**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:

Our predictive model forecasts customers at highest risk of delinquency using financial attributes and payment history. Two models were considered: **Logistic Regression** (simple) and **Random Forest Classifier** (complex).

**Recommended Model**: Random Forest Classifier – captures non-linear patterns and feature interactions, providing higher predictive accuracy.

**Top 5 Features**:

* **Missed Payments**: The strongest predictor of future delinquency.
* **Credit Utilization**: High usage indicates financial stress.
* **Debt-to-Income Ratio**: A higher ratio increases repayment risk.
* **Credit Score**: Lower scores correlate with higher delinquency likelihood.
* **Loan Balance**: Larger balances relative to income can hinder repayment.

**Conceptual Workflow**:

* **Data Preprocessing**: Clean and prepare features.
* **Model Training**: Build multiple decision trees on random subsets of data.
* **Prediction**: Each tree classifies customers.
* **Aggregation**: A majority vote determines the final “Low Risk” or “High Risk” prediction.
* **Interpretation**: Use SHAP/LIME for explainability.

### **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

The **Random Forest Classifier** is the ideal choice due to its high **accuracy** and **robustness** in handling complex financial datasets. It balances powerful predictive performance with sufficient **interpretability** (when using tools like SHAP/LIME), making it suitable for both effective risk management and regulatory compliance. Its ability to handle non-linear relationships in the data makes it more relevant for financial prediction than simpler linear models and aligns well with Geldium’s need for a reliable and actionable risk prediction tool.

### **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

**Evaluation Metrics**:

* **Precision, Recall, F1 Score**: These metrics will assess the model's accuracy, with a focus on its ability to correctly identify delinquent customers without an excessive number of false positives.
* **AUC (Area Under the Curve)**: We will use AUC to measure the model's overall discriminatory power, indicating its ability to correctly rank customers by risk.

**Bias and Fairness Checks**:

* We will evaluate the model's performance across different demographic groups (e.g., age, location) to detect any potential biases.
* If significant disparities are found, we will apply bias mitigation techniques like **re-sampling** or **re-weighting** the data to ensure fair and equitable predictions.

**Ethical Considerations**: The model should be used as a guide for risk management, not as a tool for discrimination. Customers and regulators should have **transparency** into how decisions are made, and the model should not perpetuate existing societal biases.