Customer Delinquency Prediction Model

Prepared by: ***Rehan Khan Intern****, Tata GenAI Data Analytics Program****,*** ***Date:*** *July 2025*

Executive Summary:

This report outlines a predictive model to identify customers likely to become delinquent based on financial and behavioral indicators. Using logistic regression, the model forecasts missed payment risks and supports Geldium's collections team in prioritizing early outreach. The approach emphasizes fairness, accuracy, and explainability—ensuring responsible and compliant AI-driven decision-making in the financial sector.

Visual Workflow Diagram:

Load Data → Preprocessing → Feature Selection → Train-Test Split → Train Model → Evaluate → Interpret Results

1. Model Logic (Generated with GenAI)

Model Used**: **Logistic Regression

***Purpose**:*

The model predicts whether a customer is likely to become delinquent (miss payments), using past financial behavior, credit history, and other personal information.

Steps of the Model Pipeline:

1. Load dataset and identify target column: Delinquent_Account
2. Handle missing values in: Income, Loan_Balance, etc.
3. Encode categorical variables like: Employment_Status, Credit_Card_Type
4. Normalize numerical variables
5. Select top 5 features:
    - Missed_Payments
    - Credit_Utilization
    - Debt_to_Income_Ratio
    - Loan_Balance
    - Credit_Score
6. Split data into training and testing sets
7. Train Logistic Regression model
8. Predict customer risk and evaluate performance

2. Justification for Model Choice

I chose Logistic Regression because:

- It is ideal for binary classification (Yes/No delinquency).
- It gives probability-based outputs, helping prioritize risk levels.
- It is transparent and interpretable, which is crucial in financial services.
- It works well with structured numeric and categorical data.
- It supports Geldium's goals of explainability, fairness, and ethical decision-making, especially for regulatory and business compliance.

More complex models like neural networks may be accurate, but they lack explainability, which is not suitable for finance use cases.

3. Evaluation Strategy

To make sure the model is accurate and fair, we will use:

Evaluation Metrics

- Accuracy – How many predictions were correct
- Precision – Of those predicted delinquent, how many actually were
- Recall – How many delinquent customers were caught
- F1 Score – Balance between Precision and Recall

- AUC-ROC – Ability to distinguish between risky and safe customers

Bias & Fairness Checks

- Use SHAP to explain each prediction
- Check for Demographic Parity and Disparate Impact
- Monitor whether certain groups (e.g., location, employment type) are unfairly flagged

Class Imbalance Handling

- If few customers are delinquent in data, apply Oversampling (e.g., SMOTE) or Class Weights

Ethical Use

- Use model only for supportive interventions (like early outreach), not for rejection or penalties.
- Ensure fairness, avoid discriminatory variables (like ZIP code if it's a proxy for race or income).

Future Scope:

The model can be further improved by incorporating more time-series data (monthly payment behavior), experimenting with ensemble models, or integrating real-time prediction dashboards using Power BI or Tableau.

Business Impact:

With this model, Geldium can identify risky customers early and offer them timely assistance or intervention plans. This helps reduce defaults, improve customer relationships, and meet compliance goals — enhancing both financial and ethical performance.