

Predictive Model Plan

Report:

1. Model Logic (Generated with GenAI)

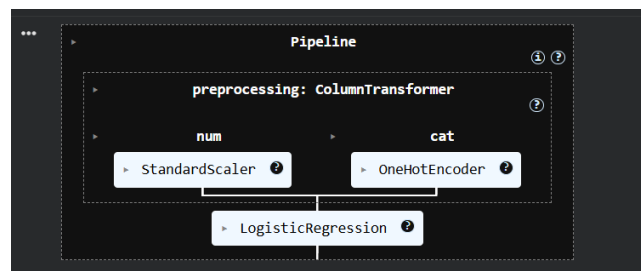
Goal: Predict whether a customer will become delinquent on credit payments using historical financial and demographic features.

Logic Overview:

- **Input:** Cleaned dataset with numeric and categorical features (income, credit utilization, employment status, etc.)
- **Preprocessing:**
 - Standardize numeric features
 - One-hot encode categorical variables
- **Model:** Logistic Regression to estimate probability of delinquency
- **Alternative model:** Decision Tree to capture non-linear relationships
- **Output:** Probability score and binary delinquency prediction

Step-by-step:

1. Load the dataset and define target (Delinquent) and features.
2. Split data into training (80%) and test (20%) with stratification to preserve class distribution.
3. Build preprocessing pipeline for numeric and categorical features.
4. Fit logistic regression model using balanced class weights to handle class imbalance.
5. Evaluate model using test data.
6. Perform fairness checks across key groups (e.g., income levels).



2. Justification for Model Choice

Why Logistic Regression?

- **Interpretability:** Coefficients indicate how each feature impacts delinquency risk.
- **Regulatory-friendly:** Transparent and explainable to auditors or compliance teams.
- **Probability output:** Allows risk-based decision making (e.g., preemptive interventions).
- **Efficiency:** Fast to train, easy to deploy, low computational cost.
- **Proven in finance:** Widely used in credit scoring and risk assessment.

Secondary Option – Decision Tree:

- Captures non-linear relationships and interactions between features.
- Can improve accuracy if logistic regression underfits, but less interpretable.

3. Evaluation Strategy

Metrics to measure performance:

Metric	Purpose
Accuracy	Overall correctness of predictions
Precision	Fraction of predicted delinquents who are truly delinquent (reduces false positives)
Recall	Fraction of true delinquents correctly identified (reduces false negatives)
F1 Score	Balance between precision and recall
ROC-AUC	Overall ranking ability across thresholds

Interpretation Plan:

- **High recall:** Ensures at-risk customers are identified for interventions.
- **High precision:** Avoids wrongly flagging reliable customers.
- **ROC-AUC:** Assesses model's discrimination power.

Fairness & Bias Considerations:

- Evaluate predictions across income, gender, and customer segments.
- Identify and mitigate bias to avoid discrimination.
- Regular monitoring to ensure model decisions are consistent and ethical.

Monitoring Plan:

- Track model drift over time using rolling performance metrics.
- Re-train periodically with new payment data.
- Inspect logistic coefficients for unexpected shifts in feature importance.