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: Given the high class imbalance in the dataset (bad loans are $< 5\%$), we evaluated model performance based on precision-recall metrics rather than accuracy or ROC AUC alone. In particular, we focused on the F1 score and PR AUC for the bad loan class to ensure that the model is both effective at catching risky applications (high recall) and not overly conservative (reasonable precision).

     ROC AUC was included for completeness, but due to class imbalance, we relied more heavily on PR AUC and class-wise classification reports.
   - Visualizations: I choose XGBoost.

     Compared to other models, XGBoost had the best overall balance between:
     - High recall on the positive class (i.e., capturing most bad loans).
     - Reasonable precision on both classes.
     - Better generalization (less overfitting than the stacking ensemble).
   - 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.
     - XGBoost: Best Performing Model Achieved strong generalization performance:
       * F1-score for class 1 (bad loans): 0.86
       * High recall: caught 77% of all bad loans on test set.
       * Tuned with GridSearchCV + scale_pos_weight to address class imbalance.
     - Fairness Evaluation Evaluated Equal Opportunity Difference (EOD):
       * Gender EOD: 0.0036
       * Race EOD: 0.0503

       Ensured model treats demographic groups fairly, an essential step in responsible lending.
     - Interpretability with SHAP Used SHAP to:
       * Visualize global feature impact (e.g., FICO, PTI, LTV).
       * Provide local explanations for individual predictions.
       * Helps justify decisions to stakeholders, ensuring transparency.

- Pitfalls and Challenges:
  (a) Highly Imbalanced Data Class 1 (bad loans) was about 95% of the data $\rightarrow$ naive models over-predicted class 1.
  SMOTE led to overfitting, especially in complex models like stacking.
  (b) Feature Quality and Missingness
  Many features were highly missing or redundant.
  Dropped several features, including demographic info (which you later had to re-merge for fairness analysis).
  (c) Model Bias Toward Positive Class
  Most models had high recall for class 1 but poor precision for class 0 (good loans).
  Risk: rejecting too many good applicants (false positives), which could harm user experience or profit margins.