**Model Logic using GenAI**

**Random Forest Model for Credit Delinquency Prediction**

Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy and robustness. It works well for credit risk assessment due to its ability to handle complex, nonlinear relationships.

**Model Summary**

* **Approach**: Uses multiple trees, each trained on different subsets of the data, to reduce overfitting and improve generalization.
* **Top 5 Input Features for Prediction**

1. **Missed Payments** – Strongest predictor, indicating historical delinquency trends.
2. **Credit Utilization Ratio** – High ratios suggest financial instability.
3. **Debt-to-Income Ratio** – Higher values indicate potential payment challenges.
4. **Credit Score** – Lower scores correlate with higher delinquency risk.
5. **Loan Balance** – Large outstanding debts can signal repayment difficulties.

**Justifying model choice**

Random Forest is an ideal model for predicting credit delinquency because it handles complex, nonlinear relationships between financial and behavioral factors. Unlike simpler models like Logistic Regression, Random Forest aggregates multiple decision trees, reducing overfitting and improving generalization. It works well with imbalanced data, ensuring delinquent cases aren’t overlooked, and provides clear feature importance rankings, making predictions interpretable for financial institutions. Given its efficiency, scalability, and robustness against noise, it’s well-suited for identifying high-risk customers based on key indicators like missed payments, credit utilization, and debt-to-income ratio

**Plan for evaluating model performance**

1. **ROC-AUC Score** – Measures the model's ability to distinguish between delinquent and non-delinquent customers. A higher AUC score indicates better performance.
2. **F1-Score** – Balances precision (avoiding false positives) and recall (avoiding false negatives), making it ideal for imbalanced datasets.
3. **Recall** – Important for financial risk models, as it ensures the model correctly identifies actual delinquent customers, minimizing missed high-risk cases