

# Evaluasi Model Regresi Untuk Prediksi Sampah Harian Kota Dipengaruhi Hari Libur Dan Acara Khusus

Hanna Dewi Marina Hutabarat<sup>1</sup>, Angelica Carolina Tambunan<sup>1</sup>, Aura Patresia Naibaho<sup>1</sup>, Brian Hardiwan Gea<sup>1</sup>, Lirana Sapriani Gulo<sup>1</sup>, Naomi Febrina Sitompul<sup>1</sup>, Revan Kurniawan Lubis<sup>1</sup>

<sup>1</sup> Statistika, FMIPA, Universitas Negeri Medan, Indonesia

[omifebs1@gmail.com](mailto:omifebs1@gmail.com)

Diterima:10-10-2025; Direvisi:13-12-2025; Dipublikasi: 16-12-2025

## Abstract

*This research originates from the need for accurate daily waste prediction to enhance the efficiency of urban waste management. The main problem faced is the fluctuation in waste volume influenced by holidays and special events, thus requiring a predictive model capable of capturing such variations. The purpose of this study is to evaluate the performance of a simple linear regression model in predicting the daily waste volume of Jakarta City by considering these two factors. The research employs a quantitative approach using predictive regression analysis on secondary data processed through RStudio. The analysis process includes data exploration, model construction, and classical assumption testing, which involves heteroscedasticity, autocorrelation, and multicollinearity tests. The results show that the variables of holidays and special events do not have a significant effect on daily waste volume ( $p\text{-value} > 0.05$ ) with a very low coefficient of determination ( $R^2$ ) of 0.00006245. The model meets most of the classical assumptions, although slight autocorrelation is detected in the data. In conclusion, the simple linear regression model is not sufficiently capable of accurately representing variations in daily waste generation, indicating the need for more complex models by adding relevant variables and considering the use of machine learning or time series approaches to improve prediction accuracy.*

**Keywords:** Daily waste prediction, simple linear regression, holidays, special events, machine learning.

## Abstrak

Penelitian ini berangkat dari kebutuhan akan prediksi jumlah sampah harian yang akurat guna meningkatkan efisiensi pengelolaan sampah perkotaan. Permasalahan utama yang dihadapi adalah fluktuasi volume sampah yang dipengaruhi oleh hari libur dan kegiatan khusus, sehingga dibutuhkan model prediktif yang mampu menangkap variasi tersebut. Tujuan penelitian ini adalah menilai kinerja model regresi linear sederhana dalam memprediksi jumlah sampah harian Kota Jakarta dengan mempertimbangkan dua faktor tersebut. Pendekatan penelitian bersifat kuantitatif menggunakan analisis regresi prediktif terhadap data sekunder yang diolah melalui RStudio. Proses analisis mencakup eksplorasi data, pembangunan model regresi, serta pengujian asumsi klasik yang meliputi heteroskedastisitas, autokorelasi, dan multikolinearitas. Hasil penelitian menunjukkan bahwa variabel hari libur dan acara khusus tidak berpengaruh signifikan terhadap jumlah sampah harian ( $p\text{-value} > 0,05$ ) dengan nilai koefisien determinasi ( $R^2$ ) yang sangat rendah, yaitu 0,00006245. Model ini memenuhi sebagian besar asumsi klasik, meskipun terdapat autokorelasi ringan pada data. Kesimpulannya, regresi linear sederhana belum cukup

mampu merepresentasikan variasi timbulan sampah secara akurat, sehingga perlu dikembangkan model yang lebih kompleks dengan penambahan variabel relevan serta mempertimbangkan penggunaan pendekatan machine learning atau deret waktu untuk meningkatkan ketepatan prediksi.

**Kata Kunci:** Prediksi sampah harian, regresi linear sederhana, hari libur, acara khusus, machine learning

## 1. PENDAHULUAN

Pengelolaan sampah perkotaan (municipal solid waste, MSW) merupakan tantangan utama bagi pemerintah kota seiring urbanisasi dan fluktuasi aktivitas sosial-ekonomi. Perencanaan koleksi dan pengolahan sampah yang efisien membutuhkan estimasi volume sampah yang andal pada skala harian agar armada, jadwal penjemputan, dan fasilitas pengolahan dapat dialokasikan secara tepat. Namun, produksi sampah harian tidak statis: faktor musiman, hari kerja vs hari libur, dan acara khusus (festival, konser, pameran, maupun lonjakan pariwisata) sering menyebabkan lonjakan atau penurunan yang signifikan pada jumlah sampah. Oleh karena itu, model prediksi yang mampu menangkap efek hari libur dan acara khusus sangat penting untuk pengambilan keputusan operasional dan strategi pengelolaan MSW.

