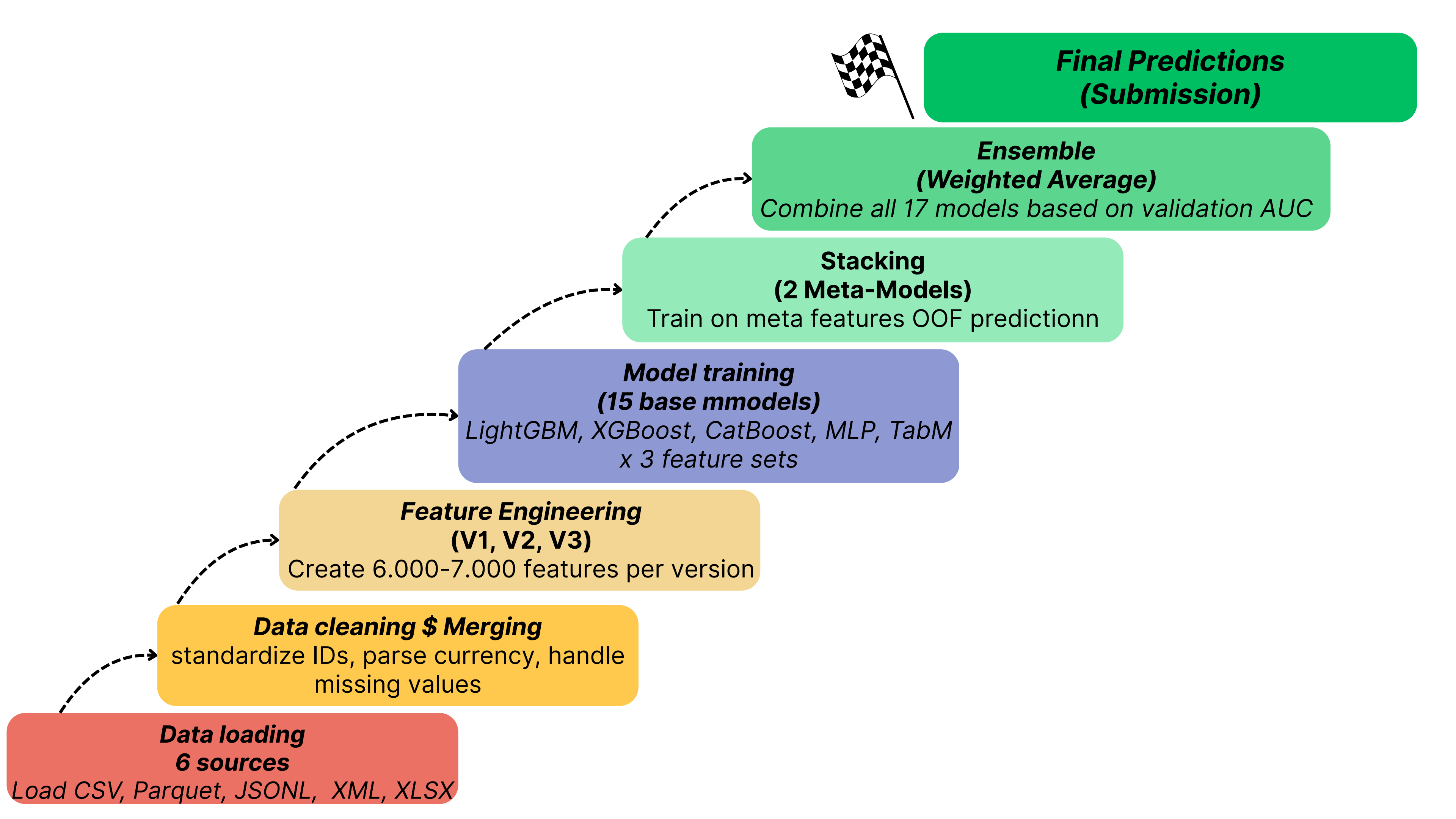




# Performance Metrics Report

November 16, 2025





# Data Cleaning & Merging

## Cleaning

parsing issues

data types

formats

