

Credit Risk Early Warning – Backtesting Report

1. Objective

This report evaluates the performance of a credit risk early warning model designed to predict future default events using historical credit card customer behavior. The model outputs a Probability of Default (PD) and maps customers into operational risk tiers (Green / Amber / Red) for early intervention.

2. Dataset & Experimental Setup

Data Source: UCI Machine Learning Repository – Default of Credit Card Clients (Taiwan). The dataset contains 30,000 anonymized credit card customers. An 80/20 train-test split was used, with the test set serving as an out-of-time proxy.

3. Models Evaluated

Baseline models include Logistic Regression and Decision Tree. Neural models include a PyTorch MLP and a Transformer encoder that models the last six months of customer credit behavior as a temporal sequence. The Transformer model demonstrated the strongest ranking performance.

4. Evaluation Metrics

Model performance was evaluated using ROC-AUC, KS statistic, PR-AUC, and Brier score, which are standard metrics in credit risk modeling.

5. Backtesting Results (Hold-out Test Set)

Metric	Value
ROC-AUC	≈ 0.77
KS Statistic	≈ 0.42
PR-AUC	≈ 0.45
Brier Score	≈ 0.16

6. Risk Tier Analysis

Predicted PDs were converted into Green, Amber, and Red risk tiers. Observed default rates increase monotonically across tiers, indicating effective risk stratification and suitability for operational deployment.

7. Stability & Monitoring

Population Stability Index (PSI) was computed for key features between training and test samples. All monitored features exhibited PSI values below common alert thresholds, suggesting no significant population drift during the backtesting period.

8. Conclusion

The backtesting results demonstrate that the proposed early warning framework produces well-calibrated PD estimates, effectively separates customers by risk, and meets standard requirements for deployment in credit risk management systems.