

Kombinasi Model ARIMA dan KNN Dalam Peramalan Harga Produk

Guntur Wijaya¹, Mardiani²

^{1,2}Universitas Multi Data Palembang, Jl.Rajawali 14, Palembang, 30113, Indonesia
gunturwijaya_2426311004@mhs.mdp.ac.id, mardiani@mdp.ac.id

Info Artikel

Riwayat Artikel:

Received 2026-01-12

Revised 2026-02-02

Accepted 2026-02-10

Abstract – This study proposes a product price forecasting model based on a hybrid ARIMA–KNN approach to support inventory decision-making at PT ABC. The novelty of this study lies in the use of a combination of time series statistical models and nonlinear learning to improve the stability of short-term price predictions on real transaction data sourced directly from the company's ERP system. The dataset used is historical cement purchase data from January 2023 to September 2025, processed through data cleaning, transformation, and monthly price aggregation stages. The results show that the hybrid ARIMA–KNN model provides more stable performance than a single ARIMA model, with an MAE value of 1604.94, an RMSE of 2299.37, and an R^2 of 0.2881. Although the model's ability to explain price variability is still limited, the hybrid approach has been proven to reduce extreme prediction fluctuations, maintain trend continuity, and produce a prediction curve that is closer to actual price movements. Different from previous studies that generally use statistical or machine learning models separately, this study shows that the integration of ARIMA and KNN can improve prediction stability in uncertain market conditions. The scientific contribution of this research lies in providing a scalable and applicable hybrid approach for price forecasting based on real industrial data, as well as being a basis for the development of advanced forecasting models based on nonlinearity and deep learning.

Keywords: ARIMA; data mining; inventory; K-Nearest Neighbor (KNN); price prediction;

Corresponding Author:

Guntur Wijaya

Email:

gunturwijaya_2426311004@mhs.mdp.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Penelitian ini mengusulkan model peramalan harga produk berbasis pendekatan hibrida ARIMA–KNN untuk mendukung pengambilan keputusan persediaan pada PT ABC. Kebaruan penelitian ini terletak pada pemanfaatan kombinasi model statistik deret waktu dan pembelajaran nonlinier untuk meningkatkan stabilitas prediksi harga jangka pendek pada data transaksi riil yang bersumber langsung dari sistem ERP perusahaan. Dataset yang digunakan berupa data historis pembelian semen periode Januari 2023 hingga September 2025 yang diproses melalui tahapan pembersihan data, transformasi, dan agregasi harga bulanan. Hasil penelitian menunjukkan bahwa model hibrida ARIMA–KNN memberikan kinerja yang lebih stabil dibandingkan model ARIMA tunggal, dengan nilai MAE sebesar 1604,94, RMSE sebesar 2299,37, dan R^2 sebesar 0,2881. Meskipun kemampuan model dalam menjelaskan variabilitas harga masih terbatas, pendekatan hibrida terbukti mampu mengurangi fluktuasi prediksi ekstrem, menjaga kontinuitas tren, dan menghasilkan kurva prediksi yang lebih mendekati pergerakan harga aktual. Berbeda dari penelitian sebelumnya yang umumnya menggunakan model statistik atau machine learning secara terpisah, penelitian ini menunjukkan bahwa integrasi ARIMA dan KNN dapat meningkatkan stabilitas prediksi pada kondisi pasar yang tidak menentu. Kontribusi ilmiah penelitian ini terletak pada penyediaan pendekatan hibrida yang terukur dan aplikatif untuk peramalan harga berbasis data industri nyata, serta menjadi dasar bagi pengembangan model peramalan lanjutan berbasis nonlinier dan deep learning.

Kata Kunci: ARIMA, data mining, K-Nearest Neighbor (KNN), persediaan, prediksi harga

I. PENDAHULUAN

Menurut UU RI no 11 tahun 2008 tentang Informasi dan Transaksi Elektronik, pemanfaatan teknologi informasi dan komunikasi di laksanakan berdasarkan asas kepastian hukum, manfaat, kehati-hatian, itikad baik dan kebebasan memilih teknologi atau netral teknologi. Dengan pesatnya perkembangan ilmu pengetahuan dan teknologi pada zaman sekarang ini, mendorong kemajuan di semua bidang kehidupan manusia [1]. Hal ini menyebabkan dunia yang tanpa sekat, sehingga semua informasi dapat di akses mudah. Dalam dunia Bisnis [2], penggunaan teknologi sangat dirasakan oleh pelaku dunia usaha. Penggunaan ERP (Enterprise Resource Planning) dari stok, harga penjualan, harga pembelian, invoice, sampai pemantauan hasil pekerjaan dapat di lakukan olehnya. Harga penjualan merupakan hal mendasar dalam melakukan proses bisnis. Untuk itu di perlukan sebuah perlakuan khusus agar pelaku bisnis bisa memperkirakan jumlah stok dan harga penjualannya.

Data harga penjualan pada masa waktu lampau dapat digunakan untuk pengambilan keputusan dalam hal bisnis atau pun dalam hal penelitian. Dengan penggunaan teknologi yang tepat data data tersebut dapat kita olah sedemikian rupa sehingga bisa menghasilkan sebuah alat yang digunakan untuk menghasilkan keputusan yang tepat [3]. Harga

penjualan ini biasanya akan berjalan berkesinambungan dalam proses bisnis perusahaan. Sistem persediaan barang atau bisa juga disebut dengan sistem inventory barang merupakan suatu sistem yang mengelola persediaan barang di Gudang [4]. Sekarang ini penggunaan teknologi persediaan barang untuk mengelola stok Gudang. Untuk menghindari kekosongan persediaan barang dan untuk memenuhi kebutuhan pembeli atas barang yang dibutuhkan, pebisnis harus dapat mengetahui perkiraan jumlah barang yang harus disediakan, untuk itu diperlukan adanya suatu peramalan atau prediksi berapa jumlah barang yang harus tersedia agar penjualan tetap berjalan lancar [5]. Dalam pengolahan persediaan barang dan harga penjualan dapat menggunakan sistem informasi. Sistem informasi yang digunakan adalah *data mining*. *Data mining* yang biasa disebut sebagai *knowledge discovery in database (KDD)* merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam himpunan data yang berukuran besar [6]. Hasil dari *data mining* [7] ini dapat kita gunakan untuk pengambilan keputusan di masa depan [8].