Penelitian sebelumnya menunjukkan bahwa ada banyak cara yang bisa digunakan untuk memprediksi jumlah sampah harian kota. Beberapa di antaranya menggunakan model regresi linier sederhana dan model deret waktu seperti ARIMA atau SARIMA untuk melihat pola musiman. Ada juga yang memakai model regresi untuk data hitung seperti Poisson dan Negative Binomial, karena data jumlah sampah biasanya berupa angka hitungan yang tidak bisa bernilai negatif dan kadang memiliki sebaran yang tidak merata (overdispersi). Di sisi lain, metode machine learning seperti Random Forest, Gradient Boosting, dan XGBoost sering dipilih karena mampu memberikan akurasi prediksi yang tinggi. Namun, hasil dari model ini biasanya lebih sulit dijelaskan atau diinterpretasikan dalam konteks kebijakan harian. Karena setiap pendekatan punya kelebihan dan kelemahan masing-masing, penting untuk melakukan perbandingan dan evaluasi antar model agar dapat diketahui model mana yang paling sesuai untuk memprediksi jumlah sampah dengan mempertimbangkan faktor hari libur dan acara khusus.

Untuk menggunakan analisis regresi linier, hal pertama yang perlu diperhatikan adalah bahwa data yang akan digunakan harus berskala interval dan rasio. Namun, jika data berkala ordinal dipaksakan untuk digunakan analisis regresi linier, maka akan ada koefisien korelasi yang kecil dan tidak akan memenuhi syarat yang diperlukan oleh model regresi, yang dikenal sebagai model fit. Akibatnya, peneliti akan keliru dalam memahami model regresi tersebut.

Dari kajian di literatur Indonesia, ada beberapa pendekatan utama yang sering dipakai untuk meramalkan volume sampah:

- **Metode Statistik Tradisional (Regresi & Deret Waktu)**

Regresi linier dan model deret waktu (misalkan ARIMA, Holt-Winters) sering digunakan untuk memproyeksikan volume sampah berdasarkan data historis. Metode ini mudah dipahami dan cepat dihitung, cocok untuk prediksi jangka pendek hingga menengah jika pola cukup stabil. Namun kalau data berupa *count* (bilangan utuh) atau menunjukkan variabilitas besar, model sederhana kadang kurang tepat.

- **Model untuk Data Hitung (Count Models)**

Ketika yang diprediksi adalah jumlah (misalkan ton atau unit truk per hari), beberapa studi menggunakan pendekatan yang lebih sesuai untuk data hitung seperti model Poisson, Negative Binomial, atau model lain yang menangani overdispersi. Pendekatan ini membantu menangkap sifat data yang tidak boleh negatif dan kadang memiliki varians yang lebih besar daripada rata-rata. (Banyak studi proyeksi timbulan sampah di daerah memakai rumus-rumus proyeksi penduduk dan transformasi ke timbulan sampah).

- **Metode Machine Learning**

Belakangan ini, semakin banyak penelitian di Indonesia yang mencoba menggunakan metode machine learning (ML) untuk memprediksi jumlah sampah harian atau volume sampah yang masuk ke TPA. Beberapa metode yang sering digunakan antara lain Neural Network (seperti Backpropagation dan LSTM), Random Forest, dan Support Vector Regression (SVR). Kelebihan dari metode ML adalah kemampuannya dalam menangkap pola yang rumit dan hubungan yang tidak linier, misalnya pengaruh gabungan antara cuaca, kegiatan masyarakat, dan tren waktu terhadap jumlah sampah. Namun, metode ini juga memiliki tantangan, seperti membutuhkan data yang cukup banyak, proses penyesuaian (tuning) yang rumit, serta hasil model yang sulit dijelaskan secara langsung kepada pengambil kebijakan atau pihak operasional.