## Missing value strategy

Median for  
continuous features

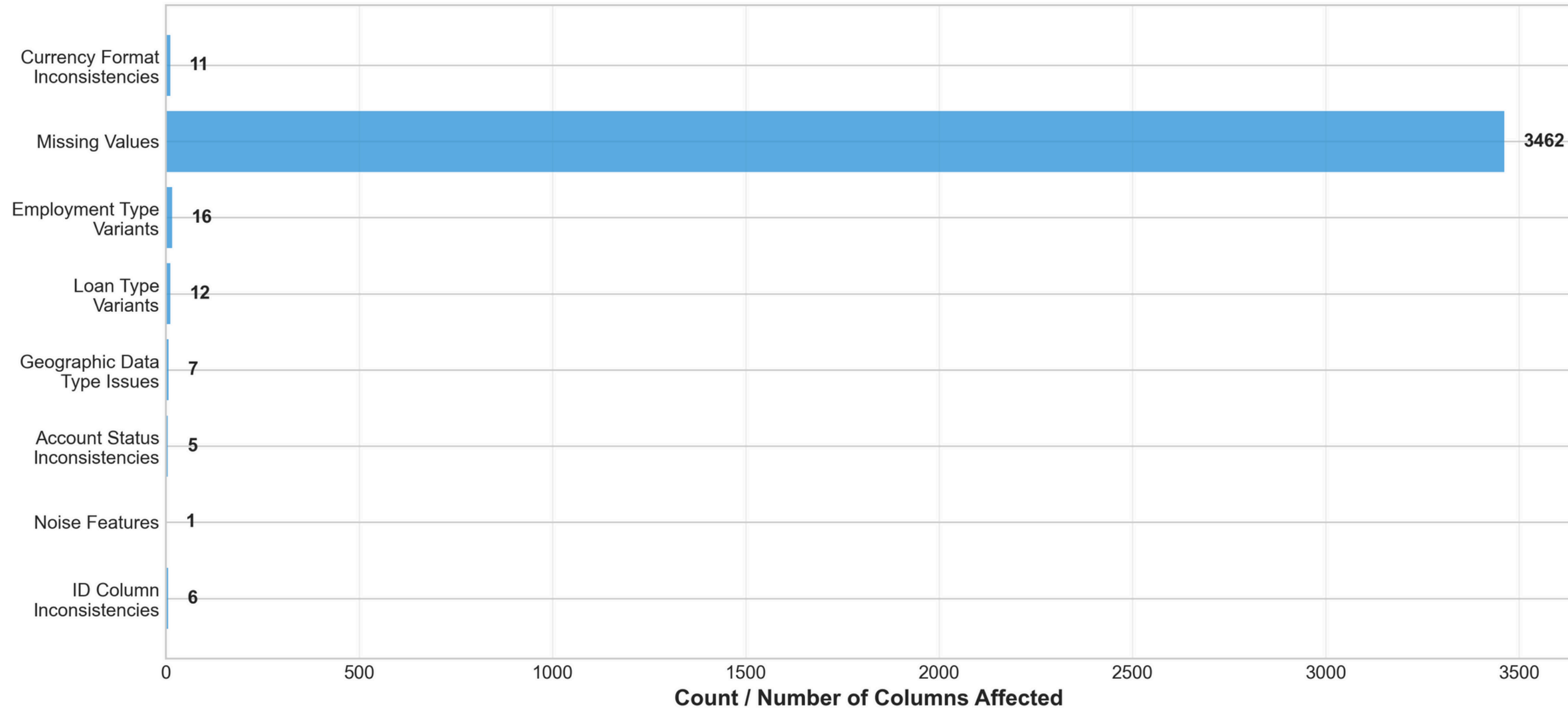
Mode for categorical  
Features

## Merging

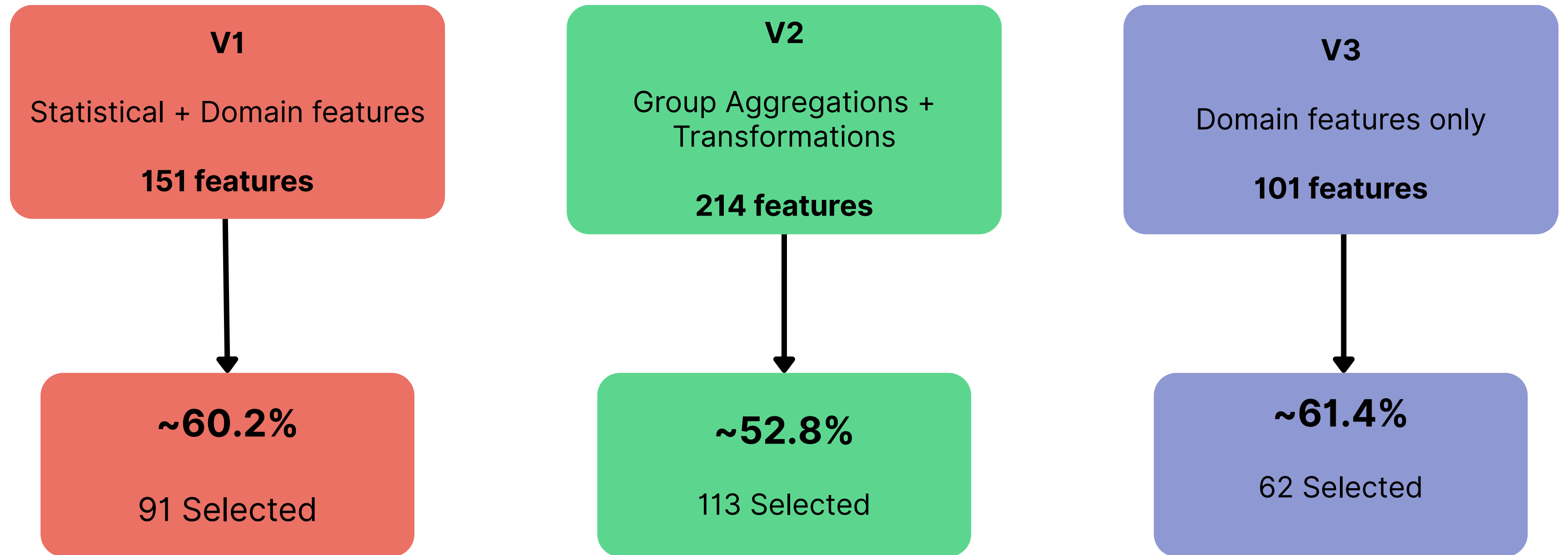
