**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)

Using Deepseek AI, I generated a predictive model using Decision Trees to estimate the risk of delinquency for any given customer. The model uses several key features from the dataset to predict one of two possible outcomes: likely delinquency or unlikely delinquency. The logic of the model is as follows:

The model chosen would be eXtreme Gradient Boosting, a powerful ensemble of decision trees geared towards credit risk prediction.

It accepts the raw customer data: Age, Income, Missed\_Payments, Credit\_utilization and Employment\_Status- imputing and cleaning it where necessary.

Put simply, it would Load the data, Create features, do pre-processing, train to optimize for catching true delinquencies, evaluate the results and explain its predictions.

# 2. Justification for Model Choice

I chose it for its ability to handle non-linear patterns, interactions and imbalanced data accurately while still having transparency through its SHAP values. The dataset is not particularly large so the decision trees shouldn’t be too hard to implement. It should be able to predict delinquency more effectively than logistic regression, while not being too dense like neural networks are. Such traits would be good for Geldium’s business needs, as the model is tailored to predict delinquency while still allowing the analytics team and the decision makers understand their data.

# 3. Evaluation Strategy

The metrics that would be used to evaluate the model’s performance would be Recall, Precision, F-1 Score, and AUC-ROC. They would be evaluated on a time-based split to avoid leaks and with stratified K-fold cross-validation. SHAP values should keep the decisions explainable by showing feature importance.

To keep things fair and unbiased, the following metrics could be used: Equal Opportunity, Disparate Impact Ratios, Calibration within Groups, Feature Importance Auditing. Additionally, the model would be monitored for any drift, have humans in the loop manually reviewing high risk cases and the model would be trained quarterly so it doesn’t become historically distorted.

Protected attributes would be respected in as great capacity as possible, while avoiding proxy discrimination. It would be a task going beyond standard compliance, one that would build trust in the brand, empower the borrower and systems that actually see the people behind the patterns. It would also involve having proper accountability and governance.

Additionally, in Geldium’s case, it would also be a matter of promoting economic inclusion, to enable responsible access to credit by using alternative data sources responsibly and using the model to asses the ability to pay on more than past borrowing behavior alone.