

# Prediksi Volume Sampah Perkotaan Berbasis Data Spasial Menggunakan Random Forest di DKI Jakarta

**<sup>1</sup>Aftor Maulana, <sup>2</sup>Andika Firmansyah, <sup>3</sup>Anastasia R Ndun, <sup>4</sup>Aidina Ristyawan**

<sup>1-4</sup>Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail : <sup>1</sup>[aftorwtf02@gmail.com](mailto:aftorwtf02@gmail.com), <sup>2</sup>[andikaboy507@gmail.com](mailto:andikaboy507@gmail.com), <sup>3</sup>[ratnandun4@gmail.com](mailto:ratnandun4@gmail.com),  
<sup>4</sup>[aidinaristyawan@unpkediri.ac.id](mailto:aidinaristyawan@unpkediri.ac.id)

*Penulis Korespondens : Aidina Ristyawan*

**Abstrak**—Pengelolaan sampah di DKI Jakarta semakin menantang akibat pertumbuhan penduduk dan perubahan pola konsumsi yang konsumtif. Hal ini menyebabkan peningkatan volume sampah yang tidak seimbang dengan kapasitas pengelolaannya. Penelitian ini bertujuan memprediksi volume sampah menggunakan algoritma *Random Forest Regressor* berbasis data spasial, seperti luas permukiman, jumlah dan kepadatan penduduk. Data dikumpulkan dari BPS, Dinas Lingkungan Hidup DKI Jakarta, serta platform pemetaan digital, mencakup lebih dari 600 wilayah administratif. Hasil model menunjukkan performa prediksi yang tinggi dengan nilai  $R^2$  sebesar 0,87, *Mean Absolute Error* sebesar 15,3 ton, dan *Root Mean Squared Error* sebesar 21,7 ton. Fitur paling berpengaruh dalam prediksi adalah luas kawasan permukiman, jumlah penduduk, dan kepadatan wilayah. Model ini berpotensi membantu perencanaan distribusi armada dan fasilitas pengelolaan sampah secara lebih efektif. Penambahan variabel seperti aktivitas ekonomi dan mobilitas penduduk direkomendasikan untuk meningkatkan akurasi di masa depan.

**Kata Kunci**— prediksi volume sampah, *Random Forest Regressor*, data spasial, perencanaan sampah.

**Abstract**—Waste management in Jakarta faces increasing challenges due to population growth and shifting consumption patterns. These changes have led to a surge in waste volume that exceeds the available management capacity. This study aims to predict waste volume using the Random Forest Regressor algorithm based on spatial data, including residential area size, population count, and population density. Data were collected from BPS, the Environmental Agency of DKI Jakarta, and digital mapping platforms, covering over 600 administrative regions. The model demonstrated strong predictive performance with an  $R^2$  value of 0.87, a Mean Absolute Error of 15.3 tons, and a Root Mean Squared Error of 21.7 tons. The most influential features were residential area size, total population, and population density. This model can support more effective planning for fleet distribution and waste management facilities. Future improvements may include adding variables such as economic activity and population mobility to enhance prediction accuracy.

**Keywords**— waste volume prediction, *Random Forest Regressor*, spatial data, waste management planning.

This is an open access article under the CC BY-SA License.



## I. PENDAHULUAN

Jakarta terus-menerus dihadapkan pada keberagaman masalah sampah yang tak kunjung mereda. Berdasarkan data BPS Provinsi DKI Jakarta tahun 2019, kepadatan penduduk Jakarta pada 2018 mencapai 15.804 jiwa/km<sup>2</sup>, meningkat dibandingkan tahun 2010 sebesar 14.506 jiwa/km<sup>2</sup>, meskipun laju pertumbuhan penduduk sedikit menurun menjadi 1,07% per tahun (BPS DKI Jakarta, 2019). Jumlah penduduk yang terus berkembang ini diikuti oleh peningkatan volume sampah yang signifikan, di mana hingga November 2018, jumlah sampah yang masuk ke TPST Bantargebang mencapai 244.813 ton/hari atau 7.164,52 m<sup>3</sup>/hari (Dinas Lingkungan Hidup DKI Jakarta, 2019).

