

SPESIFIKASI DATASET PENYISIHAN DATAVIDIA 9.0

Prediksi Harga Pangan
Setiap Provinsi di Indonesia



Dipersiapkan Oleh
Tim Problem Setter Datavidia 9.0
2025



TUJUAN PENYISIHAN

Tahap penyisihan bertujuan untuk menilai kemampuan peserta sebelum melangkah ke babak final. Dalam tahap ini, peserta diuji dalam dua aspek utama: ketepatan prediksi model dan kemampuan mengomunikasikan hasil analisis secara tertulis. Selain sebagai seleksi, tahap ini juga menjadi sarana pembelajaran bagi peserta dalam mengasah keterampilan analisis data dan penyampaian informasi secara efektif.

Pada tahap pertama penyisihan, peserta diminta untuk membuat model prediksi berdasarkan dataset yang telah disediakan. Penilaian dalam tahap ini akan sangat bergantung pada ketepatan prediksi yang dihasilkan oleh model peserta, sehingga peserta harus mampu menerapkan berbagai teknik analisis dan pemrosesan data untuk mendapatkan hasil terbaik. Peserta ditantang untuk tidak hanya memahami struktur dan karakteristik data, tetapi juga mengeksplorasi berbagai strategi pemodelan, rekayasa fitur, serta teknik optimasi yang dapat meningkatkan performa prediksi. Selain itu, kreativitas dalam memilih metode yang tepat dan efisiensi dalam mengolah data menjadi aspek penting yang turut dinilai dalam tahap ini.

Setelah itu, peserta juga akan menyusun laporan hasil temuan dan analisis terhadap data. Peserta diharapkan mampu menyusun temuannya dalam bentuk tulisan yang jelas, terstruktur, dan informatif. Penulisan ini bertujuan untuk menguji sejauh mana peserta memahami hasil analisis yang telah dilakukan serta bagaimana peserta menyampaikan wawasan yang diperoleh dari eksplorasi data dan pemodelan. Kemampuan dalam mengorganisasi informasi, menjelaskan metodologi, serta menarik kesimpulan yang relevan menjadi faktor utama dalam penilaian tahap ini.

Dengan adanya dua tahap penyisihan ini, kompetisi tidak hanya berfokus pada aspek teknis dalam membangun model prediktif, tetapi juga mendorong peserta untuk mengembangkan keterampilan komunikasi ilmiah. Dengan demikian, peserta yang berhasil melaju ke babak final adalah peserta yang tidak hanya unggul dalam analisis data, tetapi juga mampu menyampaikan temuannya dengan baik, sebagaimana yang dibutuhkan dalam dunia nyata dalam bidang data science.



LATAR BELAKANG

Fluktuasi harga bahan pangan merupakan tantangan signifikan dalam perekonomian Indonesia, memengaruhi berbagai sektor mulai dari industri hingga rumah tangga. Ketidakstabilan harga ini dapat berdampak pada kebijakan ekonomi, rantai pasokan, dan daya beli masyarakat. Sebagai contoh, usaha kuliner dan katering sering kali harus menetapkan harga tetap bagi konsumen, sementara harga bahan baku yang tidak menentu menimbulkan risiko finansial yang signifikan.

Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa pada Februari 2024, inflasi year-on-year (y-on-y) mencapai 2,75%, dengan kelompok makanan, minuman, dan tembakau mengalami kenaikan sebesar 6,36%. Komoditas seperti beras, cabai merah, dan cabai rawit menjadi penyumbang utama inflasi ini. Kenaikan harga beras, sebagai makanan pokok, sangat berdampak pada pengeluaran rumah tangga Indonesia.

Selain itu, nilai tukar rupiah terhadap dolar AS juga memengaruhi harga pangan. Pada Desember 2024, depresiasi rupiah mencapai sekitar 5,9% dibandingkan tahun sebelumnya. Depresiasi ini menyebabkan kenaikan harga bahan pangan impor seperti kedelai dan gandum yang berdampak pada produk olahannya seperti tahu, tempe, dan roti.

Dalam menghadapi tantangan ini, analisis data dan peramalan harga pangan menjadi penting. Menurut laporan McKinsey, perusahaan yang menerapkan manajemen rantai pasokan berbasis data mengalami pengurangan biaya logistik sebesar 15% dan peningkatan tingkat layanan hingga 20%.