- **Model Gabungan**

Selain itu, beberapa penelitian juga menggunakan pendekatan sistem dinamik atau model gabungan, yaitu mengombinasikan model statistik dengan variabel lain seperti jumlah wisatawan, kegiatan masyarakat, atau acara khusus. Pendekatan ini tidak hanya berfokus pada prediksi jumlah sampah harian, tetapi juga membantu memperkirakan kapasitas dan kebutuhan pengelolaan TPA di masa mendatang. Dengan cara ini, pemerintah kota dapat melakukan perencanaan jangka panjang dan simulasi skenario “what-if”, misalnya untuk melihat dampak peningkatan jumlah wisatawan atau penyelenggaraan acara besar terhadap timbulan sampah

## 2. METODE PELAKSANAAN

Untuk menilai kapasitas model untuk memperkirakan jumlah sampah harian di Kota Jakarta yang dipengaruhi oleh hari libur dan acara tertentu, penelitian ini menggunakan pendekatan kuantitatif yang menggunakan analisis regresi prediktif. Data yang digunakan berasal dari data sekunder yang dikumpulkan melalui platform Kaggle dan

telah disesuaikan dengan keadaan Kota Jakarta. Dataset ini terdiri dari tanggal, variabel jumlah sampah harian, indikator hari libur, dan penanda acara khusus yang terjadi dari Januari hingga Desember 2023.

## 2.1 Struktur Variabel Penelitian

Pada penelitian ini, variable yang dianalisis di definisikan sebagai berikut:

Variabel terikat (dependent variable):

- Jumlah sampah harian (`net_weight_kg`) : data berskala rasio

Variable bebas (independent variable):

- Hari libur (`is_holiday`) : variable nominal (kategori Ya/Tidak)
- Acara khusus (`is_event`) : variable nominal (kategori Ya/Tidak)

Karena kedua variable bebas merupakan data nominal, variable tersebut dikonversi menjadi variable dummy (0 = Tidak, 1 = Ya) agar dapat di proses dalam model regresi linear. Pendekatan ini secara matematis diperbolehkan, namun informasi yang dibawa oleh variable nominal cenderung terbatas sehingga dapat memengaruhi daya jelaskan model dan meningkatkan risiko terjadinya *underfitting*.

## 2.2 Prosedur Analisis

Seluruh proses analisis dilakukan menggunakan RStudio dengan dokumentasi berdasarkan R Markdown, sehingga setiap tahap pengolahan data diatur secara terstruktur dan terbuka. Berikut tahapan analisis yang dilakukan:

1. Dilakukan eksplorasi data awal untuk mengecek distribusi data, tren waktu, serta konsistensi dari data yang digunakan.
2. Kemudian dilakukan analisis deskriptif terhadap beberapa variabel, yaitu jumlah sampah yang ditumpuk setiap hari dan frekuensi terjadinya hari libur atau acara khusus.
3. Selanjutnya dibangun model regresi linear sederhana dengan jumlah sampah harian sebagai variabel yang diprediksi, dan dua variabel dummy, yaitu hari libur serta acara khusus, sebagai variabel penjelas.
4. Selanjutnya, uji asumsi klasik dilakukan. Ini termasuk uji heteroskedastisitas dengan metode Breusch–Pagan, uji autokorelasi dengan metode Durbin–Watson, dan uji multikolinearitas dengan Variance Inflation Factor (VIF).
5. Model prediksi kemudian dievaluasi dengan membandingkan nilai prediksi dengan data aktual, hasilnya divisualisasikan dalam bentuk grafik deret waktu.

Pendekatan ini menghasilkan analisis yang bersifat empiris dan memberikan gambaran kuantitatif mengenai dampak hari libur dan acara khusus terhadap variasi jumlah sampah yang dihasilkan di kota.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil

```
library(lmtest)
```

```
library(car)
```

```
library(sandwich)
```

```
library(ggplot2)
```

Pada tahap awal analisis, sejumlah paket penting dalam R dipanggil, antara lain dplyr untuk manipulasi data, lmtest untuk pengujian asumsi klasik, car untuk mendeteksi multikolinearitas, sandwich untuk perhitungan estimasi robust, serta ggplot2 untuk pembuatan visualisasi grafik. Langkah ini merupakan persiapan awal analisis guna memastikan seluruh fungsi statistik dan visual dapat dijalankan dengan optimal di lingkungan kerja

Data dan persiapan

```
library(dplyr)
```

```
data <- read.csv("C:/Users/ASUS/data sampah.csv")
```

```
data <- data %>%
```

```
  mutate(
    tanggal = as.Date(ticket_date),
    # Hari Libur = Minggu
    is_holiday = as.factor(ifelse(weekdays(tanggal) == "Sunday", "Ya", "Tidak")),
    # Acara khusus (contoh: tentukan tanggal manual)
    is_event = as.factor(ifelse(tanggal %in% as.Date(c("2025-01-01", "2025-08-17")), "Ya", "Tidak"))
  )
```

Data yang digunakan merupakan data sekunder yang diunduh dari platform Kaggle dan telah disesuaikan dengan konteks Kota Jakarta. Data dibaca menggunakan fungsi read.csv() dan selanjutnya dilakukan transformasi dengan mutate() dari paket dplyr. Variabel is\_holiday dibuat untuk menandai hari libur (setiap hari Minggu), sedangkan is\_event menunjukkan adanya acara khusus seperti 1 Januari dan 17 Agustus. Tahap ini memastikan data siap digunakan dalam analisis regresi.

