**Predictive Model Plan**

# 1. Model Logic (Generated with GenAI)

To predict customer delinquency, I explored two modeling options:

* **Logistic Regression**: A simple, interpretable baseline model well-suited for binary classification.
* **Random Forest Classifier**: A more advanced ensemble model that captures nonlinear relationships and handles mixed data types effectively.

**Recommended Model**: Random Forest Classifier. It balances predictive power and interpretability through feature importance.

**Top 5 input features**:

1. **Missed\_Payments** – Highly correlated with delinquency.
2. **Credit\_Utilization** – Reflects financial stress or over-dependence on credit.
3. **Credit\_Score** – Measures creditworthiness.
4. **Debt\_to\_Income\_Ratio** – Indicates repayment capacity.
5. **Recent Payment History (Month\_1 to Month\_6)** – Provides behavior trends.

**Model Workflow**:

1. Input features are preprocessed (imputed, normalized, and encoded).
2. Data is split into training and test sets using stratified sampling.
3. The Random Forest model is trained on labeled data.
4. Predictions are generated, and probability scores are output.
5. Model evaluation is conducted using classification metrics and fairness checks.

# 2. Justification for Model Choice

 **Model Selected**: **Random Forest Classifier**

 **High Accuracy**: Capable of modeling complex, nonlinear patterns, leading to better performance on structured financial data.

 **Handles Mixed Data Types**: Can effectively process both numerical and categorical variables without extensive preprocessing.

 **Interpretability via Feature Importance**: While not as transparent as logistic regression, Random Forests provide feature importance scores to explain model behavior.

 **Robust Against Overfitting**: Ensemble method reduces variance and improves generalization, especially helpful for moderately sized datasets like Geldium’s.

 **Business Suitability**: Aligns with Geldium’s need for accurate, explainable, and operationally scalable predictions in a regulated financial environment.

 **Deployment Readiness**: Easily integrated into production pipelines, supports retraining and monitoring for long-term use.

# 3. Evaluation Strategy

The model will be evaluated using the following metrics:

* **Accuracy**: Measures overall correctness of predictions.
* **Precision & Recall**: Help balance false positives and false negatives.
* **F1 Score**: Combines precision and recall to account for class imbalance.
* **ROC-AUC**: Evaluates model’s ability to distinguish between classes.
* **Fairness Checks**: Analyze performance across groups (e.g., Employment\_Status, Location) to identify bias.

Metrics will be interpreted with a focus on maximizing the F1 Score and AUC, as well as ensuring fairness through subgroup analysis. SHAP values or other explainability tools can be used to understand the model's decisions. Ethically, the model must avoid discriminatory behavior, especially in financial outcomes, and all predictions should be auditable.