Untuk menjawab tantangan tersebut, Penyisihan Datavidia 9.0 mengadakan kompetisi peramalan harga bahan pangan berbasis data. Tugas peserta adalah membangun model prediksi yang akurat dan memberikan wawasan bagi pengambilan keputusan strategis oleh pelaku industri dan pembuat kebijakan. Melalui kompetisi ini, diharapkan peserta dapat mengembangkan pemahaman mendalam tentang analisis data dalam konteks nyata dan berkontribusi dalam menciptakan solusi berbasis data untuk mengatasi permasalahan sosial dan ekonomi akibat fluktuasi harga pangan.

DESKRIPSI TUGAS

Pada penyisihan ini, Anda diminta untuk melakukan peramalan deret waktu harga bahan pangan untuk setiap provinsi. Selain meramal, Anda juga perlu membuat laporan hasil penggerjaan Anda seputar analisis, metode yang digunakan, dan sebagainya sesuai pada Guidebook Datavidia 9.0.



Sumber:

<https://www.liveplan.com/blog/forecasting/what-are-the-qualities-that-make-a-good-financial-forecast>

Secara keseluruhan, terdapat 40367 baris yang harus Anda prediksi. Itu merupakan gabungan dari 13 jenis bahan pangan, 34 provinsi, dan 92 hari. Metrik yang digunakan untuk mengukur galat pada tugas kali ini adalah mean absolute percentage error (MAPE). Formula MAPE dapat dirumuskan sebagai berikut.

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|$$

```
import numpy as np

def mape(y_true: np.ndarray, y_pred: np.ndarray) -> float:
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
```

MAPE merupakan metrik yang cukup umum digunakan pada kasus peramalan deret waktu. Metrik ini mengukur seberapa besar deviasi prediksi dan *forecast* secara proporsional. Hal ini dapat membantu kita untuk tidak perlu memikirkan perbedaan skala dari masing-masing harga.

Anda dapat melakukan submisi di Kaggle dengan jumlah maksimal submisi sebanyak 5 kali.

PENJELASAN DATASET

Bagian ini merupakan bagian penjelasan setiap dataset, mulai dari sampai dengan metadata. Bagian ini diharapkan dapat memberikan informasi yang mencukupi sehingga Anda dapat memahami informasi/konteks dari dataset dan memperkuat kredibilitas dataset itu sendiri.

Harga Bahan Pangan



Harga Bahan Pangan adalah dataset yang disiapkan oleh Badan Pangan Nasional. Badan Pangan Nasional (Bapanas) atau *National Food Agency* (NFA) merupakan lembaga pemerintah yang berada di bawah dan bertanggung jawab kepada Presiden dalam pelaksanaan urusan pangan untuk menciptakan kedaulatan pangan, ketahanan pangan, dan kemandirian pangan bagi negara.

Dataset ini merupakan dataset deret waktu yang diambil dari tanggal 1 Januari 2022 sampai dengan 31 Desember 2024. Secara keseluruhan, dataset ini berukuran $13 \times 34 \times 1096 = 484432$ sel. Kami melakukan pemisahan data train dan data test dengan tanggal pertama pada data test adalah 1 Oktober 2024. Dapat diperoleh informasi tambahan bahwa data train berisi $13 \times 34 \times 1004 = 443768$ sel dan data test berisi $13 \times 34 \times 92 = 40664$ sel.

Secara ringkas, terdapat 13 jenis bahan pangan yang harus diprediksi pada 34 provinsi untuk setiap deret waktu.



Google Trend

Google Trends

Dataset Google Trend—sesuai namanya—kami dapatkan dari Laman Google Trends. Google Trends menyediakan akses ke sampel permintaan pencarian aktual yang sebagian besar tidak difilter yang dibuat ke Google. Data tersebut dianonimkan (tidak ada yang diidentifikasi secara pribadi), dikategorikan (menentukan topik untuk permintaan pencarian) dan diagregasi (dikelompokkan bersama). Hal ini memungkinkan untuk menampilkan minat pada topik tertentu dari seluruh dunia atau hingga geografi tingkat kota.

