# Building Better Credit Scores: Machine Learning and NLP for Optimized Risk Assessment

Aman Kar akar@ucsd.edu
Daniel Mathew drmathew@ucsd.edu
Tracy Pham tnp003@ucsd.edu
Mentor: Brian Duke brian.duke@prismdata.com
Mentor: Kyle Nero kyle.nero@prismdata.com
Mentor: Berk Ustun berk@ucsd.edu

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

## Introduction

Traditional credit scoring models often disadvantage those with limited credit history, such as young adults, recent immigrants, and cash-reliant individuals. We propose a fairer, more transparent measure that leverages detailed bank transaction data. By analyzing cash flow patterns and spending behavior with advanced modeling, we develop a probability-based scoring model to more accurately assess delinquency risk and improve credit access.

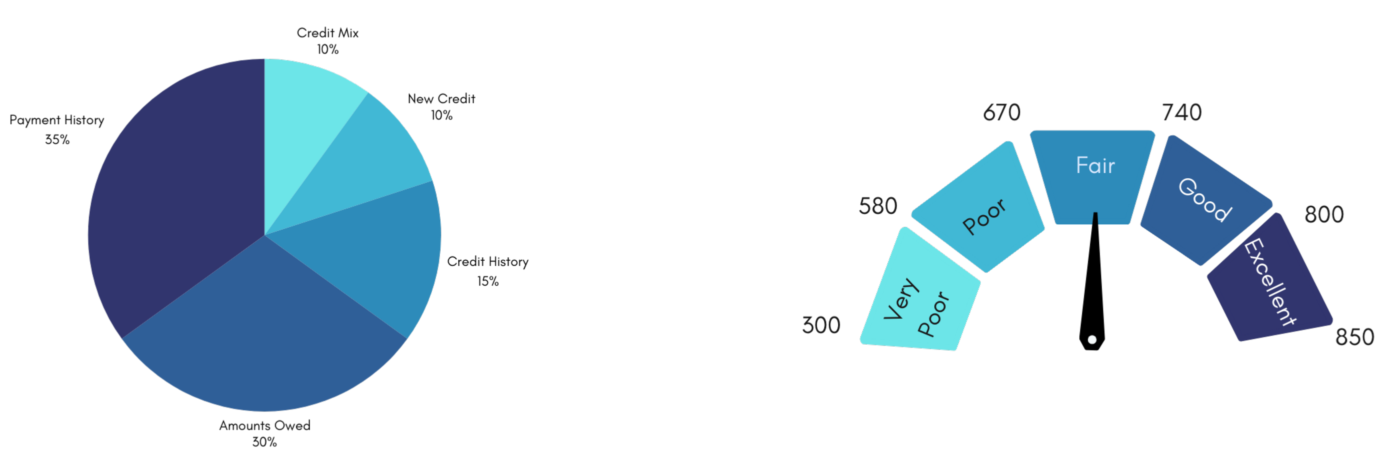


Figure 1. Traditional credit scoring model.

## Key Datasets

Our analysis utilizes four key datasets to assess creditworthiness:

- Account Data – Consumers' account types, balances, and balance dates.
- Consumer Data – Consumers' credit scores and delinquency targets.
- Transaction Data – Transactions with categories, amounts, & credit/debit indicators.
- Category Mappings – Links transaction category IDs to transaction category names.

