

Nama : Elmo Allistair

Kelas : 3KA17

NPM : 12118220

Konsep data mining

Penambangan data (*data mining*) adalah proses menemukan anomali, pola, dan korelasi dalam kumpulan data besar untuk memprediksi hasil. Dengan menggunakan berbagai teknik, Anda dapat menggunakan informasi ini untuk meningkatkan pendapatan, memotong biaya, meningkatkan hubungan pelanggan, mengurangi risiko, dan banyak lagi.

Proses *data mining* untuk menemukan koneksi dan memprediksi tren masa depan memiliki sejarah yang panjang. Terkadang disebut sebagai "penemuan pengetahuan dalam database", istilah "penggalian data" tidak diciptakan hingga tahun 1990-an. Tetapi pondasinya terdiri dari tiga disiplin ilmu yang saling terkait: statistik (studi numerik tentang hubungan data), kecerdasan buatan (kecerdasan mirip manusia yang ditampilkan oleh perangkat lunak dan / atau mesin) dan pembelajaran mesin (algoritma yang dapat belajar dari data untuk membuat prediksi). Apa yang lama adalah baru lagi, karena teknologi penambangan data terus berkembang untuk mengimbangi potensi data besar yang tak terbatas dan daya komputasi yang terjangkau.



Penerapan Data Mining

Data mining adalah inti dari upaya analitik di berbagai industri dan disiplin ilmu.

Telekomunikasi, Media & Teknologi

Perusahaan telekomunikasi, media, dan teknologi dapat menggunakan model analitik untuk memahami data pelanggan, membantu mereka memprediksi perilaku pelanggan dan menawarkan kampanye yang sangat bertarget dan relevan.

Asuransi

Dengan pengetahuan analitik, perusahaan asuransi dapat memecahkan masalah kompleks terkait penipuan, kepatuhan, manajemen risiko, dan pengurangan pelanggan. Perusahaan telah menggunakan teknik *data mining* untuk memberi harga produk secara lebih efektif di seluruh lini bisnis dan menemukan cara baru untuk menawarkan produk kompetitif ke basis pelanggan mereka yang ada.

Pendidikan

Dengan tampilan kemajuan siswa yang disatukan dan digerakkan oleh data, pendidik dapat memprediksi kinerja siswa. *Data mining* membantu pengajar mengakses data siswa, memprediksi tingkat pencapaian, dan menunjukkan siswa atau kelompok siswa yang membutuhkan perhatian ekstra.

Manufaktur

Menyelaraskan rencana pasokan dengan prakiraan permintaan sangat penting, seperti halnya deteksi dini masalah, jaminan kualitas, dan investasi dalam ekuitas merek. Produsen dapat memprediksi keausan aset produksi dan mengantisipasi pemeliharaan, yang dapat memaksimalkan waktu kerja dan menjaga jalur produksi sesuai jadwal.

Perbankan

Algoritma otomatis membantu bank memahami basis pelanggan mereka serta miliaran transaksi di jantung sistem keuangan. Penambangan data membantu perusahaan jasa keuangan mendapatkan pandangan yang lebih baik tentang risiko pasar, mendeteksi penipuan lebih cepat, mengelola kewajiban kepatuhan terhadap peraturan, dan mendapatkan pengembalian optimal atas investasi pemasaran mereka.

Proses Data Mining

Proses-proses yang biasa dilakukan oleh data mining antara lain: deskripsi, prediksi, estimasi, klasifikasi, pengelompokan dan asosiasi.

Description

Deskripsi bertujuan untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat dengan mudah dipahami oleh para ahli dalam domain aplikasi. Aturan yang dihasilkan harus mudah dipahami agar dapat meningkatkan tingkat pengetahuan dalam sistem secara efektif. Tugas deskriptif merupakan tugas data mining yang seringkali dibutuhkan dalam teknik *post processing* untuk memvalidasi dan menjelaskan hasil dari proses data mining. *Post Processing* adalah proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan.

Prediction

Prediksi mirip dengan klasifikasi, tetapi data diklasifikasikan berdasarkan perilaku atau nilai yang diprediksi di masa mendatang. Contoh tugas prediksi, misalnya untuk memprediksi penurunan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan ke depan.