Meskipun hanya sampel pencarian Google yang digunakan di Google Trends, ini sudah cukup karena Google menangani miliaran pencarian setiap hari. Memberikan akses ke seluruh dataset akan terlalu besar untuk diproses dengan cepat. Dengan mengambil sampel data, Google dapat melihat dataset yang representatif dari semua pencarian Google, sambil menemukan wawasan yang dapat diproses dalam hitungan menit setelah suatu peristiwa terjadi di dunia nyata.

Google Trends menormalisasi data pencarian untuk mempermudah perbandingan antar istilah. Hasil pencarian dinormalisasi berdasarkan waktu dan lokasi kueri dengan proses berikut: Setiap titik data dibagi dengan total pencarian di geografi dan rentang waktu yang diwakilinya untuk membandingkan popularitas relatif. Jika tidak, tempat dengan volume pencarian tertinggi akan selalu menduduki peringkat teratas. Angka-angka yang dihasilkan kemudian disesuaikan dalam rentang 0 hingga 100 berdasarkan proporsi topik terhadap semua pencarian di semua topik. Berbagai wilayah yang menunjukkan minat pencarian yang sama untuk suatu istilah tidak selalu memiliki volume pencarian total yang sama.

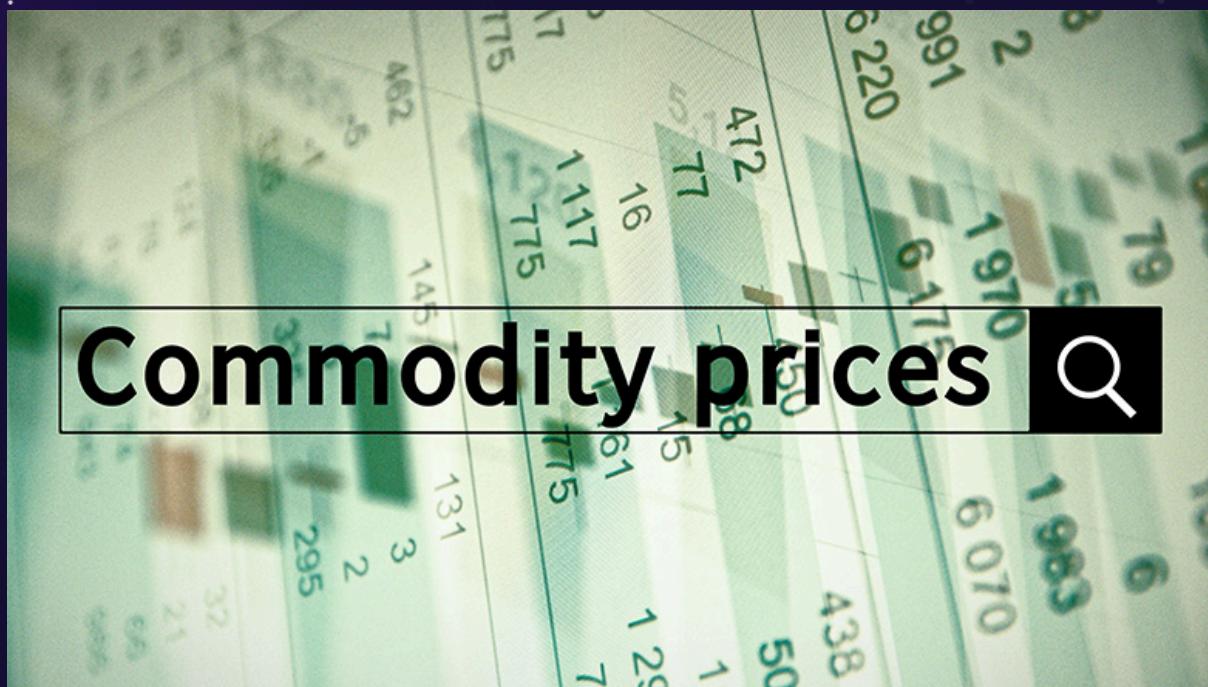
Secara garis besar, dataset ini berukuran $13 \times 35 \times 1004 = 456820$ sel yang mungkin dapat membantu Anda dalam melakukan peramalan harga bahan pangan.