## Data Exploration

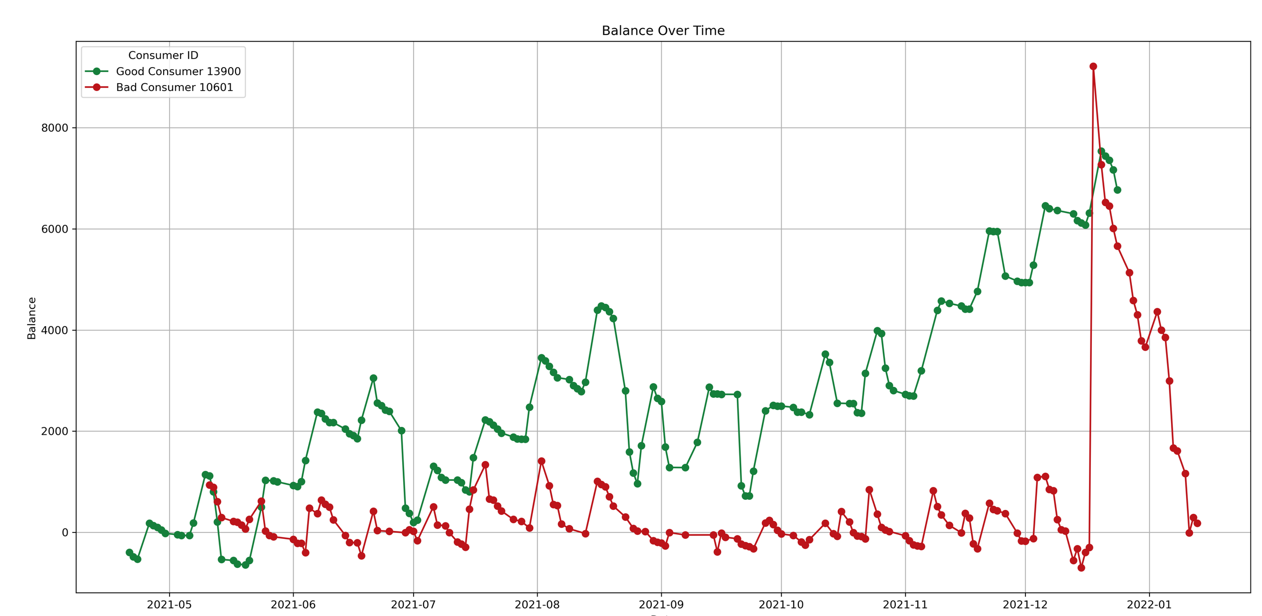


Figure 2. Balance over Time of Delinquent vs. Non-Delinquent Consumer

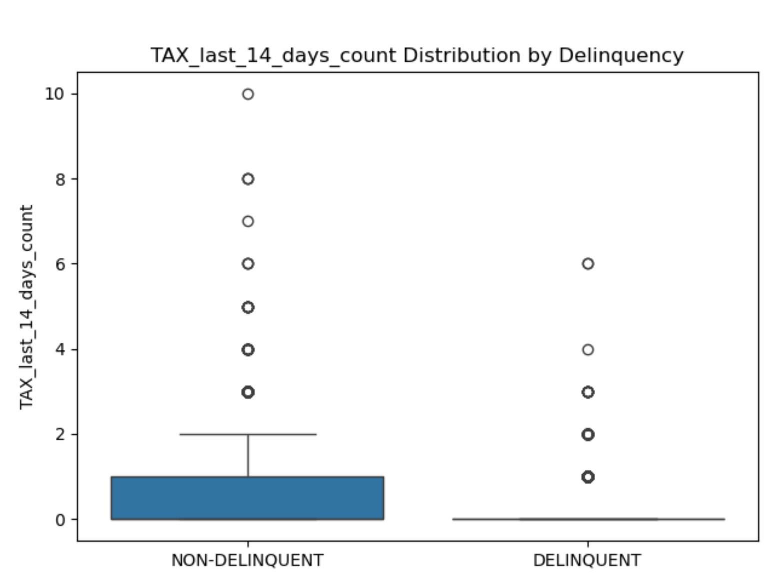


Figure 3. Tax Transactions in the Last 14 Days

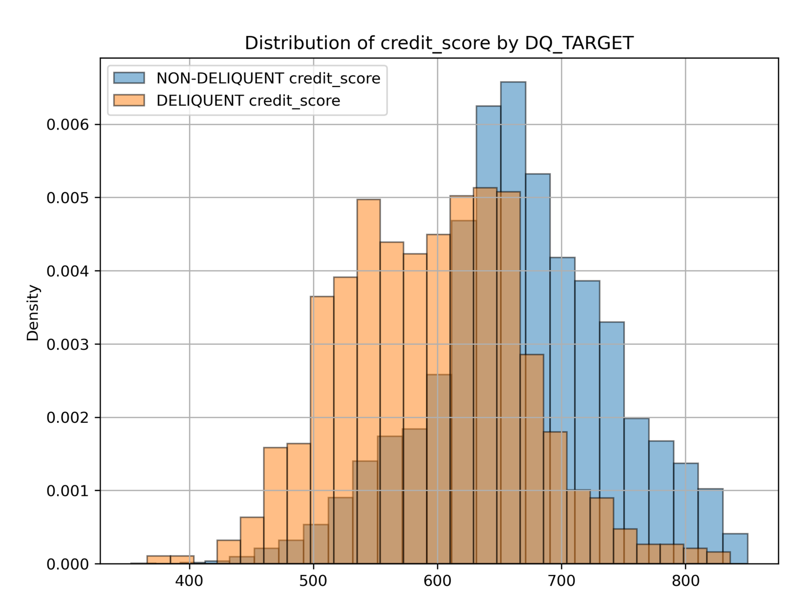


Figure 4. Distribution of Credit Score of Delinquent vs. Non-Delinquent Consumers

## Model Pipeline

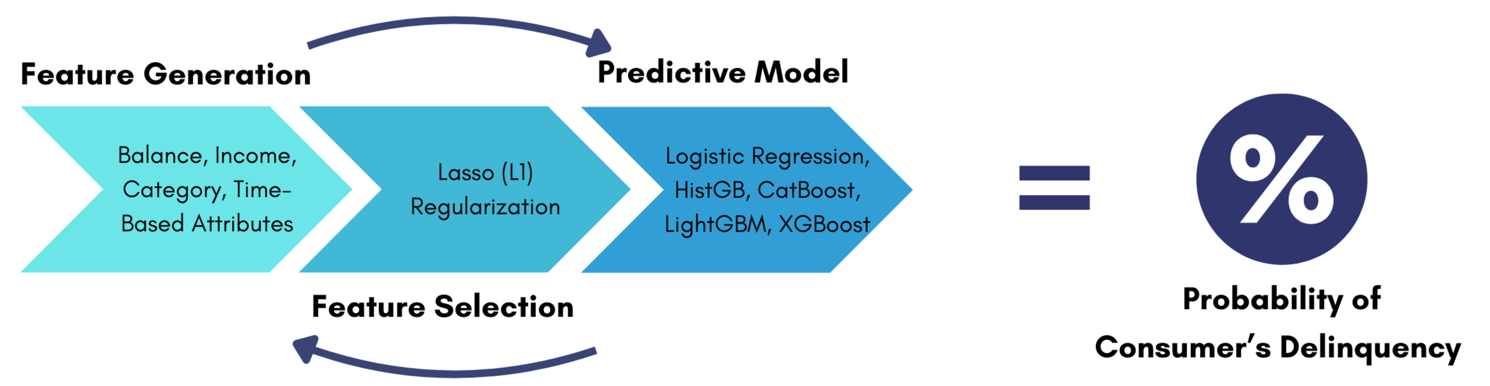


Figure 5. Cash Score Model Pipeline

*We adopted a cyclic feature generation process to iteratively improve model performance.*

Model Evaluation - We used the following metrics to assess model performance:

- ROC AUC: Measures class distinction, with higher values indicating better performance and discriminative power between positives and negatives.
- Accuracy: Proportion of correct predictions.
- Precision: Ratio of true positives to predicted positives.
- Recall: Ratio of true positives to actual positives.
- Confusion Matrix: Displays true/false positives and negatives to assess errors.

## Feature Engineering

**Feature Generation:**

- Risk Indicators: Identified high-risk behaviors by focusing on specific transaction categories (i.e. gambling, overdraft fees, buy now pay later).
- Standardization: Standardized non-categorical features to ensure consistent scaling.
- Resampling: Utilized SMOTE and undersampling techniques to redistribute training data and balance the majority and minority classes.

