# Predictive Model Plan – Geldium Delinquency Prediction

## 1. Model Logic (Generated with GenAI)

Model Type: Logistic Regression (simple & interpretable) or Gradient Boosted Trees (complex & high-performing).  
  
Workflow:  
1. Data Ingestion – Load customer demographic, financial history, and credit behavior data from Geldium’s dataset.  
2. Feature Selection – Choose top predictors such as:  
 - Credit utilization ratio  
 - Number of missed payments in past 12 months  
 - Monthly income  
 - Loan-to-income ratio  
 - Age of credit account  
3. Preprocessing – Handle missing data (imputation), normalize numeric variables, encode categorical features.  
4. Model Training – Train using historical delinquency labels (0 = non-delinquent, 1 = delinquent).  
5. Prediction – For new customers, model outputs a probability of delinquency.  
6. Risk Scoring – Probability score is converted into “Low / Medium / High” risk categories.  
  
Summary: The model predicts the likelihood that a customer will miss debt payments based on financial history and credit behavior.   
It uses key features like credit utilization and missed payments to assign a delinquency probability score, helping Geldium prioritize risk management.

## 2. Justification for Model Choice

I recommend Logistic Regression as the primary model because it is transparent, interpretable, and widely accepted in financial risk modeling,   
meeting regulatory compliance requirements. It allows Geldium to clearly explain decisions to customers and regulators while maintaining solid predictive power.   
While more complex models like Gradient Boosted Trees can improve accuracy, logistic regression offers a strong balance between performance and explainability,   
critical for trust and adoption in financial services.

## 3. Evaluation Strategy

Metrics:  
- Accuracy – overall correctness of predictions  
- Precision & Recall – balance between catching high-risk customers and avoiding false alarms  
- F1 Score – harmonic mean of precision and recall, useful for imbalanced data  
- AUC-ROC – ability to distinguish between delinquent and non-delinquent customers  
  
Bias Checks:  
- Demographic parity – ensure risk predictions are not disproportionately high for certain groups  
- Disparate impact analysis – detect unfair treatment between demographic segments  
  
Ethical Considerations:  
- Use only relevant financial and behavioral data, avoid discriminatory variables  
- Provide transparency in decisions to maintain customer trust  
  
Interpretation:  
- High AUC (>0.85) means strong separation between classes  
- F1 Score >0.75 indicates balanced precision/recall  
- If bias metrics show disparity, re-train using fairness constraints or re-sampling