
Analisis Tren dan Peramalan Timbunan Sampah padat Rumah Tangga di Kota Pekalongan Menggunakan K-Means dan ARIMA

Karina Dwi Syavira¹, Umi Meganinditya Wulandari², Abdul Razak Naufal³

^{1,2,3}Program Studi Informatika, Fakultas Sains dan Teknologi,

Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan

Email: ¹karinadwisavira@gmail.com, ²nindityaw@gmail.com, ³naufal@itsnupekalongan.ac.id

(Naskah masuk: 17 Desember 2025, diterima untuk diterbitkan: 15 April 2026)

Abstrak: Pertumbuhan penduduk dan ekonomi di Kota Pekalongan telah menyebabkan peningkatan volume sampah yang signifikan, mengakibatkan kondisi overload pada TPA Degayu dan risiko pencemaran lingkungan. Pengelolaan sampah yang reaktif dan kurangnya data presisi menghambat perencanaan infrastruktur yang efisien. Penelitian ini bertujuan untuk menganalisis karakteristik wilayah dan meramalkan volume timbunan sampah guna mendukung transisi menuju manajemen berbasis data. Metode yang digunakan mengacu pada kerangka kerja Knowledge Discovery in Database (KDD) dengan pendekatan hibrida: algoritma K-Means untuk pengelompokan (klasterisasi) wilayah berdasarkan karakteristik sampah dan demografi, serta Autoregressive Integrated Moving Average (ARIMA) untuk peramalan deret waktu (time series). Implementasi dilakukan menggunakan bahasa pemrograman Python. Hasil analisis menunjukkan bahwa model ARIMA(1, 1, 1) merupakan model terbaik dengan nilai Root Mean Square Error (RMSE) sebesar 275.023. Secara statistik, model memenuhi asumsi white noise dan distribusi normal. Namun, hasil peramalan untuk periode 2025–2026 menunjukkan tren stasioner di kisaran 3.600 ton dan belum mampu menangkap pola musiman (seasonality) yang kuat pada data aktual. Oleh karena itu, meskipun model ini valid secara statistik untuk menyerap informasi dasar, pengembangan lebih lanjut menggunakan metode SARIMA disarankan untuk meningkatkan akurasi pada data yang memiliki fluktuasi musiman ekstrem.

Kata Kunci - Sampah Padat Perkotaan; Machine Learning; K-Means; ARIMA; Peramalan

Trend Analysis and Forecasting Household Solid Waste Generation in Pekalongan City Using K-Means and ARIMA

Abstract: Population and economic growth in Pekalongan City have led to a significant increase in solid waste volume, resulting in an overload at the Degayu Landfill and environmental pollution risks. Reactive waste management and a lack of precise data hinder efficient infrastructure planning. This study aims to analyze regional characteristics and forecast waste generation volume to support a transition toward data-driven management. The methodology follows the Knowledge Discovery in Database (KDD) framework using a hybrid approach: the K-Means algorithm for clustering regions based on waste characteristics and demographics, and the Autoregressive Integrated Moving Average (ARIMA) for time series forecasting. The implementation was conducted using Python. The results indicate that the ARIMA(1, 1, 1) model is the best-fitting model, yielding a Root Mean Square Error (RMSE) of 275.023. Statistically, the model satisfies the assumptions of white noise and normal distribution. However, the forecast for the 2025–2026 period shows a stationary trend around 3,600 tons and fails to capture the strong seasonality in the actual data. Therefore, while the model is statistically valid for capturing basic information, further development using the SARIMA method is recommended to improve accuracy for data with extreme seasonal fluctuations.

Keywords - Municipal Solid Waste; Machine Learning; K-Means; ARIMA; Forecasting.

1. PENDAHULUAN

Permasalahan sampah padat perkotaan makin mendesak akibat pertumbuhan penduduk, urbanisasi, dan ekonomi. Volume sampah global diperkirakan naik tajam dari 2,24 miliar ton (2020) menjadi 3,88 miliar ton (2050), melebihi laju populasi karena pola konsumsi yang boros. Tanpa intervensi, beban ini melebihi kapasitas lingkungan, terutama di perkotaan. Infrastruktur

pengelolaan sampah masih sangat berbeda antara negara maju dan berkembang, dengan negara berkembang bergantung pada metode tidak ramah lingkungan seperti open dumping [1].

Di Indonesia, pengelolaan sampah kompleks dengan timbunan mencapai 37,37 juta ton (2022), didominasi sampah rumah tangga [2]. Masalah utama adalah rendahnya efektivitas penanganan sampah hilir, dengan 7,2-11,3 juta ton sampah tidak terkelola, berujung pada pencemaran dan risiko kesehatan [3]. Data nasional juga tidak konsisten, menyulitkan perencanaan, sehingga diperlukan model prediktif yang berbasis data lokal dan terverifikasi [4].

Provinsi Jawa Tengah, termasuk Kota Pekalongan, adalah kontributor besar sampah nasional. Pekalongan mengalami peningkatan penduduk dan volume sampah, dari 41.059 ton pada tahun 2018 menjadi 44.044 ton pada tahun 2023 [5]. TPA Degayu yang satu-satunya sudah overload, menimbulkan risiko lingkungan dan memaksa pemerintah mencari solusi teknologi pengolahan seperti Waste to Energy .

Pengelolaan saat ini masih reaktif dan linear, tak mampu mengantisipasi lonjakan volume sampah [6]. Oleh karena itu, diperlukan manajemen berbasis prediksi yang efisien, didukung data presisi untuk optimalisasi fasilitas dan anggaran [7]. Metode statistik konvensional kurang mampu menangani kompleksitas pola sampah, sedangkan machine learning (ML) menawarkan keunggulan dalam mengolah data non-linier dan musiman, meningkatkan akurasi prediksi dan kebijakan [8]. Penelitian ini mengusulkan model hibrida K-Means dan ARIMA. K-Means untuk mengelompokkan wilayah berdasarkan karakter sampah (spasial), sedangkan ARIMA untuk meramalkan volume sampah masa depan (temporal). Integrasi ini sudah terbukti efektif dalam berbagai studi di Indonesia dan Asia Tenggara, menurunkan kesalahan prediksi dan menangkap pola heterogen di kota-kota berkembang [8].

Judul penelitian ini adalah "Trend Analysis and Forecasting of Household Solid Waste Generation in Pekalongan City Based on Machine Learning Using K-Means and ARIMA Algorithms", dengan tujuan memberikan gambaran lengkap tentang karakter dan volume sampah di Pekalongan. Hasilnya diharapkan mendukung transisi ke sistem pengelolaan sampah yang lebih cerdas, efisien, dan berkelanjutan di tengah keterbatasan lahan TPA Degayu.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan dengan menggunakan pendekatan kuantitatif yang terstruktur, mengacu pada kerangka kerja standar *Knowledge Discovery in Database* (KDD) untuk memastikan validitas dan reliabilitas hasil. Alur penelitian dirancang secara sistematis mulai dari identifikasi masalah, pengumpulan data, pra-pemrosesan, pemodelan menggunakan algoritma hibrida (*K-Means* dan *ARIMA*), hingga evaluasi akurasi model. Implementasi teknis dilakukan menggunakan bahasa pemrograman *Python* versi 3.x dalam lingkungan *Jupyter Notebook*, dengan memanfaatkan pustaka *Pandas* untuk manipulasi data, *Scikit-Learn* untuk klusterisasi, dan *Statsmodels* untuk analisis deret waktu [9].

