

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)

Use a GenAI tool (e.g., ChatGPT, Gemini) to generate the logic or structure of your predictive model.

- You may include pseudo-code, a step-by-step process, or a simplified code snippet.
- Briefly explain what the model is designed to do.

Paste your GenAI-generated output below or describe the logic in your own words:

The model aims to predict the likelihood that a customer will become delinquent based on financial and behavioral data. A Logistic Regression model is selected for its simplicity, interpretability, and effectiveness for binary classification.

Model Pipeline:

1. Data Preparation:

- *Impute missing values using median for 'Income', 'Credit_Score', and 'Loan_Balance'.*
- *Standardize numerical features: 'Income', 'Credit_Score', 'Credit_Utilization', 'Debt_to_Income_Ratio', 'Loan_Balance'.*
- *Encode categorical variables: 'Employment_Status', 'Credit_Card_Type', 'Location', and monthly payment histories.*

2. Feature Selection:

- *Key features: Employment_Status, Credit_Card_Type, Loan_Balance, Missed_Payments, Credit_Utilization.*

3. Modeling:

- *Apply Logistic Regression for initial modeling.*
- *Optionally test Gradient Boosted Trees (XGBoost) as a more complex model.*

4. Output:

- *Predict delinquency probability (0-1) and classify as delinquent (Yes/No) based on threshold.*

2. Justification for Model Choice

Explain why you selected this specific model type (e.g., logistic regression, decision tree, neural network). Consider:

- Accuracy
- Transparency
- Ease of use or implementation
- Relevance for financial prediction
- Suitability for Geldium's business needs

Logistic Regression is chosen because it offers high interpretability, allowing financial institutions like Geldium to explain risk decisions to regulators, auditors, and customers. It balances accuracy with simplicity and supports transparent feature importance analysis. While complex models like Gradient Boosted Trees may provide better performance, their lack of transparency could complicate regulatory compliance and operational deployment. Logistic Regression supports Geldium's business needs for fair, transparent, and actionable credit risk predictions.

3. Evaluation Strategy

Outline how you would evaluate your model's performance. Include:

- Which metrics you would use (e.g., accuracy, precision, recall, F1 score, AUC)
- How you would interpret those metrics
- Any plans to detect or reduce bias in your model
- Ethical considerations in making predictions about customer financial behavior

The model will be evaluated on multiple dimensions:

- *Accuracy: Overall prediction correctness.*
- *Precision: Ability to avoid false positives.*
- *Recall: Ability to correctly identify actual delinquent accounts.*
- *F1 Score: Balances precision and recall.*
- *AUC-ROC: Measures discriminative ability of the model.*

Fairness checks will include subgroup analysis for Employment_Status, Credit_Card_Type, and Location to ensure no protected groups are unfairly impacted. Bias mitigation strategies may include regular monitoring of feature distributions, re-weighting, and fairness-aware modeling if discrepancies are detected. Ethical considerations include ensuring model decisions do not disproportionately affect vulnerable customer segments and maintaining compliance with financial regulations.