Mata Uang



Dataset ini berasal dari Yahoo Finance—dengan ICE sebagai data provider—yang berisi berbagai jenis mata uang yang ada di dunia. Kami hanya menyediakan beberapa mata uang saja, seperti MYRUSD, SGDUSD, THBUSD, dan USDIDR. Secara garis besar pada mata uang, terdapat $4 \times 5 \times 716 = 14320$ sel yang kami sediakan.

Harga Komoditas Global



Dataset ini kami dapatkan dari Investing.com. Kami menyediakan 6 jenis data harga komoditas sebagai berikut.

- Crude Oil WTI Futures—data perdagangan minyak mentah WTI yang diperdagangkan di pasar global.
- Natural Gas Futures—mencakup data historis gas alam.
- Newcastle Coal Futures—mencakup data batu bara termal yang dieksport (dikirim FOB) dari pelabuhan Newcastle di New South Wales, Australia.
- Palm Oil Futures—mencakup data harga minyak mentah sawit
- US Sugar #11 Futures—mencakup harga gula #11 di US.
- US Wheat Futures—mencakup harga gandum di US.

Untuk setiap dataset, terdapat kolom Price, Open, High, Low, Vol., dan Change % yang berisi penjelasan sebagai berikut.

- Price — Harga terakhir atau harga penutupan terbaru dari aset pada periode tertentu.
- Open — Harga pembukaan aset di awal periode perdagangan.
- High — Harga tertinggi yang dicapai oleh aset dalam periode tertentu.
- Low — Harga terendah yang dicapai oleh aset dalam periode tertentu.
- Vol. — Volume perdagangan, menunjukkan jumlah aset yang diperdagangkan dalam periode tertentu.
- Change % — Persentase perubahan harga dibandingkan dengan periode sebelumnya, menunjukkan kenaikan atau penurunan harga dalam bentuk persentase.



Tentang Futures

Dalam kontrak berjangka, diperlukan perhitungan khusus untuk menghitung *cost of carry*. Hal ini disebabkan karena kontrak berjangka mencakup biaya penyimpanan (*storage costs*) dari komoditas yang bersangkutan, serta tingkat suku bunga bebas risiko (*risk-free interest rate*) – yaitu tingkat pengembalian dari investasi yang tidak memiliki risiko kerugian finansial. Konsep ini bersifat hipotetis, karena semua perdagangan dan investasi selalu mengandung risiko kerugian, meskipun sangat kecil. Namun, untuk keperluan praktis, tingkat pengembalian dari obligasi pemerintah berisiko rendah sering digunakan.

Dalam perhitungan berikut, istilah *convenience yield* mengacu pada premi yang terkait dengan kepemilikan fisik suatu komoditas, bukan produk derivatif atau kontrak terkait, sebagai bagian dari kontrak berjangka.

$$F(t) = S(t)e^{(r+s-c)\times t}$$

Dengan

- $S(t)$ adalah nilai masa kini (present value)
- $F(t)$ adalah harga futures
- r adalah bunga bebas risiko (risk-free rate)
- s adalah biaya penyimpanan (persentase terhadap harga spot)
- c adalah yield konvensional
- t adalah waktu penyimpanan dalam fraksi

Sebagai contoh, misalkan harga spot emas adalah \$1.200 per ons dan biaya penyimpanan emas sebesar \$5 per ons untuk enam bulan. Kontrak futures enam bulan (0.5) untuk emas, dengan tingkat bunga bebas risiko 0,25%, adalah \$1.206,51, atau

$$1206.51 = (1200 + 5)e^{(0.0025 \times 0.5)}$$

Harga futures komoditas tidak selalu lebih tinggi daripada harga spot. Harga futures memperhitungkan ekspektasi pasokan dan permintaan serta tingkat produksi, di antara faktor lainnya. Selisih antara harga spot dan harga futures pada waktu tertentu disebabkan oleh biaya penyimpanan dan tingkat bunga.

NARAHUBUNG

Muhammad Fathur Rizky

WA: 0895392152737

LINE: fathurwithyou

Mohamad Maulana Firdaus R.

WA: 087728997015

LINE: maull04

Email

competition@arkavidia.com