Nama : M. Thoriq Aziz

NIM : 04318029

Prodi : Teknik Informatika

UTS Penggalian Data dan Analitika Bisnis

# Pengertian Data Mining

Data adalah suatu kumpulan yang terdiri dari fakta-fakta untuk memberikan gambaran yang luas mengenai dengan suatu keadaan. Melalui data, kita dapat menganalisa, menggambarkan, atau menjelaskan suatu keadaan. Data dikumpulkan melalui cara-cara tertentu lalu diolah menjadi suatu informasi yang jelas dan mudah dipahami setiap orang. Data juga dibutuhkan untuk berbagai macam kebutuhan, misalnya berbisnis, penelitian, kependudukan, dan lain-lain. Data Mining merupaka proses menambang/mengambil data dan menganalisanya yang kemudian akan menghasilkan suatu informasi,pola,pengetahuan atau apapun yang dibutuhkan dan didapat dari hasil analisa.

# Peran Data Mining

Data Mining merupakan suatu proses pengumpulan informasi penting dari sebuah data yang besar dilakukan melalui beberapa proses yang meliputi metode statistika, matematika, maupun teknologi artificial intelligence. Data Mining sendiri dapat menggambarkan suatu pengumpulan beberapa teknik yang bertujuan agar menemukan pola-pola yang belum diketahui pada data yang terkumpul. Selain itu pengumpulan data tersebut juga menjadi bagian dari Knowledge Discovery of Database (KDD) yang terdiri dari beberapa tahapan yaitu pemilihan data, pengolahan, transformasi, penerapan serta evaluasi.

# Penambangan data

Cara Mendapatkan Data

Sebagai contoh saya akan melakukan crawling dengan menggunakan keyword: “Harga

saham”, sehingga hal yang perlu melakukan:

1. Buka terminal / command prompt (cmd)

2. Pilih destinasi folder yang akan digunakan untuk menyimpan data csv

3. Jalankan command seperti ini:

$ twint -s "harga saham" -o harga-saham.csv --limit 30 --since 2018-01-01 --until 2019-

12-31 --csv --lang id

Contoh proses crawling:

• Mengenai: “jokowi atau prabowo” → tweet yang mengandung kata

“jokowi” atau “prabowo” akan diambil

• Membatasi dan mencari tweet sebanyak: ~1400

• Simpan menjadi file: .csv

• Nama (output) file: jokowi-prabowo.csv

• Ambil Tweet mulai tanggal: 20 Oktober 2018 (2018–10–20)

• Sampai dengan tanggal: 2 November 2018 (2018–11–2)

twitterscraper "jokowi OR prabowo" --output jokowi-prabowo.csv --limit 1400 --begindate 2018-

10-20 --enddate 2018-11-2 –csv

# Pemanfaatan Aplikasi Data Mining

Penambangan data bisa digunakan di berbagai sektor, Mulai dari sektor bisnis, manajemen, keuangan dan lain sebagaianya. Berikut Contoh penerapan Data mining di beberapa sektor :

1. Market Analysis dan Management

Dalam sektor pemasaran biasanya data mining digunakan untuk Pemasaran target, manajemen hubungan pelanggan (CRM), analisis pasar, cross selling, segmentasi pasar.

* Target Pemasaran, Misalnya menemukan kelompok pelanggan “model” yang memiliki karakteristik yang sama: minat,tingkat pendapatan, kebiasaan belanja, dll. atau menentukan pola pembelian pelanggan dari waktu ke waktu.
* Analysis lalu lintas pasar, Menemukan hubungan / hubungan antar produk penjualan, & prediksi berdasarkan asosiasi tersebut.
* Profiling pelanggan, Jenis pelanggan apa yang membeli produk apa (pengelompokan atau klasifikasi)
* Analisis kebutuhan pelanggan, Misalnya identifikasi produk terbaik untuk berbagai kelompok pelanggan, Memprediksi faktor apa yang akan menarik pelanggan baru, Penyediaan informasi ringkasan, Laporan ringkasan multidimensi, Informasi ringkasan statistik (kecenderungan dan variasi pusat data)

1. Corporate Analysis & Risk Management Penerapan Data mining dalam sektor perusahaan biasanya digunakan untuk prediksi, retensi pelanggan, underwriting yang lebih baik, kontrol kualitas, analisis kompetitif.

* Perencanaan keuangan dan evaluasi aset, Misalnya analisis dan prediksi arus kas, analisis klaim kontinjensi untuk mengevaluasi aset, analisis cross-sectional dan time series (rasio keuangan, tren analisis, dll.)
* Planning Perencanaan sumber daya, Misalnya merangkum dan membandingkan sumber daya dan pengeluaran
* Persaingan, Misalnya memantau pesaing dan arah pasar, mengelompokkan pelanggan ke dalam kelas dan penetapan harga berbasis kelas prosedur, dan mengatur strategi penetapan harga di pasar yang sangat kompetitif.

1. Fraud Detection & Mining Unusual Patterns Data mining juga berfungsi untuk mencari dan mendeteksi fraud pada sebuah sistem. Dengan menggunakan data mini maka akan bisa melihat dari jutaan transaksi yang masuk.

* Pendekatan: Clustering & konstruksi model untuk penipuan, analisis outlier
* Aplikasi: Layanan kesehatan, ritel, layanan kartu kredit, telecomm. Misalnya Asuransi otomatis, Pencucian uang, Asuransi kesehatan, Telekomunikasi, Analisis pola yang menyimpang dari norma yang diharapkan, Industri retail, Dll.

# Naïve Bayes

Algoritma Naïve Bayes merupakan sebuah metode klasifikasi yang berakar pada teorema bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes.

Keuntungan penggunaan adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan setimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

# Rapid Miner

Rapid miner merupakan aplikasi data science yang dapat digunakan untuk mempermudah mengelola dataset, Rapid Miner memiliki banyak set-modul yang berisi metode analitik untuk mengolah data. Salah satunya adalah metode Naïve Bayes, dengan Rapid Miner kita dapat dengan mudah mengetahui nilai probabilitas, akurasi, dll dari suatu dataset yang berisi data klasifikasi.