Dalam social economy penggunaan *data mining* sangat penting dalam prediksi penjualan. Maka *e-commerce* sangat membutuhkan prediksi tersebut untuk menentukan hubungan social dengan dampak belanja pelanggan. Oleh karena itu dengan memperluas bisnis mereka, sama saja dengan meningkatkan keuntungan. Pelaku bisnis dapat mengumpulkan informasi mengenai para pelanggannya melalui penggunaan *cookies*. Dalam penelitian [9] adalah membandingkan 5 Model Algoritma yaitu Algoritma *K-Nearest Neighbor (KNN)*, *Naive Bayes*, *Decision tree*, *Ensemble Method* dan *Linear regression* dengan menggunakan datasets Socio Economic Country untuk mendapatkan prediksi ekonomi dan pergerakan perkembangan ekonomi suatu negara dengan menggunakan semua metode algoritma *data mining*. Dalam prediksi saham *data mining* juga bisa diterapkan. Dengan menggunakan data historis saham maka hal tersebut di mungkinkan untuk dilakukan prediksi. Dalam penelitian (Akbar Rismayadi dkk) memprediksi harga saham apple dengan menggunakan metoda *machine learning support vector machine (SVM)*, *random forest* dan *K-Nearest Neighbor*. Dari hasilnya random forest dan KNN memberikan hasil yang tidak cukup baik dari SVM. Sehingga SVM merupakan algoritma yang sangat baik dalam memprediksi saham harga apple tersebut. Peramalan (*forecasting*) [11] adalah seni dan ilmu dalam memperkirakan suatu kejadian dimasa depan dan melibatkan pengambilan data historis dan memproyeksikannya ke masa mendatang dengan suatu bentuk model matematis. Penelitian [12] menggunakan metoda moving average dan exponential smoothing untuk peramalan (*forecasting*) persediaan stok barang. Biasanya jumlah persediaan barang dan harga barang itu saling keterkaitan. Berdasarkan ilmu ekonomi dimana supply dan demand itu saling berhubungan, jadi apabila sebuah perusahaan ingin menjaga kestabilan harga di pasar, maka diperlukan sebuah stabilitas ketersediaan barang. Salah satu teknik *data mining* adalah *clustering*. Penelitian [6] Clustering merupakan suatu proses pengelompokan sejumlah data atau objek ke dalam *cluster (group)* sehingga setiap dalam pada suatu cluster yang sama akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. Menggunakan K- Means dalam melihat penjualan paket data telkomsel, data yang dikelompokkan memiliki ciri-ciri yang sama berdasarkan kriteria tertentu. Salah satu aktivitas yang dilakukan untuk menganalisa data adalah klasifikasi atau pengelompokan data menjadi beberapa kategori, kelompok atau cluster. Penelitian prediksi harga crypto dengan menggunakan metoda ARIMA [13], LSTM dan Sarimax [14]. Penggunaan algoritma *time series* dalam analisis harga Bitcoin dapat memberikan beberapa manfaat. Penggunaan algoritma ini memungkinkan kita untuk mengidentifikasi pola pergerakan harga dalam data historis. *Time series* dapat mendeteksi tren, siklus, atau fluktuasi jangka pendek yang mungkin berulang dan dapat digunakan untuk membuat prediksi. Dari hasil prediksi ini akan dapat digunakan oleh para pemain saham bitcoin agar dapat mengurangi cut loss. Hasil evaluasi dari metode LSTM lebih unggul dalam menangkap dependensi jangka panjang dan pola non-linear dikarenakan hasil evaluasi yang dihasilkan model LSTM stabil di setiap koin nya.

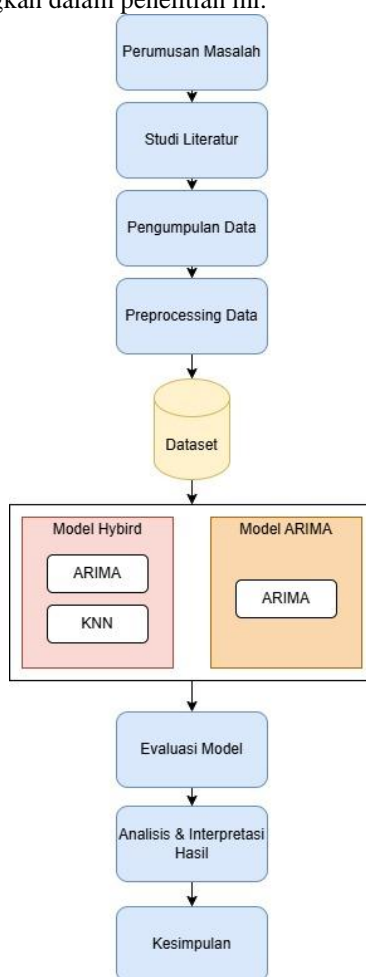
Dari hasil penelitian jurnal sebelumnya, maka ada beberapa kesamaan dalam proses pengambilan keputusan dengan menggunakan algoritma *data mining* dan prediksi (*time series*). Setelah membaca beberapa jurnal maka prediksi harga pasar untuk PT ABC dapat menggunakan algoritma KNN dan ARIMA untuk memberikan sebuah inputan yang berguna dalam menentukan arah atau model bisnis PT ABC. Dengan data historis harga dan transaksi penjualan yang tersimpan dalam sistem ERP memiliki potensi besar untuk dimanfaatkan sebagai dasar pengambilan keputusan melalui pendekatan *data mining* dan peramalan (*forecasting*). Berbagai penelitian sebelumnya menunjukkan bahwa metode peramalan berbasis deret waktu, seperti *Autoregressive Integrated Moving Average (ARIMA)*, efektif dalam menangkap pola linier dan tren jangka pendek, sementara algoritma *machine learning* seperti *K-Nearest Neighbor (KNN)* mampu memodelkan pola nonlinier pada data historis. Namun demikian, sebagian besar penelitian terdahulu masih menerapkan model ARIMA atau KNN secara terpisah, sehingga memiliki keterbatasan dalam menangani fluktuasi harga yang bersifat kompleks dan tidak stabil, khususnya pada data riil industri. Dan dengan menggunakan Bahasa pemrograman, maka *data mining* ini juga dapat dengan mudah di gunakan oleh penggunanya [15]. K-means adalah salah satu algoritma dalam *machine learning* yang digunakan untuk melakukan *clustering*, yaitu mengelompokkan data ke dalam beberapa kategori atau kelompok berdasarkan kemiripan tertentu. Maka dari pada itu, kita akan mengelompokkan stok barang dan harga penjualan untuk melihat potensi masa depan akan ketersediaan stok dan harga nya. Hasil pengelompokan bervariasi karena jumlah perubahan parameter cluster maka tantangan utama analisis *cluster* adalah jumlah *cluster* atau jumlah parameter model jarang diketahui, dan harus

ditentukan sebelum pengelompokan [6]. Ada banyak metode yang tersedia untuk memperkirakan jumlah *cluster* antara lain indeks statistik, metode berbasis varians, Teori Informasi, dan metode *goodness of fit*.

Berdasarkan celah penelitian tersebut, belum banyak kajian yang mengintegrasikan pendekatan statistik deret waktu dan pembelajaran nonlinier untuk meningkatkan stabilitas dan akurasi peramalan harga berbasis data ERP perusahaan. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model peramalan harga produk menggunakan pendekatan hibrida ARIMA–KNN pada PT ABC, dengan memanfaatkan data historis pembelian semen periode 2023–2025. Penelitian ini secara khusus bertujuan untuk (1) membandingkan kinerja model ARIMA tunggal dan model hibrida ARIMA–KNN berdasarkan metrik MAE, RMSE, dan R^2 , serta (2) menilai kemampuan pendekatan hibrida dalam meningkatkan stabilitas prediksi harga jangka pendek. Hasil penelitian ini memberikan kontribusi ilmiah dalam pengembangan model peramalan hibrida serta kontribusi praktis bagi perusahaan dalam mendukung perencanaan persediaan dan stabilisasi harga.

II. METODE

Langkah-langkah yang dilakukan dalam prediksi harga dengan metoda *ARIMA (Autoregressive integrated moving average)* [16] dan *K-Nearest Neighbor* pada perusahaan distribusi alat-alat bahan bangunan meliputi beberapa langkah sebagai berikut: *business understanding*, menentukan permasalahan pada proses bisnis PT ABC, penentuan teknik prediksi yang akan digunakan, pengumpulan data, preproses data yaitu menghilangkan data duplikat, dan membersihkan data yang tidak memenuhi syarat. Langkah selanjutnya adalah pemrosesan data dengan teknik prediksi dengan menggunakan Langkah awal yaitu menggunakan model ARIMA setelah dilakukan dengan model ARIMA maka akan dilanjutkan dengan menggunakan model KNN. Setelah selesai dengan menggunakan KNN maka akan di evaluasi model tersebut apakah sesuai dengan hasil yang diinginkan atau tidak. Selanjutnya melakukan analisis terhadap hasil dan penarikan kesimpulan dari hasil yang diperoleh. Dari alur proses tersebut dapat disajikan pada Gambar 1 yang merupakan langkah-langkah dalam penelitian ini.



Gambar 1. Metode Penelitian

A. Perumusan Masalah

Untuk tahapan ini adanya masalah yang harus di selesaikan oleh PT ABC dalam hal penyediaan stok dan harga untuk menjaga stabilitas semen yang ada di daerah distribusinya terutama daerah Sumatera Selatan.

B. Studi Literatur

Untuk tahapan ini dilakukan pencarian beberapa jurnal tentang prediksi atau forecasting harga untuk suatu jenis obyek atau barang dengan menggunakan metoda KNN dan juga ARIMA, sehingga didapatkan gambaran bagaimana menjawab permasalahan yang ada dengan menggunakan acuan jurnal tersebut.

C. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan menghimpun data historis harga dan transaksi pembelian semen yang diperlukan untuk keperluan analisis dan pemodelan prediksi. Data yang digunakan dalam penelitian ini berasal dari sistem *Enterprise Resource Planning* (ERP) PT ABC dan mencakup data penjualan serta data stok semen pada periode sebelumnya. Secara keseluruhan, dataset yang diperoleh terdiri dari 3.489 transaksi yang tercatat selama periode Januari 2023 hingga September 2025. Data tersebut memuat informasi penting seperti tanggal transaksi, jenis produk, jumlah pembelian, harga satuan, serta total nilai transaksi.

D. Preprocessing Data

Tahap preprocessing dilakukan untuk menyiapkan data agar layak digunakan dalam pemodelan prediksi. Proses ini meliputi pembersihan data dengan menghapus duplikasi dan kolom yang tidak relevan, standarisasi format tanggal transaksi, serta penyaringan data sesuai periode penelitian Januari 2023 hingga September 2025. Selain itu, dilakukan penanganan data yang tidak lengkap menggunakan KNN Imputation, serta transformasi data melalui agregasi harga bulanan dan penerapan differencing untuk memenuhi asumsi stasioneritas pada model ARIMA. Tahapan ini bertujuan menghasilkan dataset yang bersih, konsisten, dan representatif sehingga mendukung keandalan hasil peramalan.

E. Pemrosesan Data (ARIMA dan ARIMA + K-NN)

Tahap pemrosesan data dilakukan dengan menerapkan dua metode utama, yaitu *Autoregressive Integrated Moving Average* (ARIMA) dan *K-Nearest Neighbor* (KNN), pada data yang telah melalui tahap preprocessing. Dataset yang digunakan berasal dari data historis harga semen periode Januari 2023 hingga September 2025 yang telah diagregasi secara bulanan. Data tersebut kemudian dibagi menjadi data training sebesar 80% dan data testing sebesar 20% secara kronologis untuk menjaga karakteristik deret waktu dan menghindari kebocoran data. Model ARIMA digunakan sebagai model dasar untuk menangkap pola linier dan tren pada data deret waktu. Secara matematis, model ARIMA (p, d, q) dirumuskan sebagai berikut:

$$\Delta^d Y_t = c + \sum_{i=1}^p \phi_i \Delta^d Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

p menyatakan orde *autoregressive*, d tingkat diferensiasi, dan q orde *moving average*. Model ARIMA dilatih menggunakan data training untuk menghasilkan prediksi awal harga semen. Selanjutnya, metode KNN Regression digunakan untuk memodelkan pola nonlinier yang tidak dapat ditangkap oleh ARIMA. KNN bekerja dengan menghitung jarak antar data menggunakan metrik *Euclidean*, yang dirumuskan sebagai:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Nilai prediksi KNN diperoleh dari rata-rata nilai target dari sejumlah k tetangga terdekat. Dimana pada penelitian ini menggunakan $k = 3$. Penggabungan ARIMA dan KNN dilakukan dengan pendekatan hybrid, di mana nilai residual (selisih antara data aktual dan prediksi ARIMA) digunakan sebagai input pada model KNN untuk menghasilkan koreksi terhadap prediksi awal. Implementasi model ARIMA dan KNN dilakukan menggunakan Python. Pendekatan hybrid ini mampu meningkatkan stabilitas dan akurasi prediksi harga semen, sehingga hasilnya dapat digunakan sebagai dasar pendukung pengambilan keputusan bisnis PT ABC.

F. Evaluasi Model

Tahapan ini ketika sudah selesai dilakukan pemodelan menggunakan K-NN, maka akan dilakukan evaluasi model terhadap hasil yang di dapat. Dengan melakukan pengukuran performa model menggunakan tiga metrik utama: MAE (*Mean Absolute Error*), RMSE (*Root Mean Square Error*) [17] dan *R-Squared* (R^2). Evaluasi dilakukan untuk kedua model guna mengetahui perbedaan akurasi dan efisiensi. MAE mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi, yang dirumuskan sebagai:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

RMSE mengukur akar dari rata-rata kuadrat kesalahan prediksi dan dirumuskan sebagai:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Sementara itu, R^2 digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual, dengan rumus:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

Nilai MAE dan RMSE yang lebih kecil menunjukkan akurasi prediksi yang lebih baik, sedangkan nilai R^2 yang mendekati 1 menunjukkan kemampuan model yang semakin baik dalam merepresentasikan pola data. Evaluasi ini digunakan untuk membandingkan performa kedua model dan menentukan model terbaik untuk peramalan harga semen.

G. Analisis dan Interpretasi Hasil

Tahapan analisis hasil diharapkan mendapatkan hasil dari data data yang sudah di olah menggunakan KNN dan ARIMA, sehingga dapat memberikan saran atau perbaikan bagi stakeholder PT ABC guna memberikan efek yang baik bagi perkembangan PT ABC.

H. Kesimpulan

Memberikan masukan kesimpulan berdasarkan hasil eksperimen dan analisis. Dengan memberikan rekomendasi terhadap metode terbaik yang dapat diterapkan oleh PT. ABC dalam mengoptimalkan stok dan harga agar kepuasan pelanggan dapat tercapai dengan menjaga stabilitas harga di masyarakat atau pasaran.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian dan pembahasan terkait penerapan model ARIMA dan model *hybrid* ARIMA+K-NN dalam peramalan harga produk pada PT.ABC. Pembahasan difokuskan pada hasil pengolahan data, hasil *preprocessing data*, pemodelan ARIMA, model *hybird* ARIMA+K-NN, evaluasi kinerja model serta analisis perbandingan performa untuk menilai efektifitas pendekatan *hybird* dalam mendukung perencanaan persediaan dan stabilisasi harga.

A. Hasil

1) *Pengolahan Data* : Dataset yang digunakan merupakan hasil rekapitulasi pembelian semen dari beberapa pemasok, di antaranya PT Baturaja Multi Usaha, PT Jaya Abadi Kontrindo, dan mitra lain yang menjadi bagian dari rantai pasok perusahaan. Data mencakup beberapa kolom penting yaitu No. PO, Kode Supplier, Nama Supplier, Tanggal PO, Kode Barang, ITEM, QTY, UOM, Harga, Sub Total, PPN, dan Total Inc PPN. Proses pemuatan data dilakukan melalui Google Colab yang terhubung langsung ke Google Drive. Dengan perintah `drive.mount('/content/drive')`, file dataset Dataset.xlsx dibaca menggunakan pustaka pandas dan dikonversi menjadi DataFrame untuk memudahkan eksplorasi. Dari hasil pembacaan awal diperoleh total 1.544 baris data, yang kemudian ditampilkan lima baris pertama untuk memastikan struktur data sudah sesuai. Data yang digunakan dalam penelitian ini diperoleh dari sistem ERP PT. ABC berupa data historis pembelian semen baturaja dalam satuan zak ke principle selama 3 tahun (tahun 2023, 2024 dan 2025). Data yang ada tersedia di ERP tersebut berupa 3489 transaksi, data nya tidak terdapat data null, tetapi data yang terdapat di dalam ERP tersebut merupakan data campuran dari pembelian semen ke principle berupa data semen curah, zak dan jumbo bag dengan berbagai merk. Dalam penelitian ini di khususkan untuk memprediksi harga semen dalam satuan zak. Terdapat juga promo yang d berikan oleh principle bila mencapai target penjualan, yaitu bonus semen free, sehingga data tersebut masuk dalam ERP system PT ABC, oleh karena itu data tersebut di hapuskan saat mengolah data, agar harga yang di dapat adalah harga sebenarnya. Dataset yang digunakan memiliki atribut – atribut yaitu No, No.PO, Supplier, Tanggal, Item, Qty, UOM, Harga, Sub Total, PPN, Total Inc PPN. Dari atribut atribut yang terdapat dalam data set tersebut akan di lakukan pengolahan data dengan menggunakan google collabs yang d hubungan dengan google drive yang sudah berisi dataset pembelian ke supplier (pabrik) dan python.