Combined datasets using  
*customer\_id*

Validated integrity  
after merging

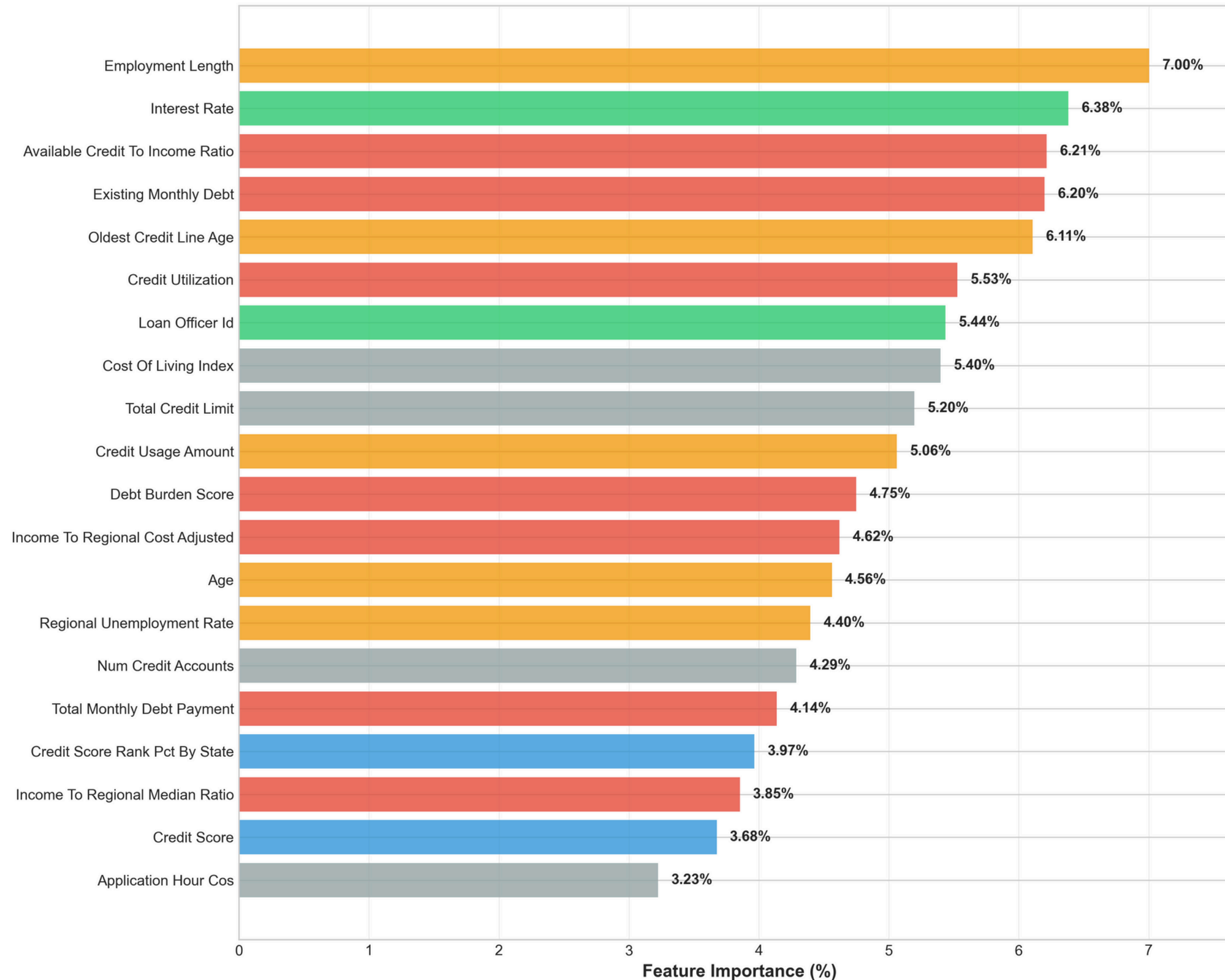
Data Quality Issues Identified and Resolved



