

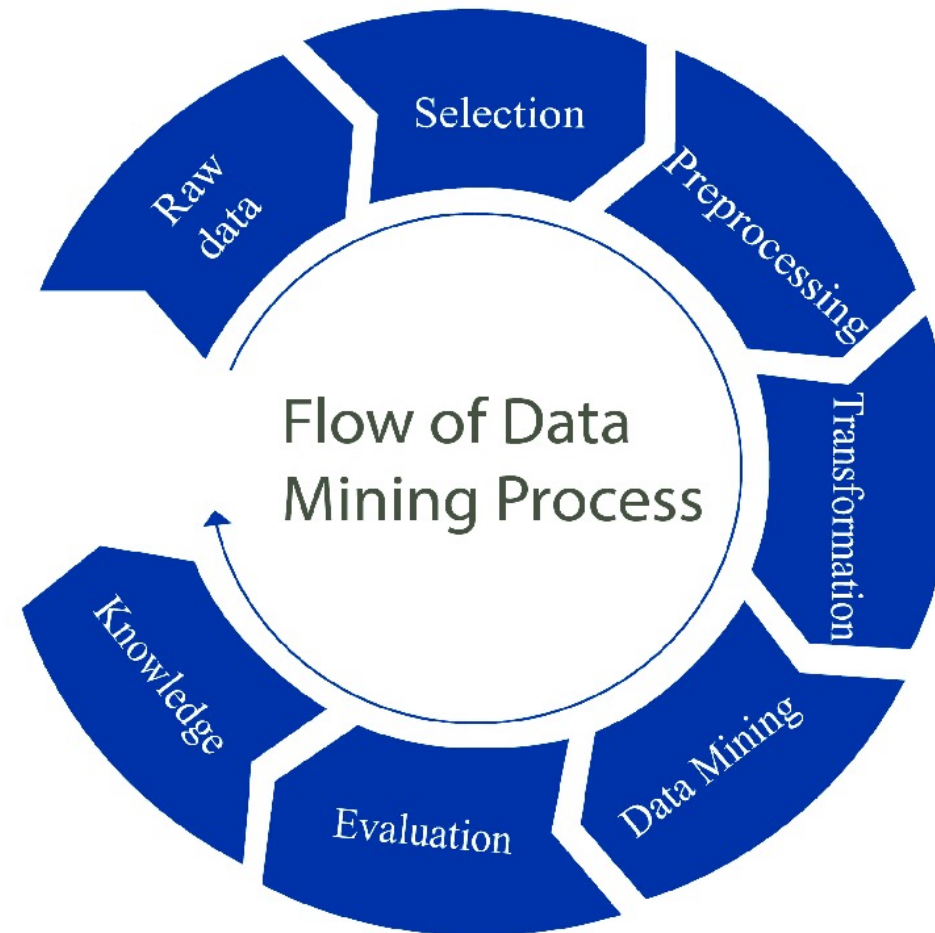
Data Mining Tasks

Data Mining 101 Introduction

Tim Pengajar:

- **Ratih Ardiati Ningrum, S.Si., M.S., M.Stat.**
- **Indah Fahmiah, S.Si., M.Stat.**





Altaf, Saud & Asad, Rimsha & Ahmad, Shafiq & Ahmed D.Sc. PhD, Iftikhar & Abdollahian, Mali & Zaindin, Mazen. (2023). A Hybrid Framework of Deep Learning Techniques to Predict Online Performance of Learners during COVID-19 Pandemic. Sustainability. 15. 11731. 10.3390/su151511731.



Data mining tasks

- Information integration and data warehouse construction
 - Data cleaning, transformation, integration, and multidimensional data model
- Data cube technology
 - Scalable methods for computing (i.e., materializing) multidimensional aggregates
 - OLAP (online analytical processing)
- Multidimensional concept description: Characterization and discrimination
 - Generalize, summarize, and contrast data characteristics, e.g., dry vs. wet region

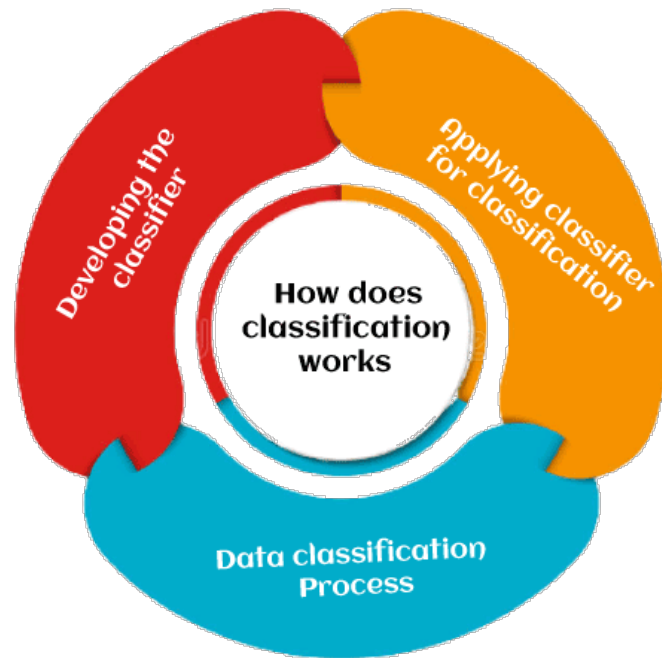


Data mining tasks: Association and Correlation Analysis

- Frequent patterns (or frequent item sets)
 - What items are frequently purchased together in your Walmart?
- Association, correlation vs. causality
 - A typical association rule
 - Diaper \rightarrow Beer [0.5%, 75%] (support, confidence)
 - Are strongly associated items also strongly correlated?
- How to mine such patterns and rules efficiently in large datasets?
- How to use such patterns for classification, clustering, and other applications?



