

AI IN ASSET MANAGEMENT

Unsupervised Learning I: Overview of Techniques

Joseph Simonian, PhD

Practitioner Brief written by Mark Fortune



Unsupervised learning is a branch of machine learning (ML) that enables financial analysts and investors to discover hidden patterns in data even absent labeled examples. Traditional financial models depend on labeled data. But in today's high-frequency, high-dimensional markets, such labels are often incomplete, outdated, or unavailable. Unsupervised learning addresses this challenge by revealing patterns, relationships, and anomalies without the need for predefined outcomes.

This chapter of *AI in Asset Management: Tools, Applications, and Frontiers* provides financial practitioners with a clear, actionable overview of unsupervised learning techniques—such as clustering, dimensionality reduction, and anomaly detection—and how they uncover hidden patterns in financial data without requiring labeled outcomes. It translates complex ML methods into practical tools for portfolio construction, regime detection, risk monitoring, and strategy refinement.

Who Should Read This Chapter?

This chapter is written for portfolio managers, quantitative analysts, risk professionals, investment strategists, and data-driven investors seeking to enhance their decision making with innovative tools from unsupervised

ML techniques to identify patterns, enhance decision making, and stay ahead in complex, data-rich markets.

Why This Chapter Matters Now

Markets are more volatile, data are more complex, and traditional models struggle to keep up. This chapter discusses how unsupervised learning helps investors find patterns and adapt—without relying on outdated assumptions or labeled data. The chapter equips investors with adaptive tools that evolve with market complexity, enabling more responsive and robust strategies.

What Does This Chapter Deliver?

This chapter is a practical guide to using unsupervised learning—such as clustering, anomaly detection, and dimensionality reduction—to uncover hidden insights, improve portfolio strategies, and enhance decision making in modern financial markets.

It demonstrates how unsupervised learning improves decision making in dynamic and uncertain markets, translates advanced machine learning methods into practical investment applications, and provides actionable techniques that investment professionals can implement immediately—no technical background required.

“Unsupervised learning methods can be invaluable in helping practitioners discover structures in financial data that may prove valuable in their portfolio and risk management effort.”

Joseph Simonian, PhD

Practical Applications

Here are six major financial use cases for unsupervised learning, including portfolio construction, anomaly detection, and regime classification.



Portfolio construction

Use clustering techniques to group assets by shared behaviors, improve diversification, reduce hidden correlation risk, and build more robust portfolios.

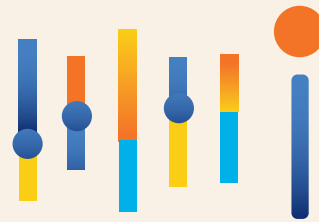
Example: Group US and international equities based on volatility and momentum trends rather than sector or geography, revealing diversification opportunities not visible through traditional lenses.



Market regime detection

Identify market regimes from macro signals to inform strategy timing and risk positioning. Detect structural or macroeconomic shifts earlier by identifying changing patterns in market and economic indicators.

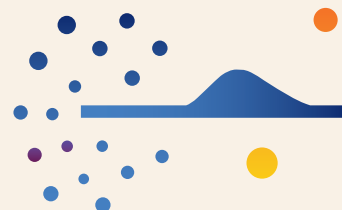
Example: Spot a shift from a growth-led to an inflation-driven regime by clustering macro data—before the shift is reflected in central bank announcements or price action.



Signal classification

Cluster trading signals by predictive performance to enhance model accuracy and reduce redundancy.

Example: Classify hundreds of alpha signals by their historical Sharpe ratios and stability, and then select only the most robust and regime-consistent signals for strategy input.



Noise reduction and factor discovery

Apply dimensionality reduction tools, such as principal component analysis (PCA) and t-SNE (t-distributed stochastic neighbor embedding), to extract essential signals from noisy datasets, uncover key drivers of asset returns, and reduce data complexity.

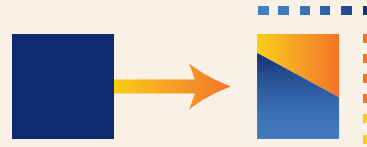
Example: Use PCA to compress 200+ macro indicators into three components explaining 90% of portfolio risk exposure, making model tuning and reporting far more interpretable.



Anomaly detection

Spot unusual transactions or risk exposures early using advanced outlier detection methods. Use such algorithms as isolation forest and local outlier factor (LOF) to detect outliers, fraud, or extreme market behavior.