2.1.1. Identifikasi Masalah

Langkah Identifikasi

Masalah utama yang diidentifikasi adalah ketidakpastian volume timbunan sampah di masa depan yang menyulitkan perencanaan alokasi armada dan kapasitas TPA Degayu di Kota Pekalongan.

a. Langkah 1

Analisis Kondisi Eksisting Kondisi TPA Degayu saat ini telah mencapai tahap kritis (*overload*) dengan ketinggian tumpukan sampah mencapai lebih dari 20 meter, yang diperparah oleh keterbatasan lahan untuk perluasan.

b. Langkah 2

Kebutuhan Perencanaan Diperlukan metode peramalan yang akurat untuk memproyeksikan volume sampah hingga tahun 2025 sebagai dasar pengambilan keputusan strategis dalam investasi teknologi *Waste-to-Energy*.

2.1.2. *Perumusan Masalah*

Variabel yang digunakan dalam penelitian meliputi volume sampah bulanan (ton), komposisi sampah (organik, anorganik), dan data demografi (jumlah penduduk) di empat kecamatan.

- 1) **Variabel Dependen** Volume timbunan sampah padat rumah tangga per bulan yang akan diprediksi menggunakan model ARIMA.
- 2) **Variabel Independen** Waktu (bulan/tahun) dan karakteristik wilayah (kepadatan penduduk, profil sampah) yang digunakan sebagai input untuk klusterisasi K-Means.

2.2 *Tahapan Pengujian*

Tahapan pengujian merupakan inti dari proses penelitian ini, di mana data mentah diolah menjadi informasi yang bermakna melalui serangkaian prosedur komputasi yang ketat.

2.2.1. *Pra-pemrosesan Data (Data Preprocessing)*

1. **Pembersihan Data (Data Cleaning)** Proses ini bertujuan untuk menghilangkan *noise* dan inkonsistensi dalam data yang dapat mendistorsi hasil analisis.
 - 1) **Penanganan Missing Values** Nilai yang hilang dalam dataset diimputasi menggunakan teknik interpolasi linier atau rata-rata (*mean*) untuk menjaga kontinuitas deret waktu yang sangat dibutuhkan oleh algoritma ARIMA agar dapat membaca pola tren dengan benar.
 - 2) **Penanganan Outlier** Baris data yang mengandung *outlier* ekstrem atau nilai yang tidak logis (misalnya nilai volume negatif atau nol) dihapus atau diganti secara selektif. Hal ini dilakukan karena keberadaan *outlier* dapat menggeser posisi *centroid* pada algoritma *K-Means* secara drastis, sehingga menghasilkan pembagian kluster yang bias. Selain itu, pada model ARIMA, *outlier* dapat menyebabkan varians residual menjadi tidak konstan (*heteroskedastisitas*), yang melanggar asumsi stasioneritas dan menurunkan akurasi peramalan secara signifikan ([10]).
2. **Transformasi Data**, Pada proses ini data disesuaikan format dan skalanya agar kompatibel dengan algoritma yang digunakan.
 - 1) **Normalisasi Data** Untuk algoritma *K-Means*, data karakteristik wilayah dinormalisasi menggunakan metode *Standard Scaler (Z-score)* agar seluruh variabel memiliki skala yang setara ($\text{mean}=0$, $\text{std}=1$). Tanpa normalisasi, variabel dengan nilai nominal besar (seperti jumlah penduduk) akan mendominasi perhitungan jarak *Euclidean*, mengabaikan variabel penting lainnya yang bernilai kecil (seperti persentase sampah) [11].
 - 2) **Uji Stasioneritas** Untuk algoritma ARIMA, data deret waktu diuji stasioneritasnya menggunakan *Augmented Dickey-Fuller (ADF) Test*. Jika nilai *p-value* > 0.05 (tidak stasioner), maka dilakukan proses pembedaan (*differencing*) untuk menstabilkan rata-rata dan varians data [12].

2.2.2. *Pemodelan (Modelling)*

1. **Klusterisasi Menggunakan K-Means** Algoritma ini digunakan untuk memetakan karakteristik wilayah.
 - a. **Penentuan K Optimal** Menggunakan metode *Elbow* untuk mencari titik belok penurunan nilai *Sum of Squared Errors (SSE)* guna menentukan jumlah kluster yang paling efisien.

b. **Proses Klustering** Algoritma mempartisi 4 kecamatan ke dalam kluster-kluster berdasarkan kemiripan fitur yang telah dinormalisasi, misalnya memisahkan wilayah dengan dominasi sampah organik tinggi dari wilayah komersial.

2. **Peramalan Menggunakan ARIMA**, dalam penelitian ini algoritma ARIMA digunakan untuk memproyeksikan volume sampah di masa depan.

a. **Identifikasi Model** Menganalisis plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*) untuk menentukan orde model sementara (p, d, q).

b. **Estimasi Parameter** Melatih berbagai kombinasi model pada data latih (*training set* sebesar 80% dari total data) dan memilih model terbaik berdasarkan nilai *Akaike Information Criterion* (AIC) terendah.

2.2.3. Evaluasi Model

1. **Pengukuran Akurasi** Validasi dilakukan dengan membandingkan hasil prediksi (*forecast*) terhadap data aktual (*testing set*) yang telah disisihkan.

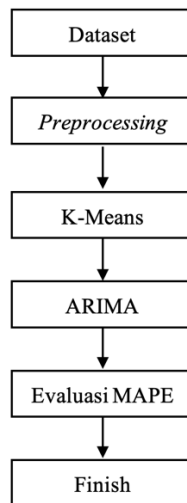
2. **Metode MAPE** Akurasi diukur menggunakan *Mean Absolute Percentage Error* (MAPE) dengan rumus:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{At - Ft}{At} \right|$$

Nilai MAPE yang rendah (di bawah 10-20%) menunjukkan bahwa model memiliki kemampuan prediksi yang baik dan layak digunakan sebagai dasar pengambilan kebijakan

2. HASIL DAN PEMBAHASAN

2.1. Pembahasan



Gambar 1. Flowchart Penelitian

Data yang digunakan dalam penelitian ini tergolong sebagai data sekunder yang diperoleh dari instansi pemerintah terkait serta publikasi resmi yang kredibel. Sumber data utama berasal dari laporan tahunan Dinas Lingkungan Hidup (DLH) Kota Pekalongan, dokumen *Kota Pekalongan Dalam Angka* yang diterbitkan oleh Badan Pusat Statistik (BPS), serta data agregat dari Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) Kementerian Lingkungan Hidup dan Kehutanan. Periode data yang diambil mencakup rentang waktu dari Januari 2018 hingga Desember 2023, yang dipilih untuk menangkap pola musiman jangka panjang serta dampak anomali seperti pandemi COVID-19 terhadap volume sampah. Penggunaan data sekunder dalam penelitian ini dinilai efisien dan memadai karena menyediakan catatan historis yang konsisten mengenai volume timbulan sampah dan variabel demografis tanpa memerlukan survei lapangan yang memakan waktu lama.

Keabsahan data ini telah diverifikasi melalui proses validasi silang (*cross-validation*) dengan tren nasional yang dilaporkan dalam studi Bank Dunia.