Peningkatan ini tidak terlepas dari perubahan pola konsumsi yang cenderung boros terhadap produk sekali pakai serta gaya hidup masyarakat urban yang konsumtif. Oleh karena itu, memproyeksikan pertumbuhan penduduk dan volume sampah di masa depan menjadi hal yang penting sebagai dasar dalam menyusun strategi pengelolaan sampah yang tepat, relevan, efektif, dan efisien. Tanpa strategi yang terarah, dampak negatif seperti polusi udara, sanitasi buruk, penurunan kualitas hidup, gangguan kesehatan masyarakat, potensi konflik sosial, serta meningkatnya beban ekonomi dan ekologi akibat pengelolaan sampah akan semakin sulit dihindari.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *Decision Tree* telah berhasil diterapkan dalam berbagai konteks sistem rekomendasi. Astuti et al. (2020) menunjukkan bahwa algoritma ID3 pada sistem rekomendasi pakaian berbasis web dapat menghasilkan akurasi tinggi berdasarkan karakteristik pengguna. Wu et al. (2019) menggabungkan *Decision Tree* dengan *deep learning* untuk merekomendasikan gaya berpakaian dan mendapatkan rata-rata rating 86,25% dari pengguna terhadap prototipe *user interface*. Agustina dan Maharani (2021) menggunakan *Decision Tree* untuk segmentasi pemasaran berbasis rating preferensi pada platform *Orebae.com*, sementara Kurniawan (2022) menerapkan model CART dalam merekomendasikan ukuran pakaian berdasarkan data fisik pengguna.

Meskipun demikian, sebagian besar penelitian tersebut belum mengintegrasikan penilaian rating pengguna secara eksplisit dalam proses klasifikasi, serta belum menghubungkan karakteristik pengguna, produk, dan preferensi secara langsung. Oleh karena itu, dibutuhkan pengembangan sistem rekomendasi yang tidak hanya menyarankan produk, tetapi juga dapat memprediksi nilai rating pengguna terhadap produk dengan lebih spesifik dan akurat.

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi pada platform e-commerce dengan pendekatan klasifikasi rating menggunakan algoritma *Decision Tree*. Penelitian ini ingin menyelidiki pola data penilaian dan preferensi pengguna terhadap produk pakaian. Dengan demikian, penggunaan model *Decision Tree* diharapkan dapat meningkatkan akurasi klasifikasi dan kualitas rekomendasi dibandingkan metode rekomendasi konvensional. Sistem yang dibangun diharapkan dapat berkontribusi dalam pengembangan sistem rekomendasi yang lebih personal, adaptif, dan berbasis data. Metode *Decision Tree* masih jarang digunakan dalam klasifikasi rating pada sistem rekomendasi jika dibandingkan dengan pendekatan seperti *collaborative filtering* dan *content-based filtering*, padahal memiliki keunggulan dari segi interpretabilitas dan efisiensi.

## II. METODE

### 2.1 Data Mining

*Data mining* atau penggalian data merupakan proses menggali informasi dalam jumlah besar untuk menemukan pola-pola dan detail penting yang tersembunyi dengan bantuan komputer sebagai alat utama. Dalam konteks pengelolaan sampah, *data mining* membantu dalam mengidentifikasi seberapa besar volume sampah yang dihasilkan, memprediksi peningkatan sampah di masa depan, serta mengkaji faktor-faktor yang memengaruhi fluktuasi volume sampah di berbagai lokasi. Teknik ini memungkinkan peneliti untuk mengolah, menganalisis, dan menampilkan data secara logis dan terstruktur, sehingga mampu menghasilkan rekomendasi dan prediksi yang lebih tepat sasaran.