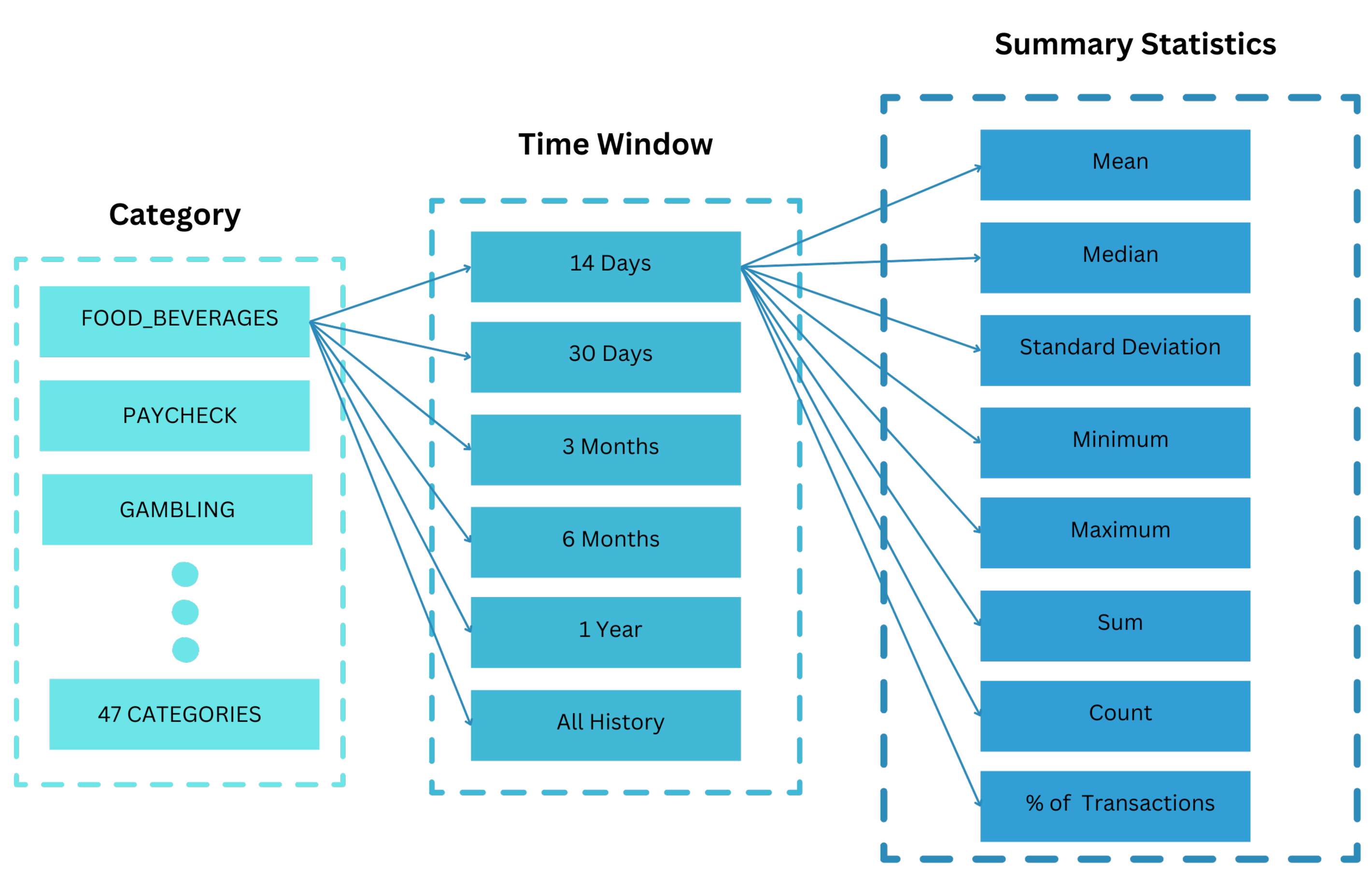


Figure 6. Category-Based Feature Generation Process

**Feature Selection:**

- Correlation Analysis: Removed features highly correlated with each other.
- Lasso (L1) Regularization: Selected top features ranked by L1 feature importance.
- Embedded Method: Utilized Random Forest to effectively rank and select features.