Tahapan penelitian ini disusun mengikuti kerangka kerja standar penambangan data (*data mining*) yang sistematis, dimulai dari pemahaman masalah hingga evaluasi. Tahapan penelitian ini disusun mengikuti kerangka kerja standar penambangan data (*data mining*) yang sistematis, dimulai dari persiapan data mentah hingga evaluasi kinerja model.

a. *Preprocessing*

Penelitian dimulai dengan menyiapkan data mentah yang bersih dan valid, termasuk menangani nilai hilang dengan imputasi atau interpolasi untuk menjaga kontinuitas deret waktu. Validasi tipe data juga dilakukan agar variabel numerik terbaca tepat. Proses ini penting untuk memastikan akurasi model machine learning. Penghapusan data dan penanganan outlier dilakukan selektif untuk menjaga keakuratan model, karena outlier bisa menggeser hasil kluster K-Means dan melanggar asumsi stasioneritas ARIMA. Data noise atau nilai tidak logis seperti negatif atau nol harus dihapus atau diimputasi agar model tidak salah belajar. Studi menunjukkan pembersihan data yang baik menurunkan error prediksi sampah. Selanjutnya, data untuk K-Means dinormalisasi dengan Standard Scaler agar variabel seimbang, sedangkan data ARIMA tidak diskalakan tapi diuji stasioneritasnya dengan ADF Test dan didifferencing jika perlu untuk stabilisasi.

b. *K-Means*

K-Means Clustering adalah metode untuk mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik ke dalam sejumlah groups (jumlah bilangan bulat positif). Ini mendefinisikan sebuah cluster oleh massa yang mewakili mean dari cluster. Setelah data bersih dan tertransformasi, algoritma K-Means diterapkan untuk mengelompokkan kecamatan di Pekalongan berdasarkan karakteristik serupa. Jumlah kluster optimal (k) ditentukan dengan metode Elbow melalui analisis SSE. Algoritma lalu mempartisi kecamatan ke kluster homogen dengan profil sampah dan demografi mirip. Hasil klusterisasi dianalisis untuk memahami profil unik tiap kelompok, seperti "Kluster Organik Tinggi" dan "Kluster Anorganik Tinggi," sebagai dasar strategi pengelolaan sampah yang lebih tepat sasaran.

c. *ARIMA*

Tahap peramalan menggunakan ARIMA diterapkan pada data deret waktu volume sampah yang sudah stasioner. Parameter model (p, d, s) ditentukan lewat analisis ACF dan PACF. Berbagai model dilatih dengan 80% data latih dan diuji pada data sisanya. Model terbaik dipilih berdasarkan AIC terendah untuk menyeimbangkan akurasi dan kesederhanaan. Model ini memproyeksikan volume sampah Pekalongan periode 2024–2025, sebagai dasar perencanaan kapasitas TPA dan armada.

d. *Evaluasi MAPE*

Evaluasi MAPE mengukur tingkat kesalahan peramalan ARIMA dibandingkan data aktual. Dalam penelitian ini, evaluasi dilakukan di akhir pemodelan untuk memastikan akurasi model hibrida (K-Means dan ARIMA) sebelum mengambil kesimpulan.

2.2 Analisis Hasil

Berdasarkan data pengelolaan sampah Kecamatan Pekalongan Barat awal 2018 pada Tabel 1. volume sampah meningkat dari 964,16 ton pada Januari menjadi 1.022,18 ton pada Mei, meski penduduk tetap stabil di 97.200 jiwa. Kenaikan ini menunjukkan peningkatan konsumsi atau aktivitas ekonomi, sehingga setiap individu menghasilkan sampah lebih banyak. Sampah organik mendominasi dengan puncak 63,1% di Februari, sementara sampah plastik konsisten sekitar 20%.

Pada Mei, sampah kategori "Lainnya" naik tajam menjadi 14,7%, kemungkinan akibat kejadian khusus atau perubahan pola buangan masyarakat.

Tabel 1. Rekapitulasi Volume dan Komposisi Sampah Kecamatan Pekalongan Barat (Januari - Mei 2018)

Tahun	Bulan	Kecamatan	Volume Sampah (ton/bulan)	Organik (%)	Plastik (%)	Kertas (%)	Lainnya (%)	Jumlah Penduduk
2018	1	Pekalongan Barat	964.16	61.7	21.0	10.5	6.8	97200
2018	2	Pekalongan Barat	974.08	63.1	19.3	8.5	9.1	97200
2018	3	Pekalongan Barat	988.15	59.2	22.2	8.8	9.9	97200
2018	4	Pekalongan Barat	1007.51	61.4	19.1	10.9	8.6	97200
2018	5	Pekalongan Barat	1022.18	58.1	18.0	9.2	14.7	97200

Tabel 2. Data Mentah Agregat (Sebelum Scaling)

Kecamatan	profile_Organik_pct	profile_Plastik_pct	profile_Kertas_pct	profile_Lainnya_pct	waste_per_capita
Pekalongan Barat	0.620007	0.201069	0.090666	0.088263	0.009785
Pekalongan Selatan	0.575150	0.230852	0.080383	0.113585	0.009918
Pekalongan Timur	0.592297	0.221441	0.080572	0.105622	0.010402
Pekalongan Utara	0.580803	0.207478	0.088597	0.123125	0.010384

Tabel 2 menyajikan data mentah agregat mengenai profil komposisi dan timbunan sampah di empat kecamatan Kota Pekalongan (Barat, Selatan, Timur, dan Utara) sebelum dilakukan proses standardisasi data (*scaling*). Berdasarkan tabel tersebut, terlihat bahwa **sampah organik** merupakan jenis sampah yang paling mendominasi di seluruh kecamatan dengan proporsi berkisar antara 57% hingga 62%, di mana Pekalongan Barat memiliki persentase sampah organik tertinggi. Sementara itu, **sampah plastik** menjadi kontributor terbesar kedua dengan proporsi sekitar 20% hingga 23%, dengan persentase tertinggi tercatat di Kecamatan Pekalongan Selatan.

Selain kedua jenis sampah utama tersebut, tabel juga merinci persentase sampah kertas dan jenis lainnya yang angkanya relatif kecil dan stabil di kisaran 8% hingga 12%. Dari sisi beban sampah per individu (*waste per capita*), data menunjukkan variasi yang tipis antarwilayah, di mana **Pekalongan Timur** memiliki tingkat timbunan sampah per orang tertinggi (0,0104), sedangkan **Pekalongan Barat** memiliki tingkat terendah (0,0097). Data ini memberikan gambaran dasar karakteristik sampah di setiap wilayah yang nantinya perlu disetarakan skalanya agar dapat dianalisis lebih lanjut secara statistik.

Tabel 3. Data Siap Cluster (Setelah Scaling)

Kecamatan	profile_Organik_pct	profile_Plastik_pct	profile_Kertas_pct	profile_Lainnya_pct	waste_per_capita
Pekalongan Barat	1.617487	-1.213412	1.210532	-1.515322	-1.226009
Pekalongan Selatan	-0.979078	1.342213	-1.007724	0.464018	-0.744208
Pekalongan Timur	0.013482	0.534683	-0.967008	-0.158404	1.018102
Pekalongan Utara	-0.651891	-0.663484	0.764200	1.209708	0.952115

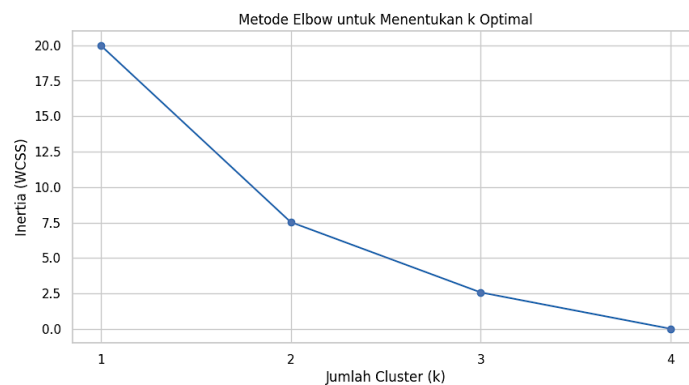
Pada Tabel 3. secara garis besar, data menunjukkan adanya perbedaan karakteristik sampah yang sangat mencolok antar-wilayah di Pekalongan. Kecamatan **Pekalongan Barat** terlihat memiliki profil yang paling "hijau" dan efisien secara statistik agregat, karena wilayah ini mencatatkan tingkat produksi sampah per kapita paling rendah dibandingkan tiga kecamatan lainnya, serta memiliki komposisi sampah organik tertinggi yang mencapai angka rata-rata 62%. Hal ini dikuatkan oleh data *scaling* yang menunjukkan skor positif sangat tinggi (1.61) untuk organik namun skor negatif drastis (-1.21) untuk plastik, yang artinya masalah utama di wilayah Barat murni adalah pengelolaan sampah basah atau sisa makanan, bukan sampah plastik.