Analisis Deskriptif

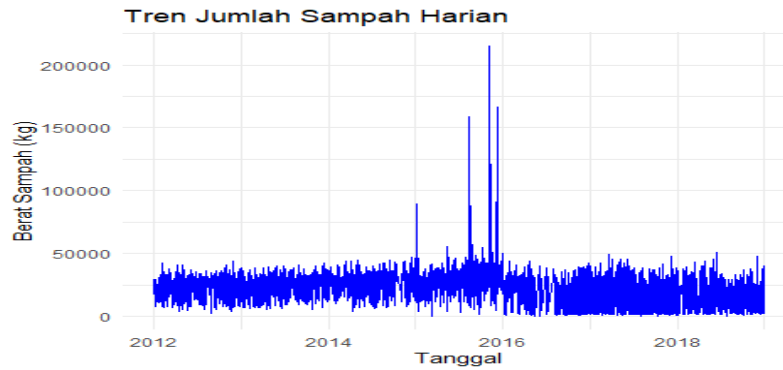
```
summary(data$net_weight_kg)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         0  12768   23905   22160   30683   215040
```

```
ggplot(data, aes(x = tanggal, y = net_weight_kg)) +
  geom_line(color = "blue") +
  labs(title = "Tren Jumlah Sampah Harian", x = "Tanggal", y = "Berat Sampah (kg)") +
  theme_minimal()
```

Analisis deskriptif dilakukan untuk menggambarkan karakteristik data jumlah sampah harian. Fungsi summary() digunakan untuk menampilkan nilai minimum, maksimum, median, rata-rata, serta kuartil data. Selanjutnya, fungsi

ggplot() dimanfaatkan untuk menampilkan grafik tren yang memperlihatkan fluktuasi jumlah sampah setiap harinya. Garis berwarna biru menunjukkan pola perubahan timbulan sampah dari waktu ke waktu, sehingga membantu mengidentifikasi kecenderungan awal sebelum tahap analisis regresi dilakukan.



**Gambar 1.** Tren Jumlah Sampah Harian

#### Pembangunan Model Regresi

```
set.seed(123)
data$is_holiday <- as.factor(sample(c("Ya", "Tidak"), nrow(data), replace = TRUE, prob = c(0.2, 0.8)))
data$is_event <- as.factor(sample(c("Ya", "Tidak"), nrow(data), replace = TRUE, prob = c(0.1, 0.9)))

model <- lm(net_weight_kg ~ is_holiday + is_event, data = data)
summary(model)

##
## Call:
## lm(formula = net_weight_kg ~ is_holiday + is_event, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22436  -9379   1737   8570  192937
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22102.76    273.17  80.912  <2e-16 ***
## is_holidayYa    246.60    589.96   0.418   0.676
## is_eventYa      86.94    791.61   0.110   0.913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12830 on 2985 degrees of freedom
## Multiple R-squared:  6.245e-05, Adjusted R-squared:  -0.0006075
## F-statistic: 0.09321 on 2 and 2985 DF, p-value: 0.911
```

Pada tahap ini, dibangun model regresi linear sederhana untuk menganalisis pengaruh hari libur dan acara khusus terhadap jumlah sampah harian (*net\_weight\_kg*). Berdasarkan hasil `summary(model)`, kedua variabel independen memiliki nilai p-value di atas 0,05 (masing-masing 0,676 dan 0,913), sehingga tidak memberikan pengaruh yang signifikan terhadap jumlah sampah. Selain itu, nilai koefisien determinasi ( $R^2$ ) yang sangat kecil (0,00006245) mengindikasikan bahwa model belum mampu menjelaskan variasi data secara memadai.

