ADA Boost With Decision TREES

Code Description:

This code builds a credit risk prediction model using **AdaBoost** with **Decision Trees** as base learners. It follows a structured workflow:

1. **Data Preprocessing:**
   Missing values in numeric columns are filled with median values, and categorical features are encoded using Label Encoding.
2. **Train-Test Split:**
   The dataset is split into training (80%) and testing (20%) sets with stratification to preserve class distribution.
3. **Hyperparameter Tuning:**
   RandomizedSearchCV is used to optimize AdaBoost hyperparameters including number of estimators, learning rate, and base tree depth, maximizing ROC AUC over 3-fold cross-validation.
4. **Model Evaluation:**
   The best model is evaluated on the test set using metrics such as accuracy, F1 score, ROC AUC, and a classification report.
5. **Feature Importance:**
   Permutation importance is calculated to interpret the model, highlighting the top features influencing predictions.
6. **Data Drift Simulation & Monitoring:**
   The code simulates data drift by altering features such as loan intent and interest rate, then calculates the Population Stability Index (PSI) to quantify prediction distribution shift.
7. **Automatic Retraining:**
   If the PSI exceeds a threshold (0.2), indicating significant drift, the model is retrained on the drifted data to maintain performance.

8. **Results Export:**
   Final predictions with actual labels and probabilities are saved and downloaded for further analysis.

Results Description:

- The AdaBoost model, after hyperparameter tuning, achieves strong predictive performance on unseen test data, reflected by high accuracy, F1 score, and ROC AUC values.
- Permutation feature importance reveals which input variables most strongly impact loan default predictions, aiding interpretability and potential domain insights.
- Simulated drift testing shows how changes in key feature distributions (e.g., loan intent and interest rates) affect model output.
- The PSI score quantifies prediction shift; if drift is significant (PSI > 0.2), retraining successfully restores model effectiveness, ensuring robustness in dynamic environments.
- Overall, the approach combines predictive accuracy with ongoing monitoring and adaptive retraining, supporting reliable credit risk assessment in production settings.