Data mining tasks: Classification



- Classification and label prediction
 - ✓ Construct models (functions) based on some training examples
 - ✓ Describe and distinguish classes or concepts for future prediction. E.g., classify countries based on (climate), or classify cars based on (gas mileage)
 - ✓ Predict some unknown class labels
- Typical methods
 - ✓ Decision trees, naïve Bayesian classification, support vector machines, neural networks, rule-based classification, pattern-based classification, logistic regression, etc.
- Typical applications:
 - ✓ Credit card fraud detection, direct marketing, classifying stars, diseases, web-pages, etc.

Source: <https://www.javatpoint.com/classification-and-predication-in-data-mining>



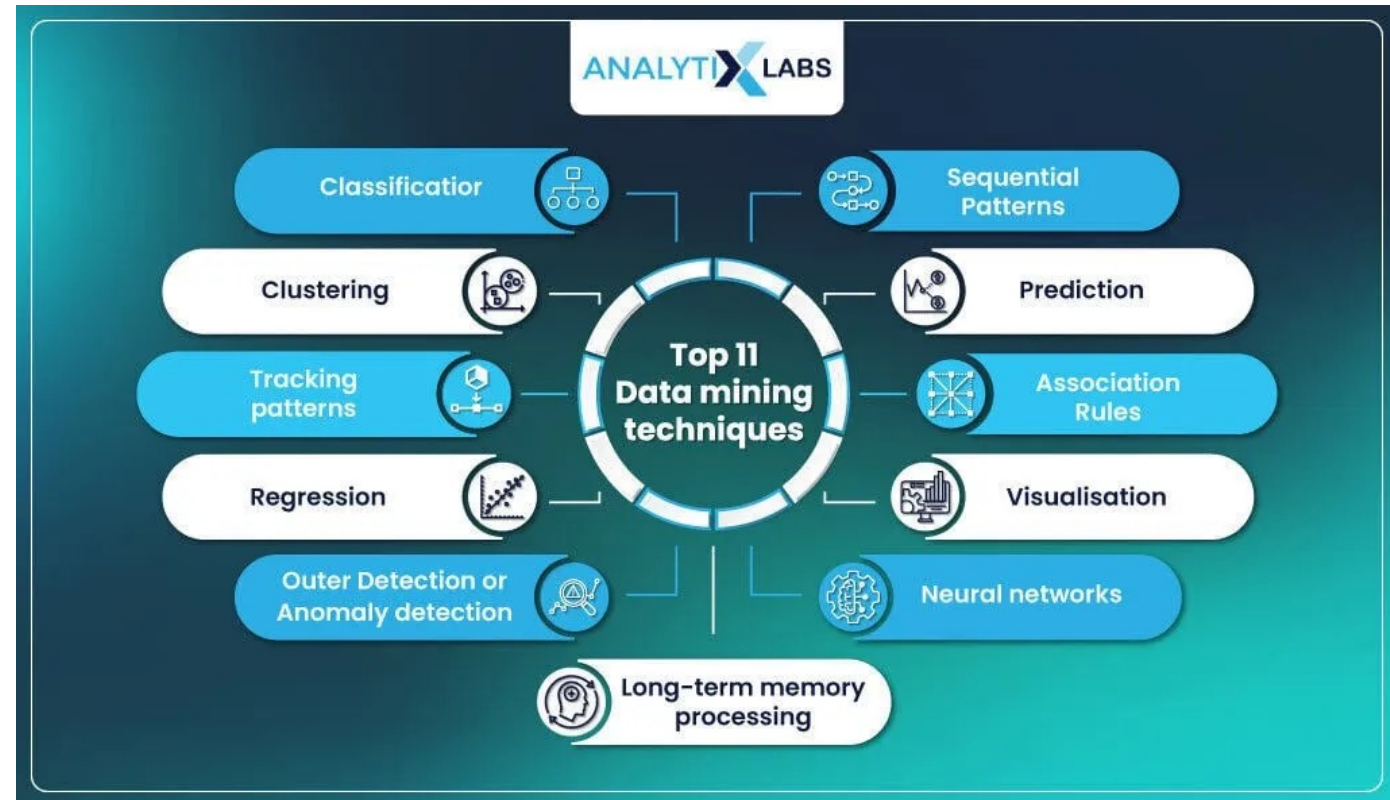
Data mining tasks: Cluster Analysis

- Unsupervised learning (i.e., Class label is unknown)
- Group data to form new categories (i.e., clusters), e.g., cluster houses to find distribution patterns
- Principle: Maximizing intra-class similarity & minimizing interclass similarity
- Many methods and applications



Data Mining Methods

1. Decision Tree Classifiers:
Used for modeling, classification
2. Association Rules:
Used to find associations between sets of attributes
3. Sequential patterns:
Used to find temporal associations in time series
4. Hierarchical clustering:
used to group customers, web users, etc



Source: <https://medium.com/>



Terima Kasih!

