TASK MODEL PLANNING

1. Model Logic (Generated with GenAI)

We aim to build a **classification model** to predict the likelihood of a customer becoming delinquent on their payments. The model follows this logic:

Step-by-Step Model Pipeline

1. **Data Preparation**
   - Load and clean the dataset (handle null values, outliers, etc.)
   - Encode categorical features using one-hot or label encoding
   - Normalize numerical fields (e.g., balance, income, due amount)
2. **Feature Engineering**
   - Create features such as `payment_ratio = paid_amount / due_amount`
   - Derive customer behavior features (e.g., frequency of late payments)
   - Encode tenure and activity signals
3. **Model Building**
   - Use a **Random Forest Classifier** to train on the labeled data
   - Split dataset into training and test (e.g., 80/20 split)
   - Optimize using cross-validation
4. **Prediction**
   - Output a delinquency probability (e.g., 0.83 = 83% chance)
   - Tag customers as "Low", "Medium", or "High" risk based on thresholds

---

2. Justification for Model Choice

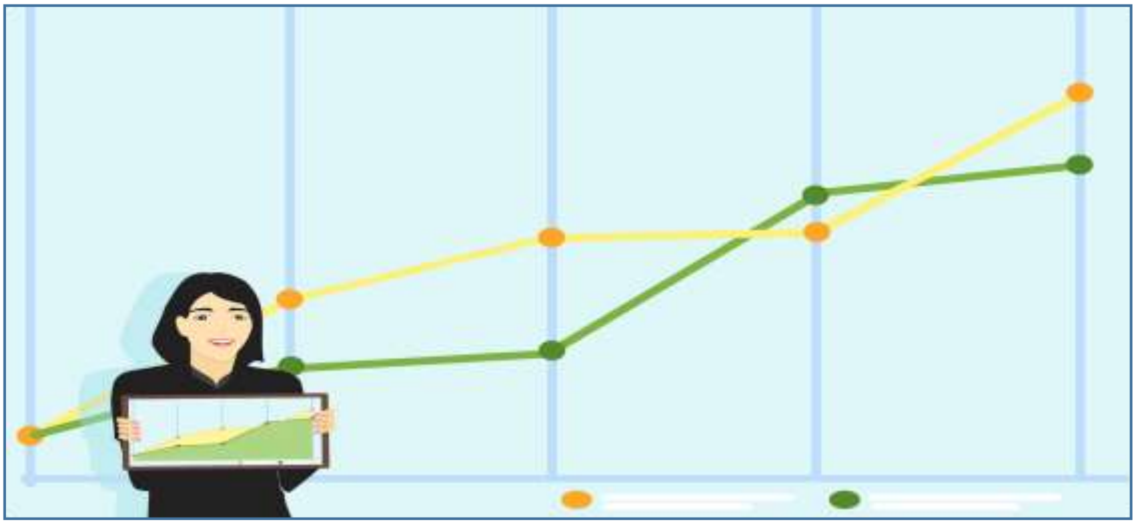
We selected the **Random Forest Classifier** because:

- ✅ **Accuracy:** Performs well with mixed-type features and handles class imbalance better than simpler models.
- ✅ **Transparency:** Allows feature importance analysis and is interpretable with SHAP values.
- ✅ **Ease of Use:** Quick to train, easy to debug, and resistant to overfitting.
- ✅ **Suitability:** Ideal for financial risk modeling, especially when both numerical (e.g., balance, tenure) and categorical (e.g., customer type) variables are involved.
- ✅ **Business Fit:** Aligns with Geldium's needs for an explainable model that provides confidence in operational decision-making.

---

3. Evaluation Strategy

To assess the model’s performance, we will use the following strategy:



📊 Evaluation Metrics

- **Accuracy**: Overall correctness of predictions
- **Precision**: Focused on correctly identifying actual defaulters (true positives)
- **Recall**: Ensures we capture as many real delinquents as possible
- **F1-Score**: Balances precision and recall
- **AUC-ROC**: Measures discriminatory power of the classifier

⚖️ Fairness & Bias Monitoring

- Monitor model bias across key attributes (e.g., age, gender, geography)
- Use fairness-aware metrics like **Equal Opportunity** or **Disparate Impact**
- Retrain model periodically to reduce drift and bias accumulation

🔐 Ethical Considerations

- Ensure **explainability** in all outputs (via SHAP or LIME)
- Do not penalize customers based on protected attributes
- Maintain **data privacy** and avoid using features that could reinforce financial discrimination

REPORT BY:

MICHEL(GENAI CONSULTANT)