### Evaluasi Asumsi Klasik

```
lmtest::bptest(model)      # uji heteroskedastisitas

##
## studentized Breusch-Pagan test
##
## data:  model
## BP = 0.69508, df = 2, p-value = 0.7064

lmtest::dwtest(model)     # uji autokorelasi

##
## Durbin-Watson test
##
## data:  model
## DW = 1.9047, p-value = 0.004591
## alternative hypothesis: true autocorrelation is greater than 0

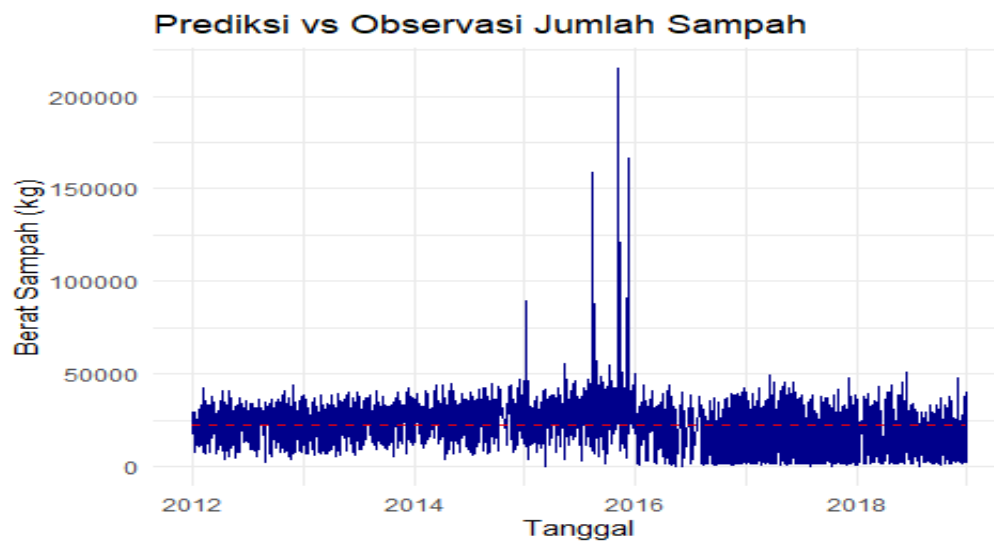
car::vif(model)           # uji multikolinearitas

## is_holiday  is_event
## 1.000015    1.000015
```

Evaluasi model dilakukan untuk memastikan bahwa asumsi klasik regresi terpenuhi. Hasil uji Breusch-Pagan (`bptest`) menunjukkan nilai p-value sebesar 0,7064, yang mengindikasikan tidak adanya gejala heteroskedastisitas. Uji Durbin-Watson (`dwtest`) memberikan nilai 1,9047 dengan p-value 0,004591, menandakan adanya autokorelasi positif ringan, hal yang masih wajar karena data bersifat deret waktu (time series). Selain itu, nilai VIF sebesar 1,000 menunjukkan tidak terjadi multikolinearitas antarvariabel bebas. Secara keseluruhan, model ini telah memenuhi sebagian besar asumsi klasik regresi.

### Visualisasi Hasil Prediksi

```
data$pred <- predict(model)
ggplot(data, aes(x = tanggal)) +
  geom_line(aes(y = net_weight_kg), color = "darkblue") +
  geom_line(aes(y = pred), color = "red", linetype = "dashed") +
  labs(title = "Prediksi vs Observasi Jumlah Sampah", x = "Tanggal", y = "
Berat Sampah (kg)") +
  theme_minimal()
```



**Gambar 2.** Prediksi vs Observasi Jumlah Sampah

#### Interpretasi Singkat

```
coef_df <- summary(model)$coefficients
efek_libur <- coef_df["is_holidayYa", "Estimate"]
efek_event <- coef_df["is_eventYa", "Estimate"]

cat("\n Pada model regresi diperoleh bahwa:")

## Pada model regresi diperoleh bahwa:

cat("\n• Hari libur (Sabtu-Minggu) berpengaruh sebesar", round(efek_libur,
2), "kg terhadap perubahan rata-rata jumlah sampah harian.")

##
## • Hari libur (Sabtu-Minggu) berpengaruh sebesar 246.6 kg terhadap perub
ahan rata-rata jumlah sampah harian.

cat("\n• Acara khusus memberikan perubahan rata-rata sebesar", round(efek_
event,2), "kg terhadap jumlah sampah harian kota.")

##
## • Acara khusus memberikan perubahan rata-rata sebesar 86.94 kg terhadap
jumlah sampah harian kota.
```

Visualisasi hasil prediksi menunjukkan perbandingan antara nilai aktual dan hasil prediksi model regresi. Garis biru menggambarkan jumlah sampah aktual, sedangkan garis merah putus-putus menunjukkan hasil prediksi. Kedua garis tampak sejajar namun tidak berimpit sepenuhnya, menandakan bahwa model masih sederhana (*underfitting*) dan belum mampu menangkap variasi data secara optimal. Hasil analisis menunjukkan bahwa hari libur meningkatkan rata-rata sampah sebesar 246,6 kg dan acara khusus sebesar 86,94 kg, tetapi keduanya tidak berpengaruh signifikan secara statistik terhadap jumlah sampah harian.

