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