**Predictive Model Plan**

# 1. Model Logic (Generated with GenAI)

The goal of the model is to predict if a customer is likely to become delinquent (1) or not (0) using historical financial and behavioral data.

**GenAI-Generated Model Logic (based on Geldium's dataset):**

**1. Load Dataset:**

Include variables like: Income, Credit\_Score, Credit\_Utilization, Missed\_Payments, Loan\_Balance, etc.

**2. Feature Selection:**

**Key features:**

• Credit\_Utilization

• Debt\_to\_Income\_Ratio

• Missed\_Payments

• Credit\_Score

• Account\_Tenure

**3. Data Preprocessing:**

• Handle missing values (e.g., impute Income)

• Encode categorical columns: Employment\_Status, Credit\_Card\_Type

• Normalize numerical features if needed

**4. Model Application:**

• Use Logistic Regression to estimate the probability of delinquency

**5. Prediction:**

• Apply a classification threshold (e.g., 0.5) to label customer as high risk (1) or not (0)

**6. Output:**

• Return predicted label and top contributing features per prediction

**7. Evaluation:**

• Accuracy, Precision, Recall, F1 Score, AUC, Confusion Matrix

# 2. Justification for Model Choice

I selected **Logistic Regression** as the modeling approach for the following reasons:

• It is ideal for binary classification, which aligns with our target variable Delinquent\_Account (0 or 1).

• It offers high transparency; coefficients directly indicate how features like Credit\_Utilization and Debt\_to\_Income\_Ratio affect risk.

• The model produces probability scores, which are useful for setting risk thresholds and tailoring intervention strategies.

• Logistic Regression is easy to implement and deploy, making it suitable for operational use at Geldium.

• Most importantly, it aligns with Geldium’s business needs for explainable and responsible financial decision-making, and is compliant with regulatory standards for fairness and auditability.

# 3. Evaluation Strategy

To evaluate the performance and fairness of the model, the following strategy will be applied:

#### Metrics to Measure Effectiveness:

• **Accuracy:** Measures how often the model is correct overall.

• **Precision:** Ensures flagged risky customers are truly at risk (reduces false positives).

• **Recall:** Measures how well the model identifies actual delinquent customers (reduces false negatives).

• **F1 Score:** A balance of precision and recall—ideal for imbalanced classes.

• **AUC-ROC:** Measures how well the model ranks high-risk vs. low-risk customers across all thresholds.

• **Confusion Matrix:** Provides breakdown of true/false positives and negatives to interpret errors.

#### Bias & Fairness Checks:

• Perform disparate impact analysis to ensure no group (based on Age, Income level, Location, or Employment\_Status) is disproportionately flagged.

• Use SHAP values to explain individual predictions and detect whether sensitive features influence risk unfairly.

• Assess demographic parity and equal opportunity fairness across subgroups.

#### Ethical Considerations:

• Use only relevant financial and behavioral variables, avoid proxies like Location that could relate to race or class.

• Ensure customer privacy is preserved and predictions are used as support for human-led interventions, not automated rejections.

• Ensure the model is continuously monitored for drift and bias over time.