Predictive Model Plan

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 predictive model is designed to classify whether a customer will become delinquent (1) or not (0) using historical financial behavior and demographic data.

GenAI-generated model logic:-
1. Load and clean dataset (handle missing values, outliers)
2. Encode categorical variables (Employment_Status, Credit_Card_Type, etc.)
3. Engineer features (e.g., Total_Late_Payments from Month_1 to Month_6)
4. Split the dataset into training and test sets (e.g., 80/20)
5. Train a Logistic Regression and Random Forest model
6. Use cross-validation to evaluate model performance
7. Compare models using precision, recall, F1, and AUC
8. Select the best model and explain predictions using SHAP

The model predicts the likelihood of delinquency based on patterns in credit utilization, missed payments, income, and other attributes.

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

We selected the Random Forest classifier for the following reasons:

- Accuracy: It consistently performs well with structured/tabular financial data.
- Transparency: It allows interpretation of feature importance, which is critical in credit risk modeling.
- Ease of use: Built-in handling of categorical and numerical features makes it implementation-friendly.
- Business Relevance: It’s robust to missing data and outliers, which are common in real-world financial records.

- Suitability: Geldium needs a model that can balance precision and recall to avoid both customer churn and financial risk.

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

We will evaluate model performance using:

- *Accuracy: Overall correctness of predictions.*
- *Precision: Important to minimize false positives (flagging non-delinquent customers as risky).*
- *Recall: Important to identify as many true delinquent cases as possible.*
- *F1 Score: Balances precision and recall — suitable for imbalanced data.*
- *AUC-ROC: Indicates model's ability to separate classes across thresholds.*

Bias & Ethics:-
- We'll monitor feature importance to ensure that sensitive attributes (e.g., Location or Employment Status) do not introduce bias.
- We'll apply fairness-aware validation and stratified sampling.
- Ethical concern: Misclassifying a customer could deny them fair access to credit, so we'll maintain transparency and human review in model deployment.