An Introduction to Artificial Intelligence (AI) in Finance

Chapter 4:
Unsupervised Learning

Universität
Münster

4. Unsupervised Learning

1. Basic Definitions
2. Clustering Techniques
3. Dimensionality Reduction

4.1 Basic Definitions

Unsupervised Learning: Class of machine learning techniques that do not require labeled input data. Unsupervised learning is used to identify patterns, groupings, or structures within the data.

(e.g., customer segmentation, market basket analysis, anomaly detection in credit card transactions)

No labels required

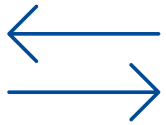
Exploratory nature

Flexibility

4.2 Clustering Techniques
Example Applications in Finance

Customer segmentation
(e.g., group banking customers based on spending behavior, product usage, and financial goals)

Portfolio Diversification Analysis
(e.g., clustering assets based on historical performance, volatility, and correlation)

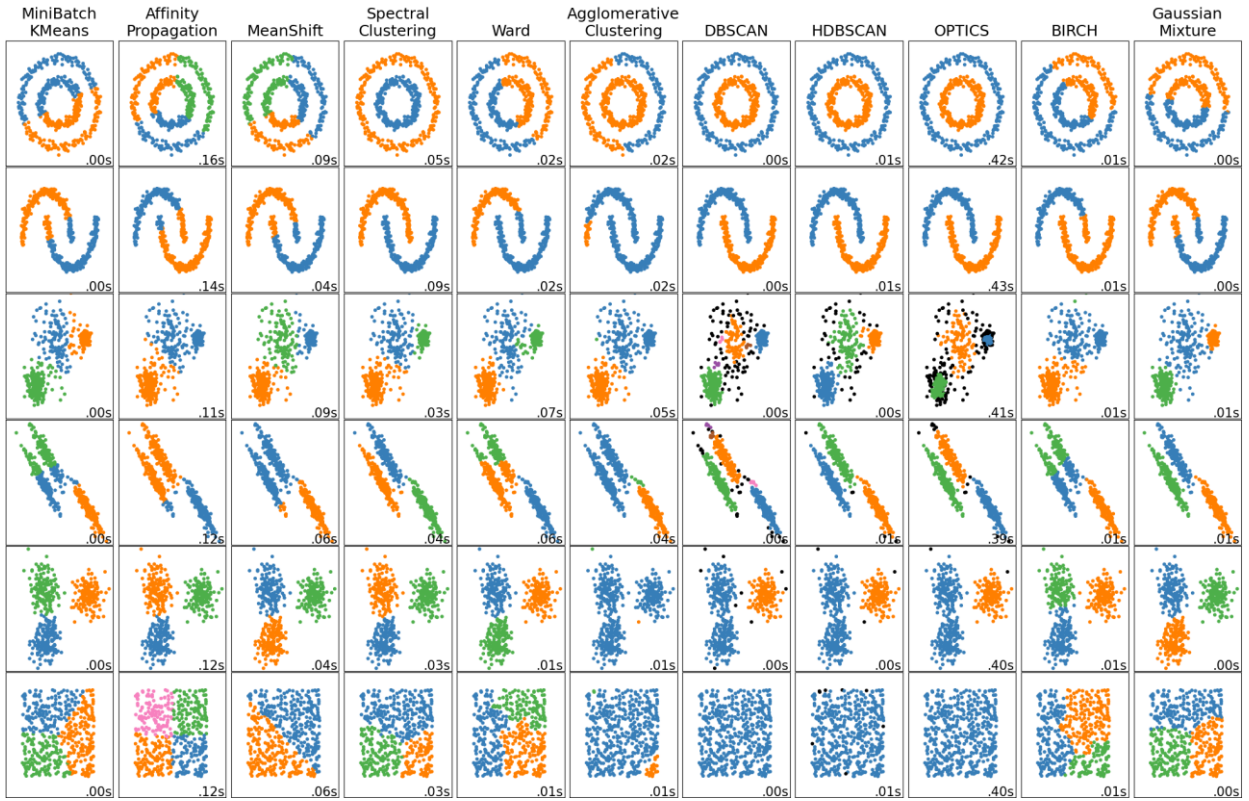


Transaction Anomaly Detection
(e.g., identifying unusual patterns or trading activity)

4.2 Clustering Techniques

Algorithms

Clustering: Algorithms for grouping a set of objects in such a way that objects in the same group (or cluster) are more similar than those in other groups.