Sebaliknya, **Pekalongan Selatan** justru menghadapi situasi darurat sampah anorganik yang serius. Data menunjukkan bahwa wilayah ini merupakan penyumbang polusi plastik terbesar dengan persentase mencapai 23%, jauh di atas rata-rata kecamatan lain. Dalam analisis statistik terstandarisasi (*scaling*), Pekalongan Selatan memiliki skor anomali positif yang tinggi untuk plastik (1.34), sehingga strategi penanganan sampah di wilayah ini tidak bisa disamakan dengan wilayah

Barat; fokus di sini haruslah pada pengurangan penggunaan plastik sekali pakai dan optimalisasi bank sampah.

Sementara itu, tantangan berbeda ditemukan di **Pekalongan Timur dan Utara**, di mana isu utamanya bukan pada jenis sampahnya, melainkan pada perilaku masyarakatnya yang cenderung boros sampah (*wasteful*). Kedua kecamatan ini memiliki angka *waste per capita* tertinggi (sekitar 0,0104 ton/orang), yang berarti setiap individu di sana membuang sampah lebih banyak daripada warga di wilayah lain. Khusus untuk Pekalongan Utara, terdapat pula anomali pada tingginya kategori sampah "Lainnya", yang mungkin mengindikasikan adanya limbah residu spesifik atau aktivitas industri kecil yang tidak terklasifikasi sebagai organik, kertas, maupun plastik.

Namun, meskipun data agregat menempatkan Pekalongan Barat sebagai wilayah dengan beban per kapita terendah, data tren bulanan dari Januari hingga Mei 2018 memberikan peringatan dini yang penting. Terlihat adanya kenaikan volume sampah yang konsisten setiap bulannya di Pekalongan Barat (naik sekitar 6% dalam 5 bulan) padahal jumlah penduduknya tetap. Fenomena ini mengindikasikan bahwa meskipun beban per orang saat ini masih rendah, tren gaya hidup masyarakat di sana sedang bergerak ke arah yang lebih konsumtif, sehingga jika tidak diintervensi, efisiensi yang dimiliki wilayah Barat bisa tergerus di masa depan.



Gambar 2. Metode Elbow (*Elbow Method*)

Penentuan pengelompokan wilayah menggunakan grafik *Metode Elbow* pada Gambar 2, yang menunjukkan bahwa titik optimal untuk membagi karakteristik sampah di empat kecamatan tersebut berada pada titik $k = 3$. Hal ini mengindikasikan bahwa meskipun terdapat empat kecamatan, pola timbulan sampah mereka sebenarnya mengerucut pada tiga profil perilaku yang berbeda, di mana ada dua kecamatan yang memiliki kemiripan karakteristik tinggi sehingga dapat disatukan dalam satu klaster, sementara dua kecamatan lainnya memiliki karakter unik yang berdiri sendiri.

Berdasarkan data profil wilayah yang telah distandarisasi (*scaling*), pembagian tiga klaster ini terlihat sangat jelas perbedaannya. **Pekalongan Barat** menonjol sebagai wilayah dengan karakteristik paling unik karena memiliki tingkat efisiensi terbaik dengan skor *waste per capita* terendah (-1,22) serta dominasi sampah organik yang sangat tinggi (skor 1,61), namun sangat minim sampah plastik. Sebaliknya, **Pekalongan Selatan** justru menjadi wilayah dengan masalah plastik terberat karena memiliki skor anomali plastik yang sangat tinggi (1,34) dengan proporsi organik yang rendah, sehingga penanganannya harus berfokus pada reduksi sampah anorganik. Sementara itu, **Pekalongan Timur dan Pekalongan Utara** cenderung masuk dalam kelompok yang sama karena keduanya mencatatkan perilaku masyarakat yang paling boros sampah dengan skor timbulan per kapita tertinggi (di atas 0,95), meskipun Pekalongan Utara memiliki catatan khusus berupa tingginya sampah residu atau kategori "Lainnya".

Namun, data tren bulanan memberikan peringatan penting bagi wilayah Pekalongan Barat yang secara agregat terlihat paling efisien tersebut. Meskipun beban per kapitanya rendah, data deret waktu dari Januari hingga Mei 2018 menunjukkan adanya tren peningkatan volume sampah total yang konsisten, bergerak naik dari 964,16 ton menjadi 1.022,18 ton. Mengingat jumlah penduduk di wilayah ini tetap stabil di angka 97.200 jiwa, kenaikan volume ini murni disebabkan oleh perubahan gaya hidup masyarakat yang semakin konsumtif setiap bulannya, ditambah dengan

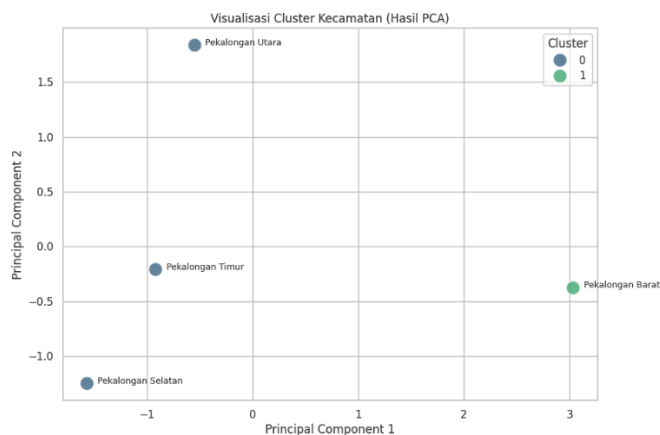
adanya anomali pada bulan Mei di mana terjadi lonjakan mendadak pada komposisi sampah kategori "Lainnya" mencapai 14,7% yang perlu diinvestigasi lebih lanjut penyebabnya.

Tabel 4. Hasil Clustering Data Sampah.

Kecamatan	Cluster	Organik (%)	Plastik (%)	Kertas (%)	Lainnya (%)	Waste per Capita (ton/orang)
Pekalongan Barat	1	0.620007	0.201069	0.090666	0.088263	0.009785
Pekalongan Selatan	0	0.575150	0.230852	0.080383	0.113585	0.009918
Pekalongan Timur	0	0.592297	0.221441	0.080572	0.105622	0.010402
Pekalongan Utara	0	0.580803	0.207478	0.088597	0.123125	0.010384

Analisis data pengelolaan sampah pada Tabel 4. menunjukkan bahwa penurunan variasi data mulai melandai secara signifikan pada titik $k = 2$, sehingga pembagian wilayah menjadi dua kluster utama dianggap sebagai langkah paling optimal. Hasil algoritma *clustering* kemudian mengonfirmasi hal ini dengan memisahkan **Pecamatan Pekalongan Barat** ke dalam kelompok tersendiri (Cluster 1), sementara **Pecamatan Pekalongan Selatan, Timur, dan Utara** digabungkan menjadi satu kelompok besar (Cluster 0) karena memiliki kemiripan tantangan pengelolaan sampah.

Pemisahan Pekalongan Barat sebagai kluster unik didasari oleh profil sampahnya yang sangat kontras dengan wilayah lain, di mana data *scaling* menunjukkan wilayah ini memiliki dominasi sampah organik tertinggi dengan skor anomali 1,61 serta tingkat efisiensi terbaik karena memiliki skor *waste per capita* terendah (-1,22). Sebaliknya, kelompok Cluster 0 menghadapi beban yang lebih berat, di mana Pekalongan Selatan mencatatkan skor tertinggi untuk polusi plastik (1,34), sedangkan Pekalongan Timur dan Utara memiliki skor positif pada *waste per capita* yang mengindikasikan bahwa penduduk di kedua wilayah tersebut cenderung lebih boros dalam memproduksi sampah dibandingkan warga Barat.



