Prediksi Harga Komoditas Bahan Pangan

1Tiffany Anastasia Jocelyn   
*IPB University*  
tiffanyanastasia@apps.ipb.ac.id2Nurul Fadillah   
*IPB University*  
nurulsacad.arc@gmail.com  
3Anisa Hayatullah   
*IPB University*  
annanisa@apps.ipb.ac.id

# Pendahuluan

## Latar Belakang Masalah

Fluktuasi harga bahan pangan merupakan tantangan signifikan dalam perekonomian Indonesia yang berdampak luas terhadap berbagai sektor, mulai dari industri hingga rumah tangga. Ketidakstabilan harga ini dapat mempengaruhi kebijakan ekonomi, rantai pasokan, serta daya beli masyarakat.

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 memiliki dampak signifikan terhadap pengeluaran rumah tangga Indonesia.

Selain faktor domestik, nilai tukar rupiah terhadap dolar AS juga berperan dalam menentukan harga bahan pangan, terutama yang bergantung pada impor. 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 pada akhirnya berdampak pada produk olahannya seperti tahu, tempe, dan roti.

Dalam menghadapi tantangan ini, analisis data dan peramalan harga pangan menjadi semakin penting. Studi menunjukkan bahwa penerapan manajemen rantai pasokan berbasis data dapat membantu mengurangi biaya logistik hingga 15% serta meningkatkan efisiensi layanan hingga 20%. Oleh karena itu, pengembangan model prediksi harga bahan pangan yang akurat dapat menjadi solusi strategis bagi pelaku industri dan pembuat kebijakan dalam mengantisipasi serta mengelola dampak fluktuasi harga pangan.

Dengan mempertimbangkan urgensi permasalahan ini, dilakukan penelitian untuk membangun model prediksi harga bahan pangan yang berbasis data. Penelitian ini bertujuan untuk mengembangkan pemahaman mendalam tentang pola fluktuasi harga, serta memberikan wawasan yang dapat mendukung pengambilan keputusan strategis dalam sektor ekonomi dan ketahanan pangan.

## Deskripsi Masalah

Fluktuasi harga bahan pangan menjadi tantangan signifikan dalam perekonomian Indonesia. Ketidakstabilan ini mempengaruhi berbagai sektor, mulai dari industri hingga rumah tangga, serta berdampak pada kebijakan ekonomi, rantai pasokan, dan daya beli masyarakat. Usaha kuliner dan industri makanan, misalnya, sering kali harus menghadapi ketidakpastian harga bahan baku yang dapat meningkatkan risiko finansial.

Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa inflasi kelompok makanan, minuman, dan tembakau mencapai 6,36% pada Februari 2024, dengan komoditas utama seperti beras, cabai merah, dan cabai rawit sebagai penyumbang utama kenaikan harga. Selain itu, depresiasi rupiah terhadap dolar AS sebesar 5,9% pada akhir 2024 turut berkontribusi pada kenaikan harga bahan pangan impor seperti kedelai dan gandum. Kondisi ini semakin menambah tekanan terhadap industri makanan dan daya beli masyarakat.

Dalam menghadapi ketidakpastian harga pangan, analisis data dan peramalan harga menjadi penting untuk membantu pelaku industri dan pembuat kebijakan dalam mengambil keputusan strategis. Dengan adanya model prediktif yang akurat, risiko akibat volatilitas harga pangan dapat dikurangi, sehingga stabilitas pasokan dan harga dapat lebih terjaga.

## Tujuan Analisis

1. Menganalisis tren harga bahan pangan berdasarkan data historis untuk memahami pola perubahan harga.
2. Mengembangkan model prediksi harga pangan guna memperkirakan pergerakan harga dalam jangka pendek maupun panjang.
3. Mengevaluasi faktor-faktor yang berkontribusi terhadap fluktuasi harga pangan, termasuk inflasi, nilai tukar mata uang, serta dinamika pasar global dan domestik.
4. Memberikan wawasan bagi pelaku industri dan pembuat kebijakan dalam menyusun strategi mitigasi risiko terhadap volatilitas harga pangan.
5. Meningkatkan efisiensi manajemen rantai pasokan melalui pendekatan berbasis data untuk mengurangi dampak lonjakan harga terhadap konsumen dan bisnis.

