

Data mining → Disiplin Ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data besar.

testing → Data yang diambil secara umum yang menghasilkan rumus

training → rumus yang didapatkan dari testing digunakan untuk mengolah data training.

Ciri-Ciri Data Mining

Algoritma Data Mining

① estimasi / Perkiraan →
• label
• numerik } hasil → rumus.

② Forecasting / Prediction →
• label
• time series
↳ nominal
• numerik
↳ main time series

③ Klasifikasi →
• label
• numerik
• nominal } hasil pda/tree

④ Klustering →
• Numerik
• Tanpa LABEL

⑤ Asosiasi →
• Boolean
• Tanpa LABEL

↳ nilai support
↳ $\frac{n}{NA}$ → total yang di data
NA → total semua.

• nilai Confident
↳ $\frac{NAA}{NA}$ → dari total di data mencari data.
NA → total yang di data.

• Split Data
↳ untuk menjadikan Data training dan testing

① estimasi
↳ Linear regression, neural network, Support Vector machine, etc.

② Forecasting
↳ Linear regression, neural network, Support vector machine, etc.

③ Klasifikasi
↳ Naive bayes, K-nearest neighbor, C.45, ID3, CART, Linear Discriminant analysis, logistic regression, etc.

④ Klustering
↳ K-means, K-medoids, self-organizing map (Som), Fuzzy C-means, etc.

⑤ Asosiasi
↳ FP-growth, A Priori, Coefficient of correlation, Chi square, etc.

* metode learning Algoritma Data mining

- ① Supervised learning → estimasi, prediksi, klasifikasi
- ② Unsupervised learning → clustering

Proses Data mining

① Himpunan data

② metode Data Mining

③ Pengetahuan

- * Formula
- * decision tree
- * korelasi
- * Rule
- * cluster

④ Evaluasi (Akurasi, Error, etc)

① estimasi

- RMSE, MSE, MAPE

② Prediction

- RMSE, MSE, MAPE

③ classification

- confusion matrix & accuracy
- ROC curve - Area under curve (AUC) →

0.90 - 1.00 → excellent

0.80 - 0.90 → good

0.70 - 80 → fair

0.60 - 70 → poor

0.50 - 60 → Failure

④ clustering

- internal = DBI, Dunn Index
- external = Confusion matrix, Rand measure, F-measure, Jaccard Index

⑤ Association:

- lift chart - lift ratio
- precision and recall (F-measure)

Data mining

created by : Fauzan.

Data mining → Disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan menemukan pola dari suatu data besar.

Pengolahan

Data \Rightarrow Informasi \Rightarrow Pengetahuan. \Rightarrow Kebijakan

Data → Representasi dari fakta

Informasi → hasil Pengolahan sebuah model, format, organisasi atau suatu perubahan bentuk dari data yang memiliki nilai tertentu.

Pengetahuan → Informasi yang sudah diorganisasikan dan diproses untuk memperoleh pemahaman, pengalaman, pembelajaran yang terakumulasi, sehingga dapat diaplikasikan dalam masalah atau proses bisnis tertentu.

Proses Data Mining

