**Predictive Model Plan – Student Template**

# **1. Model Logic (Generated with GenAI)**

This predictive model is designed to forecast the likelihood of a customer becoming delinquent based on their financial behavior and demographic information. We use a logistic regression model, suitable for binary classification problems such as “delinquent” (1) or “not delinquent” (0). This model estimates the probability of delinquency using a sigmoid function and allows us to set a risk threshold (e.g., 0.5) to classify outcomes.

**Step-by-step logic:**

1. **Data Preparation**: Handle missing values (e.g., impute Income using median; drop rows with excessive nulls).
2. **Feature Selection**: Use relevant predictors:
   * Missed\_Payments
   * Credit\_Utilization
   * Income
   * Debt\_to\_Income\_Ratio
   * Account\_Tenure
3. **Data Scaling**: Normalize numerical variables for consistency.
4. **Model Training**: Fit a logistic regression model on the training dataset.
5. **Prediction**: Generate probability scores between 0–1 for each customer.
6. **Classification**: Apply threshold (e.g., 0.5) to classify as Delinquent (1) or Not Delinquent (0).
7. **Evaluation**: Use metrics such as precision, recall, F1 score, and AUC-ROC to evaluate performance.

# **2. Justification for Model Choice**

We chose **logistic regression** because it strikes an effective balance between accuracy, interpretability, and ease of implementation. In financial services, where regulatory compliance and fairness are crucial, models must be explainable. Logistic regression offers clear insights into how each variable affects the risk of delinquency, which supports informed decision-making.

Unlike neural networks, which act as black boxes, or decision trees, which can overfit, logistic regression provides **probability-based predictions** and allows Geldium to set flexible thresholds for risk classification. It works well with Geldium’s structured dataset and supports **real-time, scalable predictions** for Collections team workflows. Overall, it is well-suited to Geldium’s need for a transparent, reliable model to prioritize interventions.

# **3. Evaluation Strategy**

To ensure the model performs well, we will evaluate it using the following metrics:

* **Accuracy** – Proportion of total correct predictions (both positive and negative).
* **Precision** – Focuses on how many customers predicted as delinquent actually are. Important to minimize false positives.
* **Recall** – Measures how many actual delinquent customers we successfully identified. Crucial for risk detection.
* **F1 Score** – Harmonizes precision and recall to give a balanced performance metric.
* **AUC-ROC** – Assesses the model’s ability to rank customers correctly by delinquency risk.

**Bias and fairness checks:**

* Conduct **demographic parity** analysis to ensure fairness across customer segments (e.g., employment type, income brackets).
* Apply **disparate impact ratio** to evaluate if the model disproportionately flags any group.
* Use tools like **SHAP** or **LIME** for interpretability and transparency in decision-making.

**Ethical considerations:**  
We ensure the model does not reinforce historical bias (e.g., against lower-income or unemployed individuals) and is used only for supportive interventions—not punitive actions. Continuous monitoring will help ensure the model stays fair, reliable, and aligned with Geldium’s ethical and business values.