**User Training Manual**

**Predictive Analytics for Loan Default Risk: A SAS Enterprise Miner Approach**

**Gelin Gangoue Mawa**

**May 11, 2025**

# 1. Introduction

This training manual provides an overview of how to navigate and interpret the key elements of the Loan Default Risk prediction project built using SAS Enterprise Miner. The project uses data mining techniques to identify borrowers likely to default on loans, enabling more informed credit risk management.

# 2. Project Objective

The primary goal is to develop predictive models that can accurately identify customers at risk of becoming delinquent on loans. This involves importing and preparing the dataset, building logistic regression and decision tree models, evaluating model performance using statistical and graphical diagnostics, and selecting the most accurate model for deployment.

# 3. Dataset Overview

Source: https://www.kaggle.com/datasets/brycecf/give-me-some-credit-dataset by Bryce Freshcorn  
Target Variable: SeriousDlqin2yrs – Indicates whether an individual experienced 90+ days delinquency in the past two years.  
Key Predictors:  
- RevolvingUtilizationOfUnsecuredLines  
- DebtRatio  
- age  
- NumberOfTimes90DaysLate, etc.

# 4. SAS Enterprise Miner Process Overview

A complete Process Flow Diagram is included as a visual guide. The data was split into training and validation sets using the Data Partition node. Models built: Logistic Regression (baseline), and Decision Tree (interpretable model).

# 5. Logistic Regression Outputs

Fit Statistics Table: Shows ASE, AIC, MISC rates for train/validate/test sets.  
Iteration Plot: Illustrates convergence based on Average Squared Error across iterations.  
Cumulative Lift Chart: Measures model performance in identifying defaults among top-scored customers.  
Relevant Images:  
- Logistic\_Regression\_Fit\_Statistics.png  
- Logistic\_Regression\_Iteration\_Plot.png  
- Logistic\_Regression\_Cumulative\_Lift.png

# 6. Decision Tree Outputs

Tree Diagram: Shows rule-based splits based on variable importance.  
Fit Statistics Table: ASE and MISC for training and validation data.  
Variable Importance Table: Indicates most predictive features.  
Leaf Statistics & Subtree Plot: Help optimize the tree depth and complexity.  
Cumulative Lift: Evaluates ranking quality of predictions.  
Relevant Images:  
- Decision\_Tree\_Diagram\_Results.png  
- Decision\_Tree\_Fit\_Statistics.png  
- Variable\_Importance\_Table.png  
- Decision\_Tree\_Leaf\_Statistics.png  
- Decision\_Tree\_Subtree\_Assessment\_Plot.png  
- Decision\_Tree\_Cumulative\_Lift.png

# 7. Interpreting the Results

Logistic Regression produced a lower ASE on validation data, suggesting slightly better generalization.  
Decision Tree provided interpretable rules (e.g., NumberOfTimes90DaysLate > threshold) useful for operational decisions.  
Both models showed strong lift in early deciles, validating their ability to prioritize high-risk individuals.

# 8. Key Takeaways

SAS Enterprise Miner enabled a systematic modeling workflow with visual diagnostics.  
Variable importance and model diagnostics supported feature relevance and model validity.  
Decision Trees offered business transparency, while logistic regression provided statistical robustness.

# 9. How to Use This Project

View the README.md file for a summary and links to download the report, diagrams, and documentation.  
Use the PDF and Word reports to understand the modeling rationale, techniques, and findings.  
Review screenshots to interpret model outputs without needing to run SAS.

# 10. Additional Files Included

- Predictive\_Analytics\_for\_Loan\_Default\_Risk.docx – Full project report with citations.  
- Process\_Flow\_Diagram.png – Overview of SAS Enterprise Miner steps.  
- Multiple .png files showing model outputs (as listed above).