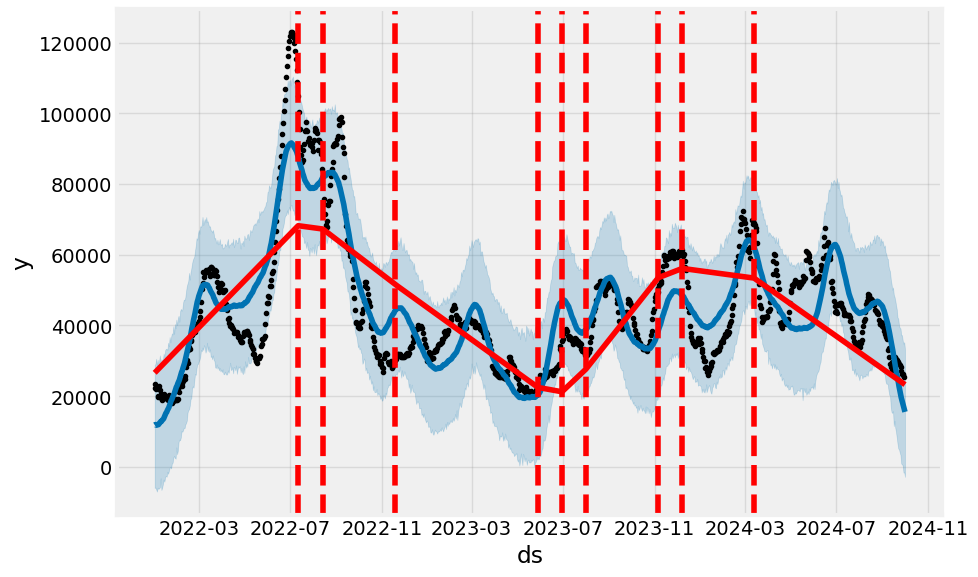
Dengan analisis ini, diharapkan solusi berbasis data dapat diterapkan untuk mengatasi tantangan yang timbul akibat fluktuasi harga pangan, sehingga stabilitas ekonomi dan kesejahteraan masyarakat dapat lebih terjaga.

# Metode Analisis

## Exploratory Data Analysis

Dalam penelitian ini, eksplorasi awal terhadap data dilakukan dengan visualisasi data deret waktu untuk memahami pola dan karakteristik data. Harga bahan pangan cabai merah keriting pada provinsi Aceh dari 1 Januari 2022 sampai 30 September 2024 memiliki tren harga yang tidak merata karena terdapat perubahan secara tiba-tiba (*abrupt change*) di beberapa titik. *Changepoint* dapat digunakan untuk membantu memprediksi perubahan mendadak yang signifikan pada data *time series*. Pendeteksian *changepoint* dapat dilakukan menggunakan *library* Prophet yang disediakan oleh Facebook. Prophet secara otomatis mendeteksi titik-titik ini untuk menangkap perubahan pola dalam data historis. Hasil deteksi *changepoint* dapat dilihat pada Gambar 1. Perubahan signifikan yang tiba-tiba ditandai oleh garis merah putus-putus.

1. Visualisasi changepoint pada data



Pada Gambar 1, tidak seluruh *changepoint* yang terdeteksi diikutsertakan ke dalam hasil visualisasi. Pada gambar, jumlah garis merah putus-putus ada sebanyak 9. Sementara itu, jumlah *changepoint* sesungguhnya yang terdeteksi pada data ada sebanyak 25 titik seperti yang dapat dilihat pada Tabel 1. Titik lonjakan perubahan mendadak yang paling ekstrem terdapat pada 12 Juli 2022. Artinya, terdapat perubahan yang signifikan pada harga cabai merah keriting di hari tersebut.