## Findings

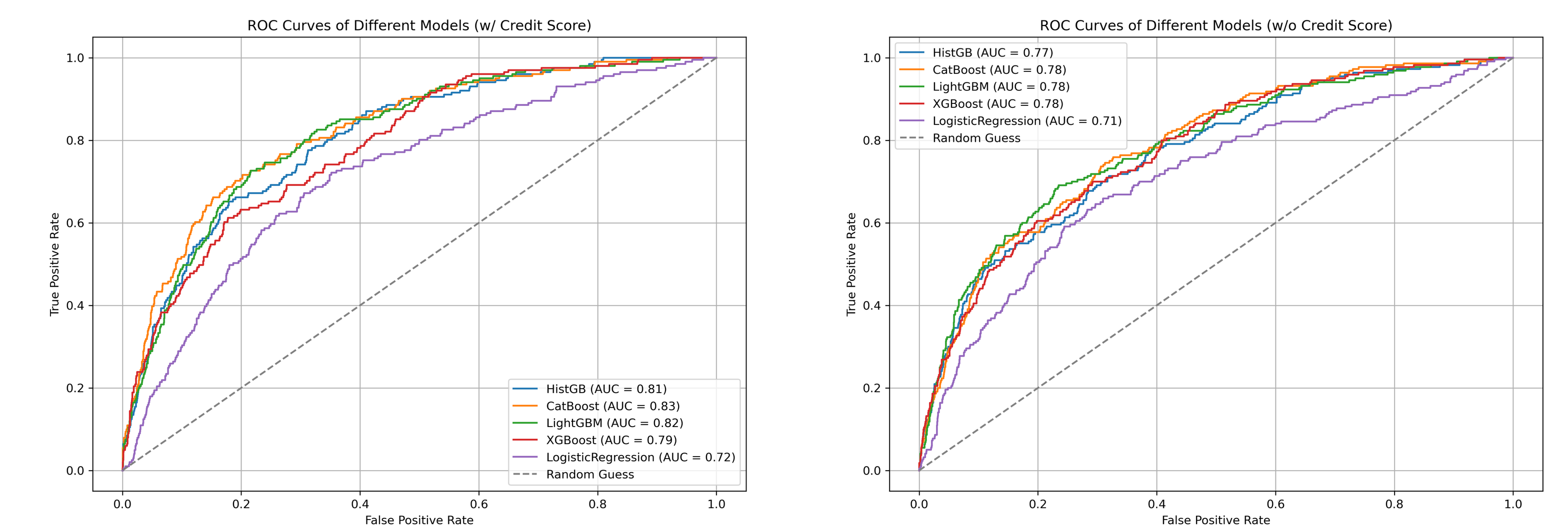


Figure 7. Comparison of ROC curves for different models (**Left**: w/ Credit Score, **Right**: w/o Credit Score )

| Model | ROC-AUC | Accuracy | Precision | Recall | F1-Score | Training | Prediction |
|---|---|---|---|---|---|---|---|
| Logistic Regression (w/o Credit Score) | 0.7079 | 0.8445 | 0.2383 | 0.2785 | 0.2568 | 1.3368 | 0.4016 |
| Logistic Regression (w/ Credit Score) | 0.7241 | 0.8571 | 0.2674 | 0.3548 | 0.3050 | 1.7175 | 0.3315 |
| LightGBM (w/o Credit Score) | 0.7796 | 0.8991 | 0.3878 | 0.0802 | 0.1329 | 4.1249 | 0.0931 |
| LightGBM (w/ Credit Score) | 0.8162 | 0.9068 | 0.4167 | 0.1382 | 0.2076 | 3.9720 | 0.0859 |
| **CatBoost (w/o Credit Score)** | **0.7704** | **0.9019** | **0.4474** | **0.0717** | **0.1236** | **38.6703** | **0.0788** |
| **CatBoost (w/ Credit Score)** | **0.8260** | **0.9170** | **0.4681** | **0.1095** | **0.1774** | **40.9512** | **0.0960** |

Table 1. Cash Score Model (w/o Credit Score vs. w/ Credit Score) Evaluation Metrics

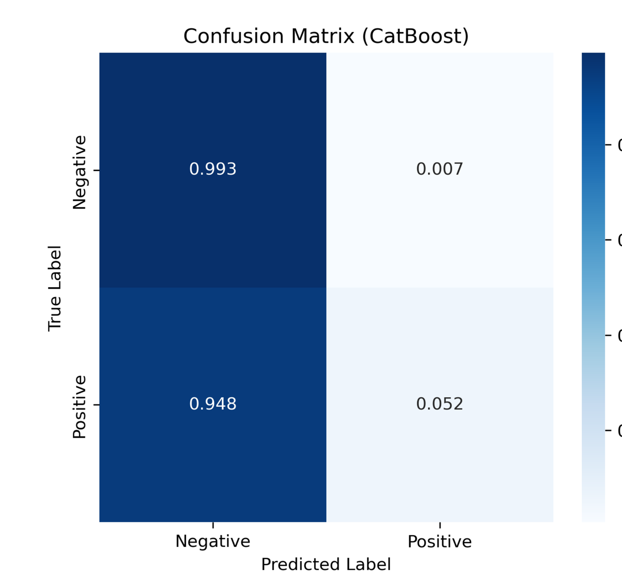


Figure 8. Confusion Matrix of CatBoost (w/o Credit Score)

