**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)

Use a GenAI tool (e.g., ChatGPT, Gemini) to generate the logic or structure of your predictive model.  
- You may include pseudo-code, a step-by-step process, or a simplified code snippet.  
- Briefly explain what the model is designed to do.

Paste your GenAI-generated output below or describe the logic in your own words:

# Step 1: Import libraries

import pandas as pd

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

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report

# Step 2: Load and clean data

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

df.fillna(df.median(), inplace=True) # Impute missing values

# Step 3: Feature selection

features = ['Credit\_Utilization', 'Missed\_Payments', 'Debt\_to\_Income\_Ratio',

'Credit\_Score', 'Employment\_Status', 'Income']

X = pd.get\_dummies(df[features], drop\_first=True)

y = df['Delinquent\_Account']

# Step 4: Train-test split

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

# Step 5: Train model

model = RandomForestClassifier(n\_estimators=100, random\_state=42)

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

# Step 6: Evaluate

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

print("Accuracy:", accuracy\_score(y\_test, y\_pred))

print(classification\_report(y\_test, y\_pred))

# 2. Justification for Model Choice

Explain why you selected this specific model type (e.g., logistic regression, decision tree, neural network). Consider:  
- Accuracy  
- Transparency  
- Ease of use or implementation  
- Relevance for financial prediction  
- Suitability for Geldium’s business needs

We selected the **Random Forest Classifier** for the following reasons:

* **High Accuracy:** It performs well on classification tasks with mixed data types
* **Transparency:** Feature importance scores help explain predictions to business teams
* **Robustness:** Handles missing data and outliers better than linear models
* **Scalability:** Suitable for large datasets and real-time scoring
* **Business Fit:** Aligns with Geldium’s need for a reliable, interpretable model to support collections decisions

# 3. Evaluation Strategy

Outline how you would evaluate your model’s performance. Include:  
- Which metrics you would use (e.g., accuracy, precision, recall, F1 score, AUC)  
- How you would interpret those metrics  
- Any plans to detect or reduce bias in your model  
- Ethical considerations in making predictions about customer financial behavior

To evaluate the model’s performance, we will use the following metrics:

* **Accuracy:** Measures overall correctness of predictions
* **Precision:** Focuses on how many predicted delinquents were actually delinquent
* **Recall:** Captures how many actual delinquents were correctly identified
* **F1 Score:** Balances precision and recall
* **AUC (Area Under Curve):** Assesses model’s ability to distinguish between classes

**Interpretation Plan:**

* A high recall ensures we don’t miss risky customers
* A balanced F1 score ensures both precision and recall are optimized
* AUC helps validate model discrimination power

**Bias Detection & Ethical Considerations:**

* Monitor prediction rates across income levels, employment types, and locations
* Avoid over-penalizing low-income or unemployed customers
* Ensure transparency in how predictions are used for outreach and interventions
* Use GenAI to simulate edge cases and test fairness