Gambar 3. Scatter Plot Visualisasi Hasil Clustering menggunakan PCA (Principal Component Analysis)

Namun, meskipun Pekalongan Barat terlihat paling ideal dalam perbandingan antar-wilayah, analisis tren bulanan memberikan peringatan dini yang penting. Data deret waktu dari Januari hingga Mei 2018 memperlihatkan bahwa volume sampah di Pekalongan Barat terus meningkat secara konsisten dari 964,16 ton menjadi 1.022,18 ton padahal jumlah penduduknya tetap stabil. Selain itu, terjadi anomali spesifik pada bulan Mei, di mana persentase sampah kategori "Lainnya" melonjak tajam menjadi 14,7%, yang menandakan adanya perubahan pola konsumsi atau kejadian luar biasa yang mengubah komposisi sampah residu di wilayah tersebut pada akhir periode pengamatan.

Gambar 3. memvisualisasikan hasil pengelompokan (clustering) empat kecamatan di Kota Pekalongan menggunakan metode Principal Component Analysis (PCA) untuk menyederhanakan data yang kompleks menjadi tampilan dua dimensi. Grafik scatter plot ini memperlihatkan pembagian wilayah menjadi dua kelompok utama yang sangat kontras. Kelompok pertama, yaitu Cluster 0 (titik biru), terdiri dari Pekalongan Utara, Pekalongan Timur, dan Pekalongan Selatan yang

posisinya cenderung berada di sisi kiri sumbu (nilai negatif pada Principal Component 1). Kedekatan posisi ketiga kecamatan ini menunjukkan bahwa mereka memiliki karakteristik komposisi dan timbulan sampah yang relatif mirip satu sama lain.

Di sisi lain, Pekalongan Barat muncul sebagai satu-satunya anggota Cluster 1 (titik hijau) dan terpisah cukup jauh di sisi kanan grafik (outlier). Posisi yang terisolasi ini mengonfirmasi temuan dari tabel data mentah sebelumnya, di mana Pekalongan Barat memang memiliki profil data yang unik seperti persentase sampah organik yang jauh lebih tinggi dan timbulan sampah per kapita yang paling rendah sehingga algoritma memisahkannya menjadi kelompok yang berbeda dari tiga kecamatan lainnya.

Tabel 5. Data agregat total Kota (*Time Series*)

Tanggal	Tahun	Bulan	Total Volume (Ton)	Organik (Idx)	Plastik (Idx)	Kertas (Idx)	Lainnya (Idx)	Penduduk
2018-01-01	2018	1	3417.17	234.9	86.3	36.5	42.2	337,600
2018-02-01	2018	2	3629.27	241.3	82.9	34.7	41.1	337,600
2018-03-01	2018	3	3609.40	231.9	91.3	36.4	40.6	337,600
2018-04-01	2018	4	3654.66	240.1	85.1	37.0	37.8	337,600
2018-05-01	2018	5	3664.20	227.8	86.6	33.5	52.2	337,600

Tabel 5. menyajikan data deret waktu (time series) bulanan untuk total sampah di seluruh kota pada periode awal tahun 2018 (Januari hingga Mei). Tabel ini berfungsi untuk memantau tren volume sampah dan komposisinya seiring berjalannya waktu.

Dari data tersebut, terlihat adanya tren peningkatan pada Total Volume Sampah, yang bergerak dari angka 3.417 ton pada bulan Januari menjadi 3.664 ton pada bulan Mei 2018. Sementara itu, kolom Penduduk menunjukkan angka yang konstan sebesar 337.600 jiwa, yang wajar karena data sensus biasanya tidak berubah setiap bulan. Kolom-kolom jenis sampah (Organik, Plastik, Kertas, Lainnya) disajikan dalam nilai indeks (Idx) atau tonase spesifik yang berfluktuasi setiap bulannya, namun tetap memperlihatkan pola konsisten di mana komponen Organik selalu memiliki nilai tertinggi (di atas 220), diikuti oleh Plastik, Lainnya, dan Kertas. Data ini krusial untuk melihat pola musiman atau kenaikan beban sampah kota dalam jangka waktu tertentu.

Tabel 6. Hasil *Clustering* & Profil Per Kecamatan (Data Mentah)

Kecamatan	Organik (Score)	Plastik (Score)	Kertas (Score)	Lainnya (Score)	Waste per Capita (Score)
Pekalongan Barat	1.617	-1.213	1.211	-1.515	-1.226
Pekalongan Selatan	-0.979	1.342	-1.008	0.464	-0.744
Pekalongan Timur	0.013	0.535	-0.967	-0.158	1.018
Pekalongan Utara	-0.652	-0.663	0.764	1.210	0.952

Tabel 6. memperlihatkan karakteristik spesifik setiap kecamatan melalui nilai skor terstandarisasi. Pekalongan Barat tampil paling kontras dengan skor tertinggi pada sampah Organik (1,617) dan Kertas, namun justru memiliki skor terendah untuk sampah Plastik dan Timbulan per Kapita. Sebaliknya, karakteristik Pekalongan Selatan didominasi oleh tingginya skor sampah Plastik (1,342), sedangkan Pekalongan Timur memimpin dalam volume Timbulan Sampah per Kapita (1,018). Terakhir, Pekalongan Utara memiliki profil unik dengan skor tertinggi pada kategori jenis sampah Lainnya (1,210).

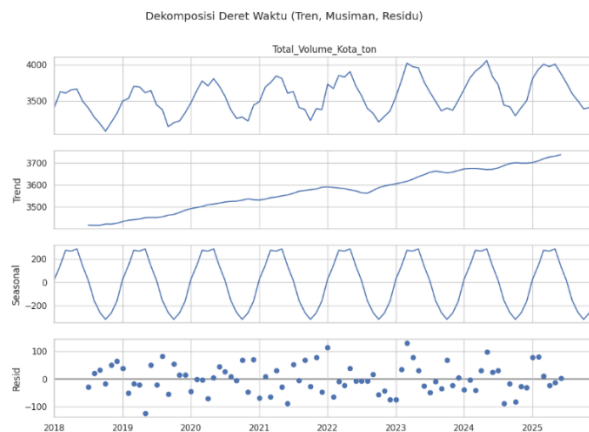
Tabel 7. Data Siap *Cluster* (Nilai *Scaling* / *Z-Score*)

Kecamatan	Organik (Score)	Plastik (Score)	Kertas (Score)	Lainnya (Score)	Waste per Capita (Score)
Pekalongan Barat	1.617	-1.213	1.211	-1.515	-1.226
Pekalongan Selatan	-0.979	1.342	-1.008	0.464	-0.744
Pekalongan Timur	0.013	0.535	-0.967	-0.158	1.018
Pekalongan Utara	-0.652	-0.663	0.764	1.210	0.952

Tabel 7. menampilkan data yang telah distandarisasi (*Z-Score*) untuk keperluan *clustering*, memperlihatkan seberapa jauh karakteristik setiap kecamatan menyimpang dari rata-rata kota. **Pekalongan Barat** tercatat sebagai wilayah paling unik dengan skor tertinggi pada sampah Organik (1,617) dan Kertas, namun justru terendah pada sampah Plastik (-1,213) dan Timbulan per Kapita (-1,226). Sementara itu, kecamatan lain memiliki karakteristik dominan masing-masing: **Pekalongan Selatan** tertinggi pada skor Plastik (1,342), **Pekalongan Timur** memimpin volume Timbulan per Kapita (1,018), dan **Pekalongan Utara** mendominasi jenis sampah Lainnya (1,210).

Analisis data tingkat kota memperlihatkan tren makro yang perlu diwaspadai, di mana total volume sampah kota mengalami kenaikan linear yang konsisten dari 3.417,17 ton pada Januari menjadi 3.664,20 ton pada Mei 2018. Kenaikan total sekitar 247 ton dalam lima bulan ini terjadi di saat jumlah penduduk kota tercatat konstan di angka 337.600 jiwa, yang mengonfirmasi bahwa laju produksi sampah per orang (*waste generation rate*) di kota ini sedang meningkat, menandakan perilaku masyarakat yang semakin konsumtif.