# FE & Selection

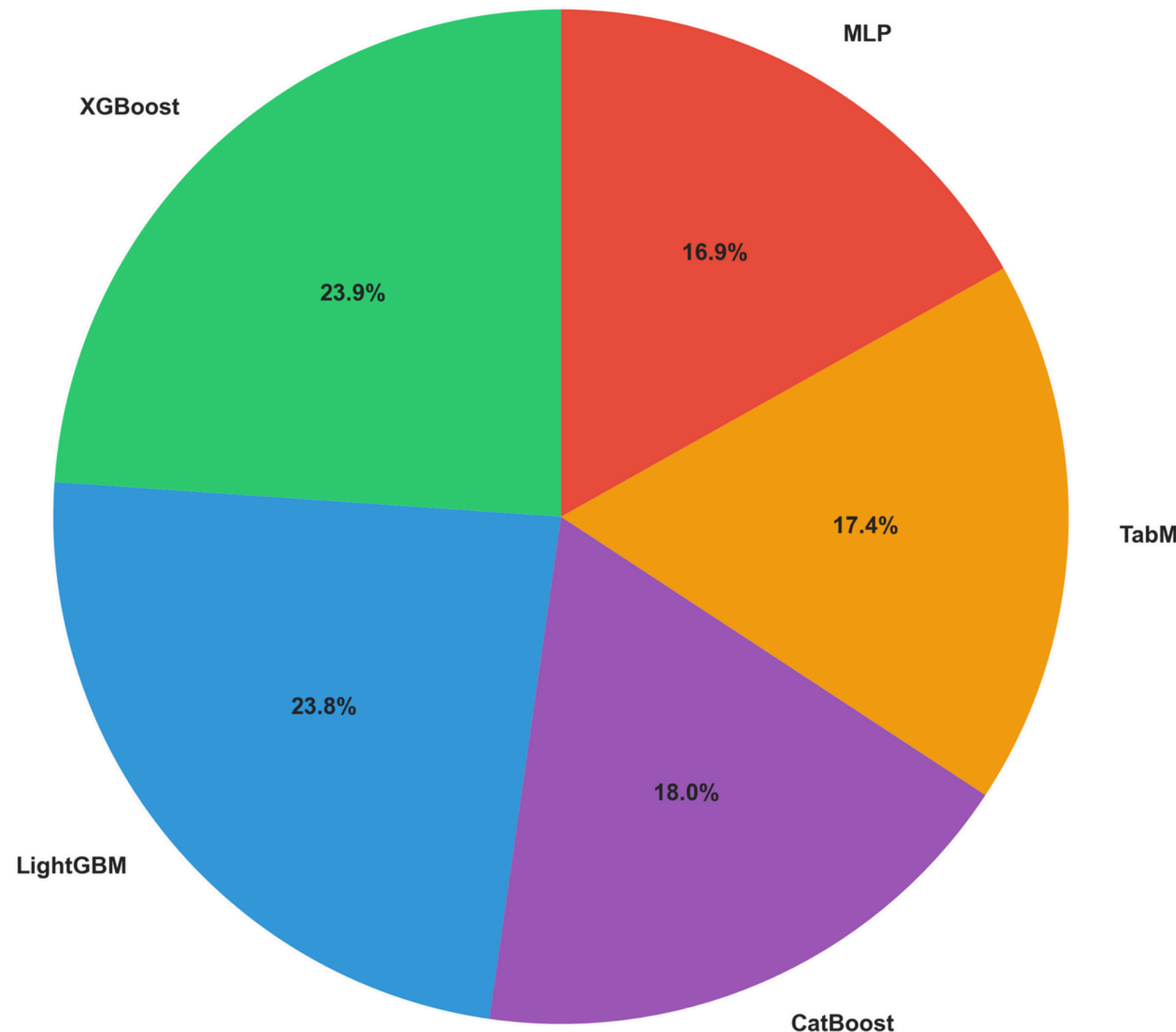


Top 20 Most Important Features



# Model Diversity for Robustness

Ensemble Weights by Models



## 1. LightGBM

Fast gradient boosting tree model.

## 2. XGBoost

High-accuracy boosting algorithm.

## 3. CatBoost

Boosting optimized for categorical data.

## 4. MLP

Artificial neural network model.