### 2.2 Klasifikasi

Klasifikasi merupakan salah satu metode dalam *data mining* yang bertujuan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan karakteristik atau atribut-atribut yang dimilikinya. Dalam konteks penelitian ini, klasifikasi digunakan untuk membedakan wilayah berdasarkan volume sampah yang dihasilkan, sehingga dapat dilakukan analisis lanjutan terhadap wilayah dengan tingkat produksi sampah yang tinggi atau rendah. Teknik ini juga relevan untuk mengenali pola-pola spesifik yang dapat membantu dalam pembuatan keputusan dan perencanaan kebijakan pengelolaan sampah.

### 2.3 K-Means

K-Means adalah algoritma klasterisasi yang digunakan untuk membagi kumpulan data ke dalam sejumlah kelompok (*cluster*) berdasarkan kemiripan data satu sama lain. Dalam penelitian ini, algoritma K-Means diterapkan untuk mengelompokkan wilayah administratif di Jakarta berdasarkan variabel seperti volume produksi sampah, pola konsumsi, dan kepadatan penduduk. Proses ini membantu mengidentifikasi karakteristik wilayah tertentu yang mungkin memerlukan pendekatan pengelolaan sampah yang berbeda.

### 2.4 Proses Data Mining

Data penelitian diperoleh dari berbagai sumber resmi, seperti Badan Pusat Statistik (BPS) DKI Jakarta dan Dinas Lingkungan Hidup Provinsi DKI Jakarta. Selain itu, koordinat geografis serta peta wilayah digital dikumpulkan dari platform pemetaan terbuka seperti OpenStreetMap (OSM) dan Google Maps API. Data historis mengenai volume sampah diperoleh dari laporan Dinas Lingkungan Hidup DKI Jakarta dan mencakup jenis serta asal sampah.

#### 1. Sumber Data Lokasi

Data lokasi meliputi informasi geografis wilayah administratif seperti kecamatan dan kelurahan yang diperoleh dari BPS DKI Jakarta dan Dinas Lingkungan Hidup Provinsi DKI Jakarta. Koordinat geografis serta peta digital wilayah diperoleh dari OSM dan Google Maps API untuk keperluan pemetaan spasial.

#### 2. Data Dimensi Area

Informasi mengenai luas wilayah setiap unit administratif (dalam kilometer persegi) dikumpulkan dari dokumen Rencana Tata Ruang Wilayah (RTRW) dan basis data GIS (Geographic

Information System) yang dikelola oleh Pemerintah Provinsi DKI Jakarta. Data ini mencakup detail seperti luas permukiman, kawasan komersial, dan fasilitas umum.

### 3. Data Volume Sampah

Data volume sampah diperoleh dari laporan resmi Dinas Lingkungan Hidup DKI Jakarta yang mencakup data harian volume sampah per wilayah dari tahun 2018 hingga 2023. Data ini juga mengandung informasi terkait jenis sampah (organik, anorganik, dan residu) serta asal sampah (rumah tangga, pasar, kawasan industri, dan sebagainya).

### 4. Metode Pengumpulan dan Validasi

Seluruh data dari berbagai sumber disatukan dan melalui proses *preprocessing* seperti normalisasi satuan, pengisian nilai hilang, dan penghapusan duplikasi. Validasi dilakukan dengan membandingkan data dari dua atau lebih sumber berbeda untuk memastikan konsistensi dan akurasi informasi.

Dengan menggabungkan data lokasi, dimensi area, dan volume sampah aktual, penelitian ini menghasilkan dataset akhir yang mencakup lebih dari 600 titik wilayah administratif. Dataset tersebut kemudian digunakan sebagai dasar pelatihan dan pengujian dalam pembangunan model prediksi volume sampah yang berbasis metode *data mining*.

## III. HASIL DAN PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa penggunaan algoritma Random Forest Regressor dalam memprediksi volume sampah di DKI Jakarta sangat efektif. Model yang dibangun berhasil mencapai nilai  $R^2$  sebesar 0,87, yang menunjukkan bahwa 87% variasi dalam volume sampah dapat dijelaskan oleh variabel-variabel yang digunakan dalam model. Ini adalah angka yang sangat baik, menunjukkan bahwa model ini dapat diandalkan untuk memprediksi volume sampah di masa depan.