Gambar 2. Dataset Transaksi Semen

2) *Preprocessing Data* : Pada bagian ini akan membahas hasil dari proses preprocessing data transaksi semen sebelum data digunakan dalam pemodelan. Preprocessing data yang dilakukan terdiri dari cleaning data, validasi data dan check nilai kosong, agregasi harga bulanan, visualisasi tren harga semen, uji stasioneritas dan identifikasi pola, dan augmented dickey fuller.

- *Cleaning Data* dilakukan untuk membuat dataset menjadi berkualitas. Pada dataset transaksi semen dilakukan beberapa proses *cleaning* yaitu menstandarkan format data (khususnya kolom Tanggal dan numerik), menghapus kolom yang tidak berisi data penting, menyaring data agar hanya mencakup periode penelitian (2023-2025), dan menghapus potensi duplikasi atau *noise* yang dapat mengganggu pembentukan pola deret waktu. Berikut hasil dari proses ini.

#	No. PO	Kode Supplier	Supplier	Tanggal PO	Kode Barang	ITEM	QTY	UOM	Harga	Sub Total	PPN	Total Inc	PPN	Unnamed: 13
0	593	PO2025/0076	SPL0005	PT Baturaja Multi Usaha	02/01/23	BRG0002	Semen Baturaja Zak @50KG	1600	Zak	58850	94160000	10357600.0	104517600.0	NaN
1	592	PO2025/0077	SPL0005	PT Baturaja Multi Usaha	03/01/23	BRG0002	Semen Baturaja Zak @50KG	1760	Zak	58850	103576000	11393360.0	114969360.0	NaN
2	573	PO2025/0101	SPL0007	PT Jaya Abadi Kontrindo	03/01/23	BRG0002	Semen Baturaja Zak @50KG	440	Zak	54500	23980000	2637800.0	26617800.0	NaN
3	591	PO2025/0078	SPL0005	PT Baturaja Multi Usaha	04/01/23	BRG0002	Semen Baturaja Zak @50KG	2040	Zak	58850	120054000	13205940.0	133259940.0	NaN
4	590	PO2025/0079	SPL0005	PT Baturaja Multi Usaha	05/01/23	BRG0002	Semen Baturaja Zak @50KG	880	Zak	58850	51788000	5696680.0	57484680.0	NaN

Gambar 3. Hasil Dataset Yang Sudah di *Cleaning*

- Validasi Data dan Check Nilai Kosong dengan menghasilkan bahwa semua kolom utama seperti *No. PO*, *Kode Supplier*, *Tanggal PO*, *Harga*, dan *PPN* memiliki nilai lengkap (tidak ada nilai hilang). Hanya kolom *Unnamed:13* yang memiliki nilai kosong sebanyak 1.542 baris, dan kolom tersebut sudah dihapus sebelumnya. Dalam konteks penelitian ini, ada tiga alasan mengapa *cleansing* dilakukan secara ketat yaitu ketepatan temporal dikarenakan data bersifat deret waktu (*time-dependent*), setiap tanggal yang hilang atau salah format dapat menyebabkan model gagal membaca urutan waktu, Konsistensi nilai Finansial karena Nilai harga, subtotal, dan PPN harus konsisten antar periode agar hasil peramalan tidak bias, dan Efisiensi Pemrosesan karena Dataset bersih mengurangi risiko error pada tahap pemodelan (ARIMA dan KNN), sehingga proses *training* dan *forecasting* berjalan lancar.
- Agregasi Harga Bulanan dilakukan dengan Langkah-langkah yaitu Pembuatan Kolom Periode : Kolom baru bernama Periode dibuat dengan mengekstraksi tahun dan bulan dari kolom Tanggal PO. Format `dt.to_period('M')` memastikan setiap nilai merepresentasikan satu bulan penuh, bukan tanggal tertentu, Perhitungan Rata-rata Harga : Data dikelompokkan berdasarkan kolom Periode menggunakan fungsi `groupby()`, lalu dihitung nilai rata-rata harga semen per bulan dengan `mean()`. Langkah ini menghasilkan satu baris data untuk setiap bulan, mewakili fluktuasi harga semen dalam periode tersebut, Konversi ke Format Timestamp : Setelah proses agregasi, kolom Periode dikembalikan ke format waktu (`datetime`) agar

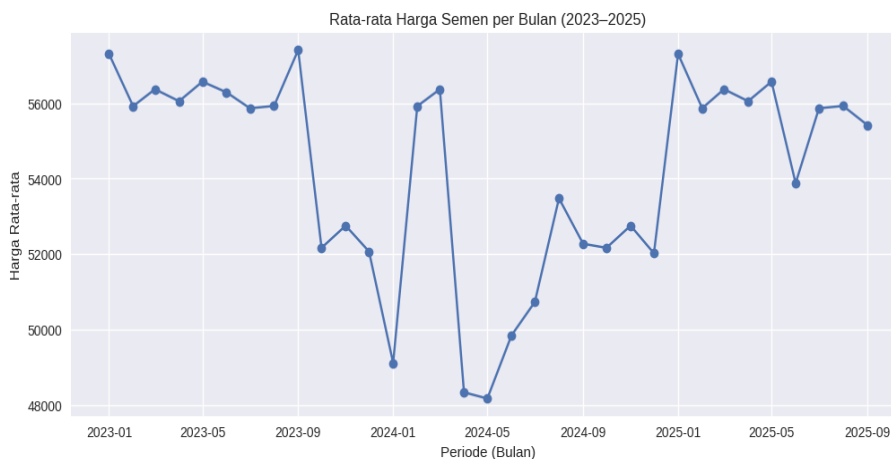
kompatibel dengan fungsi pemodelan deret waktu di tahap berikutnya dan Verifikasi Data Hasil Agregasi : Lima baris pertama hasil agregasi ditampilkan menggunakan `display()` untuk memastikan data telah terkelompok dan bernilai numerik. Dari hasil pengelompokan, diperoleh sebanyak 33 periode bulanan, dimulai dari Januari 2023 hingga September 2025. Nilai rata-rata harga semen per bulan berkisar antara Rp48.000 hingga Rp56.000 per zak (50 kg). Hasil ini menunjukkan adanya variasi harga yang cukup dinamis, dengan kecenderungan harga menurun di pertengahan 2024 dan meningkat kembali di awal 2025.

TABEL 1
 HASIL DATA AGREGASI

	Periode	Harga
0	01/01/23	57.306,410
1	01/02/23	55.920,909
2	01/03/23	56.369,605
3	01/04/23	56.052,941
4	01/05/23	56.570,754
5	01/06/23	56.287,500
6	01/07/23	55.865,753
7	01/08/23	55.926,785
8	01/09/23	57.408,632
9	01/10/23	52.166,117
10	01/11/23	52.754,000
11	01/12/23	52.068,833

Dari tabel 4.1 dapat diinterpretasikan bahwa fluktuasi harga semen cukup wajar, dengan variasi standar sekitar $\pm 4\%$ dari harga rata-rata. Pola ini mengindikasikan adanya pengaruh musiman yang dapat ditangkap secara baik oleh model ARIMA pada tahap pemodelan berikutnya.

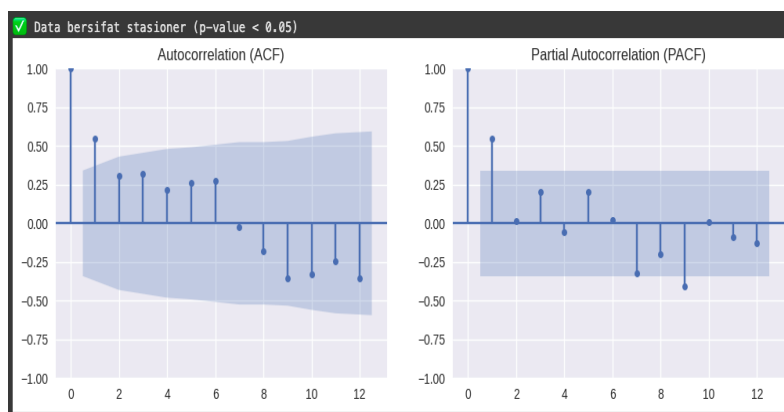
- Visualisasi Tren Harga Semen untuk melihat visualisasi tren harga semen dari data yang sudah di proses. Berikut hasil visualisasinya.



