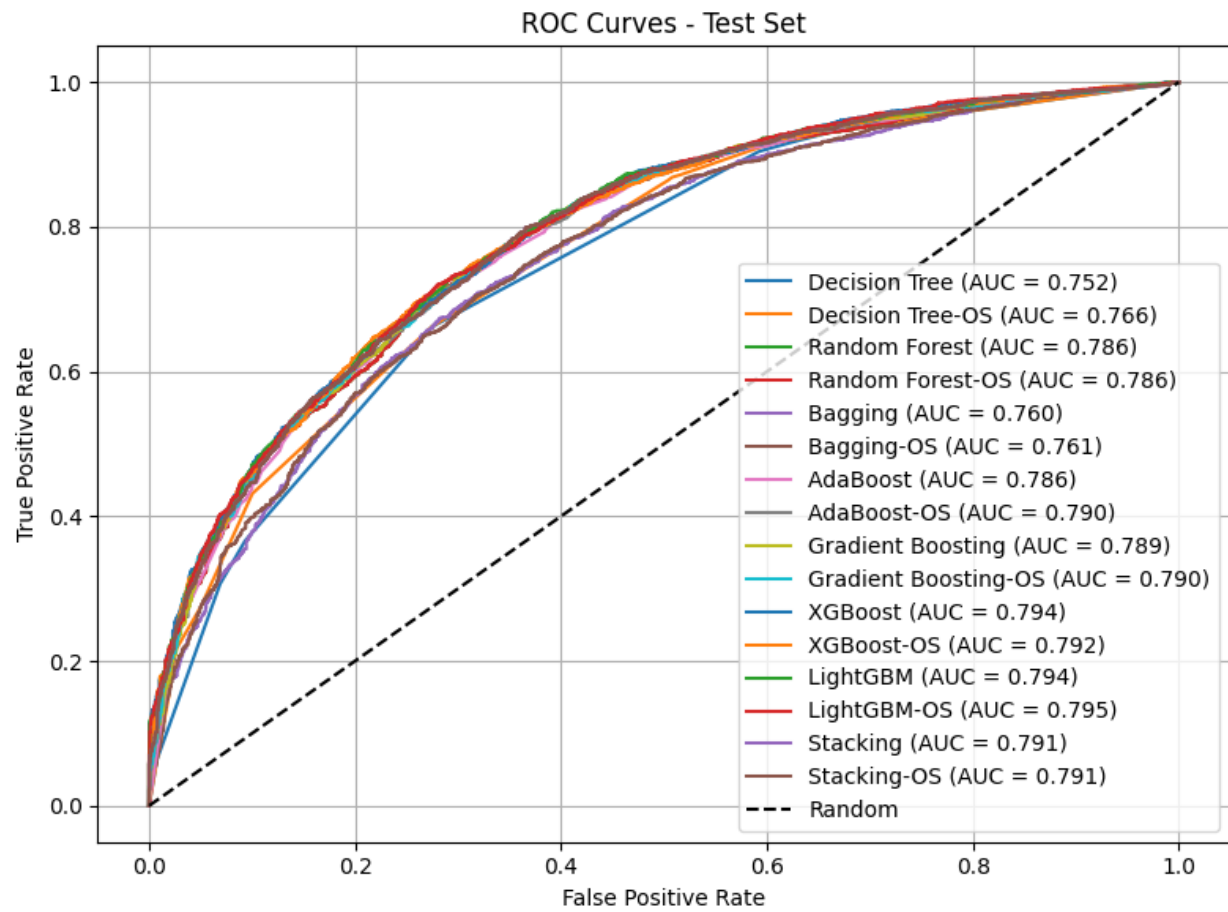| Problem Type	Binary classification (imbalanced)
Best Performance	F1 = 0.83, ROC-AUC = 0.79
Primary Metric	F1-score
Explainability	SHAP

