

Nama : Wahyu Adi Nugroho

NIM : A11.2019.12310

Kelp. : 46UG

TUGAS DATA MINING
PERTEMUAN MINGGU KE – 4

1. Sebutkan 5 peran utama data mining!
2. Algoritma apa saja yang dapat digunakan untuk 5 peran utama data mining di atas?
3. Jelaskan perbedaan estimasi dan prediksi!
4. Jelaskan perbedaan estimasi dan klasifikasi!
5. Jelaskan perbedaan klasifikasi dan klastering!
6. Jelaskan perbedaan klastering dan prediksi!
7. Jelaskan perbedaan supervised dan unsupervised learning!
8. Sebutkan tahapan utama proses data mining!

Jawaban :

1. Estimasi / Perkiraan (Estimation), Prediksi / Peramalan (Prediction / Forecasting), Klasifikasi (Classification), Klastering (Clustering), dan Asosiasi (Association).
2. Algoritma data mining yang digunakan :
 - a. Estimasi / Perkiraan (Estimation) : Linear Regression (LR), Neural Network (NN), Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM), dll.
 - b. Prediksi / Peramalan (Prediction / Forecasting) : Linear Regression (LR), Neural Network (NN), Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM), dll.
 - c. Klasifikasi (Classification) : Decision Tree CART, ID3, C4.5, Credal DT, Credal C4.5, Adaptive Credal C4.5, Naive Bayes (NB), K Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Logistic Regression (LogR), dll.
 - d. Klastering (Clustering) : K Means, K Medoids, Self Organizing Map (SOM), Fuzzy C Means (FCM), dll.
 - e. Asosiasi (Association) : FP Growth, A Priori, Coefficient of Correlation, Chi Square, dll.

Nama : Wahyu Adi Nugroho

NIM : A11.2019.12310

Kelp. : 46UG

3. Perbedaan Estimasi dan Prediksi :

- a. Estimasi nilai dari variable target ditentukan berdasarkan nilai dari variabel prediktor (atribut). Estimasi biasanya digunakan untuk menerka / memperkirakan sebuah nilai yang belum diketahui.
- b. Sebetulnya prediksi sama dengan estimasi di mana label/target/class bertipe numerik. Namun, perbedaannya adalah data yang digunakan merupakan data rentet waktu (data time series). Prediksi ini biasanya digunakan untuk meramal / memprediksi nilai masa mendatang / masa depan.

4. Perbedaan Estimasi dan Klasifikasi :

- a. Estimasi mirip dengan klasifikasi, tapi variabel targetnya berupa bilangan numerik (kontinu) dan bukan kategorikal (nominal atau diskrit).
- b. Klasifikasi adalah algoritma yang menggunakan data dengan target/class/label berupa nilai kategorikal (nominal).

5. Perbedaan klasifikasi dan klastering :

- a. Klasifikasi adalah algoritma yang menggunakan data dengan target/class/label berupa nilai kategorikal (nominal)
- b. Klastering adalah pengelompokan data , hasil observasi dan kasus ke dalam class yang mirip. Perbedaan utama algoritma klastering dengan klasifikasi adalah klastering tidak memiliki target/class/label, jadi termasuk unsupervised learning

6. Perbedaan klastering dan prediksi :

- a. Klastering adalah penegelompokan data, hasil observasi dan kasus ke dalam class yang mirip.
- b. Prediksi : Algoritma prediksi/forecasting sama dengan algoritma estimasi di mana label/target/class bertipe numerik.

7. Perbedaan supervised dan unsupervised learning :

- a. Supervised Learning adalah Machine Learning model yang mempelajari data dengan label atau target dimana evaluasi model tersebut akan berdasarkan target

Nama : Wahyu Adi Nugroho

NIM : A11.2019.12310

Kelp. : 46UG

ini. Sedangkan Unsupervised Learning adalah Machine Learning Model yang mempelajari pola data tanpa adanya target data.

- b. Model – model di Supervised Learning membutuhkan data training berupa input data dan target data yang diinginkan. Model ini dilatih untuk melakukan prediksi berdasarkan pola yang ditemukan dalam menjawab data target. Sedangkan Unsupervised Learning hanya memerlukan data input tanpa contoh target data. Model ini hanya mencoba menemukan pola dan menemukan insight penting dari data. Seringkali, proses ini disebut juga sebagai data mining.
- c. Supervised Learning menggunakan data training untuk membuat machine learning model dan model ini akan digunakan untuk diuji pada data test. Sedangkan Unsupervised Learning tidak menggunakan data training dan hanya tergantung pada data test sehingga kita tidak bisa melakukan evaluasi terhadap model.
- d. Algoritma model dari Supervised Learning adalah algoritma klasifikasi untuk memprediksi fitur kategori (yes or no) dan Regresi untuk memprediksi fitur kontinu (harga rumah, harga saham, dll). Sedangkan algoritma Unsupervised Learning seperti Clustering untuk melakukan segmentasi data (segmentasi pelanggan, segmentasi risiko, dll.) dan Dimensional Reduction.
- e. Model dari Supervised Learning dievaluasi berdasarkan dari hasil prediksi yang dilatih menggunakan Training Data dan dibandingkan hasilnya dengan prediksi oleh Test data. Sedangkan Unsupervised Learning harus di evaluasi secara subjektif untuk mengetahui apakah prediksi yang dilakukan telah sesuai karena pengukuran evaluasi secara statistik pada unsupervised learning tidak memiliki jawaban yang benar.

8. Tahapan utama proses data mining :

- a. Input (Data)
- b. Metode (Algoritma Data Mining)
- c. Output (Pola / Model)
- d. Evaluation (Akurasi, AUC, RMSE, dll)