**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 goal of this predictive model is to estimate the likelihood that a customer will become delinquent based on financial and behavioral attributes. Here is the high-level logic:

**Model Pipeline:**

1. **Data Cleaning**

* **Impute missing income values using median imputation grouped by employment status.**
* **Normalize categorical features (e.g.,Employment\_Status,Credit\_Card\_Type).**
* **Cap credit utilization at 1.0.**

1. **Feature Engineering**

* **Encode categorical variables using one-hot encoding.**
* **Aggregate payment history from Month\_1 to Month\_6 to create "Risk\_Score".**

1. **Model Training**

* **Use Logistic Regression for baseline interpretability.**
* **Evaluate against Random Forest and XGBoost for potential performance gains.**

1. **Prediction Output**

* **Predict Delinquent\_Account (0 = No, 1 = Yes).**
* **Output probability scores for each customer.**

# 2. Justification for Model Choice

**We propose using Logistic Regression as the baseline model because:**

* It offers hightransparency, making it easy to explain the influence of variables to stakeholders.
* It performs well when the target variable is binary (like delinquency).
* It is computationally efficient and interpretable, allowing Geldium to audit decisions.

**We also plan to test Random Forest and Boost as alternatives because:**

* They often provide **higher predictive accuracy** by capturing nonlinear relationships.
* XGBoost, in particular, is known for its **robustness in financial applications**.
* These models can handle missing values and outliers more gracefully.

This multi-model approach allows us to balance **performance** and **business interpretability**.

# 3. Evaluation Strategy

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

* **Accuracy**: To assess the overall correctness of predictions.
* **Precision & Recall**: Important due to the cost of false positives (flagging good customers) and false negatives (missing at-risk customers).
* **F1 Score**: Balances precision and recall.
* **AUC-ROC**: Measures how well the model separates delinquents from non-delinquents.

**Bias Detection & Mitigation:**

* Monitor model performance across different groups (e.g., age, employment status).
* Apply **rebalancing techniques** like SMOTE if the dataset is imbalanced.

**Ethical Considerations:**

* Avoid discriminatory features (e.g., race, gender).
* Ensure decisions are used to offer **supportive financial solutions**, not punitive actions.