Machine Learning Models

March 28 2025

1. introduction:
   - Project Objective:
     - Develop predictive model that accurately classifies loan applications as good or bad based on applicant attributes.
     - Ensure Interpretability so financial institutions can justify decisions to regulators and customers.
     - optimize fairness to avoid bias against certain demographic groups.
   - Goal:
   - Dataset Description:
     - Number of Features: 42
     - Target Variables: bad_flag
     - Pre-Processing: Removing columns with more than 80% missing values, capping outliers, scaling features, encoding categorical features, removing correlated features with high correlation threshold.
   - Machine Learning Models:
     - Logistic Regression
     - Random Forest
     - XGBoost
     - Stacking Model (base models are RF and XGBoost with Logistic Regression as the meta-learner).
   - Model Evaluation

     Due to the highly imbalanced nature of our dataset where bad loans account for less than 5% we prioritized precision-recall (PR) metrics over traditional accuracy and ROC AUC. While the ROC AUC of 0.81 suggests good general discrimination, it is less informative under class imbalance.

     Instead, we focused on the PR AUC and the recall for the bad loan class. In credit risk modeling, maximizing recall for bad loans is crucial to avoid approving high-risk applicants. You can tolerate rejecting a few good loans, but you can't afford to approve bad ones.
   - Visualizations: I choose Stacking Model.
     - It achieved a high recall on the minority class (bad loans), capturing 76% of all true bad loan cases on the test set.

       In credit risk modeling, recall is critical: we aim to identify as many bad loans as possible, minimizing the risk of approving high-risk borrowers.

       While precision was low, the priority was to avoid false negatives (i.e., approving risky applicants).
     - Even though overall accuracy and precision were modest, Stacking model offered a good safety net against missed bad loans, making it suitable choice given the business goal.
   - Successes and Insights:
     - Robust Modeling Pipeline Developed a complete ML pipeline: data preprocessing → feature selection → hyperparameter tuning → model evaluation.
     - Trained and compared multiple models: Logistic Regression, Random Forest, Balanced RF, XGBoost, and Stacking Ensemble.
     - Fairness Evaluation Evaluated Equal Opportunity Difference (EOD):
       * Gender EOD: 0.0266
       * Race EOD:0.0776

       Ensured model treats demographic groups fairly, an essential step in responsible lending.
     - Interpretability with SHAP:
       * Visualize global feature impact (e.g., FICO, LTV).
       * Helps justify decisions to stakeholders, ensuring transparency.
   - Pitfalls and Challenges:

(a) Highly Imbalanced Data Class 0 (good loans) was about 95% of the data → naive models over-predicted class 0.
SMOTE led to overfitting.

(b) Feature Quality and Missingness
Many features were highly missing or redundant.
Dropped several features, including demographic info (which I later had to re-merge for fairness analysis).

(c) Model Bias Toward Positive Class
All models had poor precision for class 1 (bad loans).
Risk: rejecting too many good applicants (false positives), which could harm user experience or profit margins.