**Predictive Model Plan – Student Template**

Use this template to structure your submission. You can copy and paste content from GenAI tools and build around it with your own analysis.

# 1. Model Logic (Generated with GenAI)

The predictive model is designed to estimate the probability that a customer will become delinquent (fail to pay their loan installments). The steps include:

* 1. Data Preparation:
* Import customer loan and repayment history.
* Clean missing values, normalize continuous variables (income, balance), and encode categorical data (region, loan type).
  1. Feature selection:
* Identify key predictors: credit score, credit utilization ratio, income-to-debt ratio, payment history, number of missed payments.
  1. Model training:
* Split data into training (70%) and testing (30%).
* Train baseline Logistic Regression (interpretable).
* Train advanced models like Random Forest and Gradient Boosting for higher accuracy.
  1. Prediction:
* Output: Delinquent = Yes/No.
* Also provide a probability score (0–1) for risk assessment.
  1. Business use:
* Flag high-risk customers early.
* Prioritize outreach and personalized repayment plans.

# 2. Justification for Model Choice

2.1 Logistic Regression

* Provides probability scores for decision making
* Transparent, easy to interpret → critical in finance where regulators require explainability.

2.2 Random Forest

* Handle complex relationships better than linear models.
* Improve prediction accuracy.
* Can highlight feature importance (e.g., credit utilization > late payments).

Why suitable for Geldium?

* Balances accuracy (detecting risky accounts) with interpretability (justifying decisions to regulators and customers).
* Scalable for deployment in large financial datasets.

# 3. Evaluation Strategy

1. Metrics Used
   * Accuracy → Overall performance.
   * Recall (Sensitivity) → Ability to detect most delinquent customers (important for collections).
   * Precision → Avoids flagging too many “safe” customers as risky.
   * F1 Score → Balance between recall and precision.
   * AUC-ROC → Overall risk discrimination ability.
2. Interpretation
   * If Recall is high, Geldium can catch more risky customers early.
   * If Precision is reasonable, it avoids overburdening collections with false alarms.
   * AUC > 0.80 indicates strong model performance.
3. Bias & Ethical Considerations
   * Exclude sensitive features like gender, religion, ethnicity.
   * Monitor bias by subgroup analysis (e.g., ensure model treats regions/customers fairly).
   * Provide explainability reports (using SHAP or feature importance).
   * Ensure transparency with customers: predictions should support—not penalize—vulnerable borrowers.