

Nama : Khalfani Novian Habibi

NIM : A11.2019.11906

Kelas : A11.4620

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 prediksi dan klasifikasi!
5. Jelaskan perbedaan klasifikasi dan klustering!
6. Jelaskan perbedaan klustering dan association!
7. Jelaskan perbedaan supervised dan unsupervised learning!
8. Sebutkan tahapan utama proses data mining!

Jawab:

1.

- Estimation
- Prediction
- Classification
- Clustering
- Association

2. Algoritma Data Mining :

- Estimation : Linier Regression, Neural Network, Support Vector Machine, dll
- Prediction : Linier Regression, Neural Network, Support Vector Machine, dll
- Classification : Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, LinearDiscriminant Analysis, dll
- Klustering : K-Means, K-Medoids, Self-Organizing, Map(SOM), Fuzzy C-Means, dll
- Assosiation : FP-Growth, A Priori, dll

3. Estimasi

- Algoritma estimasi mirip dengan algoritma klasifikasi, tapi variabel target adalah berupa bilangan numerik (kontinyu) dan bukan kategorikal (nominal atau diskrit)
- Estimasi nilai dari variable target ditentukan berdasarkan nilai dari variabel prediktor (atribut)
- Algoritma estimasi yang biasa digunakan adalah: Linear Regression, Neural Network, Support Vector Machine

Prediksi

- Algoritma prediksi/forecasting sama dengan algoritma estimasi di mana label/target/class bertipe numerik, bedanya adalah data yang digunakan merupakan data rentet waktu (data time series)
- Istilah prediksi kadang digunakan juga untuk klasifikasi, tidak hanya untuk prediksi time series, karena sifatnya yang bisa menghasilkan class berdasarkan berbagai atribut yang kita sediakan
- Semua algoritma estimasi dapat digunakan untuk prediksi/forecasting

4. Prediksi

- Algoritma prediksi/forecasting sama dengan algoritma estimasi di mana label/target/class bertipe numerik, bedanya adalah data yang digunakan merupakan data rentet waktu (data time series)
- Istilah prediksi kadang digunakan juga untuk klasifikasi, tidak hanya untuk prediksi time series, karena sifatnya yang bisa menghasilkan class berdasarkan berbagai atribut yang kita sediakan
- Semua algoritma estimasi dapat digunakan untuk prediksi/forecasting

Klasifikasi

- Klasifikasi adalah algoritma yang menggunakan data dengan target/class/label berupa nilai kategorikal (nominal)
- Contoh, apabila target/class/label adalah pendapatan, maka bisa digunakan nilai nominal (kategorikal) sbb: pendapatan besar, menengah, kecil
- Contoh lain adalah rekomendasi contact lens, apakah menggunakan yang jenis soft, hard atau none
- Algoritma klasifikasi yang biasa digunakan adalah: Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, etc

5. Klasifikasi

- Klasifikasi adalah algoritma yang menggunakan data dengan target/class/label berupa nilai kategorikal (nominal)
- Contoh, apabila target/class/label adalah pendapatan, maka bisa digunakan nilai nominal (kategorikal) sbb: pendapatan besar, menengah, kecil
- Contoh lain adalah rekomendasi contact lens, apakah menggunakan yang jenis soft, hard atau none
- Algoritma klasifikasi yang biasa digunakan adalah: Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, etc

Klastering

- Klastering adalah pengelompokan data, hasil observasi dan kasus ke dalam class yang mirip
- Suatu klaster (cluster) adalah koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari klaster lain
- Perbedaan utama algoritma klastering dengan klasifikasi adalah klastering tidak memiliki target/class/label, jadi termasuk unsupervised learning
- Klastering sering digunakan sebagai tahap awal dalam proses data mining, dengan hasil klaster yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan

6. Klastering

- Klastering adalah pengelompokan data, hasil observasi dan kasus ke dalam class yang mirip
- Suatu klaster (cluster) adalah koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari klaster lain
- Perbedaan utama algoritma klastering dengan klasifikasi adalah klastering tidak memiliki target/class/label, jadi termasuk unsupervised learning
- Klastering sering digunakan sebagai tahap awal dalam proses data mining, dengan hasil klaster yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan

Asosiasi

- Algoritma association rule (aturan asosiasi) adalah algoritma yang menemukan atribut yang “muncul bersamaan”
- Dalam dunia bisnis, sering disebut dengan affinity analysis atau market basket analysis
- Algoritma asosiasi akan mencari aturan yang menghitung hubungan diantara dua atau lebih atribut
- Algoritma association rules berangkat dari pola “If antecedent, then consequent,” bersamaan dengan pengukuran support (coverage) dan confidence (accuracy) yang terasosiasi dalam aturan

7. SUPERVISED LEARNING

Supervised Learning (Pembelajaran dengan Guru):

- Sebagian besar algoritma data mining (estimation, prediction/forecasting, classification) adalah supervised learning
- Variabel yang menjadi target/label/class ditentukan
- Algoritma melakukan proses belajar berdasarkan nilai dari variabel target yang terasosiasi dengan nilai dari variable prediktor

Supervised Learning : Algoritma Estimasi, Prediksi, Klasifikasi

Unsupervised Learning

(Pembelajaran tanpa Guru):

- Algoritma data mining mencari pola dari semua variable(atribut)
- Variable (atribut) yang menjadi target/label/class tidak ditentukan (tidak ada) Algoritma clustering adalah algoritma unsupervised learning

8. Tahapan utama data mining

1. Pembersihan data (data cleaning)

Dalam proses menambang emas, tahap awal yang dilakukan adalah dengan mencari batu atau lumpur yang sekiranya bisa diolah menjadi emas yang bagus. Dalam proses tersebut, pasti menemukan satu atau bahkan mencapai puluhan batu yang kurang layak untuk diproses. Sehingga batu-batu itu perlu dibuang.

Begitu juga penerapannya dalam tahapan data mining. Sebelum diproses dan dibentuk menjadi sebuah knowledge, data yang ada harus dibersihkan terlebih dahulu. Jika ada yang mengandung error, maka data-data tersebut harus dibuang. Sehingga tersisa data yang 'bagus' untuk diolah dalam tahap selanjutnya.

2. Integrasi data (data integration)

Langkah kedua dalam tahapan data mining adalah integrasi data. Setelah menemukan batu-batu yang cocok, selanjutnya penambang akan mulai mengkombinasikan untuk dijadikan batangan emas atau bentuk emas lainnya. Dalam data mining, data yang berhasil dibersihkan juga akan diintegrasikan.

3. Transformation

Seperti bentuk emas yang berbeda-beda, maka proses pembentukannya juga akan berbeda. Begitu juga dengan data, dalam tahap transformation data akan dipilih dan diubah formatnya agar sesuai dengan teknik atau metode yang dipakai. Pada tahap ini pula kualitas data mining akan terlihat.

4. Data mining

Tahapan data mining selanjutnya adalah proses penambangan data itu sendiri. Perlu penentuan metode penambangan yang tepat.

5. Evaluasi pola (pattern evaluation)

Setelah bahan mentah emas selesai diolah, maka perlu diuji juga apakah sudah sesuai dengan standar kualitasnya atau belum. Baru setelah itu bisa didistribusikan ke toko emas dan di-display kepada pelanggan.