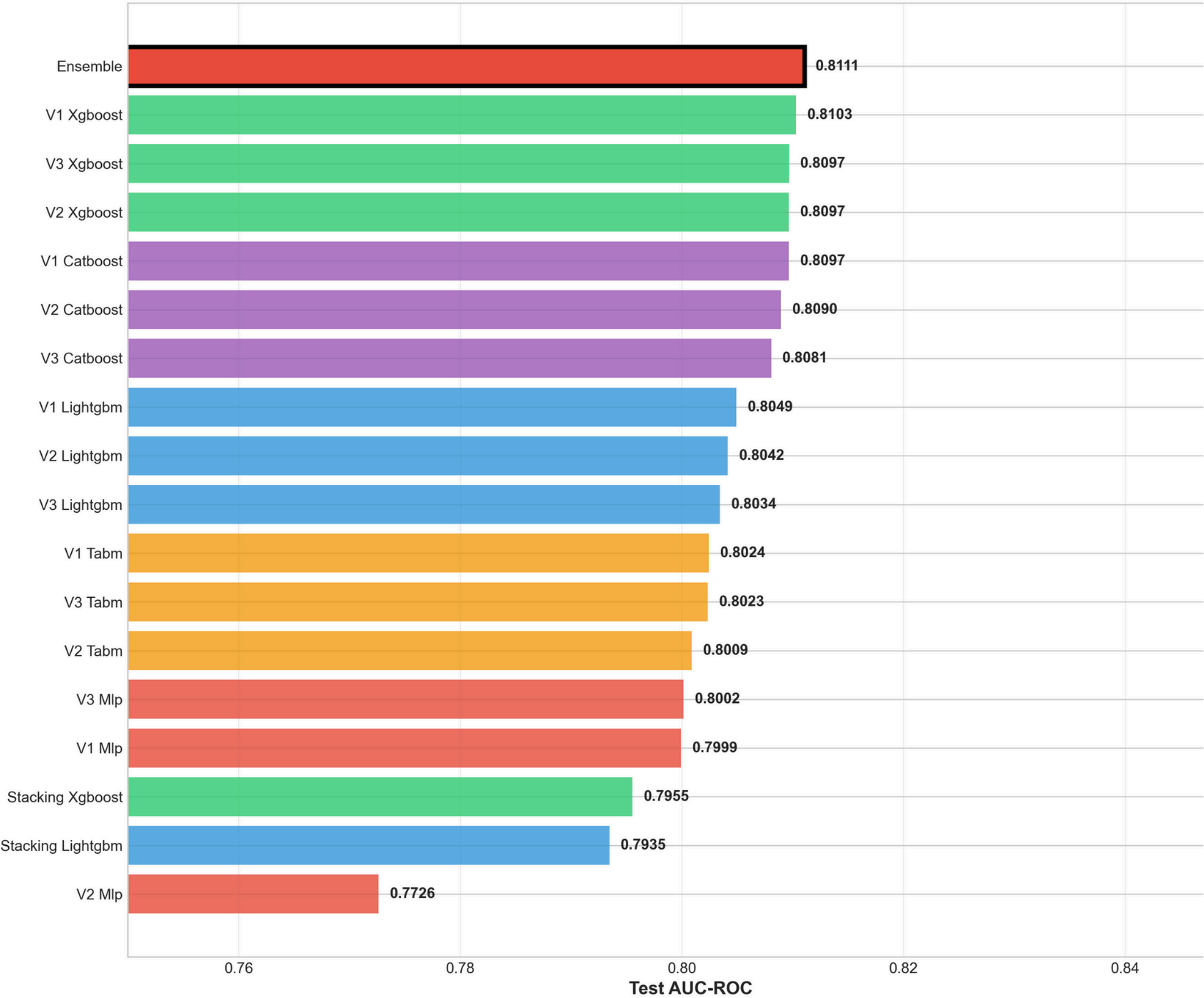
## 5. TabM

Model suited for tabular data.