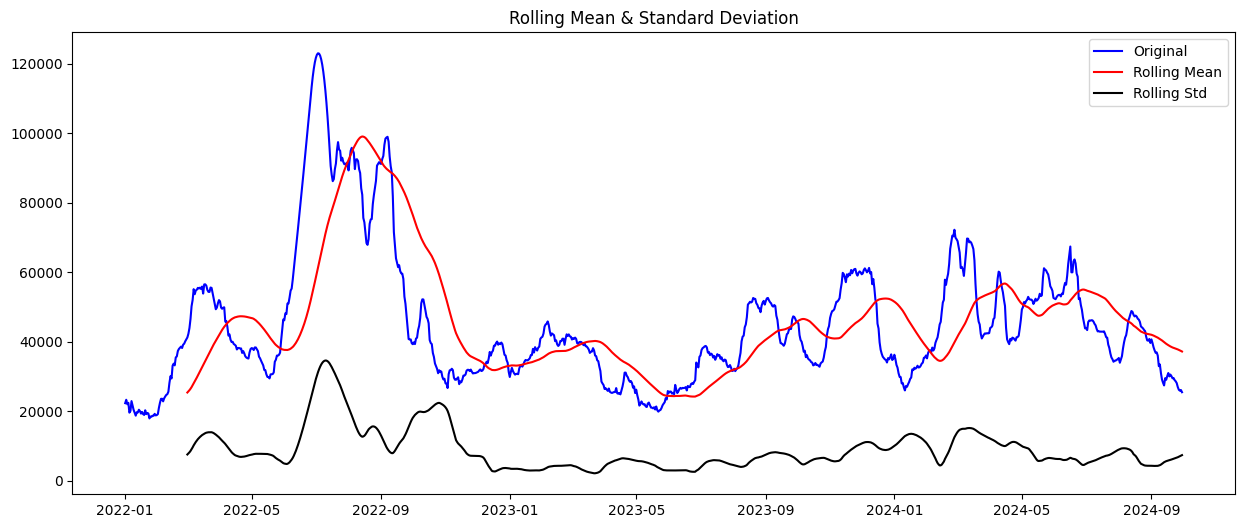
1. Changepoint yang terdeteksi pada data

| **Index** | **Date (YYYY-MM-DD)** |
| --- | --- |
| 32 | 2022-02-02 |
| 64 | 2022-03-06 |
| 96 | 2022-04-07 |
| 128 | 2022-05-09 |
| 160 | 2022-06-10 |
| 192 | 2022-07-12 |
| 225 | 2022-08-14 |
| 257 | 2022-09-15 |
| 289 | 2022-10-17 |
| 321 | 2022-11-18 |
| 353 | 2022-12-20 |
| 385 | 2023-01-21 |
| 417 | 2023-02-22 |
| 449 | 2023-03-26 |
| 481 | 2023-04-27 |
| 513 | 2023-05-29 |
| 545 | 2023-06-30 |
| 577 | 2023-08-01 |
| 610 | 2023-09-03 |
| 642 | 2023-10-05 |
| 674 | 2023-11-06 |
| 706 | 2023-12-08 |
| 738 | 2024-01-09 |
| 770 | 2024-02-10 |
| 802 | 2024-03-13 |

Eksplorasi data dilanjut dengan tahapan pengecekan *stationarity* data. *Stationarity* adalah komponen yang penting pada analisis *time series*. *Stationarity* merujuk kepada kondisi di mana *statistical properties* (sifat statistik) suatu data *time series*, seperti nilai rata-rata, varians, dan *autocorrelation*, konsisten dari waktu ke waktu. Data yang stasioner dapat memudahkan dalam membedakan komponen *trend* (pola jangka panjang dalam data)dan *seasonality* (pola yang berulang pada interval waktu tertentu).

Untuk menguji apakah data bersifat stasioner, dilakukan uji statistik Dickey-Fuller (ADF *test*). Berdasarkan hasil tes ADF, dapat disimpulkan apakah data bersifat stasioner dengan melihat nilai *p-value* dan *test statistic* dibandingkan dengan *critical value*. Dalam menguji *stationary* data, dibangun sebuah fungsi yang dapat menghitung *rolling mean* dan *rolling standard deviation* yang membantu melihat apakah rata-rata dan varians data berubah seiring waktu dan menampilkan uji statistik untuk menarik kesimpulan ADF *test*. Pada harga komoditas cabai merah provinsi Aceh, hasil visualisasi data, *rolling mean*, dan *rolling standard deviation* dapat dilihat pada Gambar 2.