Gambar 4. Visualisasi Harga Semen

Grafik dalam gambar 3 memperlihatkan tiga karakteristik utama yang ditunjukkan dari x,y (periode dalam bulan dan harga rata) yaitu Stabilitas harga pada awal 2023, menunjukkan kondisi pasar yang normal, Penurunan harga tajam pada awal 2024, kemungkinan disebabkan oleh berkurangnya aktivitas proyek atau penyesuaian stok distributor, dan Kenaikan kembali pada awal 2025, mengindikasikan peningkatan permintaan dan aktivitas konstruksi nasional. Pola fluktuatif ini menegaskan bahwa data harga semen memiliki komponen tren dan musiman (*trend and seasonality*), yang sesuai untuk dimodelkan menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*).

- Uji Stasioneritas dan Identifikasi Pola dilakukan sebelum membangun model peramalan, langkah penting yang harus dilakukan adalah memastikan bahwa data deret waktu bersifat stasioner. Stasioneritas berarti bahwa nilai rata-rata, varians, dan kovarians dari data tidak berubah terhadap waktu. Dengan kata lain, pola fluktuasi data hanya bergantung pada jarak antarperiode (*lag*), bukan pada waktu absolutnya. Model Autoregressive Integrated Moving Average (ARIMA) sangat bergantung pada asumsi bahwa data yang dianalisis bersifat stasioner. Jika data tidak stasioner, maka hasil prediksi model akan bias dan kurang akurat. Oleh karena itu, tahap ini berfungsi untuk memastikan bahwa transformasi data hasil agregasi bulanan telah memenuhi prasyarat matematis tersebut. Dalam konteks penelitian prediksi harga semen di PT ABC, stasioneritas menjadi kunci karena harga semen memiliki fluktuasi periodik akibat perubahan biaya logistik, permintaan proyek konstruksi, dan kebijakan distribusi, Model ARIMA memerlukan kestabilan statistik agar dapat memprediksi tren jangka pendek dengan baik, dan dengan memastikan stasioneritas, peneliti dapat memisahkan komponen acak dari tren jangka panjang yang lebih sistematis. Sehingga, uji stasioneritas menjadi dasar untuk menentukan apakah perlu dilakukan *differencing* (pada parameter “d” dalam ARIMA) atau tidak. Dalam gambar 4.11 di bawah ini dapat di jelaskan bahwa Uji stasioneritas (biasanya uji ADF – Augmented Dickey-Fuller test) menunjukkan p-value < 0,05, sehingga data sudah stasioner tidak ada tren atau perubahan varians seiring waktu. Untuk grafik ACF yang ada di sebelah kiri menunjukkan hubungan antar nilai data sekarang dengan nilai sebelumnya (*lag*). Jika grafik bar (batang) yang signifikan (keluar dari daerah biru), berarti nilai tersebut berkorelasi kuat dengan masa lalu. Dari grafik lag 1 masih positif besar (sekitar 0.5) lalu menurut perlahan dan menjadi negatif setelah lagi 7 sampai 8. Pola menurun bertahap sering menunjukkan model MA (moving average). Untuk grafik PACF yang ada di sebelah kanan menunjukkan hubungan langsung antara nilai sekarang dan lag tertentu setelah efek lag sebelumnya dihilangkan. Dari grafik hanya lag 1 yang jelas signifikan, sementara lag setelahnya kecil dan dalam batas (area biru). Sehingga model ini cocok untuk p,d,q menggunakan 1,0,1 karena datanya stasioner. Penjelasan grafik tersebut untuk sumbu x adalah menunjukkan jarak waktu antara observasi sekarang dengan observasi masa lalu (*lag*) dan sumbu y menunjukkan besarnya kekuatan hubungan (korelasi) antara nilai saat ini dengan nilai masa lalu.



Gambar 5. Uji Stasioner

- *Augmented Dickey Fuller* dilakukan setelah dilakukan pengujian stasioneritas terhadap data harga semen bulanan menggunakan metode Augmented Dickey-Fuller (ADF), diperoleh hasil sebagai berikut.

TABEL 2
HASIL ADF

Parameter	Nilai
ADF Statistic	-3,0350
p-value	0,0318
Critical Value (1%)	-3,6535
Critical Value (5%)	-2,9572
Critical Value (10%)	-2,6176

Uji ADF dilakukan dengan tujuan untuk mengetahui apakah data deret waktu sudah stasioner, yaitu memiliki rata-rata, variansi, dan kovarians yang konstan sepanjang waktu. Dasar pengambilan keputusan dalam uji ADF

adalah H_0 (Hipotesis nol): Data tidak stasioner (memiliki *unit root*) atau H_1 (Hipotesis alternatif): Data stasioner (tidak memiliki *unit root*). Sedangkan kriteria keputusannya Jika $p\text{-value} < 0,05$, maka H_0 ditolak \rightarrow data stasioner atau Jika $p\text{-value} \geq 0,05$, maka H_0 diterima \rightarrow data tidak stasioner. Berdasarkan hasil yang ada maka Nilai $p\text{-value} = 0,0318 < 0,05$, sehingga H_0 ditolak, Artinya, data harga semen sudah bersifat stasioner pada taraf signifikansi 5% dan Nilai ADF Statistic ($-3,0350$) juga lebih kecil dari Critical Value 5% ($-2,9572$), yang semakin memperkuat kesimpulan bahwa deret waktu telah stasioner. Dengan demikian, data tidak memerlukan proses differencing ($d = 0$) dalam model ARIMA, karena tren jangka panjang dan variansnya sudah stabil.

3) *Pemodelan ARIMA dan Model Hybrid ARIMA+K-NN* : Model Autoregressive Integrated Moving Average (ARIMA) merupakan salah satu pendekatan statistik yang paling umum digunakan untuk memprediksi data deret waktu (*time series forecasting*). Berikut ini adalah simulasi perhitungan manual untuk proses ARIMA yaitu Pada data sebelum nya data sudah stasioner ($p\text{-value} = 0.0318 < 0.05$), karena nilai p nya lebih kecil dari 0.05 maka tidak perlukan lagi differencing, sehingga $d = 0$. Model arima ini di bangun dari tiga komponen p (autoregressive – AR) mengukur seberapa besar pengaruh nilai masa lalu terhadap nilai saat ini, d (Differencing) untuk menunjukkan seberapa banyak selisih (differencing) yang perlu dilakukan agar data menjadi stasioner dan q (Moving Average – MA) untuk menangkap pola kesalahan (error) dari periode sebelumnya untuk memperbaiki prediksi berikutnya. Dari gambar 4.11 acf dan pacf menunjukkan nilai 1 sehingga p, d, q nya menggunakan 1,0,1 adalah model arimanya. Dengan p = jumlah lag dari komponen Autoregressive (AR), d = tingkat differencing untuk membuat data stasioner, q = jumlah lag dari komponen Moving Average (MA). Rumus manual ARIMA.

