Nama : Muhammad Faiq Raihan Dhiyaulhak

Nim : A11.2021.13833

Matkul : Data Mining(Tugas Pertemuan 4)

===============================================================================

1. 5 peran data mining diantaranya

* Identifikasi pola dan hubungan: Data mining dapat membantu mengidentifikasi pola dan hubungan antara data yang kompleks dan besar. Misalnya, dalam analisis penjualan ritel, data mining dapat membantu mengidentifikasi pola pembelian pelanggan, preferensi produk, dan faktor lain yang dapat mempengaruhi perilaku konsumen.
* Klasifikasi dan prediksi: Data mining juga dapat membantu dalam klasifikasi data ke dalam kelompok-kelompok atau kategori-kategori yang berbeda, dan dalam membuat prediksi berdasarkan data yang telah dikumpulkan. Contoh penggunaan data mining dalam klasifikasi dan prediksi termasuk di bidang pemasaran untuk memprediksi peluang konversi pelanggan atau di bidang kesehatan untuk memprediksi risiko penyakit tertentu.
* Pemrosesan dan analisis data yang cepat: Data mining memungkinkan pemrosesan data yang cepat dan efisien, yang membantu dalam pengambilan keputusan dan analisis data yang lebih baik dan lebih akurat. Misalnya, dalam analisis log data web, data mining dapat membantu memproses jumlah data yang besar dalam waktu yang singkat.
* Identifikasi anomali: Data mining dapat membantu mengidentifikasi anomali atau data yang tidak biasa yang mungkin perlu diinvestigasi lebih lanjut. Contoh penggunaan data mining dalam identifikasi anomali termasuk di bidang keamanan siber untuk mengidentifikasi aktivitas yang mencurigakan atau di bidang keuangan untuk mengidentifikasi aktivitas keuangan yang tidak biasa.
* Pengambilan keputusan: Data mining dapat membantu pengambilan keputusan dengan menyajikan informasi yang relevan dan signifikan dari data yang telah dikumpulkan. Misalnya, dalam analisis data kredit, data mining dapat membantu memperkirakan kemampuan seseorang untuk membayar pinjaman berdasarkan data historis dan profil kredit mereka.

1. Algoritma yang digunakan pada peran data mining diatas
2. Klasifikasi:

* Decision Trees
* Random Forests
* Naive Bayes
* K-Nearest Neighbors (KNN)
* Support Vector Machines (SVM)

1. Regresi:

* Linear Regression
* Logistic Regression
* Polynomial Regression
* Ridge Regression
* Lasso Regression

1. Klastering:

* K-Means
* Hierarchical Clustering
* DBSCAN
* Mean Shift
  + Fuzzy C-Means

1. Asosiasi:

* Apriori
* FP-Growth
* Eclat
* CARMA

1. Pemodelan Anomali:

* Isolation Forest
* Local Outlier Factor (LOF)
* One-class SVM
* Clustering-based Anomaly Detection
* Angle-based Outlier Detection

1. Perbedaan estimasi dan prediksi

Estimasi adalah proses penggunaan data historis untuk membuat perkiraan tentang parameter populasi yang tidak diketahui. Dalam hal ini, parameter populasi dapat berupa rerata, variansi, proporsi, atau parameter lainnya yang mewakili sifat atau karakteristik dari populasi. Estimasi berguna dalam statistik inferensial dan biasanya melibatkan pengujian hipotesis dan interval kepercayaan.

prediksi adalah proses membuat perkiraan tentang nilai variabel target yang belum diketahui atau belum terjadi. Prediksi berguna dalam machine learning dan sering melibatkan pengembangan model prediktif berdasarkan data historis yang digunakan untuk membuat prediksi tentang hasil yang belum terjadi pada data baru. Model prediktif dapat berupa model regresi, model klasifikasi, atau model lainnya yang dibangun dengan menggunakan teknik machine learning.

perbedaan utama antara estimasi dan prediksi adalah bahwa estimasi digunakan untuk membuat perkiraan tentang parameter populasi yang tidak diketahui, sedangkan prediksi digunakan untuk membuat perkiraan tentang nilai variabel target yang belum diketahui.

1. Perbedaan klasifikasi dan perdiksi

Klasifikasi adalah suatu teknik di mana kita mengelompokkan data ke dalam beberapa kelas berdasarkan fitur atau atribut tertentu. Tujuan dari klasifikasi adalah untuk membangun model yang dapat memprediksi kelas baru untuk setiap data baru yang diberikan. Contohnya adalah mengklasifikasikan email sebagai spam atau bukan spam.

Prediksi adalah teknik di mana kita memperkirakan nilai suatu variabel target berdasarkan variabel input lainnya. Tujuan dari prediksi adalah untuk membangun model yang dapat memprediksi nilai yang tepat atau perkiraan yang baik untuk variabel target baru berdasarkan input yang diberikan. Contohnya adalah memprediksi harga rumah berdasarkan ukuran, lokasi, dan faktor lainnya.

Perbedaan utama antara klasifikasi dan prediksi adalah fokusnya. Klasifikasi berfokus pada pengelompokan data ke dalam kelas, sedangkan prediksi berfokus pada memperkirakan nilai variabel target. Namun, keduanya menggunakan teknik dan algoritma yang serupa untuk membangun model. Klasifikasi dapat digunakan sebagai bagian dari prediksi, misalnya dalam memprediksi kelas target dari data baru yang diberikan.

