**Predictive Model Plan – Student Template**

Use this template to structure your submission. You can copy and paste content from GenAI tools and build around it with your own analysis.

# 1. Model Logic (Generated with GenAI) The predictive model aims to forecast which customers are at risk of delinquency based on their credit and financial behavior. The ultimate goal is to classify customers as either 'delinquent' or 'non-delinquent' by analyzing various input features. Predictive Modeling Pipeline Outline: \*\*Data Collection and Ingestion:\*\* Collect and ingest customer data from multiple sources in a structured format. \*\*Data Cleaning and Preprocessing:\*\* Handle missing values, detect outliers, scale features, and convert data types appropriately. \*\*Feature Engineering and Selection:\*\* Generate new features, encode categorical variables, and select relevant features using correlation analysis and model-based importance. \*\*Model Selection and Training:\*\* Use LightGBM and split the data appropriately. Perform hyperparameter tuning to improve model accuracy. \*\*Model Evaluation and Validation:\*\* Use cross-validation and various metrics for validation. \*\*Model Deployment and Monitoring:\*\* Deploy in production with continuous performance and drift monitoring. \*\*Model Options Considered:\*\* Logistic Regression (simple, interpretable) vs. LightGBM (complex, high-performing). Final recommendation: LightGBM. \*\*Top 5 Features:\*\* Missed\_Payments, Credit\_Utilization, Credit\_Score, Debt\_to\_Income\_Ratio, and Income. \*\*Workflow Summary:\*\* From ingestion to feature processing to LightGBM prediction and probability thresholding for final delinquency classification.

Use a GenAI tool (e.g., ChatGPT, Gemini) to generate the logic or structure of your predictive model.  
- You may include pseudo-code, a step-by-step process, or a simplified code snippet.  
- Briefly explain what the model is designed to do.

Paste your GenAI-generated output below or describe the logic in your own words:

[Insert GenAI model logic here]

# 2. Justification for Model Choice Gradient Boosting Machine (LightGBM) is selected for its ability to handle complex, non-linear interactions among financial features, which are common in delinquency prediction. It provides excellent accuracy, handles imbalanced datasets well, and scales efficiently for real-time applications. Although not as interpretable as Logistic Regression, it can still offer transparency via SHAP values and feature importances, fulfilling regulatory needs. Given Geldium’s requirement for both predictive strength and operational scalability, LightGBM strikes the ideal balance, supporting fairness monitoring, regulatory explainability, and efficient deployment.

Explain why you selected this specific model type (e.g., logistic regression, decision tree, neural network). Consider:  
- Accuracy  
- Transparency  
- Ease of use or implementation  
- Relevance for financial prediction  
- Suitability for Geldium’s business needs

[Insert your justification here]

# 3. Evaluation Strategy Evaluation involves measuring both predictive performance and fairness. \*\*Accuracy and Reliability Metrics:\*\* ROC AUC, Precision, Recall, F1-Score, and the Confusion Matrix. AUC measures overall discrimination ability, precision minimizes false positives, and recall ensures high-risk individuals are not missed. \*\*Bias and Fairness Checks:\*\* Statistical Parity Difference, Equal Opportunity Difference, and Average Odds Difference will assess whether predictions differ unfairly across demographic groups. \*\*Mitigation Strategies:\*\* Resampling, reweighting, adversarial debiasing, and threshold tuning. \*\*Interpretation for Improvement:\*\* Use confusion matrix analysis, SHAP values, and segment performance reviews. Continually monitor for data or concept drift and retrain the model when necessary. Ethical care will be taken to avoid unfair penalization or systemic bias, especially in financial decisions.

Outline how you would evaluate your model’s performance. Include:  
- Which metrics you would use (e.g., accuracy, precision, recall, F1 score, AUC)  
- How you would interpret those metrics  
- Any plans to detect or reduce bias in your model  
- Ethical considerations in making predictions about customer financial behavior

[Insert your evaluation plan here]