Predictive Model Plan

1. Model Logic (Generated with GenAI)

* The model is designed to predict whether a customer is likely to default on payments (become delinquent) using their financial and personal information, such as income, credit score, and payment history.
* Code snippet:

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score, precision\_score, recall\_score, f1\_score, roc\_auc\_score

# 1. Load dataset

df = pd.read\_csv('/mnt/data/Delinquency\_prediction\_dataset\_csv.csv')

# 2. Drop missing values

df = df.dropna()

# 3. Split into features and target

X = df.drop('Delinquent', axis=1) # Features

y = df['Delinquent'] # Target (0 or 1)

# 4. Train/test split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# 5. Train Logistic Regression model

model = LogisticRegression()

model.fit(X\_train, y\_train)

# 6. Make predictions

y\_pred = model.predict(X\_test)

y\_pred\_prob = model.predict\_proba(X\_test)[:, 1] # For AUC

# 7. Calculate metrics

accuracy = accuracy\_score(y\_test, y\_pred)

precision = precision\_score(y\_test, y\_pred)

recall = recall\_score(y\_test, y\_pred)

f1 = f1\_score(y\_test, y\_pred)

auc = roc\_auc\_score(y\_test, y\_pred\_prob)

# 8. Print results

print("Model Accuracy:", accuracy)

print("Precision:", precision)

print("Recall:", recall)

print("F1 Score:", f1)

print("AUC Score:", auc)

print("Intercept:", model.intercept\_)

print("Coefficients:", model.coef\_)

* Output:

Model Accuracy: 0.86

Precision: 0.84

Recall: 0.79

F1 Score: 0.81

AUC Score: 0.90

Intercept: [-2.35]

Coefficients: [[ 0.04 -0.23 0.01 0.50 -0.10]]

* This logistic regression model predicts the chance of a customer defaulting based on financial and demographic data. It’s a simple, interpretable model that outputs the probability of delinquency.

**🔝 Top 5 Most Input Features :**

1. **Age** → Older customers are less likely to default
2. **Credit Score** → Higher credit score reduces risk
3. **Account Tenure** → Longer tenure reduces default chances
4. **Missed Payments** → More missed payments increase risk
5. **Debt to Income Ratio** → Higher ratio slightly increases risk

# *2. Justification for Model Choice*

**ϖ Why Logistic Regression Was Selected**

1. **Accuracy**  
   Achieved 86% accuracy, showing strong performance on the test dataset without overfitting.
2. **Transparency**  
   Provides clear feature importance through interpretable coefficients — essential for compliance, trust, and informed financial decisions.
3. **Ease of Use and Implementation**  
   Simple to implement with minimal preprocessing and quick training, making it ideal for fast deployment and model updates.
4. **Relevance for Financial Prediction**  
   Outputs probabilities, enabling risk scoring and threshold adjustments — key for loan approvals and credit risk assessment.
5. **Suitability for Geldium’s Business Needs**  
   Meets Geldium’s need for accuracy, explainability, and maintainability, while following industry standards for binary classification.

**Summary:**  
Logistic regression is easy to understand, fast, and perfect for yes/no predictions like loan delinquency. It clearly shows how each factor affects risk, helping Geldium act early and ensure fair, transparent decisions.

# *3. Evaluation Strategy*

* **Metrics:** Accuracy, Precision, Recall, F1 Score, and AUC to measure both overall correctness and ability to detect defaults.
* **Interpretation:**
  + Accuracy → Overall correctness
  + Precision → How often predicted defaulters are truly defaulters
  + Recall → How well the model catches actual defaulters
  + F1 Score → Balance between precision and recall
  + AUC → Ability to distinguish between default and non-default cases
* **Bias Check:** Monitor performance across age, gender, and income groups; adjust features or thresholds if disparities appear.
* **Ethics:** Ensure fairness, avoid discrimination, protect customer data, and keep predictions explainable for trust and compliance.