Rumus Manual ARIMA

$$\Delta^d Y_t = c + \phi_1 \Delta^d Y_{t-1} + \dots + \phi_p \Delta^d Y_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \tag{6}$$

Dari hasil fitting dari program pyton terdapat : $c = 52.317$, $\phi = 0.724$, $\theta_1 = -0.462$, Maka persamaan model arima untuk $p\ d\ q = 1,0,1$ adalah

$$Y_t = 52.317 + 0.724 Y_{t-1} - 0.462 a_{t-1} \tag{7}$$

Misalkan kita memiliki data sebagai berikut :

TABEL 3
CONTOH HARGA SEMEN PERBULAN

Bulan	Harga (Rp)	Error (ε)
Jan	50.000	-
Feb	51.000	-200
Mar	52.000	100
Apr	?	?

$$\begin{aligned} Y_4 &= 52.317 + 0.724(52.000) - 0.462(+100) \\ &= 52.317 + 37.648 - 46.200 \\ &= 43.765 \end{aligned}$$

Secara matematis melalui proses python model ARIMA dinyatakan dalam tiga komponen utama, yaitu Autoregressive (AR, p) — hubungan antara nilai saat ini dengan satu atau lebih nilai masa lalu, Integrated (I, d) — proses differencing untuk menghilangkan sebuah trend menjadikan data stasioner, dan Moving Average (MA, q) — hubungan antara nilai saat ini dengan kesalahan (residual) masa lalu.

```

SARIMAX Results
-----
Dep. Variable:      Harga      No. Observations:      33
Model:             ARIMA(1, 0, 1)  Log Likelihood         -381.675
Date:              Tue, 21 Oct 2025  AIC                          611.398
Time:              12:49:40      BIC                      617.336
Sample:            - 0          HQIC                       613.364
Covariance Type:   opg
-----
coef    std err      z      P>|z|    [0.025    0.975]
-----
const    5.422e+04  1185.019   45.752   0.000   5.19e+04  5.65e+04
ar.L1     0.6478      0.410     1.579    0.114   -0.156    1.452
ma.L1    -0.1419      0.442    -0.321    0.748   -1.009    0.725
sigma2    5.439e+06   1.57e+06   3.461    0.001   2.36e+06  8.52e+06
-----
Ljung-Box (L1) (Q):      0.01  Jarque-Bera (JB):      7.11
Prob(Q):                 0.93  Prob(JB):              0.03
Heteroskedasticity (H):  0.98  Skew:                  -0.79
Prob(H) (two-sided):     0.98  Kurtosis:              4.63
    
```

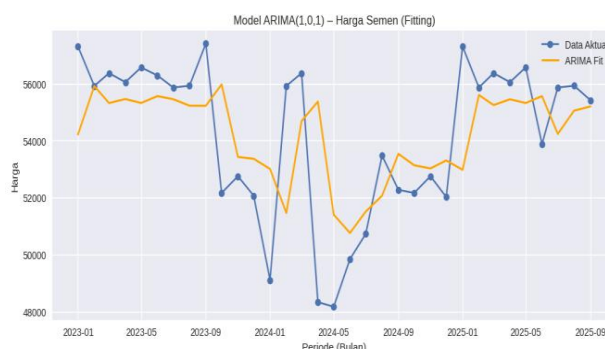
Gambar 6. Hasil dari Pemodelan Dengan Python

Konstanta (const = 54.220) maka nilai ini merepresentasikan rata-rata level harga semen bulanan, yaitu sekitar Rp54.220 per zak. Konstanta ini menjadi titik keseimbangan (mean level) ketika pengaruh harga bulan sebelumnya dan error masa lalu bernilai nol. Dengan demikian, nilai ini menunjukkan harga dasar semen dalam periode pengamatan 2023–2025.

Parameter AR(1) = 0,6478 menunjukkan pengaruh positif antara harga semen bulan sebelumnya terhadap harga bulan saat ini. Nilai koefisien 0,6478 berarti sekitar 64,78% variasi harga semen bulan ini dipengaruhi oleh harga bulan lalu, sedangkan sisanya dipengaruhi oleh faktor lain. Nilai $p\text{-value} = 0.114 (>0,05)$ menunjukkan bahwa pengaruhnya tidak signifikan secara statistik, tetapi tetap memiliki peran struktural dalam model.

Parameter MA(1) = -0,1419 menunjukkan hubungan antara kesalahan prediksi bulan lalu (error) dengan harga bulan ini. Nilai negatif berarti bahwa jika pada bulan sebelumnya terjadi kesalahan positif (harga aktual lebih tinggi dari prediksi), maka pada bulan berikutnya model cenderung mengoreksi ke bawah. Namun, nilai $p\text{-value} = 0,748$ menunjukkan bahwa komponen MA tidak signifikan secara statistik dalam menjelaskan variasi harga semen.

$\text{Sigma}^2 = 5.439\text{e}+06$ Nilai ini menggambarkan varians error model (noise). Semakin kecil nilai sigma^2 , semakin kecil tingkat kesalahan model. Dalam kasus ini, nilainya masih cukup moderat, menandakan bahwa model ARIMA(1,0,1) mampu menangkap fluktuasi harga dengan akurasi yang cukup baik.

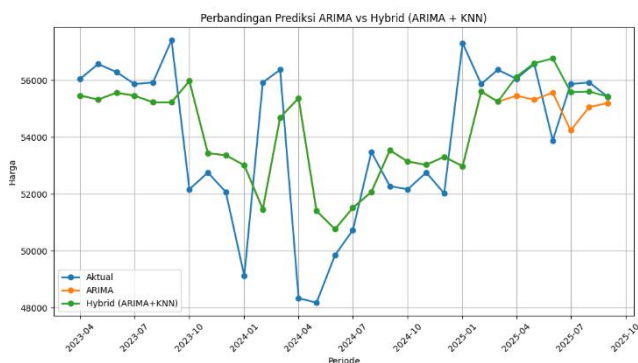


Gambar 7. Grafik Pemodelan ARIMA

Dari hasil grafik tersebut yg menunjukkan sumbu x sebagai periode (bulan) dan sumbu y sebagai harga dapat diinterpretasikan bahwa harga semen diprediksi mengalami penurunan bertahap. Ini mencerminkan fase normalisasi pasar setelah kenaikan harga yang terjadi pada pertengahan 2025, tidak terdapat lonjakan ekstrem dalam prediksi harga. Hal ini mengindikasikan bahwa faktor musiman dan permintaan konstruksi masih dalam kondisi stabil dan model ARIMA (1,0,1) mampu merepresentasikan hubungan linier antar periode dengan baik, meskipun belum mampu menangkap fluktuasi non-linier yang mungkin disebabkan oleh faktor eksternal seperti biaya energi atau distribusi.

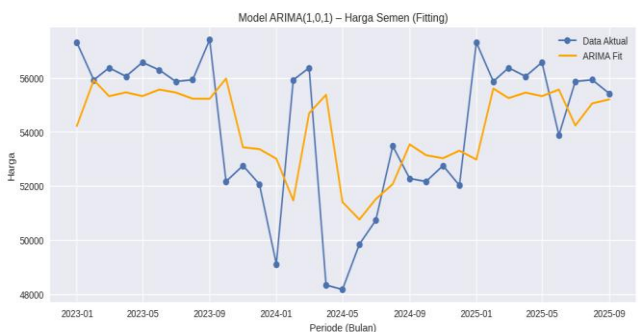
Model *Hybrid ARIMA + KNN* merupakan pendekatan integratif yang menggabungkan kekuatan model statistik dan algoritma *machine learning*. Pendekatan ini berangkat dari pemahaman bahwa ARIMA unggul dalam menangkap hubungan linier dan jangka pendek, sedangkan K-Nearest Neighbors (KNN) efektif dalam mengenali pola non-linier dan kompleks pada data residual. Dalam konteks prediksi harga semen PT ABC dataset berisi data data deret waktu (*time series*) seperti harga semen per hari / minggu / bulan. Tetapi harga semen juga di pengaruhi oleh factor-factor ekonomi dan operasional seperti harga bahan bakar, biaya logistik, musim pasar. Sebagian dari pola harga tersebut bersifat linear dan musiman (bisa diestimasi dari trend dan pola masa lalu) namun sebagian lain non-linier (tiba-tiba berubah karena factor eksternal). Model ARIMA mampu merepresentasikan tren utama fluktuasi harga, namun belum sepenuhnya dapat memprediksi perubahan yang disebabkan oleh faktor-faktor eksternal seperti variasi permintaan proyek, biaya distribusi, atau perubahan harga bahan bakar. Oleh karena itu, komponen KNN digunakan untuk mempelajari sisa kesalahan (residual) yang dihasilkan ARIMA dan memperbaikinya, sehingga diperoleh prediksi akhir yang lebih akurat dan adaptif.

