

Nama : puspitayanti

Nim : 22131002

Tugas

Analisis Ekplorasi Data Pengertian, Jenis dan Pentingnya

Matematikawan Amerika John Tukey Awalnya mengembangkan Analisis Ekplorasi Data pada 1970-an. Sampai saat ini, teknik ADE terus menjadi metode yang banyak digunakan dalam proses penemuan data. Di luar pemodelan formal atau pengujian hipotesis, membuka pintu lebar untuk pemahaman yang lebih baik tentang variabel kumpulan data dan hubungannya. Hal ini juga membantu untuk menentukan apakah Teknik Statistik yang telah dipertimbangkan untuk analisis data sesuai atau tidak.

Definisi Analisis Ekplorasi Data

Eksplorasi data adalah cara untuk mengenal data sebelum mengolahnya. Melalui **survei** dan investigasi, kumpulan data berukuran besar disiapkan untuk analisis yang lebih mendalam dan terstruktur.

Mengutip laman *Alteryx*, eksplorasi memungkinkan pemahaman yang lebih mendalam tentang kumpulan data sehingga memudahkan navigasi dan penggunaan data di kemudian hari.

Semakin baik seorang analis mengetahui data yang mereka kerjakan, semakin baik pula analisis mereka.

Eksplorasi yang sukses dimulai dengan pikiran terbuka, mengungkapkan jalur baru penemuan, dan membantu

mengidentifikasi serta menyempurnakan pertanyaan dan masalah analitik di masa depan.

Analisis ekplorasi data banyak digunakan oleh Ilmuwan Data saat menganalisis dan menyelidiki kumpulan data, merangkum karakteristik utama data hingga metode visualisasi. Ini membantu Ilmuwan Data untuk menemukan Pola Data, anomali Spot, pengujian hipotesis, dan atau asumsi.



Jadi secara sederhana dapat didefinisikan sebagai metode yang membantu Data Scientist menentukan cara terbaik untuk memanipulasi sumber data yang diberikan untuk mendapatkan jawaban yang dibutuhkan sebagai tujuan.

Seberapa penting Analisis Data Eksplorasi adalah Ilmu Data

Tujuan utama ADE adalah untuk membantu melihat lebih dalam pada kumpulan data sebelum membuat asumsi apa pun, mengidentifikasi kesalahan yang jelas, mendapatkan pemahaman

yang lebih baik tentang pola di dalam kumpulan data, mencari tahu outlier dan/atau kejadian anomali, dan yang tak kalah pentingnya, untuk menemukan hubungan yang menarik antara variabel.

Analisis Data Eksplorasi sangat penting untuk Analisis Data di arena Ilmu Data. Pertama, ADE digunakan untuk memastikan hasil yang dihasilkan ilmuwan Data valid dan dapat diterapkan untuk tujuan yang diinginkan. Kedua, ADE membantu para pemangku kepentingan untuk memastikan bahwa mereka selalu mengajukan pertanyaan yang tepat. Ini juga membantu menjawab pertanyaan tentang deviasi standar, variabel kategoris, dan interval kepercayaan. Akhirnya, setelah ADE selesai dan wawasan diambil, fitur-fiturnya kemudian dapat digunakan untuk analisis atau pemodelan data yang lebih canggih, termasuk pembelajaran mesin.

Jenis Analisis Data Eksplorasi

Ada empat jenis ADE:

Non-grafis univariat:

Univariate Non Graphical adalah bentuk paling sederhana dari analisis data. di sini hanya terdiri dari satu variabel. Menjadi variabel tunggal, itu tidak berurusan dengan penyebab atau hubungan. Sebaliknya, tujuan utama dari analisis univariat adalah untuk menggambarkan data dan menemukan pola di dalamnya.