### 3.2 Pembahasan

Analisis regresi linear sederhana menunjukkan bahwa terhadap data jumlah sampah harian di Kota X, diperoleh bahwa variabel *hari libur* dan *acara khusus* tidak memberikan pengaruh yang signifikan terhadap banyaknya sampah yang dihasilkan setiap hari. Nilai *p-value* untuk kedua variabel masing-masing sebesar 0,676 pada *hari libur* dan 0,913 pada *acara khusus*, kedua nilai tersebut secara jelas lebih tinggi daripada tingkat signifikansi 0,05. Secara statistik, temuan ini menunjukkan adanya perbedaan jumlah sampah antara hari kerja, hari libur, dan hari dengan acara khusus tidak cukup kuat untuk dijadikan dasar dalam membangun model prediksi.

Nilai koefisien determinasi ( $R^2$ ) yang sangat rendah, yaitu hanya sebesar 0,00006245 (atau sekitar 0,006%), semakin memperkuat hasil tersebut. Artinya, hanya sekitar 0,006% variasi jumlah sampah harian yang mampu dijelaskan oleh variabel *hari libur* dan *acara khusus*, lebih dari 99,9% variasi tetap dipengaruhi oleh faktor-faktor lain yang tidak tercakup dalam model, seperti kondisi cuaca, tingkat kepadatan penduduk, kebiasaan konsumsi masyarakat, maupun efektivitas pengelolaan sampah oleh pihak berwenang. Selain itu, rendahnya  $R^2$  juga dipengaruhi oleh tipe variable bebas yang digunakan dalam penelitian ini. Hari libur dan acara khusus merupakan variable kategorik nominal yang dikodekan menjadi variable dummy. Variable nominal hanya membawa informasi dua kategori (Ya/Tidak) sehingga kontribusi variasinya terhadap model regresi linear sangat terbatas. Ketika variable penjelas tidak mengandung variasi kuantitatif yang cukup, model regresi linear cenderung menghasilkan koefisien kecil, *p-value* besar, dan performa prediksi yang rendah. Dengan demikian, struktur tipe data variable bebas turut menjadi penyebab utama lemahnya model.

Hasil pengujian asumsi klasik menunjukkan bahwa model regresi memenuhi sebagian besar asumsi dasar. Pengujian heteroskedastisitas menggunakan metode Breusch-Pagan menghasilkan nilai *p* sebesar 0,7064, yang menunjukkan tidak adanya masalah heteroskedastisitas. Demikian pula, pengujian multikolinearitas menunjukkan Variance Inflation Factor (VIF) sebesar 1.000 untuk kedua variabel, yang menunjukkan tidak ada korelasi linier yang tinggi antara variabel independen. Namun, pengujian autokorelasi menggunakan metode Durbin-Watson menghasilkan nilai DW sebesar 1,9047 dengan nilai *p* sebesar 0,004591, yang menunjukkan sedikit autokorelasi positif pada data. Hal ini diharapkan karena data yang digunakan adalah data deret waktu, di mana jumlah sampel pada satu hari seringkali terkait dengan hari sebelumnya.

Secara visual, grafik yang membandingkan hasil prediksi dengan data aktual mengindikasikan bahwa garis prediksi (merah putus-putus) hampir sejajar dengan garis observasi (biru tua), namun tidak menggambarkan fluktuasi yang nyata seperti pada data sebenarnya. Ini menunjukkan bahwa model regresi yang diterapkan masih terlalu sederhana (*underfitting*) dan belum mampu menangkap variasi kompleks yang memengaruhi jumlah sampah harian di lapangan.

