Predictive Model Plan – Student Template

1. Model Logic (Generated with GenAI)

Model Type: Gradient Boosting Machine (LightGBM or XGBoost)

How It Works ?

A Gradient Boosting Machine builds many weak prediction models (usually decision trees) one at a time. Each new tree fixes the mistakes made by previous trees.

Step-by-Step Process:

* **Start Simple**
  Begin with a basic prediction like the average delinquency rate.

* **Find the Mistakes**
  For each customer, calculate the difference between what actually happened and what the model predicted. These differences are called residuals.

* **Build a Tree to Fix Mistakes**
  Train a new decision tree to predict these residuals. This tree learns to correct the current errors.

* **Add the Fix**
  Add this new tree's predictions to the overall prediction (adjusted by a learning rate).

* **Keep Going**
  Repeat steps 2-4 for a set number of rounds or until the model stops improving. Each round makes the model better at avoiding errors.

* **Make Final Prediction**
  Add up predictions from all trees. Convert this sum into a probability using a mathematical function like sigmoid.

* **Make the Decision**
  Apply a cutoff (usually 0.5) to turn probabilities into yes/no decisions about delinquency.

Purpose

This model sorts customers into two groups: likely to become delinquent or unlikely to become delinquent. It uses financial and behavioral data to make these predictions. Banks use this for risk management and decision making.

2. Justification for Model Choice

Geldium needs to predict which customers might stop paying their loans. Getting this wrong costs money and hurts the business.

Gradient Boosting Machines work better than simpler models for this job. They spot complex patterns in customer data that other models miss. For example, they can see how credit scores, income, debt, and payment history work together in ways that aren't obvious.

Better predictions mean Geldium can focus on the right customers and make smarter decisions. This protects the company's money and keeps the loan business healthy.

3. Evaluation Strategy

Testing a delinquency model is tricky because most customers pay on time. We need metrics that work well even when the data is unbalanced.

Key Metrics:

- **Accuracy** measures how often the model gets it right overall. We'll track this but won't rely on it alone. A model could look accurate just by predicting everyone pays on time.
- **F1 Score** balances finding actual problem customers with avoiding false alarms. A high F1 score means we catch real delinquent customers without flagging too many good customers by mistake.
- **AUC-ROC (*Area Under the Receiver Operating Characteristic Curve*)** shows how well the model ranks customers by risk. Higher scores (closer to 1.0) mean the model can sort customers better, helping Geldium focus on the riskiest accounts first.
- **Fairness Checks** make sure the model treats different groups equally. We'll check if prediction rates are similar across demographics and if we catch problems at similar rates for all groups.
- **Spotting Unfair Predictions:** If the model performs worse for certain age groups or income levels, that's a red flag. We need to fix bias before using the model.

Preventing Bias

- **Before Training**

  We'll use techniques like **SMOTE** to balance the data better. This helps the model learn about problem customers without ignoring them. We'll also check if any features might unfairly represent protected groups.

- **During Training**

  We'll adjust model settings to handle unbalanced data. Tools like XGBoost and LightGBM have built-in options for this.

- **After Training**

  If fairness tests show problems, we'll adjust decision thresholds for different groups. For borderline cases, humans will review the decisions.

Ethical Responsibilities

- **Transparency**

  Gradient boosting models are complex, but we'll use tools like SHAP to explain important decisions. Customers deserve to understand why they were rejected.

- **Fairness**

  We'll actively watch for and fix bias. Loan decisions should be based on financial risk, not demographics.

- **Privacy**

  Customer financial data stays protected according to all regulations.

- **Accountability**

  Clear rules about who's responsible for the model. Systems in place to catch and fix problems.

- **Human Oversight**

  Humans review important decisions, especially loan denials. Automated systems support decisions but don't make them alone.

The END goal is building a model that's accurate, fair, and trustworthy. This protects both Geldium and its customers.