Portfolio Management and Risk Optimization Weekly Report (D64–D68)

Machine Learning Applications in Credit Risk, Unsupervised Segmentation, and Systemic Risk Propagation

Executive Summary

This report provides a consolidated overview of analytical work conducted between Days 64 and 68, covering Machine Learning–driven credit risk modeling, unsupervised risk segmentation, and systemic risk propagation analysis.

Key achievements for the week include:

- Establishing ML frameworks for risk identification and prediction.

- Building and benchmarking credit models using Logistic Regression, Random Forest, and XGBoost.

- Extracting meaningful borrower segments via K-Means clustering.

- Constructing a conceptual Bayesian Network to analyze systemic contagion across interconnected markets.

Logistic Regression emerged as the strongest and most interpretable credit risk model. Unsupervised segmentation reinforced supervised insights, and systemic simulation illustrated how equity shocks propagate to FX, credit spreads, and bond markets. The progression from micro- to macro-level modeling supports comprehensive portfolio risk optimization.

Introduction

The objective of this weekly report is to summarize the progression of tasks that span Machine Learning applications in risk management. The work focuses on three pillars:

1. **Micro-Level Risk:** Predicting individual borrower creditworthiness.

2. **Unsupervised Discovery:** Identifying hidden patterns and risk clusters.

3. **Macro-Level Systemic Risk:** Modeling contagion across financial markets.

This holistic analytical framework enhances both portfolio stability monitoring and stress-testing capabilities.

Day-by-Day Analysis

Day 64 — Foundations of Machine Learning for Risk Management

Day 64 centered on establishing the theoretical foundation for Machine Learning in risk management and outlining its primary applications.

Key ML Tasks in Risk Management

- **Risk Classification:** Predicting default probabilities and assigning risk levels.

- **Risk Clustering:** Segmenting customers by financial behavior and risk characteristics.

- **Systemic Risk Propagation:** Modeling contagion pathways across networks of financial entities.

Core Challenges Identified

- Data quality issues such as bias and missingness.

- Interpretability requirements for regulatory compliance.

- Dynamic environments where risk profiles and patterns shift over time.

Day 65 — Supervised Credit Risk Modeling (Baseline Models)

Work on Day 65 focused on supervised credit risk modeling using the **German Credit dataset** (1,000 applicants), which exhibited a clear **class imbalance**:

- **Good credit:** 70%

- **Bad credit:** 30%

Model Performance

Logistic Regression

- Accuracy: 0.7833

- ROC-AUC: **0.8011**

- Strength: High interpretability

Random Forest

- Accuracy: 0.7667

- ROC-AUC: 0.7932

- Strength: Captures non-linearities

Key Observations

- Logistic Regression achieved the highest AUC and remains the most regulatory-friendly.

- Both models struggled with recall for "bad credit" cases (LR: 0.53, RF: 0.40).

- Important predictors included savings/checking account status, loan purpose (education), and credit duration.

Day 66 — Model Enhancement with XGBoost

Day 66 introduced a more advanced, ensemble-based model: **XGBoost**.

Model Setup

- Learning rate: 0.1

- Max depth: 7

Performance Results

- Accuracy: 0.7233

- ROC-AUC: 0.7558

Comparative Summary

Model	AUC	Interpretability	Complexity
Logistic Regression	**0.8011**	High	Low
Random Forest	0.7932	Moderate	Medium
XGBoost	0.7558	Low	High

Key Takeaways

- Simpler models outperformed XGBoost due to the dataset's small size and structure.

- XGBoost did not yield gains without deeper hyperparameter tuning.

- Logistic Regression remains the preferred model for transparent, regulated scoring systems.

Day 67 — Unsupervised Segmentation & Anomaly Detection

Day 67 focused on uncovering hidden behavioral patterns through unsupervised methods.

K-Means Clustering (Risk Segmentation)

- Optimal clusters: **2**

- Silhouette Score: **0.309** (moderate separation)

Cluster 0 — High-Risk (41.67% default rate)

- Long credit durations

- High loan amounts

- Unstable employment

- Limited property ownership

Cluster 1 — Moderate-Risk (29.86% default rate)

- Smaller loan amounts

- Shorter durations

- More stable borrower characteristics

DBSCAN (Anomaly Detection)

- All 1000 applicants labeled anomalies due to restrictive parameters.

- Indicates dataset heterogeneity and need for parameter tuning.

Day 68 — Systemic Risk Propagation via Bayesian Networks

Day 68 expanded the scope from individual borrower risk to macro-financial contagion analysis.

Bayesian Network Nodes

- Equity Index

- FX Rate

- Bond Yields

- Credit Spreads

Causal Flow:
Equity → FX → Bond Yields → Credit Spreads

Equity Shock Scenario (~–20%)

Node	Adverse State	Prior	Posterior	Δ Change	Interpretation
FX	Depreciate	23%	**85%**	+62%	Strong capital outflow response
Credit Spreads	Widen	8.45%	**45.05%**	+36.6%	Credit conditions deteriorate sharply
Bond Yields	Rise	14.68%	**5.15%**	–9.53%	Flight to safety lowers yields

Insights

- FX and credit spreads serve as early, sensitive indicators of systemic stress.

- Bond market reaction shows classic risk-off behavior.

- The model demonstrates how shocks cascade across markets.

Key Insights of the Week

1. **Simple models outperform complex ones** on small structured datasets.

2. **Interpretability is vital** for regulatory credit risk environments.

3. **Unsupervised segmentation confirmed high-risk borrower clusters.**

4. **Systemic modeling reveals contagion pathways** crucial for macro stress testing.

5. **Integrated micro-to-macro analytical frameworks** provide stronger risk visibility.

Conclusion

The analytical work from Days 64–68 successfully integrated supervised, unsupervised, and systemic risk modeling approaches. Logistic Regression provided the best balance of accuracy and interpretability for credit risk modeling. Clustering revealed meaningful risk segments, while Bayesian Network simulations captured broader market contagion effects.

Next steps include:

- Addressing class imbalance using SMOTE or cost-sensitive learning,

- Conducting deeper ensemble model tuning,

- Applying SHAP for advanced explainability,

- Expanding systemic stress-testing with additional macroeconomic variables.

These enhancements will strengthen both predictive performance and regulatory compliance, supporting a robust portfolio risk management strategy.