# Predictive Model Plan – Task 2

## 1. Model Logic (Generated with GenAI)

Based on the dataset and insights from EDA, I propose using a Logistic Regression model to predict customer delinquency. Logistic regression is suitable because the target variable (Delinquent\_Account) is binary (0 = Not Delinquent, 1 = Delinquent).  
  
The model would take input features such as Income, Missed\_Payments, Credit\_Utilization, Debt\_to\_Income\_Ratio, and Account\_Tenure to calculate the probability of delinquency. The steps include:  
1. Data Ingestion: Load the dataset.  
2. Feature Selection: Use top predictive variables.  
3. Data Preprocessing: Handle missing values and normalize data.  
4. Model Training: Fit logistic regression on the labeled data.  
5. Prediction: Output a probability score and classify as delinquent if probability > 0.5.  
6. Evaluation: Use metrics such as precision, recall, F1 score, and AUC-ROC.

Top 5 Input Features:

* - Income
* - Missed\_Payments
* - Credit\_Utilization
* - Debt\_to\_Income\_Ratio
* - Account\_Tenure

## 2. Justification for Model Choice

Logistic Regression is chosen for its simplicity, interpretability, and suitability for binary classification. It helps estimate the likelihood of a customer being delinquent while providing transparency into how each feature affects the outcome. For financial institutions like Geldium, this model allows easy explanation to stakeholders, supports regulatory compliance, and can be quickly deployed and monitored. Although more complex models like neural networks may offer higher accuracy, logistic regression strikes a good balance between performance and explainability—critical for responsible decision-making in credit risk.

## 3. Evaluation Strategy

To evaluate the model’s performance, the following strategy will be applied:  
  
- Accuracy: Measures overall correct predictions.  
- Precision: Ensures flagged customers are truly delinquent.  
- Recall: Captures the maximum number of actual delinquent customers.  
- F1 Score: Balances precision and recall.  
- AUC-ROC: Measures the model's ability to separate classes.  
- Confusion Matrix: Visual analysis of prediction types.  
  
Bias Detection:  
Fairness checks such as demographic parity and disparate impact will be conducted to ensure that no customer group is unfairly disadvantaged. GenAI tools will help identify unusual patterns, but results will be validated using statistical fairness metrics.  
  
Ethical Consideration:  
All predictions will be reviewed with human oversight to avoid unfair treatment, and predictions will be monitored continuously for unintended bias.