

Anggota Kelompok:

- Elmo Allistair
- Pandu Herdiansyah

Tools untuk data analisis

1. Tableau

Tableau adalah alat visualisasi data yang kuat dan paling cepat berkembang yang digunakan dalam Industri Intelijen Bisnis. Ini membantu dalam menyederhanakan data mentah dalam format yang sangat mudah dimengerti. Tableau membantu membuat data yang dapat dipahami oleh para profesional di tingkat mana pun dalam suatu organisasi. Ini juga memungkinkan pengguna non-teknis untuk membuat dasbor yang disesuaikan. Analisis data dapat dilakukan dengan sangat cepat dengan tool Tableau dan visualisasi yang dibuat berupa dashboard dan lembar kerja.

2. Python

Untuk analisis data dan komputasi interaktif serta visualisasi data, Python pasti akan membuat perbandingan dengan bahasa dan alat pemrograman open source dan komersial lainnya yang digunakan secara luas, seperti R, MATLAB, SAS, Stata, dan lainnya. Dalam beberapa tahun terakhir, dukungan Python yang ditingkatkan untuk pustaka (seperti pandas dan scikit-learn) telah menjadikannya pilihan populer untuk tugas analisis data. Dikombinasikan dengan keseluruhan kekuatan Python untuk rekayasa perangkat lunak tujuan umum, ini adalah pilihan yang sangat baik sebagai bahasa utama untuk membangun aplikasi data.

3. Microsoft Power BI

Microsoft Power BI adalah alat analisis dan visualisasi data yang digunakan oleh banyak ahli teknologi untuk mengungkap wawasan ke dalam kumpulan data. Power BI menawarkan kemampuan *data warehouse* termasuk persiapan data, penemuan data, dan dasbor interaktif.

4. Spreadsheet

Spreadsheet adalah aplikasi komputer untuk organisasi, analisis, dan penyimpanan data dalam bentuk tabel. Spreadsheet memungkinkan kita menganalisis data dalam jumlah besar untuk mendapatkan wawasan baru tentang proses atau bisnis. Hal yang membuat spreadsheet unik adalah kemampuannya untuk menghitung nilai menggunakan rumus matematika dan data dalam sel.

5. Apache Spark

Apache Spark adalah platform komputasi cluster yang dirancang agar cepat dan serbaguna, di sisi kecepatan, Spark memperluas model MapReduce yang populer untuk mendukung lebih banyak jenis penghitungan secara efisien, termasuk kueri interaktif dan *stream processing*. Apache Spark telah mencapai kinerja tinggi untuk data batch dan streaming, menggunakan penjadwal DAG yang canggih, pengoptimal kueri, dan mesin eksekusi fisik.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) yang merupakan proses menyelidiki kumpulan data, menjelaskan subjek, dan memvisualisasikan hasilnya. EDA adalah pendekatan analisis data yang menerapkan berbagai teknik untuk memaksimalkan wawasan tertentu ke dalam kumpulan data, mengungkapkan struktur yang mendasarinya, mengekstrak variabel signifikan, mendeteksi pencilan dan anomali, menguji asumsi, mengembangkan model, dan menentukan parameter terbaik untuk estimasi di masa mendatang.

EDA memungkinkan kita untuk memvisualisasikan data untuk memahaminya serta membuat hipotesis untuk analisis lebih lanjut. Analisis eksplorasi berpusat pada pembuatan sinopsis data atau wawasan untuk langkah selanjutnya dalam proyek data mining. Komponen kunci dari EDA adalah meringkas data, analisis statistik data, dan visualisasi data.

Pada dasarnya, ada empat langkah yang terlibat dalam analisis, yaitu:

1. Pendefinisian masalah, meliputi menentukan tujuan analisis, menguraikan peran dan tanggung jawab, memperoleh status data, dan melakukan analisis biaya.
2. Persiapan data, langkah ini melibatkan metode untuk menyiapkan dataset sebelum melakukan analisis aktual, meliputi pendefinisian sumber data, menentukan skema dan tabel, memahami karakteristik utama data, membersihkan data, menghapus data yang tidak relevan, dan membagi data menjadi beberapa bagian
3. Analisis data, merupakan tahap paling penting yang meliputi meringkas data, menemukan korelasi dan hubungan antar data, mengembangkan model prediktif, mengevaluasi model, dan menghitung akurasi.
4. Pengembangan dan representasi hasil, langkah ini merupakan tahap penyajian data kepada target audiens dalam bentuk grafik, ringkasan, tabel, dan diagram.

Contoh kasus model data mining Supervised dan Unsupervised

1. Supervised

Metode *supervised* adalah metode yang mencoba menemukan hubungan antara atribut masukan (terkadang disebut variabel independen) dan atribut target (terkadang disebut sebagai variabel dependen). Hubungan yang ditemukan direpresentasikan dalam struktur yang disebut sebagai model.

Contoh dari *supervised learning* adalah:

- 1) Prediksi harga rumah
- 2) Klasifikasi hewan
- 3) Prediksi cuaca
- 4) Klasifikasi teks
- 5) Prediksi rute perjalanan

2. Unsupervised

Metode *Unsupervised* adalah metode yang mencoba menemukan struktur tersembunyi pada data yang tidak berlabel. Karena contoh yang diberikan kepada pelajar tidak berlabel, tidak ada kesalahan atau sinyal penghargaan untuk mengevaluasi solusi potensial.

Contoh dari *unsupervised learning* adalah:

- 1) Menemukan segmen pelanggan
- 2) Mengurangi kompleksitas suatu masalah

DAFTAR PUSTAKA

McKinney, Wes. 2018. Python for Data Analysis, **Data Wrangling with Pandas, NumPy, and IPython**. Sebastopol: O'Reilly Media, Inc

Mukhiya, Suresh Kumar; Ahmed, Usman. 2020. **Hands-On Exploratory Data Analysis with Python**. Birmingham: Packt Publishing Ltd.