Predictive Model Plan – Student Template

1. Model Logic (Generated with GenAI)

GenAI Tool Used: *ChatGPT*

Model Logic / Structure

The objective of this model is to ***predict whether a customer is likely to become delinquent (Yes/No)*** *based on historical loan and repayment data.*

Step-by-Step Model Logic:

1. *Load the customer delinquency dataset.*
2. *Identify the target variable:*
    a. ***Delinquent_Flag (Yes / No)***
3. *Select important input features:*
    a. *Days Past Due (DPD)*
    b. *Number of missed EMIs*
    c. *Loan amount*
    d. *EMI amount*
    e. *Income*
    f. *Employment status*
    g. *Loan tenure and interest rate*
4. *Clean the data by handling missing values and ensuring correct data types.*
5. *Split the dataset into training and testing sets.*
6. *Apply* ***Logistic Regression*** *to learn patterns between customer behavior and delinquency.*
7. *Generate predictions indicating* ***high risk or low risk*** *of delinquency for each customer.*

Model Purpose (Simple Explanation):

The model analyzes past customer behavior and loan details to estimate the ***probability of future delinquency****, enabling early risk identification and proactive intervention.*

2. Justification for Model Choice

The selected model for delinquency prediction is ***Logistic Regression****.*

Reasons for Selection:

- ***Accuracy:***
  Logistic Regression performs well for binary classification problems such as delinquent vs non-delinquent customers.

- ***Transparency & Explainability:***
  The model clearly shows how each variable (DPD, missed EMIs, income, etc.) impacts delinquency risk, which is critical for financial decision-making.
- ***Ease of Implementation:***
  It is simple to implement, maintain, and explain to non-technical stakeholders.
- ***Relevance for Financial Prediction:***
  Logistic Regression is widely used in ***credit risk and default prediction*** *across banking and financial institutions.*
- ***Suitability for Geldium's Business Needs:***
  Geldium requires an interpretable, reliable, and regulator-friendly model to support fair customer treatment and proactive collection strategies.

3. Evaluation Strategy

Performance Metrics Used:

- ***Accuracy:***
  Measures overall correctness of predictions.
- ***Precision:***
  Indicates how many predicted delinquent customers were actually delinquent.
- ***Recall:***
  Measures how well the model identifies actual delinquent customers (very important for risk prevention).
- ***F1 Score:***
  Balances precision and recall, especially useful if delinquent cases are fewer.
- ***ROC-AUC Score:***
  Evaluates the model's ability to distinguish between delinquent and non-delinquent customers.

Interpretation of Metrics:

- *High recall ensures fewer risky customers are missed.*
- *A good AUC score indicates strong separation between high-risk and low-risk customers.*
- *Balanced precision avoids unnecessary intervention on low-risk customers.*

Bias Detection & Reduction:

- *Avoid over-reliance on sensitive attributes such as gender or location.*
- *Monitor prediction outcomes across different customer groups.*
- *Regularly validate the model to ensure fair treatment.*

Ethical Considerations:

- *Predictions should support* ***early assistance****, not customer discrimination.*

- *Model decisions must be explainable and transparent.*
- *Customer data privacy and responsible AI usage must be strictly followed.*