### 3.1 Hasil Prediksi

Penelitian ini menggunakan model regresi berbasis algoritma Random Forest Regressor untuk memprediksi volume sampah berdasarkan data lokasi (kecamatan/kelurahan) dan dimensi area (luas wilayah dalam  $km^2$ , luas kawasan organisasi, dan komersial). Dataset dibagi menjadi 80% data latih dan 20% data uji. Berikut adalah metrik evaluasi model:

Tabel 1. Hasil Prediksi

Metrik Evaluasi	Hasil
Mean Absolute Error (MAE)	15,3 ton
Root Mean Squared Error (RMSE)	21,7 ton
$R^2$ Score	0.87

Performa model menunjukkan cukup baik dengan nilai  $R^2$  sebesar 0.87, yang berarti model mampu menjelaskan 87% variasi volume sampah berdasarkan variabel input.

### 3.2. Analisis Hasil

Model Random Forest menunjukkan performa yang lebih baik dibandingkan regresi linier biasa ( $R^2 = 0.63$ ). Hal ini menunjukkan bahwa hubungan antara luas wilayah dan volume sampah bersifat non-linier dan kompleks, yang dapat ditangkap lebih baik oleh model berbasis ensemble seperti Random Forest.

1. Berdasarkan analisis kepentingan fitur, fitur paling signifikan dalam prediksi volume sampah adalah:

- Luas kawasan industri
- Jumlah penduduk per wilayah (data tambahan)
- Luas total wilayah
- Kepadatan penduduk

Area organisasi dan kepadatan penduduk memiliki korelasi positif yang kuat terhadap volume sampah. Wilayah dengan organisasi padat dan luas cenderung menghasilkan volume sampah lebih tinggi.

2. Kesalahan Distribusi:

Kesalahan prediksi cenderung lebih tinggi pada wilayah yang memiliki aktivitas non-permukiman dominan, seperti kawasan industri atau perdagangan besar. Hal ini menunjukkan bahwa faktor aktivitas ekonomi juga berpengaruh terhadap volume sampah, namun belum sepenuhnya tercakup dalam model.

### 3.3 Pembahasan

Hasil penelitian ini menunjukkan bahwa data spasial (lokasi dan luas wilayah) dapat digunakan secara efektif untuk memprediksi volume sampah di wilayah perkotaan. Model prediktif ini dapat digunakan oleh Dinas Lingkungan Hidup sebagai alat bantu perencanaan pengangkutan sampah, dengan menyesuaikan kapasitas armada berdasarkan estimasi volume per wilayah. Namun, untuk meningkatkan akurasi di masa depan, disarankan untuk menambahkan variabel lain seperti:

- Tingkat ekonomi masyarakat
- Jumlah fasilitas umum (pasar, mall, sekolah)
- Data mobilitas penduduk harian

Dengan menambahkan variabel-variabel ini, model diharapkan dapat memberikan prediksi yang lebih akurat dan relevan, sehingga dapat membantu pemerintah dalam merencanakan strategi pengelolaan sampah yang lebih baik dan lebih efisien. Secara keseluruhan, penelitian ini memberikan kontribusi yang signifikan terhadap pengelolaan sampah di DKI Jakarta. Dengan menggunakan data spasial dan algoritma Random Forest, model ini tidak hanya memberikan prediksi yang akurat tetapi juga memberikan wawasan yang berharga bagi pemerintah dalam merencanakan strategi pengelolaan sampah yang lebih baik di masa depan.

## IV. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma Random Forest Regressor dapat digunakan secara efektif untuk memprediksi volume sampah di DKI Jakarta berdasarkan data

spasial, seperti lokasi wilayah administratif dan dimensi area. Model yang dikembangkan mencapai nilai  $R^2$  sebesar 0,87, yang menunjukkan bahwa 87% variasi dalam volume sampah dapat dijelaskan oleh variabel-variabel yang digunakan. Metrik evaluasi lainnya, seperti Mean Absolute Error (MAE) sebesar 15,3 ton dan Root Mean Squared Error (RMSE) sebesar 21,7 ton, juga menunjukkan performa yang baik dari model ini.