1. Perbedaan klasifikasi dan klastering

Klasifikasi adalah suatu metode dalam pembelajaran mesin yang dilakukan untuk memprediksi kelas atau label pada data baru berdasarkan data pelatihan yang telah diberi label sebelumnya. Klasifikasi dapat digunakan untuk membangun model prediktif yang dapat memprediksi label atau kelas dari data baru. Contoh dari aplikasi klasifikasi adalah sistem pendeteksi spam email, pengenalan suara, atau identifikasi objek pada gambar.

Sementara itu, klastering adalah suatu metode dalam pembelajaran mesin yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok atau klaster yang serupa berdasarkan kemiripan atau kesamaan fitur. Tujuan dari klastering adalah untuk menemukan pola atau struktur di dalam data tanpa memerlukan data latih yang telah diberi label. Contoh dari aplikasi klastering adalah segmentasi pelanggan, analisis citra, atau pemetaan wilayah berdasarkan tingkat kriminalitas.

Dalam ringkasan, klasifikasi berfokus pada prediksi label atau kelas dari data baru berdasarkan data pelatihan yang diberi label, sedangkan klastering berfokus pada mengelompokkan data ke dalam kelompok-kelompok yang serupa berdasarkan kemiripan atau kesamaan fitur.

1. Perbedaan klastering dan prediksi

Klastering (clustering) adalah teknik pengelompokan data atau objek ke dalam beberapa kelompok (cluster) berdasarkan kesamaan karakteristik atau sifat tertentu, seperti jarak atau kemiripan antara objek-objek tersebut. Dalam klastering, tujuannya adalah untuk menemukan kelompok-kelompok yang berbeda dari data tanpa mengetahui klasifikasi yang sebenarnya atau label. Klastering sering digunakan dalam analisis data dan pemrosesan citra, misalnya dalam segmentasi citra.

Sementara itu, prediksi (prediction) adalah teknik untuk membangun model dari data yang sudah diketahui untuk memprediksi nilai yang tidak diketahui atau untuk mengklasifikasikan objek ke dalam kelas-kelas tertentu berdasarkan data yang sudah ada. Dalam prediksi, data dilabeli atau diberi klasifikasi sebelumnya, dan tujuannya adalah untuk membangun model yang dapat digunakan untuk memprediksi atau mengklasifikasikan data baru berdasarkan pola atau informasi yang terdapat pada data yang sudah diketahui.

1. Perbedaan supervised dan unsupervised learning

Dalam supervised learning, model mesin belajar untuk memetakan input ke output yang sudah diketahui. Artinya, model diberi masukan (input) dan keluaran (output) yang diharapkan, dan ia belajar untuk membuat hubungan antara keduanya. Pada dasarnya, supervised learning bertujuan untuk mengembangkan model yang dapat mempelajari pola dari data latih yang sudah dilabeli dan kemudian mengklasifikasikan atau memprediksi label dari data uji baru. Contoh aplikasi supervised learning adalah klasifikasi gambar dan prediksi harga rumah.

Di sisi lain, dalam unsupervised learning, model mesin tidak diberikan data latih yang sudah dilabeli, dan harus belajar menemukan pola dalam data tanpa bimbingan. Tujuan dari unsupervised learning adalah menemukan struktur tersembunyi dalam data, seperti pengelompokan (clustering) dan reduksi dimensi (dimensionality reduction). Contoh aplikasi unsupervised learning adalah analisis klaster dan pengelompokan pelanggan.

Dalam supervised learning, model mesin dapat diuji dan diukur dengan menggunakan data latih dan data uji, sedangkan dalam unsupervised learning, tidak ada data uji karena tidak ada label. Oleh karena itu, evaluasi dalam unsupervised learning lebih sulit dan harus dilakukan secara subjektif.

1. Sebutkan tahapan utama dalam proses data mining

* Pemahaman bisnis (Business Understanding): Tahap ini adalah tahap awal yang sangat penting dalam proses data mining. Pada tahap ini, perlu dipahami tujuan bisnis atau masalah yang ingin dipecahkan, serta data apa yang diperlukan untuk menjawab masalah tersebut.
* Pemahaman data (Data Understanding): Pada tahap ini, dilakukan eksplorasi data untuk memahami karakteristik data yang digunakan. Hal ini dilakukan untuk mengidentifikasi kualitas data, data yang hilang atau tidak lengkap, serta hubungan antar variabel dalam data.
* Persiapan data (Data Preparation): Tahap ini melibatkan pengolahan data, seperti pembersihan data, transformasi data, integrasi data, dan pemilihan fitur data. Tujuannya adalah untuk menghasilkan data yang berkualitas dan dapat digunakan untuk proses selanjutnya.
* Pemodelan (Modeling): Tahap ini melibatkan penggunaan algoritma data mining untuk mengidentifikasi pola atau hubungan yang tersembunyi dalam data. Beberapa teknik yang sering digunakan pada tahap ini adalah clustering, association rule, decision tree, dan neural network.
* Evaluasi (Evaluation): Tahap ini melibatkan evaluasi model yang dibuat untuk memastikan bahwa model yang dihasilkan sesuai dengan tujuan bisnis dan dapat digunakan untuk pengambilan keputusan.
* Implementasi (Deployment): Tahap terakhir dalam proses data mining adalah mengimplementasikan model yang sudah dibuat ke dalam sistem bisnis atau proses yang digunakan. Hal ini bertujuan untuk memastikan model yang dihasilkan dapat digunakan secara efektif dalam praktik bisnis.