# Predictive Modeling Plan for Delinquency Risk – Geldium

Date: July 17, 2025

## Step 1: Generated Model Logic

Chosen Model: Logistic Regression  
  
Model Pipeline Overview:  
1. Feature Selection: Use key financial variables such as:  
 - Payment History  
 - Credit Utilization Rate  
 - Debt-to-Income Ratio  
 - Income Stability  
 - Recent Credit Activity  
  
2. Data Preprocessing:  
 - Handle missing values (mean imputation for income, median for utilization).  
 - Normalize numerical features.  
 - Encode categorical features if needed.  
  
3. Model Training:  
 - Apply logistic regression to estimate probability of delinquency.  
 - Threshold: Customers with probability > 0.5 classified as "high risk".  
  
4. Output:  
 - Binary classification: Delinquent / Non-Delinquent  
 - Risk score (0 to 1 probability)  
  
Top 5 Features Selected:  
- Payment History  
- Credit Utilization  
- Debt-to-Income Ratio  
- Income  
- Employment Type

## Step 2: Model Justification

Logistic regression is ideal for Geldium’s needs because it is simple, interpretable, and reliable for binary classification tasks like predicting credit delinquency. It allows decision-makers to understand the impact of each variable, which is critical for transparency and regulatory compliance. While more complex models (like neural networks) may provide higher accuracy, they are less explainable. For Geldium, logistic regression offers a balance between performance and explainability, supporting responsible AI use in financial services.

## Step 3: Evaluation Strategy

To evaluate the model, we will use the following metrics:  
- Accuracy: General performance across all predictions.  
- Precision: How many predicted delinquents were correct.  
- Recall: How many actual delinquents were identified.  
- F1 Score: Balance between precision and recall.  
- AUC-ROC: Model’s ability to distinguish risk levels.  
- Confusion Matrix: Diagnostic tool to visualize misclassifications.  
  
Bias & Fairness Checks:  
- Analyze predictions across demographics (e.g., income groups, regions).  
- Use fairness metrics like demographic parity and disparate impact analysis.  
- Apply SHAP values for explainability.  
  
If the model performs poorly or shows signs of bias, we will:  
- Adjust features (feature engineering)  
- Rebalance the dataset  
- Tune hyperparameters or explore alternative models