1. Analisis ADF tes pada data



Uji statistik mendapatkan hasil *p-value* = 0.011783. Dalam hal ini, *p-value* bernilai kurang dari 0.05, sehingga hipotesis nol (H0) dapat ditolak, artinya data stasioner. Selain melihat dari nilai *p-value*, kestasioneran pada data dapat disimpulkan dengan membandingkan nilai *test statistic* dan *critical value*. ADF *test* menghasilka nilai *tes statistic* sebesar -3.376604 dan *critical value* 1% sebesar -3.436966, *critical value* 5% sebesar -2.864461, dan *critical value* 10% sebesar -2.568325. *Test statistic* lebih kecil dari *critical value* 5% (-3.376604 < -2.864461), yang berarti H₀ dapat ditolak dengan tingkat signifikansi 5%, mengindikasikan bahwa data sudah cukup stasioner.

## Data Preprocessing

1. Menangani Data yang Hilang pada Harga Komoditas

Dalam penelitian ini, data harga komoditas pertanian dikompilasi dari berbagai sumber terpercaya. Namun, terdapat beberapa nilai yang hilang (*missing values*) pada periode waktu dan jenis komoditas tertentu. Secara keseluruhan, sekitar 5% dari total data harga komoditas mengalami missing values, dengan persentase yang bervariasi tergantung pada jenis komoditas dan periode waktu. Berdasarkan informasi yang tersedia mengenai proses pengambilan data, nilai yang hilang ini diasumsikan disebabkan oleh kendala teknis selama proses scraping, seperti kegagalan dalam memuat konten dinamis (*dynamic content loading*) atau perubahan struktur halaman situs web yang menyebabkan data tidak berhasil diekstraksi. Selain itu, beberapa data harga mungkin tidak tersedia akibat faktor musiman atau ketersediaan produk yang terbatas di pasar. Oleh karena itu, dalam penelitian ini diasumsikan bahwa data hilang mengikuti pola *Missing at Random* (MAR), karena kehilangan data dapat dijelaskan oleh variabel-variabel yang dapat diamati selama proses pengambilan data, meskipun proses scraping tersebut dilakukan oleh pihak lain.

Solusi yang dipilih untuk mengatasi masalah ini metode imputasi dalam penelitian ini adalah Bayesian Imputation. Pendekatan Bayesian dipilih karena kemampuannya dalam menangani ketidakpastian dan memanfaatkan informasi historis untuk memperkirakan nilai yang hilang. Penelitian ini memanfaatkan library PyMC dari Python sebagai alat imputasi berbasis pendekatan Bayesian. PyMC memungkinkan penanganan data hilang secara alami dengan menganggap nilai-nilai yang hilang sebagai variabel laten (latent variables). Melalui proses inferensi Bayesian, model tidak hanya memperkirakan nilai yang hilang tetapi juga memperhitungkan ketidakpastian dari estimasi tersebut, sehingga hasil analisis menjadi lebih robust[2].

Selain itu, PyMC memungkinkan penggunaan informasi historis harga komoditas sebagai prior distribution. Dengan adanya prior harga historis, model dapat memperkirakan nilai yang hilang berdasarkan pola harga yang pernah terjadi sebelumnya. Hal ini membuat imputasi lebih realistis dan tidak semata-mata bergantung pada data parsial yang tersedia. Misalnya, jika harga komoditas cenderung naik pada musim tertentu, informasi ini dapat dijadikan prior untuk memperkirakan nilai yang hilang pada periode yang sama. Dengan demikian, untuk data harga komoditas yang memiliki missing value dan ketidakpastian tinggi, PyMC merupakan metode yang tepat karena kemampuannya dalam memodelkan ketidakpastian dan mengintegrasikan pengetahuan awal melalui prior.

1. Menangani Data yang Hilang pada Harga Komoditas

Dalam penelitian ini, untuk menangani missing value pada variabel prediktor, metode yang digunakan adalah interpolasi spline untuk data yang hilang pada awal deret, dan interpolasi linear untuk data hilang berikutnya. Pemilihan metode ini didasarkan pada beberapa pertimbangan yang relevan dengan karakteristik data.