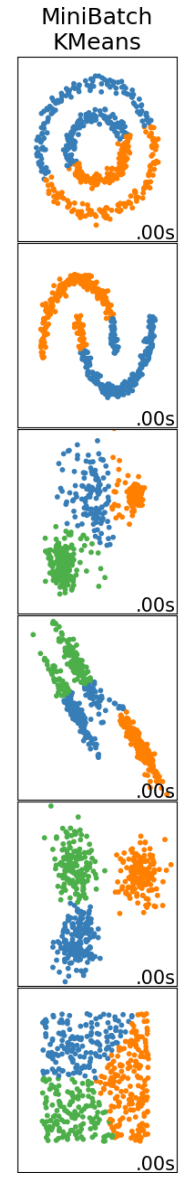
https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html

4.2 Clustering Techniques
Algorithms

K-Means Clustering: Algorithm that partitions the dataset into K clusters by minimizing the variance within each cluster.

1. Choose *K* initial centroids.
2. Assign each data point to the nearest centroid.
3. Recalculate centroids based on the mean of assigned points.
4. Repeat until convergence.

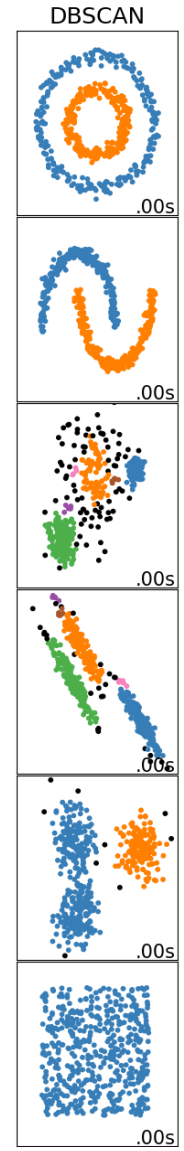
- ⊕ Simple to understand & implement
- ⊖ Requires pre-specification of *K*
- ⊖ Sensitive to outliers



4.2 Clustering Techniques
Algorithms

Density-Based Spatial Clustering of Applications with Noise (DBSCAN): Algorithm that partitions the dataset into clusters by identifying areas of high density separated by areas of low density.

- A point is considered a core point if it has at least a minimum number of points within a specified radius ε.
- Points in the neighborhood of core points are considered part of the same cluster.

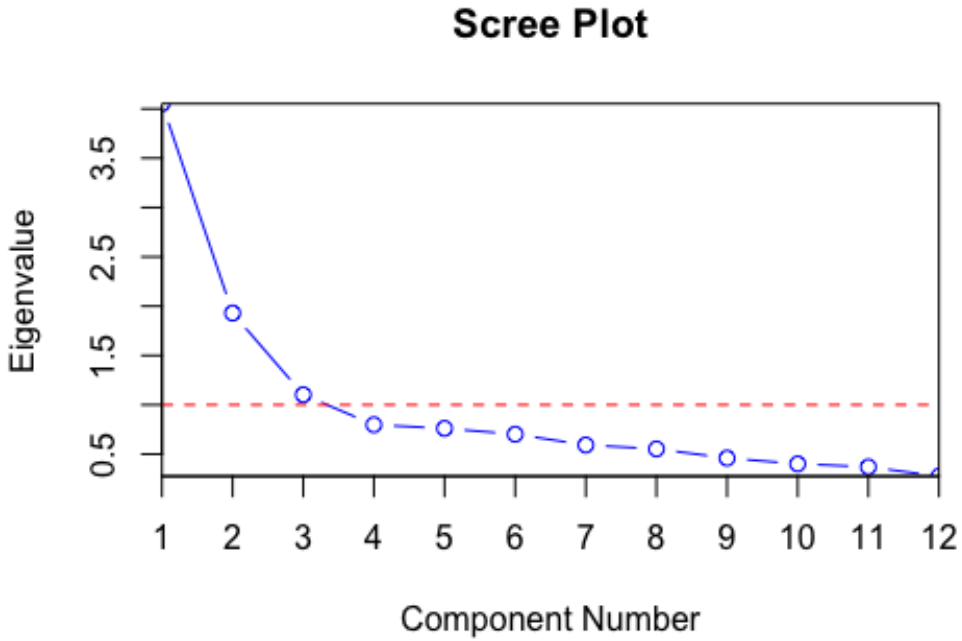


- ⊕ Able to identify arbitrarily shaped clusters
- ⊕ Robust to outliers
- ⊖ Sensitive to parameter choice

4.3 Dimensionality Reduction

Dimensionality Reduction: The process of reducing the number of random variables under consideration while keeping relevant information.

Principal Component Analysis (PCA):
Transforming original variables into a new set of uncorrelated variables that each capture the maximum variance possible.



https://commons.wikimedia.org/w/index.php?curid=75715167

E4 – Interpreting Clustering Results

The provided plot visualizes the outcome of a K-means clustering analysis on customer financial data, revealing four distinct clusters based on two key dimensions: spending and savings.

(1) Which cluster might represent customers with high (low) credit risk?

(2) Argue which customers you would rather offer a coupon to save 10% at their next purchase.

(3) You realize that customers with low savings are the majority. What would you recommend to improve your business outlook?