Tahapan evaluasi model bertujuan untuk menilai tingkat akurasi dan efektivitas pendekatan Hybrid ARIMA + KNN dalam memprediksi harga semen. Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan data aktual menggunakan beberapa metrik statistik utama, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Koefisien Determinasi (R²)*. Dengan visualisasi metrik perbandingan untuk memberikan gambaran yang lebih intuitif dan komprehensif mengenai performa masing-masing model, yaitu ARIMA tunggal dan Hybrid ARIMA + KNN. Dengan menggunakan grafik batang (*bar chart*), peneliti dapat melihat secara jelas perbedaan kinerja dari setiap metrik evaluasi (MAE, RMSE, dan R²), sehingga analisis kuantitatif yang sebelumnya bersifat numerik dapat diterjemahkan menjadi bentuk visual yang lebih mudah dipahami. Visualisasi ini juga menjadi bagian penting dalam validasi hasil, karena membantu memastikan bahwa peningkatan performa yang dihasilkan model hybrid tidak hanya terjadi secara numerik, tetapi juga terlihat secara pola visual yang konsisten antar metrik.



Gambar 8. Grafik Perbandingan Model ARIMA vs Model Hybrid

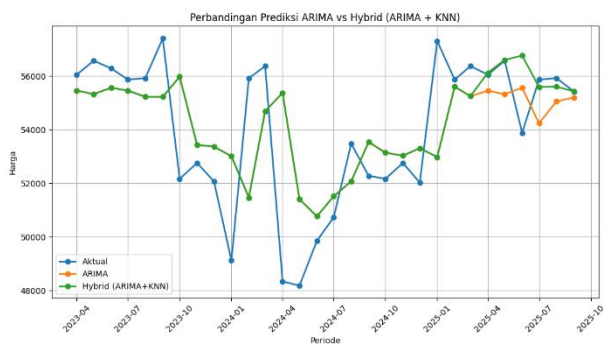
Berdasarkan hasil evaluasi model KNN terhadap residual error dari model ARIMA, diperoleh nilai MAE sebesar 1.604,94 dan RMSE sebesar 2.299,37. Nilai tersebut menunjukkan bahwa model KNN mampu mempelajari pola kesalahan (residual) dengan cukup baik, di mana rata-rata deviasi prediksi terhadap nilai aktual berada di bawah Rp 1.600. Hasil ini mengindikasikan bahwa KNN berhasil menangkap pola non-linier yang tidak dapat dijelaskan oleh model ARIMA. Dengan demikian, penerapan KNN sebagai model pelengkap (post-processing) terbukti dapat meningkatkan kemampuan generalisasi model gabungan ARIMA + KNN, khususnya pada data harga semen dengan fluktuasi moderat. Model Autoregressive Integrated Moving Average (ARIMA) digunakan sebagai model dasar (baseline) dalam penelitian ini. Model ini dipilih karena mampu menangani pola data deret waktu yang memiliki tren dan komponen autokorelasi.



Gambar 9. Grafik ARIMA vs Data Aktual

Pada gambar 9 grafik yang pada sumbu x adalah bulan dan sumbu y adalah harga menunjukkan bahwa hasil prediksi ARIMA relatif mengikuti tren data aktual, meskipun terdapat perbedaan pada beberapa titik waktu tertentu. Hal ini disebabkan oleh fluktuasi harga semen yang bersifat acak dan tidak sepenuhnya linier.

Setelah model ARIMA menghasilkan prediksi awal, tahap berikutnya adalah menerapkan model K-Nearest Neighbor (KNN Regression) sebagai model *refinement*. Model *K-Nearest Neighbor (KNN)* merupakan salah satu algoritma pembelajaran non-parametrik (*non-parametric learning algorithm*) yang digunakan baik untuk klasifikasi maupun regresi. Dalam konteks penelitian ini, KNN digunakan sebagai *regressor* untuk mempelajari dan memprediksi residual hasil model ARIMA — yaitu selisih antara harga aktual dan harga yang diprediksi oleh model linier sebelumnya.



Gambar 10. Grafik ARIMA vs Aktual vs Hybrid

Hybrid (ARIMA + KNN) dengan MAE : 1604,94, RMSE : 2299,37 dan R^2 : 0,2881. Pada gambar 10 grafik arima vs aktual vs hybrid pada sumbu x menunjukkan periode (bulan) dan sumbu y menunjukkan harga terlihat bahwa hasil prediksi gabungan (*Predicted ARIMA + KNN*) ditampilkan dengan garis hijau, sedangkan data aktual dengan garis biru. Secara visual, model kombinasi masih menunjukkan deviasi dari data aktual, namun sudah membentuk pola koreksi yang lebih halus dibandingkan model ARIMA tunggal.

4) *Evaluasi Model* : Model hybrid ARIMA + KNN dibangun untuk menggabungkan kekuatan model statistik (ARIMA) dan algoritma non-linear (KNN). ARIMA menangani pola tren linier pada data historis harga semen, sedangkan KNN bertugas memperbaiki sisa kesalahan (residual error) yang tidak dapat dijelaskan oleh model ARIMA. Pada gambar 9 Menunjukkan bahwa model hybrid mampu mengikuti arah pergerakan data aktual lebih baik dibanding ARIMA tunggal, khususnya pada periode fluktuasi harga antara Mei–September 2025. Berdasarkan hasil pengujian terhadap data periode April–September 2025, diperoleh nilai metrik sebagai berikut.

TABEL 4
TABEL METRIK MAE, RMSE DAN R^2

Metrik	Nilai	Interpretasi
MAE (<i>Mean Absolute Error</i>)	1604,94	Nilai MAE menunjukkan bahwa secara rata-rata prediksi model memiliki deviasi sekitar 1.600 satuan dari nilai aktual. Dalam konteks data dengan rentang 48.000–57.000, nilai ini mengindikasikan bahwa model berada pada tingkat akurasi menengah, namun masih terdapat kesalahan yang cukup signifikan pada beberapa periode.
RMSE (<i>Root Mean Square Error</i>)	2299,37	RMSE yang lebih tinggi dibanding MAE menunjukkan bahwa terdapat kesalahan prediksi besar (outliers) pada beberapa titik data. RMSE memberikan penalti lebih tinggi terhadap error ekstrem, sehingga nilai ini mengindikasikan bahwa model belum optimal dalam menangkap fluktuasi harga yang tajam.
R^2 (Koefisien Determinasi)	0,2881	Nilai R^2 sebesar 0,2881 berarti model hanya mampu menjelaskan 28,81% variabilitas data aktual. Secara ilmiah, ini menunjukkan bahwa kemampuan model dalam merepresentasikan pola data masih rendah, dan sebagian besar variasi harga tidak tercapture oleh model ARIMA maupun hybrid.

