Predictive Model Plan Report for Customer Delinquency Risk

Geldium Analytics Team

July 29, 2025

Executive Summary

This report outlines a predictive model plan to identify customers at risk of delinquency using logistic regression. The model leverages key features such as Credit Utilization, Missed Payments, Income, Debt-to-Income Ratio, and Account Tenure to predict delinquency likelihood. Logistic regression was chosen for its interpretability and efficiency, aligning with Geldiums need for transparent and responsible AI solutions. The evaluation strategy emphasizes precision, recall, F1 score, and AUC to ensure robust performance, with a focus on fairness through bias audits across demographic segments.

1 Model Logic

The predictive model employs logistic regression to estimate the probability of a customer becoming delinquent, with a binary outcome: 1 (likely delinquent) or 0 (not delinquent). The model uses the following features:

- Credit Utilization: Percentage of available credit used.
- Missed Payments: Number of missed payments in the past 6 months.
- Income: Customers annual income.
- Debt-to-Income Ratio: Ratio of debt to income.
- Account Tenure: Duration of the customers account with Geldium.

Model Workflow

The model follows a structured process:

1. Load the customer dataset.
2. Select features: Credit Utilization, Missed Payments, Income, Debt-to-Income Ratio, Account Tenure.
3. Define the target variable: Delinquent Account (binary).
4. Split data into training (70%) and testing (30%) sets.

5. Fit the logistic regression model on the training data.
6. Predict outcomes on the test set and evaluate using classification metrics.

2 Justification for Model Choice

Logistic regression was selected for its strengths in binary classification tasks, particularly in financial services where interpretability is critical. Key reasons include:

- **Transparency**: Logistic regression provides clear coefficient outputs, enabling stakeholders to understand each features impact on delinquency risk.
- **Simplicity**: The model is computationally efficient, requiring minimal resources and enabling rapid iteration.
- **Baseline Performance**: It delivers robust performance for credit risk analysis, making it suitable for Geldiums goal of identifying at-risk customers responsibly.
- **Stakeholder Communication**: The models interpretable outputs facilitate clear communication with business leaders and compliance teams.

This choice aligns with Geldiums objectives of implementing explainable, data-driven solutions to mitigate delinquency risk while maintaining regulatory compliance.

3 Evaluation Strategy

The models performance will be assessed using the following metrics:

- **Accuracy**: Overall correctness of predictions.
- **Precision**: Ensures minimal false positives, avoiding unnecessary interventions for low-risk customers.
- **Recall**: Ensures most high-risk customers are identified, maximizing intervention coverage.
- **F1 Score**: Balances precision and recall for a comprehensive performance measure.
- **AUC**: Evaluates the models ability to distinguish between delinquent and non-delinquent customers across probability thresholds.

Fairness and Bias Mitigation

To ensure fairness, prediction patterns will be analyzed across demographic segments (e.g., Employment Status, Location). Any disparities will trigger model reassessment or rebalancing. Ethical considerations include:

- Avoiding proxy bias by carefully selecting features that do not inadvertently reflect structural inequities.
- Maintaining transparency by documenting model logic and evaluation results.
- Clearly communicating how model outputs influence business decisions to stakeholders.

4 Conclusion

The proposed logistic regression model provides a transparent and efficient solution for predicting customer delinquency risk. By leveraging key features and a robust evaluation strategy, the model supports Geldiums goals of proactive risk management and responsible AI use. Ongoing bias audits and clear communication ensure fairness and compliance, positioning the model as a valuable tool for data-driven decision-making.