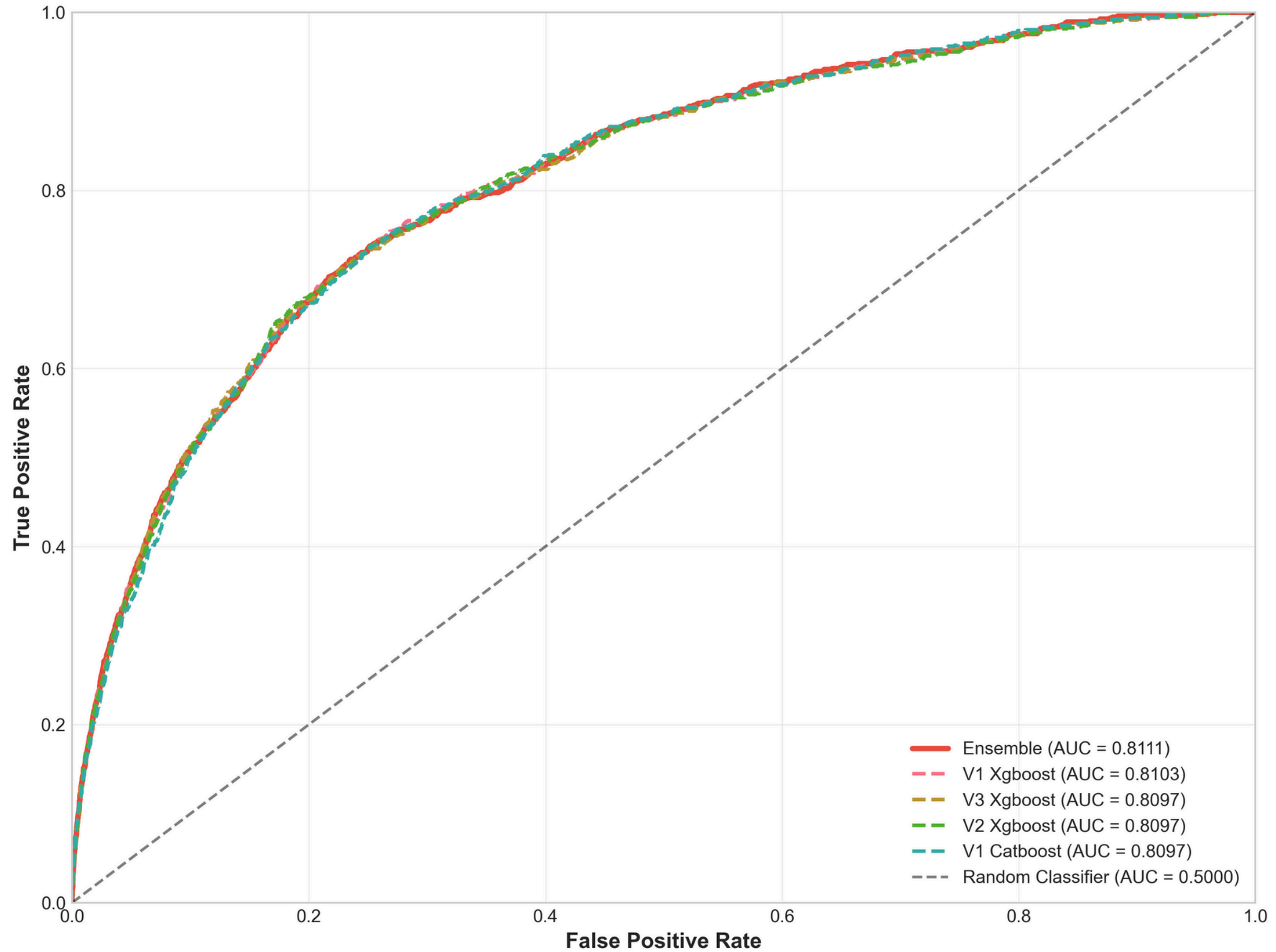
**Each model trained on 3 different feature sets (V1,V2,V3)**

