Task 2

✅ 1. Model Logic (Generated with GenAI)

Selected Model Type: Random Forest Classifier

Model Goal: To predict whether a customer is likely to become delinquent based on historical financial behavior and profile data.

GenAI-Based Model Logic:

1. Input customer financial data (e.g., Income, Credit_Utilization, Missed_Payments)
2. Encode categorical data and fill missing values
3. Use Random Forest to train on historical data with Delinquent_Account as the target
4. Generate predictions as risk scores between 0 and 1
5. Rank customers by risk to guide Collections team interventions

Top 5 Input Features:

- Missed_Payments
- Credit_Utilization
- Debt_to_Income_Ratio
- Income
- Employment_Status

✅ 2. Justification for Model Choice

Random Forest was chosen for its strong predictive power and ability to handle both categorical and numerical data.
It is less prone to overfitting than a decision tree and more flexible than logistic regression when capturing nonlinear patterns in financial behavior.
While neural networks may offer greater accuracy in very large datasets, they are less interpretable, making Random Forest a more practical choice for Geldium's operational needs.
The model is explainable (via feature importance or SHAP values), scalable, and aligns with the financial services sector's need for accuracy, transparency, and compliance.

✅ 3. Evaluation Strategy

Evaluation Metrics:

- **Accuracy:** Overall correctness of predictions
- **Precision:** How many predicted delinquents were actually delinquent
- **Recall:** How many actual delinquents were successfully identified
- **F1 Score:** Balanced measure between precision and recall
- **ROC-AUC:** Measures model's ability to separate delinquent vs. non-delinquent customers

Bias and Fairness Checks:

- Compare precision/recall across demographics (e.g., income levels, employment status)
- Use fairness libraries like Fairlearn or Aequitas
- Avoid including proxy variables (like ZIP code) that may indirectly reflect protected attributes

Ethical Considerations:

- Ensure predictions are explainable and auditable
- Avoid reinforcing historical bias in past credit decisions
- Clearly communicate how risk predictions are used in interventions

---