

Lembar Latihan

Latihan 1

Anda memiliki dataset mengenai penjualan tiket di sebuah bioskop sepanjang tahun. Dataset ini memiliki beberapa kolom berikut:

- tanggal_film (Tanggal film ditayangkan)
- judul_film (Nama film)
- durasi_film (Durasi film dalam menit)
- kapasitas_auditorium (Kapasitas tempat duduk auditorium)
- tiket_terjual (Jumlah tiket film yang terjual)
- harga_tiket (Harga tiket)

Namun, dataset ini memiliki beberapa masalah: beberapa nilai hilang atau tidak sesuai, serta adanya duplikasi. Anda harus membersihkan dataset ini agar dapat digunakan untuk analisis lebih lanjut.

Latihan 2

Anda diberi dataset yang mencakup informasi seputar perjalanan pelanggan selama satu bulan, seperti:

- tanggal_waktu (Tanggal dan waktu mulai perjalanan)
- jarak (Jarak perjalanan dalam kilometer)
- durasi (Durasi perjalanan dalam menit)
- harga (Harga perjalanan)
- driver_rating (Rating yang diberikan oleh pelanggan kepada pengemudi)
- customer_rating (Rating yang diberikan oleh pengemudi kepada pelanggan)

Anda diminta untuk membuat visualisasi yang efektif dari data tersebut untuk mengeksplorasi pola dan hubungan antara fitur-fiturnya.

Latihan 3

Anda bekerja di divisi Human Resources sebuah perusahaan dan diberi dataset yang mencakup informasi tentang karyawan, termasuk:

- umur (Umur karyawan)
- jenis_kelamin (Jenis kelamin karyawan)
- pendidikan (Tingkat pendidikan karyawan)
- lama_bekerja (Lama bekerja di perusahaan dalam tahun)
- gaji (Gaji karyawan)

Tugas Anda adalah menggunakan statistik deskriptif dan inferensial untuk menganalisis dataset ini dan memahami pola dan hubungan antara fitur-fiturnya.

Latihan 4

Anda diberi dataset mengenai pelanggan sebuah perusahaan kartu kredit. Dataset ini mencakup informasi seperti:

- pelanggan_id (ID unik untuk setiap pelanggan)
- usia (Usia pelanggan)
- jenis_kelamin (Jenis kelamin pelanggan)
- pendapatan (Pendapatan tahunan pelanggan)
- jml_kartu_kredit (Jumlah kartu kredit yang dimiliki pelanggan)
- pengeluaran_bulanan (Pengeluaran bulanan rata-rata pelanggan)

Anda diminta untuk membuat model machine learning yang akan memprediksi apakah seorang pelanggan akan tertarik pada penawaran kartu kredit baru. Anda harus menyiapkan lembar kerja yang meliputi proses pengembangan model, termasuk pemilihan fitur, pelatihan, validasi, dan evaluasi.

Latihan 5

Kami telah menyediakan contoh data produk di Sephora. Dari data ini, kami ingin tahu wawasan apa yang bisa Anda dapatkan. Beberapa bagian dari wawasan harus terkait dengan model pembelajaran mesin dan visualisasi data.

Data: <https://binapertiwi.link/filesdatascientisttest>

Semua pekerjaan Anda harus diupload ke repositori Git Anda sendiri, termasuk hasil dari wawasan Anda dalam format apa pun seperti Power Point, PDF, dll.