Data Mining

Content

- Objective
- Process
  - Gathering data
  - Pre-process
  - Data mining
    - Anomaly detection
    - Association
    - Clustering
    - Classification
    - Regression
    - Summarization
  - Modeling
  - Evaluate
- Reference

Objective

- To extract the patterns from a massive data or a large data set.
- Then create a model out of the extraction of the patterns.

Process

- Gathering data
- Pre-process
- Data mining
- Modeling
- Evaluate

Gathering data

- Collect raw data directly from the real world.
- Or just investigate from the Internet.

Pre-processing

- Organizes data and filters redundant data.
- Transforms raw data to understandable format for computers.

Data mining

- Anomaly detection
- Association
- Clustering
- Classification
- Regression
- Summarization

Anomaly detection

- Detect the unusual data from date sets.
- Includes outliers, novelties, noise, deviations and exceptions.

Association

- Find the relationships or dependencies between variables.
- Usually used to analyze the relationship or frequencies of customers' purchasing habits.

Clustering

- Group the similar objects together.
- Every members within a subset has some similar attributes.
- K-means.

Classification

- To identify which category a new instance or instances belongs.
- Major difference from cluster is that classification uses supervised learning technique.

Regression

- To find a function that has a least error from the data set.
- Which can estimate a new data from the future.

Summarization

- To extract the most important or relevant information from data sets.

Modeling

- After applying any method from data mining you get model.

Evaluation

- After exporting the model test it with some test data sets.
- With the test result you can evaluate if the model meets the requirements.

Reference

- https://www.inside.com.tw/article/6755-what-is-data-mining
- https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa
- https://en.wikipedia.org/wiki/Data_mining
- https://techdifferences.com/difference-between-classification-and-clustering.html