# Predictive Model Plan – Student Submission

## 1. Model Logic (Generated with GenAI)

Chosen Model Type: Logistic Regression (Simple) and XGBoost (Complex)  
  
Preferred Model: Logistic Regression (for transparency and compliance)  
  
Model Workflow:  
1. Data Ingestion  
 Load dataset including features such as income, credit utilization, number of missed payments, age, and credit limit.  
  
2. Feature Engineering  
 - Normalize numerical features like income and credit utilization.  
 - Create binary flags for past delinquencies.  
 - One-hot encode categorical variables (e.g., employment status, education level).  
 - Possibly derive new features, e.g., income-to-limit ratio.  
  
3. Feature Selection  
 Use correlation analysis and model-based importance metrics to retain the most predictive features.  
  
4. Model Training  
 - Split data into train/test (80/20).  
 - Train a logistic regression model using relevant features.  
 - Apply regularization (L1/L2) to prevent overfitting.  
  
5. Prediction Output  
 - The model outputs a probability score (0 to 1) representing the likelihood of delinquency.

## 2. Justification for Model Choice

Logistic regression is an excellent fit for predicting credit delinquency due to its high interpretability, ease of implementation, and proven effectiveness in financial risk modeling. It provides clear coefficients that help explain the impact of each feature, aligning with Geldium’s need for transparency and regulatory compliance. While more complex models like XGBoost offer higher predictive power, logistic regression balances performance with the need for clear, auditable decisions in customer financial profiling. It is also computationally efficient and well-suited to handle binary classification problems like delinquency prediction.

## 3. Evaluation Strategy

The model’s performance will be evaluated using:  
- Accuracy: Overall correctness of predictions.  
- Precision & Recall: To balance false positives (e.g., wrongly flagging a good customer) and false negatives.  
- F1 Score: Harmonic mean of precision and recall, especially valuable with class imbalance.  
- ROC-AUC: To measure model discrimination power across thresholds.  
  
Fairness & Bias Checks:  
- Compare performance metrics across sensitive groups (e.g., gender, income level).  
- Use fairness metrics like Demographic Parity and Equal Opportunity.  
- Apply bias mitigation techniques if necessary, such as reweighting or adversarial debiasing.  
  
Ethical Considerations:  
- Ensure the model does not reinforce systemic biases.  
- Maintain transparency in model decision-making.  
- Provide clear, customer-understandable reasons for adverse decisions.  
  
This evaluation strategy ensures the model is not only accurate but fair and compliant with financial regulations.