Main Stacked Ensemble Model

Step-by-Step Explanation

1. **Data Loading & Preprocessing**
    - Dataset is loaded and categorical features are label-encoded.
    - Missing values in the target variable loan_status are dropped.
    - Numeric features are standardized for model stability.
    - Dataset is split into stratified train and test sets (70/30) to preserve class balance.
2. **Population Stability Index (PSI) Calculation**
    - PSI measures distribution shifts of individual features between training and test sets.
    - High PSI values indicate feature drift, meaning the test data distribution deviates significantly from training.
    - The top features by PSI are printed to flag potential data drift.
3. **Y-Drift (Target Distribution Shift)**
    - The default rate (loan_status mean) in train and test sets is compared.
    - A significant difference signals potential shifts in borrower risk behavior over time.
4. **Adversarial Validation**
    - A logistic regression model is trained to distinguish train vs test samples.
    - If this adversarial model achieves high AUC (>0.6), it indicates the two datasets come from different distributions — a strong sign of drift.
5. **Hyperparameter Tuning with Optuna for XGBoost**
    - Optuna efficiently searches for optimal XGBoost hyperparameters using cross-validated AUC as the objective.

- This step results in a finely tuned XGBoost model that balances complexity and generalization.

6. **SHAP-based Feature Selection**
    - SHAP values are computed on the trained XGBoost model to quantify feature importance.
    - The top 30 most impactful features are selected, focusing the final model on stable, informative inputs.
7. **Stacked Ensemble & Calibration**
    - Three base models — Logistic Regression, Random Forest, and XGBoost — are combined via stacking with Ridge Classifier as meta-learner.
    - The stacked model's output probabilities are calibrated using Platt scaling (sigmoid method) for well-calibrated risk scores.
8. **Performance Metrics on Test Set**
    - Final evaluation metrics include AUC (~model discrimination), Brier Score (probability accuracy), Precision, F1 Score, and a full classification report.
    - These metrics quantify how well the model predicts defaults in unseen data.
9. **Model Versioning & Retraining Trigger**
    - If PSI, adversarial AUC, or Y-drift exceed preset thresholds, retraining is automatically triggered on the latest test data.
    - The retrained model is saved as a new version for traceability.
    - Otherwise, the current model is saved as the stable version.
10. **Online Learning with River (Incremental Updates)**
    - A River pipeline performs online logistic regression with incremental learning on test data points, simulating real-time adaptation to new borrower data.
    - Online AUC is reported as a continuous performance indicator.

Key Results & Insights

- **Drift Detection:**
  The combined use of PSI, adversarial validation, and Y-drift monitoring provides a

comprehensive picture of distributional changes at both feature and target levels. This multi-pronged approach helps detect even subtle drifts that may degrade model performance in production.

- **Model Performance:**
  The final stacked, calibrated ensemble achieves strong predictive power and well-calibrated probabilities, essential for credit risk decision-making where risk estimates impact financial outcomes directly.
- **Feature Stability & Explainability:**
  SHAP-based feature selection ensures interpretability and focuses the model on stable predictors, enhancing trustworthiness for compliance and regulatory needs.
- **Automated Retraining:**
  Auto-triggered retraining keeps the model up-to-date without manual intervention, enabling scalability across regions or time periods.
- **Real-time Adaptability:**
  Incorporating River for online learning means the system can adapt incrementally, further improving resilience in dynamic environments like lending markets.

Business Impact & Strategic Fit

This final credit risk model is much more than a standard predictive classifier — it is a **self-aware, adaptive credit scoring engine** designed for Zest AI's global deployment:

- Ensures **fair, accurate lending decisions** across diverse borrower populations and evolving economic conditions.
- Reduces operational risk by proactively detecting model degradation.
- Maintains **regulatory compliance** via explainable, stable feature selection.
- Enables **cost-efficient scalability** through automated retraining and online learning.
- Aligns perfectly with Zest AI's mission to democratize credit through **transparent, data-driven risk assessment.**