**Predictive Model Plan**

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)

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:

The XGBoost model operates by building an ensemble of decision trees sequentially, with each new tree correcting the errors of its predecessors. This iterative process allows it to progressively refine its predictions, especially focusing on instances that were previously misclassified, leading to a highly accurate and robust final model.

Based on the trained model, the top 5 most influential input features for predicting delinquency are:

Missed\_Payments

Income

Month\_4\_Missed

Month\_4\_Late

Loan\_Balance

# 2. Justification for Model Choice

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

In summary, the Gradient Boosting model was chosen for its unparalleled ability to deliver high predictive accuracy in a domain characterized by complex and non-linear relationships, which is critical for the demanding business needs of financial delinquency prediction. While acknowledging the inherent trade-off in transparency compared to simpler models, its performance advantages, combined with the availability of sophisticated explainability tools, make it the optimal choice for ensuring robust and actionable risk forecasts.

# 3. Evaluation Strategy

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

Our evaluation strategy is meticulously designed to ensure the model's performance is not only accurate but also fair and ethically sound. We will primarily utilize **Precision, Recall, F1-score, ROC AUC, and the Precision-Recall Curve** to assess predictive accuracy, given the critical need to correctly identify delinquent accounts while managing false positives and negatives effectively in an imbalanced dataset. These metrics will be interpreted by understanding their direct implications for financial risk: high Recall ensures few true delinquents are missed, while high Precision minimizes unnecessary interventions for non-delinquent customers. Beyond accuracy, a crucial component involves rigorously detecting and reducing bias by identifying protected customer groups (e.g., based on age or location) and systematically calculating fairness metrics like **Equal Opportunity** across these groups. Any significant disparities will trigger the application of specific bias mitigation techniques (pre-processing, in-processing, or post-processing) to correct for unfair outcomes. This comprehensive approach is underpinned by ethical considerations to prevent discriminatory predictions about customer financial behavior, ensuring equitable treatment and maintaining customer trust in a sensitive domain.