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:

[*This predictive model is a supervised binary classification system designed to estimate the probability that a customer will become delinquent. Following data cleaning and missing value imputation, numerical features are standardized and categorical features are encoded. The dataset is split into training and testing sets using stratified sampling to preserve the class distribution of delinquent and non-delinquent customers. A Gradient Boosting model (such as XGBoost or LightGBM) is trained on historical customer records to learn complex relationships between financial behavior and delinquency outcomes. The model outputs a risk probability score for each customer, which is used to rank customers from low to high risk for targeted intervention.*]

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

[*Gradient Boosting was selected because it delivers strong predictive accuracy on structured financial datasets and effectively captures non-linear relationships between customer behavior and financial risk. Compared with simpler models such as logistic regression, it provides superior ranking power for identifying the highest-risk customers, which is critical for early intervention. The model also supports explainability through feature importance and SHAP values, enabling transparency in decision-making. Its robustness to outliers and ability to handle mixed data types make it highly suitable for Geldium's delinquency prediction task and business objectives*]

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

[*Model performance will be assessed using a combination of predictive accuracy metrics, business impact measures, and fairness indicators. Core accuracy metrics will include Accuracy, Precision, Recall, F1 Score, and AUC-ROC. Recall will be prioritized to ensure that most delinquent customers are correctly identified, while Precision will be monitored to limit unnecessary customer interventions. AUC-ROC will be used to evaluate the model's overall discriminative ability.*

Fairness and bias will be evaluated by comparing model outcomes across sensitive customer segments such as Location, Employment_Status, and Age groups. Metrics such as Disparate Impact Ratio, Equal Opportunity Difference (Recall gap), and False Positive Rate Parity will be applied. Any detected bias will be mitigated through feature review, threshold adjustment, or reweighting techniques. Ethical considerations include transparency, data privacy protection, and avoiding systematic discrimination in financial decision-making.]