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

Modeling & Evaluation

Confusion Matrix Analysis

Confusion Matrices - Decision Tree

ROC Curve Evaluation

ROC Curves - Test Set

Legend:
- Decision Tree (AUC = 0.752)
- Decision Tree-OS (AUC = 0.766)
- Random Forest (AUC = 0.786)
- Random Forest-OS (AUC = 0.786)
- Bagging (AUC = 0.760)
- Bagging-OS (AUC = 0.761)
- AdaBoost (AUC = 0.786)
- AdaBoost-OS (AUC = 0.790)
- Gradient Boosting (AUC = 0.789)
- Gradient Boosting-OS (AUC = 0.790)
- XGBoost (AUC = 0.794)
- XGBoost-OS (AUC = 0.792)
- LightGBM (AUC = 0.794)
- LightGBM-OS (AUC = 0.795)
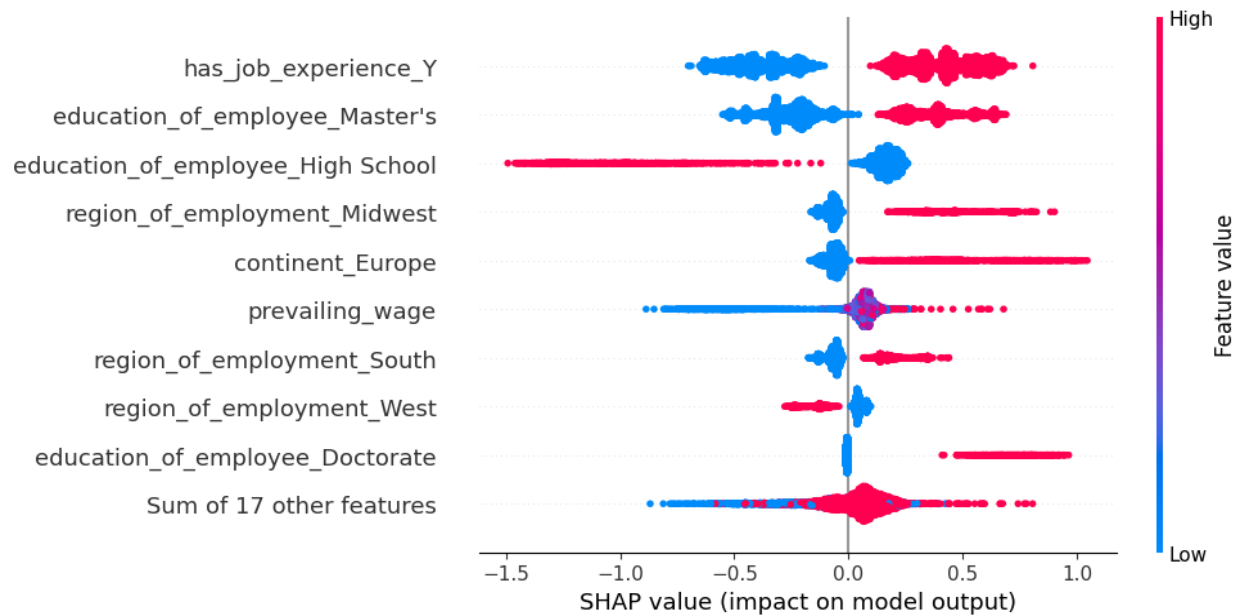- Stacking (AUC = 0.791)
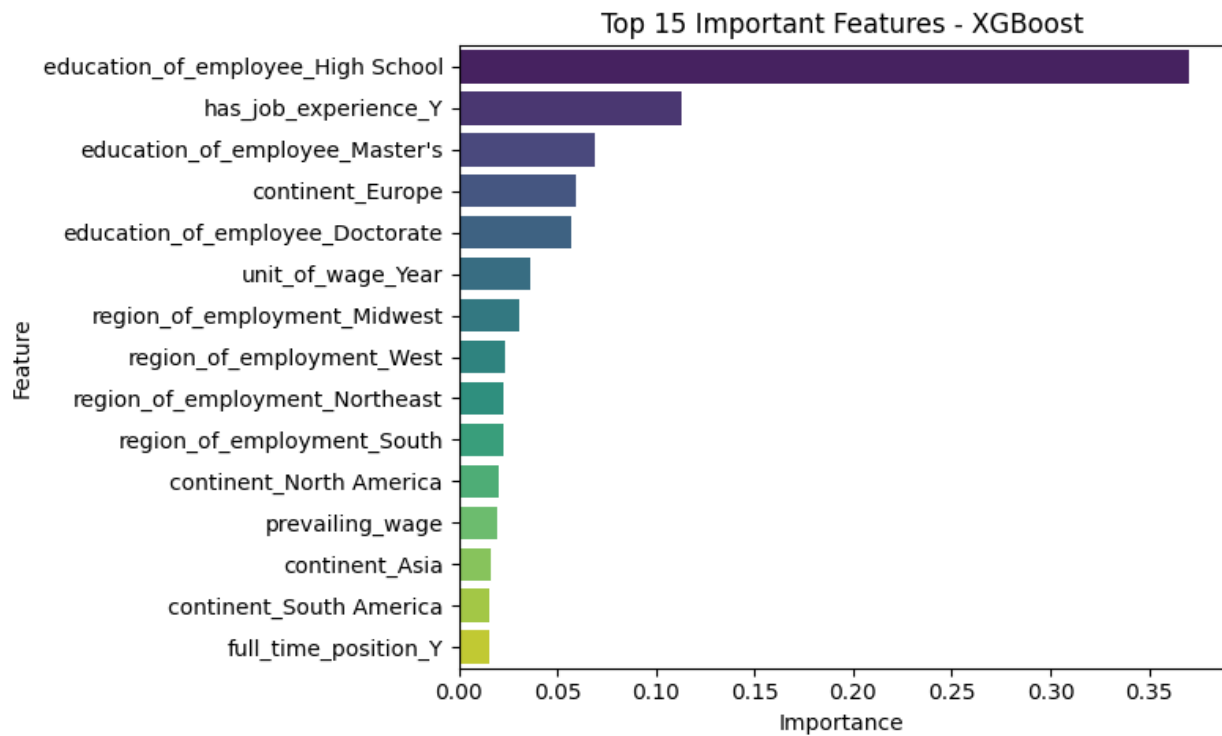- Stacking-OS (AUC = 0.791)
- Random

Key Takeaways

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

Explainability & Recommendations

SHAP Summary

Feature Importance

Business Recommendations

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

Data Source & Usage

Dataset provided by Great Learning as part of the Post Graduate Program in Data Science and Business Analytics. Used strictly for academic and portfolio demonstration purposes. License applies only to code and documentation; data rights remain with the original provider.