Pertama, interpolasi mampu memperkirakan nilai yang hilang berdasarkan tren data yang tersedia. Dengan menggunakan titik-titik data yang teramati di sekitarnya, interpolasi menangkap pola pergerakan alami dari data, sehingga estimasi yang dihasilkan lebih akurat dibandingkan metode sederhana seperti imputasi rata-rata atau median yang tidak mempertimbangkan dinamika antar variabel. Misalnya, pada variabel USD/IDR Close, Price Crude Oil, Change % Crude Oil, Price Natural Gas, Change % Natural Gas, Price Newcastle Coal, Change % Newcastle Coal, Price Palm Oil, Change % Palm Oil, Price US Sugar, Change % US Sugar, Price US Wheat, dan Change % US Wheat, interpolasi dapat menjaga kesinambungan pola kenaikan atau penurunan nilai yang terjadi dari waktu ke waktu, sejalan dengan fluktuasi pasar global.

Kedua, interpolasi sangat sesuai untuk data kontinu, terutama data time series seperti harga harian atau persentase perubahan harga. Variabel-variabel seperti harga komoditas energi dan pangan (misalnya Crude Oil, Natural Gas, Newcastle Coal, Palm Oil, US Sugar, dan US Wheat) serta nilai tukar USD/IDR merupakan data yang nilainya berubah secara bertahap mengikuti pergerakan pasar. Dengan interpolasi, pengisian data hilang dilakukan secara halus dan alami tanpa menciptakan lonjakan atau penurunan drastis yang tidak realistis, seperti yang mungkin terjadi jika menggunakan metode seperti forward-fill atau backward-fill[5].

Ketiga, interpolasi efektif digunakan ketika jumlah data hilang relatif sedikit dan tersebar (tidak berkumpul pada satu periode panjang). Karena missing value dalam data ini tersebar dan jumlahnya tidak dominan dibandingkan total observasi, interpolasi dapat mengisi kekosongan data tanpa mengubah distribusi asli dari variabel-variabel tersebut. Hal ini penting untuk menjaga validitas hasil analisis lanjutan seperti model prediksi harga atau analisis hubungan antar komoditas.

Oleh karena itu, metode interpolasi spline dan linear dipilih untuk menjaga keutuhan pola data dan akurasi estimasi, terutama untuk variabel-variabel USD/IDR Close, harga dan persentase perubahan Crude Oil, Natural Gas, Newcastle Coal, Palm Oil, US Sugar, dan US Wheat, yang merupakan data kuantitatif dan sensitif terhadap dinamika pasar.

1. Split Data

Penelitian ini juga melakukan pembagian data (split data) dilakukan untuk memisahkan data menjadi dua bagian, yaitu data latih (training data) dan data uji (testing data). Pembagian ini bertujuan untuk melatih model pada sebagian data, kemudian mengukur kinerja model pada data yang belum pernah dilihat sebelumnya guna menguji kemampuan generalisasi model. Proporsi pembagian data yang digunakan adalah 80:20, di mana sebanyak 80 persen data digunakan sebagai data latih untuk proses pelatihan dan pembentukan model prediksi dan 20 persen data sisanya digunakan sebagai data uji untuk mengevaluasi performa model.

Dengan pendekatan ini, model yang dikembangkan diharapkan mampu melakukan prediksi pada data baru secara lebih andal dan tidak overfitting terhadap data latih.

## Pemodelan dan Forecasting

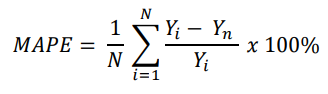
Analisis *forecasting* yang dilakukan adalah jenis analisis multivariat. *Forecasting multivariate time series* didefinisikan sebagai peramalan yang menggunakan lebih dari satu kriteria atau variabel yang berubah dari waktu ke waktu [6].

Untuk melakukan peramalan harga bahan pangan, digunakan pendekatan Multi-Output Regression dengan algoritma Random Forest Regressor. Pemilihan model ini didasarkan pada kemampuannya dalam menangani hubungan non-linear dalam data, kemampuannya untuk melakukan *forecasting* pada data yang tidak stasioner, dan keunggulannya dalam menangani variabel input yang kompleks.

Model MultiOutputRegressor digunakan untuk menangani prediksi beberapa variabel target sekaligus, yakni harga pangan di 34 provinsi. Model dasar yang digunakan adalah RandomForestRegressor dengan 100 pohon keputusan (n\_estimators=100) dan nilai random\_state=42 untuk memastikan hasil yang konsisten. Model yang telah dilatih kemudian digunakan untuk memprediksi harga bahan pangan di data uji (X\_test).