Analisis lebih lanjut mengungkapkan bahwa fitur-fitur seperti luas kawasan organisasi, jumlah penduduk, dan kepadatan penduduk memiliki pengaruh yang signifikan terhadap volume sampah. Selain itu, distribusi error prediksi menunjukkan bahwa wilayah dengan aktivitas non-permukiman, seperti kawasan industri, memerlukan perhatian lebih dalam model ini.

Dengan demikian, model prediktif ini dapat menjadi alat bantu yang berguna bagi Dinas Lingkungan Hidup dalam merencanakan transportasi dan pengelolaan sampah secara lebih efektif. Untuk meningkatkan akurasi prediksi di masa mendatang, disarankan untuk mempertimbangkan variabel tambahan seperti tingkat ekonomi masyarakat, jumlah fasilitas umum, dan mobilitas data penduduk. Penelitian ini memberikan kontribusi penting dalam pengelolaan sampah di DKI Jakarta dan dapat menjadi dasar untuk penelitian lebih lanjut di bidang ini.

## DAFTAR PUSTAKA

- [1] A. Kahfi, "Overview of Waste Management," *Jurisprud. Dep. Law, Fac. Sharia Law*, vol. 4, no. 1, p. 12, 2017.
- [2] M. Iqbal, R. M. Mulyadin, K. Ariawan, and S. Subarudi, "Analisis Implementasi Kebijakan Pengelolaan Sampah Di Provinsi Dki Jakarta," *J. Anal. Kebijak. Kehutan.*, vol. 19, no. 2, pp. 129–140, 2022, doi: 10.20886/jakk.2022.19.2.129-140.
- [3] R. Mohamad Mulyadin, K. Ariawan, and M. Iqbal, "Conflict of Waste Management in Dki Jakarta and Its Recomended Solutions," *J. Anal. Kebijak. Kehutan.*, vol. 15, no. 2, pp. 179–191, 2018, doi: 10.20886/jakk.2018.15.2.179-191.
- [4] M. R. Diani, D. Haniifah, and F. R. Dianty, "Analisis proyeksi pertumbuhan penduduk dan volume sampah DKI Jakarta terhadap dampak yang ditimbulkan," *J. Waste Sustain. Consum.*, vol. 1, no. 1, pp. 27–45, 2024, doi: 10.61511/jwsc.v1i1.2024.691.
- [5] B. A. R. Apriyeni, *Sejarah dan Perkembangan Awal Informasi Geospasial*. 2023.
- [6] C. Puspawati and B. Besral, "Pengelolaan Sampah Berbasis Masyarakat di Kampung Rawajati Jakarta Selatan," *Kesmas Natl. Public Heal. J.*, vol. 3, no. 1, p. 9, 2008, doi: 10.21109/kesmas.v3i1.237.
- [7] M. J. Bastiaans, "New Class of Uncertainty Relations for Partially Coherent Light.," *13th Congr. Int. Comm. Opt. Opt. Mod. Sci. Technol. Conf. Dig.*, pp. 638–639, 1984, doi: 10.1364/josaa.1.000711.
- [8] A. Fitria Salsabella, B. Suprianto, and M. Syariffuddien Zuhrie, "Analisis Pengeringan Sampah Pada Sungai Kawasan Jakarta Pusat Dengan Sistem Inferensi Fuzzy Takagi-Sugeno Berbasis Matlab," *J. Tek. Elektro*, vol. 11, no. 2, pp. 288–296, 2022.
- [9] D. Ardiatma, P. Lestari, and M. Chaerul, "Real data mapping of DKI Jakarta waste generation using the K-mean Clustering method at final disposal Bantargebang," *E3S Web Conf.*, vol. 485, 2024, doi: 10.1051/e3sconf/202448502015.
- [10] D. Ruwandara, M. Jajuli, and A. Rizal, "Analisis Algoritma K-Means Clustering Untuk Daerah Penyebaran Sampah di Kota Bekasi," *JOINS (Journal Inf. Syst.)*, vol. 6, no. 1, pp. 56–63, 2021, doi: 10.33633/joins.v6i1.4085.