Grafis univariat

Metode non-grafis tidak dapat memberikan gambaran data yang lengkap. Oleh karena itu, metode grafis diperlukan di sini. Jenis umum grafik univariat adalah:

- *Plot batang dan daun*: Ini menunjukkan semua nilai data dan bentuk distribusi.
- *Histogram plot batang*: di mana setiap batang mewakili frekuensi (jumlah) atau proporsi (jumlah/jumlah total) kasus untuk rentang nilai.
- *Plot kotak*: secara grafis menggambarkan ringkasan lima angka dari minimum, kuartil pertama, median, kuartil ketiga, dan maksimum.

Multivariasi non-grafis

Data multivariat muncul dari lebih dari satu variabel. Umumnya, teknik ADE non-grafis multivariat menunjukkan hubungan antara dua atau lebih variabel data melalui tabulasi silang atau statistik.

Grafis Multivarian

Data multivariasi menggunakan grafik saat menampilkan hubungan antara dua atau lebih Kumpulan Data. Grafik yang paling sering digunakan adalah diagram batang atau diagram batang yang dikelompokkan dengan masing-masing grup mewakili satu level dari salah satu variabel dan setiap batang dalam grup mewakili level variabel lainnya.

Jenis umum lainnya dari grafik multivariat meliputi:

- *Scatter plot*: Digunakan untuk memplot titik data pada sumbu horizontal dan vertikal untuk menunjukkan seberapa besar satu variabel dipengaruhi oleh variabel lain.

- *Bagan multivariat*: Adalah representasi grafis dari hubungan antara faktor dan respons.
- *Run chart*: Adalah grafik garis dari data yang diplot dari waktu ke waktu.
- *Bagan gelembung*: Adalah visualisasi data yang menampilkan beberapa lingkaran (gelembung) dalam plot dua dimensi.
- *Peta panas*: Adalah representasi grafis dari data di mana nilai digambarkan dengan warna.

Eksplorasi Data Vs Penambangan Data

Dalam ilmu data, ada dua metode utama untuk mengekstraksi data dari sumber berbeda, yakni eksplorasi data dan penambangan data.

Eksplorasi data adalah proses yang dilakukan oleh pebisnis untuk memahami tren dan pola data, dan dilakukan secara lebih luas.

Sementara penambangan data atau *data mining* yaitu proses yang lebih spesifik, biasanya dilakukan oleh para profesional data. Analisis data membuat aturan dan parameter asosiasi untuk memilah kumpulan data yang sangat besar dan mengidentifikasi pola serta tren masa depan.

Mengutip situs *Techtarget*, biasanya eksplorasi data dilakukan terlebih dahulu untuk menilai hubungan antar variabel. Kemudian penambangan data dimulai. Melalui proses ini, model data dibuat untuk mengumpulkan wawasan tambahan dari data.

Cara Kerja Eksplorasi Data



Eksplorasi data biasanya dilakukan dengan 3 langkah, yakni:

1. Memahami Variabel

Dasar setiap analisi data dimulai dengan memahami variabel. Melihat lebih detail katalog data, deskripsi bidang, dan metadata dapat memberikan wawasan tentang apa yang diwakili oleh setiap

bidang dan membantu menemukan data yang hilang atau tidak lengkap.

2. Deteksi Outlier

Outlier atau anomali dapat menggagalkan analisis dan mendistorsi realitas kumpulan data, oleh karenanya penting untuk diidentifikasi sejak dini.

Visualisasi data, metode numerik, rentang interkuartil, dan pengujian hipotesis adalah cara paling umum mendeteksi *outlier*. Plot kotak, histogram, atau plot sebar, misalnya, memudahkan untuk menemukan titik-titik yang jauh di luar rentang standar, sementara skor-z dapat menginformasikan seberapa jauh titik data dari rata-rata.

Setelah ditemukan, seorang analis dapat menyelidiki, menyesuaikan, menghilangkan, atau mengabaikan *outlier*. Apa pun pilihannya, keputusan tersebut harus dicatat dalam analisis.

3. Periksa Pola dan Hubungan

Merencanakan kumpulan data dalam berbagai cara memudahkan untuk mengidentifikasi dan memeriksa pola serta hubungan antar **variabel**.