Hasil ini mendukung temuan dari penelitian sebelumnya yang menyatakan bahwa volume sampah kota sangat dipengaruhi oleh faktor sosial-ekonomi, tingkat aktivitas masyarakat, serta kondisi iklim dan cuaca musiman. Oleh sebab itu, penelitian lanjutan disarankan untuk mengembangkan model yang lebih komprehensif, misalnya melalui regresi berganda dengan tambahan variabel seperti populasi harian, curah hujan, atau tingkat aktivitas ekonomi. Alternatif lainnya adalah menggunakan pendekatan *machine*

*learning* seperti *Random Forest* atau metode peramalan deret waktu (*Time Series Forecasting*) seperti ARIMA atau LSTM agar mampu menghasilkan prediksi yang lebih presisi dan kontekstual.

Secara keseluruhan, hasil evaluasi ini menegaskan bahwa meskipun regresi linear dapat menjadi langkah awal dalam memahami hubungan antara *hari libur*, *acara khusus*, dan jumlah sampah harian, model ini masih belum cukup kuat untuk dijadikan dasar dalam penyusunan kebijakan prediktif. Oleh karena itu, diperlukan pengembangan model dengan variabel yang lebih luas serta penyesuaian terhadap autokorelasi agar dapat mendukung perencanaan pengelolaan sampah perkotaan yang lebih akurat, efisien, dan berbasis pada data empiris.

#### 4. SIMPULAN

Untuk mengetahui apakah hari libur dan acara khusus memengaruhi jumlah sampah harian kota, penelitian ini menggunakan pendekatan kuantitatif dengan menggunakan teknik analisis regresi prediktif. Berdasarkan hasil analisis regresi linear sederhana, ditemukan bahwa nilai p-value hari libur sebesar 0,676 dan nilai p-value acara khusus sebesar 0,913, masing-masing berada di atas tingkat signifikansi 0,05, menandakan bahwa kedua variabel bebas tidak memiliki pengaruh yang signifikan. Selain itu, faktor lain seperti cuaca, kepadatan penduduk, kebiasaan konsumsi masyarakat, dan dinamika aktivitas harian warga memengaruhi sebagian besar variasi dalam jumlah sampah harian. Namun, nilai koefisien determinasi ( $R^2$ ) yang sangat kecil, yaitu 0,00006245, menunjukkan bahwa model hanya dapat menjelaskan sekitar 0,006% dari variasi dalam jumlah sampah harian.

Menurut uji asumsi klasik, model tidak menunjukkan gejala heteroskedastisitas dan multikolinearitas. Namun, mengingat data bersifat deret waktu, ada sedikit autokorelasi positif yang wajar. Hasil prediksi menunjukkan bahwa garis prediksi dan data aktual tidak berimpit sepenuhnya. Ini menunjukkan bahwa model masih bersifat underfitting dan tidak dapat menangkap kompleksitas perubahan yang terjadi pada data timbulan sampah harian.

Keterbatasan model ini bukan hanya berasal dari pengaruh lemah hari libur dan acara khusus, tetapi juga dari jenis variabel independen yang bersifat nominal. Meskipun variabel-variabel tersebut telah dikodekan menjadi variabel dummy, informasi yang dibawanya terlalu sederhana sehingga tidak cukup kuat untuk menjelaskan variasi jumlah sampah harian. Hal ini turut berkontribusi terhadap rendahnya nilai  $R^2$  dan performa prediksi model.

Oleh karena itu, penelitian selanjutnya harus mengembangkan model prediksi yang lebih komprehensif dengan menambahkan variabel kuantitatif berskala interval atau rasio yang lebih representatif. Variabel-variabel ini termasuk intensitas aktivitas masyarakat, mobilitas harian, dan parameter lingkungan lainnya yang relevan. Selain itu, penggunaan metode analisis yang lebih maju, seperti regresi berganda, *Random Forest*, atau pendekatan peramalan deret waktu (ARIMA atau LSTM), dapat menghasilkan hasil prediksi yang lebih akurat. Metode-metode ini dapat digunakan sebagai dasar untuk perencanaan pengelolaan sampah kota yang lebih efisien, akurat, dan berbasis data.

## 5. REKOMENDASI

Model regresi linier sederhana tidak efektif dalam menjelaskan variasi sampah harian karena dampak hari libur dan acara khusus lemah, dan variabel independen nominal, bahkan ketika dikodekan sebagai variabel dummy, memiliki daya penjas yang rendah. Oleh karena itu, penelitian selanjutnya harus menyertakan variabel kuantitatif pada skala interval atau rasio, seperti curah hujan, mobilitas penduduk, atau intensitas aktivitas masyarakat, untuk membuat model prediksi lebih representatif.

