

Ujian Tengah Semester

Note

- Sebelum membaca dan mengerjakan soal awalilah dengan berdoa dan membaca basmallah.
- Kerjakanlah soal-soal di bawah ini dengan **JUJUR DAN SPORTIF** !
- Bacalah seluruh soal dengan seksama
- Pilihlah soal yang lebih mudah untuk dikerjakan
- Dataset dapat didownload [disini](#)
- Gunakan RMarkdown untuk menjawab soal ini

Soal 1

Masalah

Sarah adalah seorang **manajer penjualan** di wilayah tertentu untuk perusahaan yang menjual bahan bakar minyak untuk pemanas rumah. Akhir-akhir ini, **hasil pemasaran semakin buruk, padahal biaya untuk kegiatan pemasaran justru semakin besar**. Sarah merasa perlu mencari tahu apa saja perilaku dan **faktor yang bisa memengaruhi permintaan masyarakat terhadap minyak pemanas rumah**. Ia tahu bahwa banyak hal yang bisa memengaruhi seberapa banyak orang menggunakan minyak pemanas. Sarah percaya bahwa dengan mempelajari hubungan antara faktor-faktor tersebut, ia bisa lebih mudah memantau dan memprediksi permintaan minyak pemanas, serta bisa membuat strategi pemasaran yang lebih tepat di masa depan.

Tujuan

Mempelajari hubungan antara berbagai faktor yang memengaruhi jumlah konsumsi minyak pemanas rumah.

Untuk menjawab pertanyaannya, Sarah meminta bantuan untuk membuat correlation matrix dari enam atribut yang diduga saling berhubungan. Dengan menggunakan data milik perusahaan yang sebagian besar berasal dari database konsumen, dataset yang terdiri dari variabel-variabel berikut:

1. **Insulation:** Skor kepadatan insulasi (peredam panas) rumah, dengan skala 1 sampai 10. Skor 1 berarti insulasi sangat buruk, sedangkan skor 10 berarti insulasi sangat baik.
2. **Temperature:** Rata-rata suhu udara luar di sekitar rumah selama satu tahun terakhir, diukur dalam derajat Fahrenheit.
3. **Heating_Oil:** Jumlah total minyak pemanas yang dibeli oleh pemilik rumah selama satu tahun terakhir.
4. **Num_Occupants:** Jumlah total orang yang tinggal di dalam rumah tersebut.
5. **Avg_Age:** Rata-rata usia para penghuni rumah.
6. **Home_Size:** Ukuran rumah, diberi nilai dari 1 sampai 8. Semakin tinggi nilainya, semakin besar ukuran rumah.

Ikuti langkah-langkah berikut untuk menyelesaikan masalah ini:

Langkah 1 Buat Correlation Matrix

- Gunakan data yang terdiri dari atribut berikut: **Heating_Oil**, **Insulation**, **Temperature**, **Num_Occupants**, **Avg_Age**, dan **Home_Size**.
- Tampilkan correlation matrix dalam bentuk tabel dan/atau visualisasi seperti **heatmap** agar mudah dianalisis.
- Tujuan: melihat hubungan antar variabel, khususnya yang berhubungan dengan **Heating_Oil**.

Langkah 2 Interpretasi Correlation Matrix

- Jelaskan hasil correlation matrix Sebutkan dan jelaskan variabel mana saja yang memiliki korelasi signifikan (baik positif maupun negatif) dengan jumlah konsumsi **Heating_Oil**.
- Gunakan bahasa sederhana agar mudah dimengerti, serta sertakan kesimpulan.

Langkah 3 Scatter Plot 1

- Buat scatter plot dengan sumbu X adalah **Heating_Oil** dan sumbu Y adalah **Avg_Age** (usia rata-rata penghuni rumah).

- Gunakan warna (color) pada titik-titik scatter plot berdasarkan intensitas nilai Heating_Oil (misalnya menggunakan gradasi warna atau color scale).
- Tujuan: melihat apakah ada pola antara usia penghuni dan konsumsi minyak pemanas.
- Jelaskan hasil dari Scatter Plot 1

Langkah 4: Scatter Plot 2

- Buat scatter plot dengan sumbu X adalah Temperature dan sumbu Y adalah Insulation.
- Gunakan warna (color) pada titik-titik scatter plot berdasarkan intensitas nilai Heating_Oil (misalnya menggunakan gradasi warna atau color scale).
- Tujuan: melihat apakah suhu lingkungan dan kualitas insulasi rumah memengaruhi jumlah minyak yang dikonsumsi.
- Jelaskan hasil dari Scatter Plot 2

Langkah 5 Rekomendasi untuk Sarah

Berdasarkan hasil analisis korelasi dan visualisasi scatter plot, berikan rekomendasi strategi pemasaran kepada Sarah. Misalnya, target pemasaran yang lebih tepat, penawaran promosi ke rumah tertentu, atau pendekatan komunikasi kepada pelanggan dengan karakteristik tertentu. Rekomendasi harus logis dan relevan dengan temuan data.

SOAL 2

Prediksi Jumlah Konsumsi Minyak Pemanas Rumah Sarah, yang sebelumnya sukses dengan bantuan saran yang anda berikan, kini sarah **dipromosikan menjadi VP Marketing**. Ia bertanggung jawab atas ratusan marketer yang tersebar di berbagai wilayah. Salah satu tantangan yang dihadapinya adalah **Data HeatingOil.csv bersifat rahasia dan hanya dapat diakses oleh Sarah (VP)**. Para marketer tidak boleh langsung melihat dataset tersebut, namun tetap diminta untuk memperkirakan kebutuhan minyak calon pelanggan baru.

Sarah ingin masing-masing marketer membuat proses yang dapat mengestimasi kebutuhan konsumsi minyak dari client yang mereka approach, dengan menggunakan model yang sebelumnya dihasilkan oleh Sarah, **meskipun tanpa mengakses dataset (HeatingOil.csv)**

Asumsikan bahwa data **HeatingOil-Marketing.csv** adalah data calon pelanggan yang berhasil di approach oleh salah satu marketingnya. Untuk mengatasi ini, Sarah ingin

membuat model regresi prediktif, menyimpannya, dan mendistribusikan model tersebut ke semua marketer.

Yang harus dilakukan Sarah adalah membuat proses untuk:

- Membaca dataset: `HeatingOil.csv`
- Membuat **model regresi** untuk memprediksi konsumsi minyak (`Heating_Oil`).
- **Menyimpan model** agar dapat digunakan oleh marketer tanpa harus membagikan data `HeatingOil.csv`.

Yang harus dilakukan Marketer adalah membuat proses untuk:

- Membaca **model** yang sudah disimpan oleh Sarah
- Membaca data calon pelanggan baru: `HeatingOil-Marketing.csv`
- **Menerapkan model** ke data calon pelanggan baru untuk memprediksi kebutuhan minyak pemanas rumah