B. Pembahasan Hasil

Berdasarkan hasil evaluasi, dapat disimpulkan bahwa model ARIMA mampu menangkap tren harga semen dengan cukup baik, namun belum mampu merepresentasikan dinamika fluktuatif harga yang bersifat non-linier. Kombinasi dengan model KNN memberikan peningkatan metodologis, meskipun secara empiris hasil prediksi belum optimal. Pendekatan hybrid ARIMA + KNN tetap relevan karena secara konseptual berhasil menggabungkan dua kekuatan utama ARIMA untuk menangani pola linier dan tren jangka panjang sedangkan K-NN mencoba menangkap pola non-linier dari residual error. Nilai MAE dan RMSE yang berada pada kisaran 1600–2300 mengindikasikan bahwa model memiliki kesalahan prediksi rata-rata sekitar 3–4% dari nilai harga rata-rata semen (sekitar Rp52.500). Ini berarti model masih cukup baik untuk estimasi jangka pendek namun belum stabil untuk fluktuasi musiman. Nilai R^2 sebesar 0,2881 berarti model hanya dapat menjelaskan 28,81% dari variasi data aktual. Ini menandakan bahwa sebagian besar pola dalam data tidak berhasil ditangkap oleh model yang digunakan. Dalam analisis ilmiah, nilai R^2 di bawah 0,5 umumnya dianggap kurang memadai untuk model prediksi *time series*, terutama jika digunakan untuk tujuan forecasting operasional atau pengambilan keputusan. Jika dibandingkan dengan penelitian hybrid serupa, seperti pendekatan ARIMA–KNN dan ARIMA–machine learning pada studi peramalan harga dan permintaan, hasil penelitian ini menunjukkan kecenderungan yang konsisten, yaitu peningkatan stabilitas prediksi meskipun peningkatan nilai R^2 tidak signifikan. Perbedaan utama penelitian ini terletak pada penerapan model hibrida pada data transaksi riil berbasis ERP industri, bukan pada dataset publik atau simulasi, sehingga kontribusi penelitian bersifat applied research yang relevan dengan kebutuhan praktis perusahaan. Dari sisi implikasi bisnis, model hibrida ARIMA–KNN mampu memberikan gambaran tren harga yang lebih stabil, sehingga dapat digunakan oleh manajemen PT ABC sebagai dasar dalam perencanaan persediaan, pengendalian biaya, dan penentuan strategi harga, khususnya pada periode permintaan tinggi. Dengan demikian, novelty penelitian ini tidak hanya terletak pada kombinasi metode, tetapi pada implementasi dan validasi model hybrid dalam konteks industri nyata

IV. SIMPULAN

Penelitian ini menunjukkan bahwa pendekatan peramalan harga menggunakan model hibrida ARIMA–KNN mampu menghasilkan prediksi yang lebih stabil dibandingkan model ARIMA tunggal. Model ARIMA efektif dalam menangkap tren linier harga semen, sementara KNN berperan dalam memodelkan pola nonlinier melalui pembelajaran residual, sehingga meningkatkan kualitas prediksi jangka pendek. Meskipun nilai koefisien determinasi (R^2) yang diperoleh masih tergolong moderat, hasil evaluasi menunjukkan bahwa pendekatan hibrida mampu mengurangi fluktuasi prediksi ekstrem dan menghasilkan kurva prediksi yang lebih mendekati data aktual. Dari sisi

praktis, hasil penelitian ini dapat dimanfaatkan oleh PT ABC sebagai alat bantu dalam perencanaan persediaan dan pengendalian harga, khususnya pada periode permintaan tinggi dan kondisi pasar yang tidak stabil. Sebagai rekomendasi penelitian selanjutnya, disarankan untuk mengembangkan model peramalan yang lebih kompleks dengan mengintegrasikan pendekatan deep learning atau menambahkan variabel eksternal seperti faktor musiman, permintaan pasar, dan kondisi ekonomi makro guna meningkatkan akurasi dan daya generalisasi model.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Universitas Multi Data Palembang atas dukungan dan fasilitas yang telah diberikan sehingga penelitian ini dapat dilaksanakan dan diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] B. Jie, E. Eric, D. Mervyn, V. Anggrianto, and K. Kelvin, "Pemanfaatan Dan Dampak Penggunaan Teknologi Informasi Pada Bidang Sosial," *Journal of Information System and Technology*, vol. 4, no. 2, pp. 392–397, Jul. 2023, doi: 10.37253/joint.v4i2.6298.
- [2] R. D. Tri Wulandari, N. Nurmalitasari, and H. Permatasari, "Prediksi Harga Saham Pt Bank Central Asia Tbk Dengan Menggunakan Algoritma Autoregressive Integrated Moving Average (Arima)," *Infotech: Journal of Technology Information*, vol. 10, no. 2, pp. 173–178, Nov. 2024, doi: 10.37365/jti.v10i2.278.
- [3] Naomi Dada Kodi, Gergorius Kopong Pati, and Agustina P. Setiawi, "Klasifikasi Data Mining Prediksi Penjualan dengan Metode Appriori," *Jurnal Penelitian Teknologi Informasi dan Sains*, vol. 2, no. 3, pp. 186–193, Sep. 2024, doi: 10.54066/jptis.v2i3.2405.
- [4] H. Handayani, K. U. Faizah, A. Mutiara Ayulya, M. F. Rozan, D. Wulan, and M. L. Hamzah, "Perancangan Sistem Informasi Inventory Barang Berbasis Web Menggunakan Metode Agile Software Development Designing A Web-Based Inventory Information System Using The Agile Software Development Method."
- [5] B. Mulyono, "Volume 6 Issue 2 Februari 2023 Jurnal Kolaboratif Sains Prediksi Rentet Waktu Penjualan Barang Menggunakan Algoritma Backpropagation Prediction of *Time series* of Goods Sales Using the Backpropagation Algorithm," 2023. [Online]. Available: <https://jurnal.unismuhpalu.ac.id/index.php/JKS>
- [6] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [7] D. Setiadi, "Analisis Prediksi Harga Beras Berbasis Kualitas Menggunakan Algoritma K-Nearest Neighbor".
- [8] O. A. Alghanam, S. N. Al-Khatib, and M. O. Hiari, "Data Mining Model for Predicting Customer Purchase Behavior in e-Commerce Context," 2022. [Online]. Available: www.ijacsa.thesai.org
- [9] E. Mardiani *et al.*, "Membandingkan Algoritma Data Mining Dengan Tools Orange untuk Social Economy," *Digital Transformation Technology*, vol. 3, no. 2, pp. 686–693, Nov. 2023, doi: 10.47709/digitech.v3i2.3256.
- [10] A. Akbar Rismayadi, R. Wahyudi Febrianto, A. Rachmat Raharja, and I. Hariyanti, "Ifani Hariyanti Perbandingan Kinerja Metode Machine Learning SVM, Random Forest, dan KNN dalam Prediksi Harga Saham Apple."
- [11] M. Junaidi and F. Achmadi, "Analisis Prediksi Kinerja Perusahaan Menggunakan Rasio Profitabilitas *Time series* Dan Algoritma Neuro-Fuzzy," *Jurnal Ilmiah Pendidikan Teknik dan Kejuruan*, vol. 12, no. 1, p. 65, Jan. 2019, doi: 10.20961/jiptek.v12i1.30260.
- [12] E. Sefry *et al.*, "Analisis Peramalan Persediaan Barang Menggunakan Metode Moving Average Dan Exponential Smoothing Pada CV. Sanjaya Bangun Pratama."
- [13] S. Agustini Sinaga, "Implementasi Metode Arima (Autoregressive Moving Average) Untuk Prediksi Penjualan Mobil," *Journal Global Tecnology Computer*, vol. 2, no. 3, pp. 102–109, 2023.
- [14] "Cryptocurrency Price Predictor: Aplikasi Prediksi Harga Crypto Dengan Perbandingan Metode ARIMA, LSTM Dan SARIMAX Archels Ramadhany Salsabila a1 , I Made Agus Dwi Suarjaya a2 , Wayan Oger Vihikan a3."
- [15] R. F. Dalimunthe and R. A. Putri, "Data Mining on Women's Clothing Sales in Market Places with the K-Means Clustering Algorithm," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 2, p. 476, Aug. 2024, doi: 10.24014/ijaidm.v7i2.31384.
- [16] A. Aryanusa and S. Zahara, "Terbit online pada laman web jurnal: <http://ejournal.unim.ac.id/index.php/submit> SUBMIT (Jurnal Ilmiah Teknologi Informasi dan Sains) Analisis Prediksi Harga Bitcoin Dengan Menggunakan Metode Arima Analysis Of Bitcoin Price Prediction Using Arima Method," vol. 4, no. 1, pp. 15–18, 2026, [Online]. Available: <http://ejournal.unim.ac.id/index.php/submit>
- [17] T. O. Hodson, "Root-mean-square error (RMSE) or Mean Absolute Error (MAE): when to use them or not," Jul. 19, 2022, *Copernicus GmbH*. doi: 10.5194/gmd-15-5481-2022.