

Rizqi Bachtiar Utuyansyah

A11.2022.14068

Tugas

1. Sebutkan 5 peran utama data mining !
 - Klasifikasi : Mengkategorikan data ke dalam kelompok-kelompok yang telah ditentukan berdasarkan fitur-fitur tertentu.
 - Asosiasi : Mencari pola di mana dua atau lebih item saling berkaitan dalam sebuah database.
 - Klasterisasi : Mengelompokkan data berdasarkan kesamaan, tanpa menggunakan kategori yang telah ditentukan sebelumnya.
 - Prediksi : Menggunakan data historis untuk memprediksi hasil atau tren masa depan.
 - Deteksi Anomali : Mencari data yang menyimpang dari norma atau pola yang diharapkan.

2. Algoritma apa saja yang dapat digunakan untuk 5 peran utama data mining di atas ?
 - Klasifikasi
 - o Pohon Keputusan, seperti algoritma C4.5, CART
 - o Mesin Vektor Pendukung (Support Vector Machines, SVM).
 - o Jaringan Saraf Tiruan (Neural Networks).
 - o Naive Bayes.
 - o K-Nearest Neighbors (KNN).
 - Asosiasi
 - o Apriori: Untuk menemukan itemsets yang sering muncul bersama-sama dalam database transaksi.
 - o FP-Growth: Alternatif yang lebih cepat dari Apriori untuk menemukan itemsets yang sering.
 - o Eclat: Varian lain dari algoritma asosiasi.
 - Klasterisasi
 - o K-Means.
 - o Hierarchical Clustering
 - o DBSCAN (Density-Based Spatial Clustering of Applications with Noise).
 - o Model Campuran Gaussian (Gaussian Mixture Models, GMM).
 - Prediksi
 - o Regresi Linear dan Regresi Logistik.
 - o Random Forest.
 - o Gradient Boosting Machines (seperti XGBoost dan LightGBM).
 - Deteksi Anomali
 - o One-Class SVM.
 - o Isolation Forest.
 - o Local Outlier Factor (LOF).
 - o Autoencoders

3. Jelaskan perbedaan estimasi dan prediksi !

- Estimasi
 - Fokus pada menentukan nilai atau parameter terbaik berdasarkan data yang ada.
 - Bersifat deterministik dan berfokus pada data saat ini atau historis.
- Prediksi
 - Fokus pada mengantisipasi nilai atau kejadian di masa depan berdasarkan model yang dibangun dari data historis.
 - Bersifat probabilistik dan berorientasi pada data yang belum diamati atau masa depan.

4. Jelaskan perbedaan estimasi dan klasifikasi !

- Estimasi
 - Berkaitan dengan menentukan atau memperkirakan nilai kontinu untuk suatu entitas berdasarkan data yang tersedia.
 - Contohnya meliputi memprediksi harga rumah, perkiraan penjualan, atau suhu di hari berikutnya.
- Klasifikasi
 - Berkaitan dengan menempatkan entitas ke dalam salah satu dari sekumpulan kategori atau kelas yang telah ditentukan sebelumnya.
 - Contohnya meliputi mengidentifikasi apakah email adalah spam atau bukan, atau mengkategorikan pelanggan ke dalam segmen tertentu berdasarkan perilaku belanja mereka.

5. Jelaskan perbedaan klasifikasi dan klastering !

- Klasifikasi
 - Prosedur pembelajaran yang diawasi (supervised learning) di mana model dibangun berdasarkan dataset dengan label yang telah diketahui.
 - Tujuannya adalah untuk memprediksi label kategori atau kelas dari data baru berdasarkan atribut-atributnya.
 - Contohnya meliputi memprediksi apakah seorang pelanggan akan membeli produk atau tidak, atau mengidentifikasi apakah email adalah spam atau bukan.
- Klastering
 - Prosedur pembelajaran yang tidak diawasi (unsupervised learning) di mana data dikelompokkan berdasarkan kesamaan atribut tanpa menggunakan label yang telah diketahui sebelumnya.
 - Tujuannya adalah untuk menemukan struktur atau pola tersembunyi dalam data.
 - Contohnya meliputi segmentasi pasar berdasarkan perilaku pelanggan atau mengelompokkan dokumen berdasarkan topik tanpa pengetahuan awal tentang kategori topik tersebut.

6. Jelaskan perbedaan klustering dan prediksi !

- Klustering
 - Merupakan teknik pembelajaran tidak diawasi (unsupervised learning).
 - Bertujuan untuk mengelompokkan data menjadi kumpulan atau kluster berdasarkan kesamaan fitur, tanpa menggunakan informasi yang telah diberi label.
 - Fokus pada identifikasi struktur atau pola tersembunyi dalam data.
 - Contohnya termasuk segmentasi pasar atau pengelompokan artikel berita berdasarkan kesamaan konten.
- Prediksi
 - Biasanya melibatkan teknik pembelajaran yang diawasi (supervised learning), meskipun bisa juga melibatkan metode tidak diawasi dalam konteks tertentu.
 - Bertujuan untuk mengantisipasi dan memproyeksikan hasil atau nilai dari entitas berdasarkan model yang telah dibangun dari data historis.
 - Fokus pada proyeksi kejadian atau nilai di masa depan atau data yang belum diamati.
 - Contohnya termasuk meramalkan harga saham, cuaca, atau prediksi penjualan produk di masa depan.

7. Jelaskan perbedaan supervised dan unsupervised learning !

- Supervised Learning
 - Data yang digunakan untuk pelatihan model memiliki label atau keluaran yang diketahui.
 - Tujuan utamanya adalah untuk membangun model yang dapat memprediksi label atau keluaran berdasarkan fitur masukan.
 - Contohnya meliputi klasifikasi (misalnya, memprediksi apakah email adalah spam atau bukan) dan regresi (misalnya, memprediksi harga rumah).
- Unsupervised Learning
 - Data yang digunakan tidak memiliki label atau keluaran yang diketahui.
 - Tujuan utamanya adalah untuk menemukan struktur atau pola tersembunyi dalam data.
 - Contohnya meliputi klustering (misalnya, segmentasi pasar berdasarkan perilaku pelanggan) dan reduksi dimensi (misalnya, analisis komponen utama).

8. Sebutkan tahapan utama proses data mining !

- Pemahaman Bisnis: Menentukan tujuan proyek dan pemahaman kebutuhan bisnis.
- Pemahaman Data: Eksplorasi awal dan pemeriksaan kualitas data.
- Pembersihan dan Persiapan Data: Melakukan seleksi, pembersihan, dan transformasi data.
- Pemodelan: Memilih teknik dan membangun model dengan algoritma yang sesuai.
- Evaluasi: Menilai dan memvalidasi kualitas model.
- Penerapan: Mengintegrasikan model ke dalam sistem dan menggunakan model untuk keputusan atau prediksi.