

Soal Kompetisi: Prediksi dan Kategorisasi Perubahan Harga Bahan Pokok di DKI Jakarta

Tema: Perdagangan

Jenis Kasus: Regresi dan Clustering

Latar Belakang

Setiap harinya harga-harga bahan pokok yang mencakup beras, gula, garam, bumbu dapur, dll. mengalami kenaikan dan penurunan, tergantung pasar atau hukum permintaan dan penawaran, serta musiman menjelang hari raya. Tidak terkecuali di ibukota, DKI Jakarta, harga-harga bahan pokok juga mengalami hal yang serupa.

Dataset

Anda disediakan dataset yang dapat diunduh dari tautan public berikut:

<http://data.jakarta.go.id/dataset/data-harga-bahan-pokok-di-provinsi-dki-jakarta-tahun-2018>

Data terdiri dari 12 file berbentuk csv dari bulan Januari 2018 hingga Desember 2018 berupa harga-harga harian bahan pokok. Data masih belum 100% bersih dan rapih, sehingga dibutuhkan cleansing dan formatting.

Data dictionary dari dataset tersebut adalah sbb:

tanggal : Tanggal harian

komoditi : Komoditi

harga : Harga

Aturan main:

Sebelum melakukan modeling, peserta diharapkan dapat membersihkan data, memvisualkan data, dan melakukan *exploratory data analysis* sehingga mampu memberikan *insight-insight* yang minimal dapat menjawab pertanyaan-pertanyaan berikut,

1. Apakah terlihat ada pengaruh musim (*seasonality*) terhadap perubahan harga komoditi terhadap waktu apabila digunakan dataset satu tahun 2018 saja?
2. Jika ada pengaruh musim, komoditi manakah yang sangat terpengaruh?
3. Perubahan harga paling tinggi terjadi saat kapan?
4. Adakah event yang secara bersamaan terjadi perubahan harga paling besar?
5. Mengapa terjadi hal yang demikian? Adakah kejadian bersamaan yang dapat dianalisis dari beberapa portal berita mempengaruhi harga?

Setelah itu Anda diminta memilih dan mengerjakan **salah satu** dari topik berikut:

1. **Model** prediksi **salah satu** harga komoditi bahan pokok (misalkan beras medium) apabila hanya diketahui harga-harga bahan pokok lainnya.
2. **Model clustering** untuk mengkategorisasi **seluruh** komoditi bahan pokok berdasarkan **perubahan** harga bahan pokok sepanjang 2018.

Peserta mengumpulkan Jupyter Notebook dan presentasi

1. Jupyter notebook file dengan bahasa pemrograman Python (file dengan ekstensi .ipynb)
2. Presentasi (file dengan ekstensi pptx atau ppt)
3. Seluruh file digabungkan dan di zip
4. Dikumpulkan melalui email ds-academy@compfest.id
5. Deadline pengumpulan file pada tanggal 28 Juli 2019 pukul 07.00 WIB

Kriteria penilaian

1. **Code.** Kode mudah dibaca dan rapi, termasuk dalam melakukan *data cleansing* dan *data formatting* dan visualisasi.
2. **Creativity.** Kreativitas pada pengerjaan *data cleansing*, *formatting*, *feature engineering*, penyeleksian tipe model, serta proses bagaimana melakukan validasi silang dan pembagian data training dan testing (khususnya untuk model #1).
3. **Performansi model.** Performansi model disarankan menggunakan *Root Mean Square Error* (RMSE) atau *Rsquare* untuk model #1, sedangkan untuk model *clustering* disarankan menggunakan *Dunn's index* atau *similarity matrix*.
4. **Presentasi.** Presentasi dinilai dari sistematika dan pemaparan insight, kejelasan visualisasi data, dan pemaparan model yang disajikan.