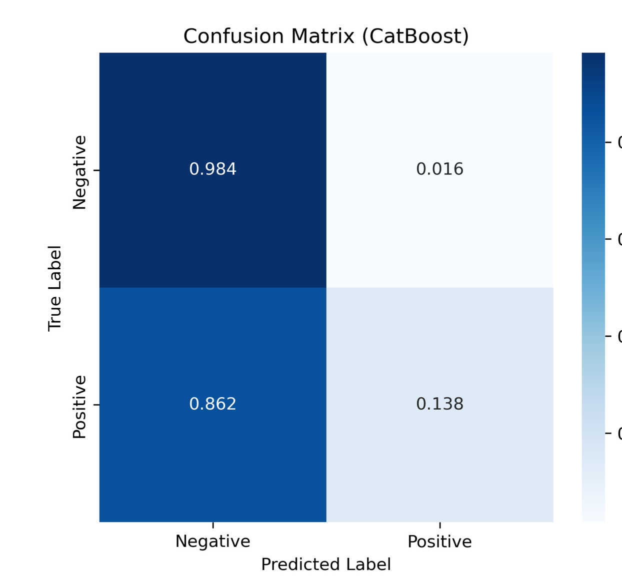


Figure 9. Confusion Matrix of CatBoost (w/ Credit Score)

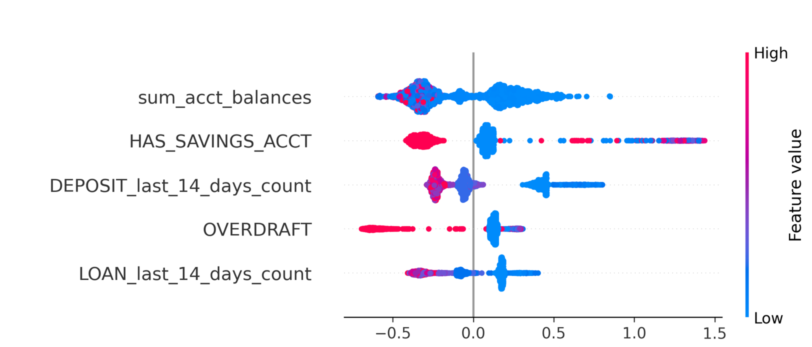


Figure 10. Top SHAP Values

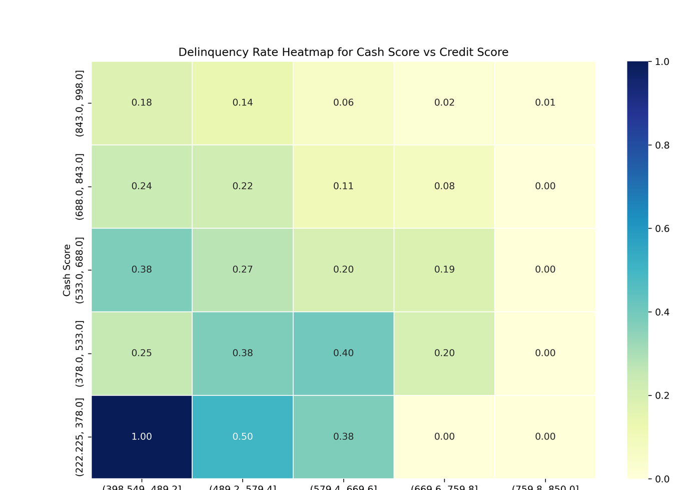


Figure 11. Delinquency Rate Heatmap

## Next Steps

- Feature Engineering: Optimize aggregated feature metrics on categories and time windows, and implement clustering algorithms to select optimal features.
- Model Refinement: Attempt deep learning models with extended hyperparameter tuning sessions to extract more complex patterns and improve predictive power.
- Bias & Fairness: Evaluate for potential biases in predictions across demographics and implement fairness constraints to ensure equitable credit assessments.