**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 Gen AI)

Below is a **sample code** for building a **credit risk prediction model** using **Logistic Regression**, which is a great interpretable baseline for binary classification tasks like predicting **credit delinquency**.

Key Features and their contribution -

Income Customer’s income level (affects affordability)

Credit Utilization - Ratio of credit used to available credit (risk indicator)

Missed Payment- Past missed payments (behavioral risk)

Debt to Income Ratio - Debt burden relative to income (financial stress)

Account Tenure - Length of account history (credit maturity indicator)

General workflow of the model -

* **Input**: Raw dataset with demographic and credit data.
* **Preprocessing**: Clean missing values, encode categories.
* **Feature Engineering**: Select key features and normalize if needed.
* **Modeling**: Train logistic regression to estimate delinquency probability.
* **Output**: A score between 0 and 1 indicating the probability of being delinquent.

Pseudo code -

* Load and Inspect Dataset
* Handle Missing Values
* Encode Categorical Variables
* Scale Numerical Features
* Split Data into Training and Test Sets
* Select and Initialize Logistic Regression Model
* Train the Model on Training Data
* Evaluate Model Performance on Test Data
* Interpret Model Coefficients
* Check for Bias and Fairness Across Groups
* Generate Predictions on New Data
* Deploy Model for Production Scoring

The above logistic regression model is designed to predict the likelihood that a customer will become credit delinquent—that is, fail to meet their debt repayment obligations. It uses key financial and behavioral features such as income, credit utilization, missed payments, and debt-to-income ratio to estimate a probability score between 0 and 1. This score reflects the customer’s risk level, helping financial institutions like Geldium make informed decisions about loan approvals, risk mitigation, and portfolio monitoring.

2. Justification for Model Choice -

I selected logistic regression for its balanced strength across accuracy, interpretability, and operational efficiency—critical factors for regulated financial environments. It was chosen for predicting credit delinquency at Geldium due to its strong balance between performance, interpretability, and operational ease. While complex models like decision trees or neural networks may offer slight accuracy gains, logistic regression delivers reliable results on structured financial data where relationships are often linear. Its transparency is critical for regulatory compliance (e.g., FCRA, RBI, GDPR), allowing Geldium to justify decisions and explain outcomes to customers. The model is lightweight, easy to deploy, and provides probabilistic outputs that support risk-based decision strategies. Given Geldium’s need for scalable, auditable, and efficient systems, logistic regression offers the ideal foundation for a responsible and adaptable risk prediction solution.

# 3. Evaluation Strategy

To assess how well the model predicts credit delinquency, a combination of the following metrics would be best suitable - To evaluate the model's performance, I rely on a mix of key metrics. **Accuracy** provides a basic overview of correct predictions, though I’m cautious with it in imbalanced datasets. **Precision** helps ensure I don’t mistakenly flag low-risk customers, while **recall** ensures I capture actual delinquents. When there’s a trade-off, I use the **F1 score** to balance both. **AUC-ROC** helps me assess how well the model ranks customers by risk.

If **precision is low**, I know the model may be unfairly rejecting creditworthy individuals. If **recall is low**, I risk missing genuinely high-risk customers. A strong **AUC** (above 0.80) tells me the model is good at separating risk levels. I use these metrics to adjust thresholds depending on business priorities—whether the goal is risk mitigation or customer inclusion.

Ultimately, I would use these insights to adjust thresholds based on business priorities—e.g., minimizing false negatives for high-risk loans or false positives in sensitive segments.

For ensuring fairness and robustness, these can be done for monitoring bias, analyzing model predictions across sensitive attributes (e.g., gender, location, employment type) to detect discrepancies in false positive/negative rates. Using SMOTE or stratified sampling to address class imbalances and prevent the model from learning biased patterns. Integrate tools like disparate impact ratio or equal opportunity difference to quantify any systemic bias. If evidence of bias is found, I would retrain the model with fairness constraints or apply post-processing corrections to adjust outcomes accordingly.

Predicting financial behavior carries ethical responsibilities, especially in contexts where decisions can affect customers’ credit access, financial stability, or livelihood. Key principles include transparency, consent and privacy, supporting human oversight.