Predictive Model Plan – Student Template

1. Model Logic (Generated with GenAI)

Chosen Model Type: Gradient Boosting Machine (e.g., LightGBM or XGBoost)

Logic/Structure: A Gradient Boosting Machine for delinquency prediction operates by iteratively building an ensemble of weak prediction models, typically decision trees. Each new tree in the sequence attempts to correct the errors made by the previous trees.

Step-by-step Process:

1. **Initialize Base Model:** Start with an initial simple prediction (e.g., the average delinquency rate).
2. **Calculate Residuals:** For each instance, calculate the "residual" (the difference between the actual delinquency status and the current prediction). These residuals are the errors that the next tree will try to predict.
3. **Build New Tree:** Train a new weak decision tree to predict these residuals. This tree learns to correct the errors of the ensemble so far.
4. **Update Predictions:** Add the predictions of this new tree (scaled by a learning rate) to the ensemble's overall prediction.
5. **Repeat:** Steps 2-4 are repeated for a specified number of iterations (or until performance stops improving). Each iteration refines the model's ability to minimize prediction errors.
6. **Final Prediction:** The final prediction is the sum of the predictions from all individual trees in the ensemble, transformed (e.g., using a sigmoid function for binary classification) to output a probability of delinquency.
7. **Thresholding:** A classification threshold (e.g., 0.5) is applied to these probabilities to assign a binary outcome (delinquent or not delinquent).

Model Purpose: The model is designed to accurately classify customers as either delinquent or non-delinquent based on their financial and behavioral features, enabling proactive risk management and informed decision-making.

2. Justification for Model Choice

For Geldium, a financial institution, accurately predicting customer delinquency is paramount for minimizing financial losses, optimizing risk management, and ensuring sustainable growth. Our chosen model, Gradient Boosting Machines, directly aligns with these goals by offering superior predictive performance over simpler models. Its ability to effectively capture intricate, non-linear relationships within the customer data, including the subtle interplay between credit score, income, debt, and payment history, allows Geldium to identify high-risk accounts with greater precision. This enhanced accuracy means more effective risk mitigation strategies, more targeted interventions, and ultimately, a healthier loan portfolio, directly contributing to Geldium's financial stability and operational efficiency.

3. Evaluation Strategy

Evaluating the delinquency prediction model requires a comprehensive approach that assesses both its predictive power and its fairness, especially given the imbalanced nature of the target variable.

Metrics to Use:

- **Accuracy:** The proportion of all correctly classified instances (both delinquent and non-delinquent).
- **F1 Score:** The harmonic mean of Precision and Recall, providing a balance between correctly identifying positive cases and avoiding false alarms. We will specifically focus on the F1-score for the positive class (delinquent).
- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** Measures the model's ability to distinguish between the positive (delinquent) and negative (non-delinquent) classes across all possible classification thresholds.
- **Fairness Checks:** Metrics to assess if the model exhibits bias across different demographic or protected groups.

How to Interpret Those Metrics:

- **Accuracy:** Will be monitored but not solely relied upon, as high accuracy can be misleading in imbalanced datasets (e.g., a model predicting no delinquency could still have high accuracy due to the majority class).
- **F1 Score:** A high F1 score for the 'delinquent' class indicates that the model is effectively balancing the identification of actual delinquent customers (Recall) with minimizing false alarms (Precision). This is crucial for practical application in risk management.
- **AUC-ROC:** A higher AUC-ROC value (closer to 1.0) signifies that the model has strong discriminative power, meaning it can effectively rank customers by their likelihood of delinquency, allowing Geldium to prioritize risk management efforts.
- **Fairness Checks:** We will examine metrics such as **Demographic Parity** (ensuring similar prediction rates across groups) and **Equalized Odds** (ensuring similar true positive rates across groups). If, for example, the model has a significantly lower true positive rate for a particular age group or income bracket, it indicates bias that needs to be addressed.

Plans to Detect or Reduce Bias in the Model:

1. **Pre-processing:**
   - **Resampling:** Consider techniques like SMOTE (Synthetic Minority Over-sampling Technique) to address the class imbalance of the

`Delinquent_Account` variable, which can implicitly reduce bias against the minority class.
    - **Data Exploration for Proxies:** Continuously analyze features for potential proxies of protected attributes, even if direct protected attributes are not used.
2. **In-processing:**
    - **Algorithmic Parameters:** Utilize any bias-mitigation parameters available within the chosen Gradient Boosting library (e.g., `scale_pos_weight` in XGBoost/LightGBM to handle class imbalance, which can also help prevent bias against the minority class).
3. **Post-processing:**
    - **Threshold Adjustment:** Experiment with different classification thresholds for different groups if fairness metrics indicate disparate impact, to balance performance and fairness across groups.
    - **Reject Option Classification:** For ambiguous cases near the decision boundary, consider human review to prevent biased automated decisions.

Ethical Considerations in Making Predictions About Customer Financial Behavior:

Ethical considerations are paramount. We must ensure:

- **Transparency:** While Gradient Boosting Machines are less interpretable, efforts will be made to use feature importance and post-hoc explanation techniques (e.g., SHAP values) to understand the key drivers of predictions, especially for rejected applications.
- **Fairness:** Actively monitor and mitigate bias across different demographic segments to ensure equitable treatment and prevent discriminatory lending practices. Decisions should be based on risk, not on protected attributes.
- **Privacy:** Adhere strictly to data privacy regulations regarding customer financial information.
- **Accountability:** Establish clear lines of accountability for model development, deployment, and monitoring, with mechanisms for auditability and redress if biased outcomes occur.
- **Human Oversight:** Maintain human oversight in critical decision-making processes, particularly for loan denials, to prevent over-reliance on automated predictions and allow for contextual judgment.