

Predictive Model Plan

1. Model Logic (Generated with GenAI)

Using ChatGPT, I generated a predictive model using logistic regression to estimate the likelihood of a customer becoming delinquent. The model uses key features such as `Credit_Utilization`, `Missed_Payments`, `Income`, `Debt_to_Income_Ratio`, and `Account_Tenure` to predict a binary outcome: 1 if the customer is likely to become delinquent, and 0 otherwise.

Pseudo-code:

1. Load dataset
2. Select features: `['Credit_Utilization', 'Missed_Payments', 'Income', 'Debt_to_Income_Ratio', 'Account_Tenure']`
3. Define target variable: `'Delinquent_Account'`
4. Split data into training and testing sets
5. Fit logistic regression model
6. Predict and evaluate using classification metrics

2. Justification for Model Choice

I chose logistic regression because it is widely used for binary classification problems and is highly interpretable. In financial services, model transparency is crucial, and logistic regression offers clear coefficient outputs to explain each predictor's influence.

The model is simple to implement, does not require large computational resources, and provides strong baseline performance for credit risk analysis. It allows for quick iteration and stakeholder communication, making it an ideal fit for Geldium's goal of responsibly identifying at-risk customers.

3. Evaluation Strategy

To evaluate the model, I would use accuracy, precision, recall, F1 score, and AUC. Precision and recall are particularly important: precision ensures we avoid unnecessary interventions for low-risk customers, while recall helps identify most high-risk customers.

The **F1** score balances both, and **AUC** evaluates the model's ability to distinguish between classes across thresholds.

To check for bias, I would examine prediction patterns across demographic segments (e.g., `Employment_Status` or `Location`) to ensure fairness.

Any strong disparities would prompt model reassessment or rebalancing. Ethical considerations include avoiding proxy bias, maintaining transparency, and clearly communicating how model outputs influence decisions.