Estimation

Estimasi hampir sama dengan prediksi, kecuali variabel target diperkirakan lebih ke arah numerik daripada arah kategori. Model dibangun dengan menggunakan catatan pelengkap yang memberikan nilai variabel target sebagai nilai prediksi. Selanjutnya pada review selanjutnya nilai estimasi variabel target didasarkan pada nilai variabel prediktif. Misalnya, perkiraan tekanan darah sistolik pasien rumah sakit didasarkan pada usia pasien, jenis kelamin, berat badan, dan kadar natrium darah. Hubungan tekanan darah sistolik dengan nilai variabel prediktif dalam proses pembelajaran akan menghasilkan model estimasi.

Classification

Klasifikasi adalah proses menemukan model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas. Klasifikasi melibatkan proses pemeriksaan ciri-ciri suatu

objek dan memasukkan suatu objek ke dalam salah satu kelas yang telah ditentukan sebelumnya.

Clustering

Clustering adalah pengelompokan data tanpa didasarkan pada kelas data tertentu ke dalam kelas objek yang sama. Cluster adalah kumpulan record yang memiliki kesamaan satu sama lain dan memiliki ketidakcocokan dengan record di cluster lain. Tujuannya untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok. Semakin besar kesamaan objek dalam sebuah cluster dan semakin besar perbedaan di setiap cluster, semakin baik kualitas analisis cluster.

Association

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul pada suatu waktu. Dalam dunia bisnis lebih sering disebut analisis keranjang belanja (*market basket analysis*). Tugas asosiasi berusaha mengungkap aturan untuk mengukur hubungan antara dua atribut atau lebih.

Tahapan Data Mining

Sebagai rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap:

Pembersihan data

Pembersihan data adalah proses menghilangkan kebisingan dan data yang tidak konsisten atau data yang tidak relevan. Secara umum, data yang diperoleh baik dari database perusahaan maupun hasil eksperimen memiliki konten yang tidak sempurna seperti data hilang, data tidak valid atau juga hanya salah ketik. Selain itu, terdapat pula atribut data yang tidak relevan dengan hipotesis data mining yang dimilikinya. Data yang tidak relevan juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi kinerja teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

Integrasi data (*Data Integration*) Integrasi data merupakan penggabungan data dari berbagai basis data menjadi satu basis data baru. Tidak jarang data yang dibutuhkan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file

teks. Integrasi data dilakukan pada atribut atribut yang mengidentifikasi entitas unik seperti atribut nama, tipe produk, nomor pelanggan dan lain-lain. Integrasi data perlu dilakukan dengan hati-hati karena kesalahan dalam integrasi data dapat menghasilkan hasil yang menyimpang bahkan menyesatkan tindakan nantinya. Misalnya jika integrasi data berdasarkan tipe produk ternyata menggabungkan produk dari kategori yang berbeda, maka akan terjadi korelasi antar produk yang sebenarnya tidak ada.

Pemilihan Data (*Data Selection*)

Data yang ada di database seringkali tidak semuanya digunakan, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Misalnya, kasus yang meneliti tentang faktor kecenderungan orang untuk membeli dalam kasus analisis keranjang pasar, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan.

Transformasi Data (*Data Transformation*)

Data diubah atau digabungkan ke dalam format yang sesuai untuk diproses di data mining. Beberapa metode data mining memerlukan format data khusus sebelum dapat diterapkan. Misalnya, beberapa metode standar seperti analisis asosiasi dan pengelompokan hanya dapat menerima input data kategoris. Oleh karena itu data berupa bilangan numerik kontinu perlu dibagi menjadi beberapa interval. Proses ini sering disebut dengan transformasi data. Proses penambangan, ini adalah proses utama ketika metode diterapkan untuk menemukan pengetahuan yang berharga dan tersembunyi dari data.

Evaluate Patterns (*Evaluasi Pola*)

Pada tahap ini hasil teknik data mining berupa pola khas serta model prediktif dievaluasi untuk menilai apakah hipotesis yang ada benar-benar tercapai. Apabila ternyata hasil yang diperoleh tidak sesuai dengan hipotesis terdapat beberapa alternatif yang dapat diambil seperti membuat umpan balik untuk perbaikan proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil tersebut sebagai hasil yang tidak diharapkan yang mungkin berguna.