Untuk memetakan strategi penanganan yang tepat, analisis statistik menggunakan Metode Elbow mengindikasikan bahwa karakteristik sampah di empat kecamatan tersebut paling optimal dibagi menjadi dua kelompok utama, karena penurunan variasi data terlihat melandai setelah titik $k = 2$. Berdasarkan hasil algoritma tersebut, wilayah kota terbelah menjadi dua realitas berbeda: Cluster 1 yang hanya diisi oleh Pekalongan Barat, dan Cluster 0 yang menggabungkan Pekalongan Selatan, Timur, dan Utara karena memiliki kemiripan tantangan beban sampah.



Gambar 4. Dekomposisi Deret Waktu (*Time Series Decomposition*)

Gambar 4 membedah data *Total_Volume_Kota_ton* dari tahun 2018 hingga 2025 menjadi tiga komponen utama untuk memahami perilaku sampah kota secara mendalam. Komponen **Trend** memperlihatkan kenaikan beban sampah yang konsisten dalam jangka panjang, bergerak naik dari kisaran 3.400 ton pada 2018 hingga melampaui 3.700 ton di tahun 2025, menandakan pertumbuhan volume sampah yang terus menerus. Di sisi lain, komponen **Seasonal** (Musiman) mengungkap adanya pola gelombang yang sangat teratur dan berulang setiap tahun, mengindikasikan bahwa lonjakan dan penurunan volume sampah terjadi pada bulan-bulan tertentu secara dapat diprediksi. Sisa fluktuasi yang tidak dijelaskan oleh tren maupun musim ditangkap dalam plot **Resid** (Residu), yang berupa titik-titik acak di sekitar garis nol.

Pekalongan Barat dipisahkan sendiri karena memiliki profil "terbaik" dan paling unik dibandingkan kecamatan lain; data scaling menunjukkan wilayah ini memiliki dominasi sampah organik yang ekstrem dengan skor anomali +1,61 serta tingkat efisiensi tertinggi karena memiliki skor *waste per capita* terendah (-1,22). Sebaliknya, kelompok Cluster 0 menghadapi masalah yang lebih kompleks, di mana Pekalongan Selatan menjadi titik pusat masalah anorganik dengan skor plastik tertinggi (+1,34), sementara Pekalongan Timur dan Utara teridentifikasi sebagai wilayah dengan penduduk yang paling boros membuang sampah (skor volume per kapita tinggi), ditambah adanya anomali limbah residu spesifik di wilayah Utara.

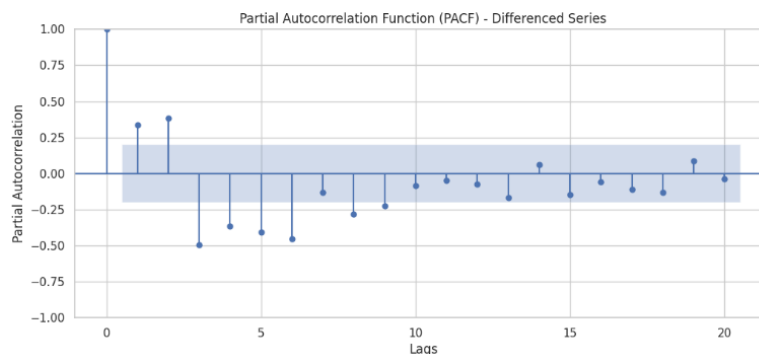
Analisis tren waktu menunjukkan peningkatan volume sampah kota yang konsisten dari Januari hingga Mei 2018 meskipun jumlah penduduk statis, yang mengonfirmasi adanya kenaikan

perilaku konsumtif masyarakat, termasuk lonjakan anomali sampah residu di Pekalongan Barat pada bulan Mei. Menggunakan *Metode Elbow*, wilayah dikelompokkan menjadi dua kluster utama: **Pekalongan Barat** (Cluster 1) yang berdiri sendiri sebagai profil paling efisien dengan dominasi organik tertinggi dan minim plastik, serta gabungan **Pekalongan Selatan, Timur, dan Utara** (Cluster 0) yang menghadapi beban lebih berat; di mana Selatan mengalami masalah plastik tertinggi, sedangkan Timur dan Utara paling boros secara per kapita. Dari sisi validasi teknis, uji *Augmented Dickey-Fuller* (ADF) menunjukkan data asli memiliki tren tidak stabil dengan p -value 0,96 ($>0,05$), namun setelah melalui proses *differencing* $d = 1$ data terkonfirmasi menjadi stasioner dengan p -value yang sangat kecil $6,9 \times 10^{-9}$ menjadikannya valid dan siap untuk pemodelan prediksi (*forecasting*) yang akurat.

Tabel 8. Validasi Teknis (Uji Stasioneritas ADF)

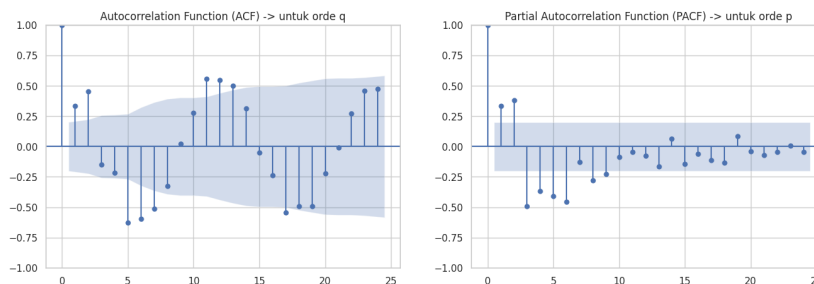
Tahap Data	Test Statistic	P-Value	Kesimpulan ($p < 0.05$)	Status Data
Data Asli	0.067629	0.963780	Tidak	Tidak Stasioner (Tren tidak stabil, belum siap prediksi)
Data Differencing ($d = 1$)	-6.594950	$6,9 \times 10^{-9}$	Ya	Stasioner (Stabil, siap untuk forecasting)

Tabel 8. merangkum hasil uji teknis Augmented Dickey-Fuller (ADF) untuk menentukan apakah data sudah cukup stabil untuk dilakukan peramalan (*forecasting*). Pada kondisi Data Asli, statistik menunjukkan bahwa data tersebut **Tidak Stasioner** dengan nilai P-Value yang sangat tinggi (0,963), yang berarti data masih memiliki tren yang berubah-ubah dan belum layak diproses. Namun, setelah dilakukan teknik transformasi **Differencing** ($d = 1$), status data berubah menjadi Stasioner dengan P-Value yang sangat kecil ($6,9 \times 10^{-9}$), menandakan bahwa data kini telah stabil dan siap digunakan untuk pemodelan prediksi masa depan.



Gambar 5. Plot PACF (*Partial Autocorrelation Function*)

Pada Gambar 5 dapat dijelaskan bahwa secara makro, tren volume sampah Kota Pekalongan meningkat konsisten dari Januari hingga Mei 2018 meskipun jumlah penduduk statis, yang mengonfirmasi adanya lonjakan perilaku konsumtif masyarakat. Pemetaan wilayah menggunakan *Metode Elbow* membagi kota menjadi dua realitas kontras: **Pekalongan Barat** (Cluster 1) sebagai profil paling efisien dengan dominasi organik tertinggi, berbanding terbalik dengan **Cluster 0** (Selatan, Timur, Utara) yang terbebani masalah polusi plastik tinggi dan volume sampah per kapita yang boros. Validasi teknis melalui uji ADF memastikan data siap untuk prediksi masa depan hanya setelah proses *differencing* ($d = 1$) karena data asli terbukti tidak stasioner. Terakhir, plot PACF mendeteksi signifikansi statistik yang kuat pada **Lag 1 hingga 3**, yang secara spesifik merekomendasikan penggunaan model **ARIMA** dengan komponen *AutoRegressive* (AR) orde 2 atau 3 sebagai metode peramalan yang paling akurat.