**(15 base models + 2 meta models)**

Model Performance Comparison (All 17 Models + Ensemble)



ROC Curve Comparison - Ensemble vs Top Individual Models



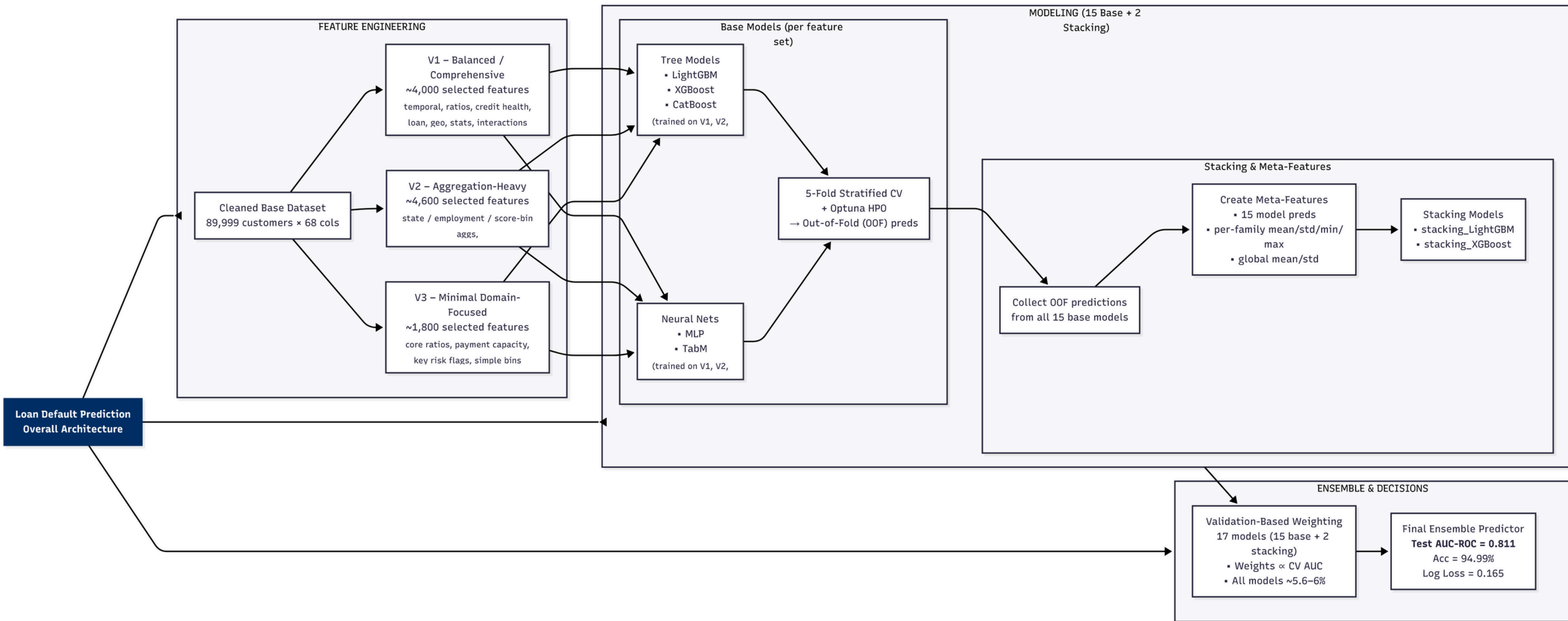




81,1%

Final Test AUC

- ✓ **Final Test Set Performance:** AUC-ROC = 0.811064
- ✓ **Cross-Validation Performance:** AUC-ROC = 0.805022
- ✓ **Test Set Accuracy:** 94.99%
- ✓ **Test Set Log Loss:** 0.165026



# Our Team

**Javohir** - Data Analyst & Engineer

**Dilyorbek** - AI/ML developer

**Maruf** - Data Engineer

**Mirjahon** - Data Engineer & Analyst

**Aziz** - AI/ML Engineer

Thank you