Selain itu, disarankan untuk menggunakan metode analitik yang lebih sesuai untuk pola data yang kompleks, seperti regresi linier berganda, Random Forest, XGBoost, atau metode deret waktu seperti ARIMA dan LSTM, yang dapat menangkap pola musiman dan autokorelasi. Studi selanjutnya juga harus menggunakan periode data yang lebih panjang dan memvalidasi model pada wilayah atau kerangka waktu yang berbeda untuk meningkatkan akurasi dan membuat model lebih dapat diterapkan untuk perencanaan pengelolaan sampah perkotaan yang lebih luas.

## 6. REFERENSI

- Ali, M., & Ina, A. Y. (2018). Pengangkutan Sampah Dengan Model Dinamis, Studi Kasus Sumba Barat Daya. *Jurnal Envirotek*, 10(2).
- Arminarahmah, N. (2025). Prediksi Pola Membuang Sampah Rumah Tangga Di Lahan Rawa Menggunakan Machine Learning. *Journal of Innovation Research and Knowledge*, 4(10), 7309-7314.
- Artika, I., & Chaerul, M. (2020). Model sistem dinamik untuk evaluasi skenario pengelolaan sampah di Kota Depok. *Jurnal Wilayah dan Lingkungan*, 8(3), 261-279.
- Fairuzsyifa, A. I., & Nugroho, Y. S. (2024). Analisis regresi linier berganda pengaruh minat calon mahasiswa di Universitas Muhammadiyah Surakarta menggunakan Python. *J. Inform. Polinema*, 10(2), 265-272.
- Hanif, M., Wibowo, R. T. J., & Shofa, M. J. (2024). Penerapan Model Sistem Dinamis Pada Sistem Pengelolaan Sampah Di TPAS Cilowong, Kota Serang. *Integrasi: Jurnal Ilmiah Teknik Industri*, 9(2), 88-98.
- Herlando, M. R., Wijoyo, S. H., & Akbar, M. A. (2025). Analisis Perbandingan Machine Learning untuk Memprediksi Harga Minyak Dengan Regresi Linear dan Support Vector Regression. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 9(6).
- Hutagalung, I. P. (2022). Analisis regresi data panel dengan pendekatan common effect model (CEM), fixed effect model (FEM) dan random effect model (REM)(Studi Kasus: IPM Sumatera Utara Periode 2014–2020). *FARABI: Jurnal Matematika Dan Pendidikan Matematika*, 5(2), 217-226.

- Kusuma, M. D. H., & Hidayat, S. (2024). Penerapan model regresi linier dalam prediksi harga mobil bekas di India dan visualisasi dengan menggunakan Power BI. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(2), 1097-1110.
- Ningsih, S., & Dukalang, H. H. (2019). Penerapan metode suksesif interval pada analisis regresi linier berganda. *Jambura Journal of Mathematics*, 1(1), 43-53.
- Nurfatmala, N., Mallongi, A., & Birawida, A. B. (2018). Model Dinamis Dalam Memprediksi Timbulan Sampah Rumah Tangga Di Kota Baubau. *Jurnal Kesehatan Masyarakat Maritim*, 1(3).
- Prasetyo, M. H., Lihawa, F., & Baderan, D. W. K. (2024). Potensi Model Sistem Dinamik dalam Sistem Pengelolaan Sampah Perkotaan. *Jurnal Wilayah, Kota dan Lingkungan Berkelanjutan*, 3(2), 274-286.
- Royinda, A., & Pamuji, F. Y. (2025). Perbandingan Metode Regresi Linier dan Single Moving Average dalam Peramalan Timbulan Sampah di Kota Malang. *Jurnal Informatika Polinema*, 11(3), 299-304.
- Rusvinasari, D., & Risnanto, A. S. (2024). Rancangan prediksi volume sampah TPA Kota Semarang dengan pendekatan sistem dinamik. *Journal of Data Science Theory and Application*, 3(1), 14-22.
- Saswita, R., Nurbaity, N., & Anggeni, U. (2024). Deteksi Perubahan Berat Badan Akseptor KB Hormonal dengan Menggunakan Model Regresi Linear Berganda. *Oksitosin: Jurnal Ilmiah Kebidanan*, 11(1), 45-55.
- Simbolon, V. A., & Horiza, H. (2023). Prediksi Tingkat Timbulan Sampah 5 Tahun Mendatang (2023-2027) di TPA Ganet Kota Tanjungpinang. *Sulolipu: Media Komunikasi Sivitas Akademika dan Masyarakat*, 23(2), 303-310.