## Evaluasi Kinerja

Evaluasi kinerja dilakukan untuk menguji atau mengevaluasi kebenaran dan mengukur kinerja model yang dibuat. Teknik pengujian yang digunakan adalah *mean absolute percentage error* (MAPE). MAPE didefinisikan seperti pada persamaan 1.

 … (1)

Semakin rendah nilai MAPE maka semakin baik model yang dihasilkan karena semakin tinggi akurasinya [7].

MAPE dipilih karena menyediakan metrik kesalahan berbasis persentase yang intuitif sehingga sangat efektif untuk kompetisi ini, saat prediksi harga komoditas yang akurat sangat penting. Mengingat hubungan ekonomi dan pasar sebagai faktor dari fluktuasi harga komoditas, diharapkan peserta mengembangkan model yang secara konsisten menghasilkan persentase kesalahan yang lebih kecil.

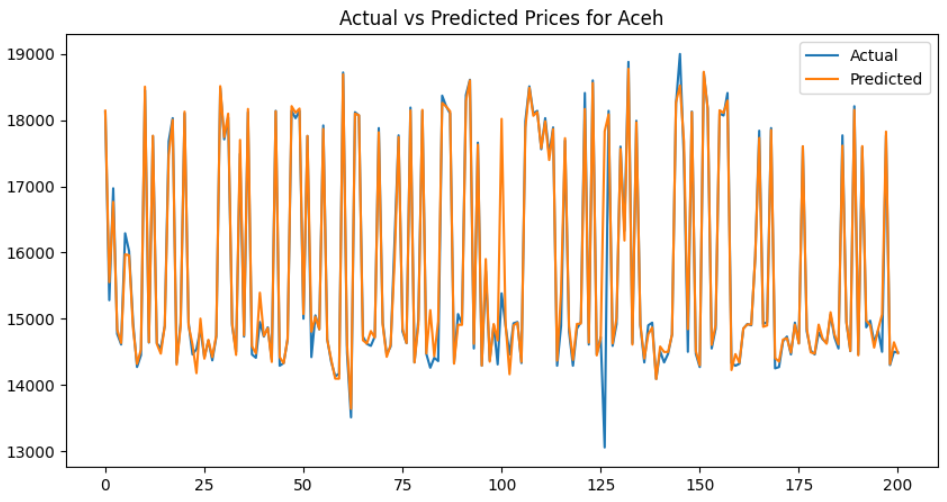
# Hasil dan Pembahasan

Jenis analisis deret waktu yang dilakukan adalah analisis *multivariate*, yakni analisis menggunakan beberapa atribut sebagai prediktor untuk melakukan peramalan pada data deret waktu. Ada banyak variabel yang dapat menjadi prediktor untuk harga bahan pangan. Dalam memilih atribut prediktor, kami memilah data yang disediakan oleh panitia Arkavidia 9.0. Dari data-data yang disediakan, kami memilih untuk menggunakan harga keenam komoditas global (*global commodity price*) dan nilai tukar Rupiah ke US Dollar. Data harga komoditas global memiliki kolom *price* yang menunjukkan harga pasar pada akhir perdagangan, *open* yang menunjukkan harga pembukaan pasar saat mulai beroperasi, *high* yang menunjukkan harga tertinggi yang dicapai oleh suatu komoditas pada hari tersebut, *low* yang menunjukkan harga terendah yang dicapai oleh suatu komoditas pada hari tersebut, *vol.* yang menunjukkan jumlah kontrak suatu komoditas yang diperdagangkan pada hari itu, dan *change %* yang menunjukkan persentase perubahan harga (*closing price*) dibandingkan dengan harga penutupan hari sebelumnya. Adapun pada dataset Mata Uang, terdapat dataset nilai tukar mata uang Malaysian Ringgit ke US Dollar, Singapore Dollar ke US Dollar, Thai Bath ke US Dollar, dan Rupiah Indonesia ke US Dollar. Kami memutuskan untuk hanya menggunakan nilai tukar Rupiah Indonesia ke US Dollar karena setelah melakukan eksplorasi menggunakan *heatmap* untuk mengetahui korelasi *pearson* dari harga komoditas bahan pangan dengan konversi mata uang, didapati bahwa hanya variabel konversi Rupiah Indonesia ke US Dollar lah yang memengaruhi harga komoditas bahan pangan, yakni beras medium, beras premium, dan gula konsumsi seperti yang dapat dilihat pada Gambar 3. Kolom yang terdapat pada dataset mata uang adalah *open* yang menunjukkan harga pertama pada hari tersebut, *high* yang menunjukkan harga tertinggi yang dicapai pada hari tersebut, *low* yang menunjukkan harga terendah yang dicapai pada hari tersebut, *close* yang menunjukkan harga terakhir sebelum pasar ditutup, *adj close* yakni harga penutupan yang disesuaikan dan bernilai sama dengan *close* pada mata uang, dan *volume* yang menunjukkan jumlah unit yang diperdagangkan pada hari tersebut. Kami hanya memilih kolom *close* karena merupakan representasi yang paling akurat untuk mewakili kurs mata uang pada hari tersebut. Kami memutuskan untuk tidak menggunakan dataset Google Trend karena banyaknya data yang hilang (*missing value*) dan sulitnya mencari data kovariat lampau maupun kovariat masa depan dari data di internet.