Example: Detect an unusually large volume spike in a low-liquidity bond using LOF—flagging it for manual review before it skews the portfolio's daily P&L attribution.



Synthetic data generation

Use generative AI models to create realistic financial scenarios for backtesting and stress testing and to validate strategies under varied conditions.

Example: Use a variational autoencoder (VAE) to generate hypothetical trading days simulating liquidity crunches, and then assess how existing strategies would behave under these synthetic stress conditions.

Practitioner Toolkit

The following provides a guide for how practitioners in key financial roles can apply unsupervised learning techniques.

Applications of Unsupervised Learning by Role

Role	Key Techniques	Primary Applications	Main Benefits
Portfolio manager	Clustering, PCA, t-SNE, regime detection	Asset grouping, diversification, regime-aware allocation	Build adaptive, diversified portfolios beyond traditional classifications
Quantitative analyst/strategist	Signal clustering, autoencoders, VAEs, synthetic data	Signal evaluation, data compression, scenario generation	Enhance research quality and test strategies in rare conditions
Risk manager	Isolation forest, LOF, PCA, clustering	Anomaly detection, risk factor simplification, concentration risk analysis	Improve early warning systems and stress-testing flexibility
Compliance/surveillance	LOF, DBSCAN, clustering	Fraud detection, behavior clustering, anomaly scoring	Detect irregular behavior in real time with fewer false positives
Data scientist/AI team member	PCA, independent component analysis (ICA), hybrid modeling, cluster evaluation	Model preprocessing, latent feature engineering, cluster validation	Bridge theory and production with ML-aligned investment workflows

Implementation

Unsupervised learning techniques integrate seamlessly into existing workflows. Clustering enhances asset grouping. Anomaly detection reinforces risk monitoring. Dimensionality reduction simplifies complex datasets. Practitioners can adopt these tools incrementally to enhance their analytical toolkit.

Metrics That Matter

From the many measures available, this chapter highlights three metrics as especially critical for practitioners:

- **Silhouette score:** This helps assess the *internal quality* of a clustering solution—essential when you do not have ground truth labels. It allows you to validate how well assets, signals, or macro indicators are grouped.
- **Adjusted Rand Index:** This allows comparison between two clustering outcomes—helpful for testing algorithm variations or stability over time. It allows you to track the consistency of clustering over changing market conditions or across different modeling choices.
- **Explained variance (using PCA):** This provides insight into how much of the underlying data structure is captured by each component—essential for dimensionality reduction. It allows you to identify major drivers of yield curve movement or compress large sets of macro variables.

Glossary

Anomaly detection: Identifies rare or abnormal observations; critical for risk and fraud detection.

Adjusted Rand Index (ARI): Measures the similarity between two clustering results, corrected for chance. Scores range from -1 to 1 (1 = identical groupings).

Clustering: Groups similar data points; useful for asset grouping and regime detection.

Dimensionality reduction: Simplifies complex data by isolating the most relevant features.

Principal component analysis: Transforms data into uncorrelated components explaining variance.

Silhouette score: Measures clustering quality (how well separated and tight clusters are).

Related Content

Spyrou, Alex, and Brian Pisaneschi. 2024. "Practical Guide for LLMs in the Financial Industry." CFA Institute Research and Policy Center. <https://rpc.cfainstitute.org/research/the-automation-ahead-content-series/practical-guide-for-llms-in-the-financial-industry>.

Tait, James. 2025. "Synthetic Data in Investment Management." CFA Institute Research and Policy Center. <https://rpc.cfainstitute.org/research/reports/2025/synthetic-data-in-investment-management>.

Wilson, Cheryll-Ann. 2025. "Creating Value from Big Data in the Investment Management Process: A Workflow Analysis." CFA Institute Research and Policy Center. <https://rpc.cfainstitute.org/research/reports/2025/creating-value-from-big-data-in-the-investment-management-process>.

Wilson, Cheryll-Ann. 2025. "Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders." CFA Institute Research and Policy Center. <https://rpc.cfainstitute.org/research/reports/2025/explainable-ai-in-finance>.



Find *AI in Asset Management: Tools, Applications, and Frontiers* and more on the CFA Institute Research and Policy Center website: <https://rpc.cfainstitute.org>.

Read the full chapter