Gambar 6. Pemodelan ARIMA

Pada Gambar 6 dapat dilihat bahwa kota pekalongan mengalami kenaikan volume sampah konsisten dari Januari hingga Mei 2018 meski populasi stagnan, yang mengonfirmasi peningkatan perilaku konsumtif serta adanya anomali lonjakan residu di Pekalongan Barat pada bulan Mei. Analisis *Clustering* membagi wilayah menjadi dua realitas: **Pekalongan Barat** (Cluster 1) yang efisien dengan dominasi organik, kontras dengan **Cluster 0** (Selatan, Timur, Utara) yang menghadapi beban berat berupa polusi plastik tinggi dan volume per kapita yang boros. Validasi teknis melalui uji ADF memastikan data hanya siap diprediksi setelah proses *differencing* ($d = 1$) agar stasioner secara statistik. Terakhir, plot PACF merekomendasikan model prediksi **ARIMA** dengan komponen *AutoRegressive* (AR) orde 2 atau 3, mengingat adanya korelasi data yang signifikan terhadap pola 2-3 bulan sebelumnya.

Tabel 9. Informasi Model & Statistik Kecocokan (*Goodness of Fit*)

Kategori	Nilai	Kategori	Nilai
Dep. Variable	Total_Volume_Kota_ton	No. Observations	84
Model	ARIMA(1, 1, 1)	Log Likelihood	-519.004
Date	Thu, 11 Dec 2025	AIC	1044.007
Time	09:30:57	BIC	1051.264
Sample	01-01-2018 - 12-01-2024	HQIC	1046.923
Covariance Type	opg		

Tabel 9. merinci spesifikasi teknis dari model peramalan yang terpilih, yaitu **ARIMA(1, 1, 1)**, yang dibangun berdasarkan analisis terhadap 84 titik data observasi dari periode Januari 2018 hingga Desember 2024. Kualitas dan validitas model ini diukur melalui statistik kecocokan (*Goodness of Fit*) dengan nilai **AIC (1044.007)** dan **BIC (1051.264)**, yang menunjukkan tingkat efisiensi model dalam memetakan pola data historis untuk keperluan prediksi masa depan.

Tabel 10. Koefisien (*Coefficients*)

Parameter	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.5859	0.278	2.104	0.035	0.040	1.132
ma.L1	-0.2595	0.349	-0.744	0.457	-0.943	0.424
sigma2	1.585e+04	2903.448	5.460	0.000	1.02e+04	2.15e+04

Tabel 10. menjabarkan nilai koefisien yang menentukan kekuatan pengaruh setiap komponen dalam model prediksi. Parameter **AutoRegressive (ar.L1)** terbukti memiliki pengaruh yang signifikan secara statistik dengan nilai *P-Value* sebesar **0,035** (di bawah ambang batas 0,05), yang mengindikasikan bahwa pola data masa lalu berperan kuat dalam menentukan tren saat ini. Sebaliknya, parameter **Moving Average (ma.L1)** memiliki *P-Value* yang tinggi (**0,457**), menandakan bahwa komponen tersebut tidak memberikan kontribusi yang signifikan secara statistik terhadap model.

Tabel 11. Diagnostik Residual

Uji Statistik	Nilai	Uji Statistik	Nilai
Ljung-Box (L1) (Q)	1.29	Jarque-Bera (JB)	1.08
Prob(Q)	0.26	Prob(JB)	0.58
Heteroskedasticity (H)	1.18	Skew	0.05
Prob(H) (two-sided)	0.66	Kurtosis	2.45

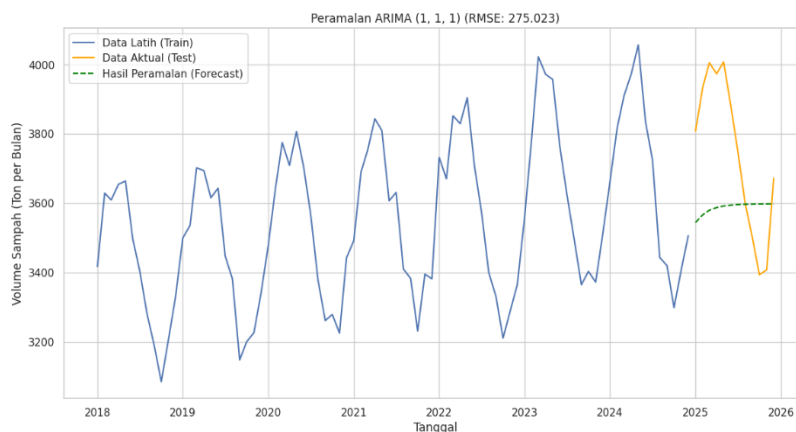
Tabel 11. menyajikan hasil uji diagnostik residual untuk memastikan bahwa model prediksi yang dibangun valid dan sehat secara statistik. Hasil uji Ljung-Box dengan probabilitas 0,26 ($>0,05$) mengonfirmasi bahwa error model bersifat acak (white noise) dan tidak memiliki autokorelasi, yang berarti model telah berhasil menangkap semua pola data yang ada. Validitas ini diperkuat oleh uji Jarque-Bera (Prob 0,58) dan **Heteroskedasticity** (Prob 0,66) yang menunjukkan bahwa residual terdistribusi secara normal dan memiliki varians yang stabil, sehingga hasil peramalan dapat dipercaya.

a) Evaluasi Signifikansi Parameter Model

Model ARIMA(1, 1, 1) yang diterapkan pada variabel Total_Volume_Kota_ton menggunakan 84 observasi (data bulanan dari 2018 hingga awal 2024). Berdasarkan tabel koefisien, komponen Auto Regressive (ar.L1) memiliki nilai p-value sebesar 0.035 (di bawah 0.05), yang menunjukkan bahwa pengaruh lag masa lalu signifikan secara statistik terhadap nilai saat ini. Sebaliknya, komponen Moving Average (ma.L1) memiliki p-value 0.457 (jauh di atas 0.05), yang mengindikasikan bahwa komponen ini tidak signifikan secara statistik dan mungkin tidak memberikan kontribusi berarti pada model. Nilai sigma2 (varians error) sangat signifikan ($P < 0.001$). AIC (1044.007) dan BIC (1051.264) dapat digunakan sebagai pembandingan jika Anda ingin mencoba model lain (misalnya ARIMA(1,1,0) karena MA tidak signifikan).

b) Pemeriksaan Diagnostik Sisaan (Residual)

Secara keseluruhan, diagnostik residual menunjukkan performa model yang baik dan memenuhi asumsi klasik. Uji Ljung-Box menghasilkan Prob(Q) sebesar 0.26, yang berarti hipotesis nol tidak ditolak; artinya, tidak ada autokorelasi yang tersisa pada residual (residual bersifat acak/white noise). Uji Heteroskedasticity dengan Prob(H) 0.66 menunjukkan varians residual konstan (homoskedastisitas), sehingga tidak ada masalah volatilitas yang berubah seiring waktu. Terakhir, uji Jarque-Bera memiliki Prob(JB) 0.58, yang mengonfirmasi bahwa distribusi residual adalah normal (didukung oleh nilai Skewness yang mendekati nol yaitu 0.05). Kesimpulannya, meskipun komponen MA tidak signifikan, model ini secara statistik menghasilkan residual yang valid dan memenuhi syarat untuk peramalan.



Gambar 7. Evaluasi Hasil Peramalan (Forecast Evaluation)

Grafik pada Gambar 7. memvisualisasikan hasil peramalan volume sampah (Ton per Bulan) menggunakan model ARIMA(1, 1, 1) dengan tingkat kesalahan RMSE sebesar 275.023. Garis biru (Data Latih) yang membentang dari tahun 2018 hingga awal 2025 memperlihatkan pola musiman (seasonality) yang sangat kuat dan berulang, dengan fluktuasi volume sampah yang berkisar antara terendah di sekitar 3.100 ton hingga puncak tertinggi yang melebihi 4.000 ton pada tahun 2024. Pola ini menunjukkan adanya siklus tahunan yang konsisten pada data historis.