1. Heatmap korelasi harga bahan pangan dengan konversi mata uang



*Hyperparameter* n\_estimators dalam Random Forest Regressor menentukan jumlah pohon keputusan (decision trees) yang akan digunakan. Berdasarkan eksperimen *hyperparameter tuning* yang dilakukan dengan pilihan nilai n\_estimator = 50, 100, 200, diperoleh nilai n\_estimator yang memberikan hasil prediksi paling baik adalah n\_estimator = 100 dengan indikator nilai MAPE yang menunjukkan angka paling minimum untuk nilai n\_estimator tersebut. MAPE yang diperoleh untuk gula pasir di daerah Aceh adalah sebesar 0.063.

1. Perbandingan data asli dengan nilai prediksi

Gambar di atas menampilkan perbandingan antara harga aktual dan harga prediksi untuk komoditas tertentu di wilayah Aceh. Garis berwarna biru menunjukkan harga aktual, sementara garis oranye merepresentasikan hasil prediksi yang dihasilkan oleh model yang digunakan. Secara umum, pola fluktuasi harga pada kedua garis memiliki kemiripan, yang mengindikasikan bahwa model mampu menangkap tren pergerakan harga dengan cukup baik. Namun, terdapat beberapa titik di mana harga prediksi mengalami deviasi dari harga aktual. Hal ini dapat disebabkan oleh berbagai faktor, seperti keterbatasan model dalam mengakomodasi lonjakan harga yang ekstrem atau adanya variabel eksternal yang tidak disertakan dalam proses pelatihan model. Meskipun demikian, model masih dapat memberikan gambaran yang cukup representatif terhadap tren harga di wilayah tersebut.

1. Hadeed, S. J., O’Rourke, M. K., Burgess, J. L., Harris, R. B., & Canales, R. A. (2020). Imputation methods for addressing missing data in short-term monitoring of air pollutants. The Science of the Total Environment, 730, 139140. <https://doi.org/10.1016/j.scitotenv.2020.139140>
2. Junger, W., & De Leon, A. P. (2014). Imputation of missing data in time series for air pollutants. Atmospheric Environment, 102, 96–104. https://doi.org/10.1016/j.atmosenv.2014.11.049
3. Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics, 15(1). https://doi.org/10.1186/1471-2105-15-276
4. Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2021b). Machine learning advances for time series forecasting. Journal of Economic Surveys, 37(1), 76–111. https://doi.org/10.1111/joes.12429
5. Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., & Dubrawski, A. (2023). NHITS: Neural Hierarchical Interpolation for Time series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 37(6), 6989–6997. <https://doi.org/10.1609/aaai.v37i6.25854>
6. Putera R. B. R., Hendery. (2022). Multivariate Time Series Forecasting pada Penjualan Barang Retail dengan Recurrent Neural Network. Jurnal Inovtek Polbeng - Seri Informatika, 7(1), 71–82. <https://doi.org/10.35314/isi.v7i1.2398>.
7. Nur N., Wajidi F., Sulfayanti, Wildayani. (2023). Implementasi algoritma *random forest regression* untuk memprediksi hasil panen padi di Desa Minanga. Jurnal Komputer Terapan, 9(1), 58-64. doi: 10.35143/jkt.v9i1.5917