Pada periode pengujian (2025-2026), terdapat perbedaan perilaku yang mencolok antara data aktual dengan hasil prediksi. Garis oranye (Data Aktual) menunjukkan volatilitas ekstrem berupa lonjakan tajam mendekati 4.000 ton yang diikuti penurunan drastis hingga di bawah 3.400 ton. Sebaliknya, garis putus-putus hijau (Hasil Peramalan) cenderung bergerak mendatar (flat) dan stabil di kisaran angka 3.600 ton. Hal ini mengindikasikan bahwa meskipun model ARIMA(1, 1, 1) mampu menangkap tren rata-rata (mean), model ini kurang responsif terhadap fluktuasi ekstrem atau perubahan mendadak yang terjadi pada data aktual terakhir.

3. KESIMPULAN

Penerapan metode ARIMA(1, 1, 1) untuk meramalkan volume sampah menghasilkan nilai RMSE (Root Mean Square Error) sebesar 275.023. Berdasarkan uji signifikansi parameter, komponen Auto Regressive (AR) berpengaruh signifikan ($p < 0.05$), namun komponen Moving Average (MA) ditemukan tidak signifikan ($p > 0.05$). Hasil peramalan pada periode pengujian (2025-2026) menunjukkan tren yang cenderung mendatar (stasioner) di kisaran 3.600 ton. Secara statistik, model ini memenuhi asumsi klasik dengan baik. Diagnostik residual menunjukkan bahwa sisaan model bersifat acak (white noise) berdasarkan uji Ljung-Box ($Prob(Q) > 0.05$) dan berdistribusi normal berdasarkan uji Jarque-Bera. Hal ini menandakan bahwa model telah berhasil menyerap informasi dasar dari data tanpa menyisakan pola autokorelasi pada sisaan. Kelemahan utama dari model ini adalah ketidakmampuannya menangkap pola musiman (seasonality) yang sangat kuat pada data historis. Grafik peramalan menunjukkan garis yang datar, sementara data aktual memiliki fluktuasi gelombang yang signifikan (naik-turun tajam). Akibatnya, model menjadi underfit dan kurang responsif terhadap lonjakan volume sampah yang ekstrem pada data aktual, sehingga akurasi berkurang untuk perencanaan jangka pendek yang membutuhkan presisi tinggi. Untuk meningkatkan akurasi peramalan di masa depan, disarankan melakukan pengembangan berikut:

1. **Penerapan Metode SARIMA:** Mengingat adanya pola siklus tahunan yang jelas, disarankan beralih ke model SARIMA (*Seasonal ARIMA*) untuk menangkap unsur musiman yang terlewatkan oleh ARIMA biasa.
2. **Evaluasi Ulang Parameter:** Mempertimbangkan untuk menghilangkan komponen MA (karena tidak signifikan) atau melakukan *hyperparameter tuning* ulang untuk menemukan kombinasi $(p, d, q)(P, D, Q)$ yang lebih optimal.
3. **Penambahan Variabel Eksogen:** Jika data tersedia, metode SARIMAX dapat digunakan dengan menambahkan variabel eksternal (seperti jumlah penduduk atau kegiatan ekonomi) yang mungkin mempengaruhi volume sampah.

UCAPAN TERIMA KASIH

Puji syukur *Alhamdulillah* penulis panjatkan ke hadirat **Allah SWT** atas segala limpahan rahmat, hidayah, dan karunia-Nya, sehingga penulis diberikan kesehatan dan kelancaran untuk menyelesaikan penelitian yang berjudul "*Trend Analysis and Forecasting of Household Solid Waste Generation in Pekalongan City Based on Machine Learning Using K-Means and ARIMA Algorithms*" ini dengan baik. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada **Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan** atas fasilitas akademik yang diberikan. Apresiasi setinggi-tingginya juga penulis sampaikan kepada **Dinas Lingkungan Hidup (DLH) Kota Pekalongan**. Kerja sama dan penyediaan data timbulan sampah serta data operasional TPA Degayu sangat esensial bagi pemodelan K-Means dan ARIMA dalam penelitian ini. Terima kasih penulis

sampaikan kepada **Ibu Umi Meganinditya Wulandari, S.Kom, M.Kom** selaku pembimbing yang telah memberikan arahan dan ilmu yang sangat berharga. Teristimewa, penulis mempersembahkan rasa terima kasih yang mendalam kepada **kedua orang tua tercinta**, serta kakak saya atas segala doa yang tiada putus, kasih sayang, serta dukungan moril dan materiil yang luar biasa. Ketulusan dan pengorbanan merekalah yang menjadi motivasi terbesar penulis untuk dapat menyelesaikan studi dan penelitian ini. Penulis juga mengucapkan terima kasih kepada *special person* (FM) yang selalu hadir memberikan dukungan, doa, dan semangat hingga penelitian ini dapat diselesaikan.

DAFTAR PUSTAKA

- [1] A. Abdeljaber, S. Al Smadi, M. Abu Talib, and M. Abdallah, "Comparative analysis of machine learning and conventional methods for waste generation forecasting," *Cleaner Engineering and Technology*, vol. 27, p. 100992, 2025.
- [2] Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan, "7,2 juta ton sampah di Indonesia belum terkelola dengan baik," Jul. 28, 2024. [Online]. Available: <https://www.kemenkopmk.go.id>
- [3] Kutnadi, "Pemkot Pekalongan-investor Malaysia sepakat bangun sampah jadi energi," *Antara News*, Jan. 12, 2025. [Online]. Available: <https://www.antaranews.com>
- [4] Kementerian Lingkungan Hidup dan Kehutanan, "Data timbulan sampah nasional tahun 2022," *Sistem Informasi Pengelolaan Sampah Nasional (SIPSN)*, 2023. [Online]. Available: <https://sipsn.menlhk.go.id>
- [5] T. A. Fauziyah, "Timbulan sampah Jateng terbanyak ketiga se-Indonesia, hasilkan 5,5 juta sampah setiap tahun," *Kompas.com*, Jan. 3, 2025. [Online]. Available: <https://www.kompas.com>.
- [6] S. Kaza, L. C. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC, USA: World Bank, 2018.
- [7] N. Sriploy and K. Lertpocasombut, "Application of ARIMA model for forecasting municipal solid waste generation in Thesis Municipality, Thailand," *Journal of Material Cycles and Waste Management*, vol. 22, pp. 123–132, 2020.
- [8] S. D. Latif, N. A. B. Hazrin, M. K. Younes, A. N. Ahmed, and A. Elshafie, "Investigating machine learning algorithms for municipal solid waste generation forecasting," *Environmental Science and Pollution Research*, vol. 31, pp. 12345–12356, 2024.
- [9] V.-N. Huynh, C. H. Nguyen, and T. M. Hoang, "A novel hybridization of ARIMA, ANN and K-Means for time series forecasting," *International Journal of Knowledge and Systems Science*, vol. 8, no. 2, pp. 36–51, 2017.
- [10] D. I. Mulyana and R. Franido, "Segmentasi citra grayscale dengan metode K-Means clustering gerak tangan bahasa isyarat Indonesia," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 11, no. 4, pp. 573–582, 2022, doi: 10.30591/smartcomp.v11i4.4243.
- [11] Badan Pusat Statistik Kota Pekalongan, *Kota Pekalongan dalam Angka 2024*. Pekalongan, Indonesia: BPS Kota Pekalongan, 2024.
- [12] Z. Ceylan, "Estimation of municipal waste generation of Turkey using socio-economic indicators by Bayesian optimization tuned Gaussian process regression," *Waste Management & Research*, vol. 38, no. 8, pp. 840–850, 2020.