

Prediksi Indikator Sosial Ekonomi Indonesia Menggunakan Regresi Linear: Studi Kasus Multisektor

Sheril Lestari¹, Lilis Roudhotul Jannah², Muhammad Kamil Fatih³, Halili Maar⁴

^{1,2,3,4,5} Program Studi Teknik Informatika, Universitas Pamulang, Tangerang, Indonesia

Corresponding Author: sherillestari169@gmail.com, lilisrj7@gmail.com, kamilpati05@gmail.com, dosen02957@unpam.ac.id

Published: July, 2025

ABSTRAK

Penerapan analisis data pada sektor strategis nasional penting untuk mendukung pengambilan keputusan berbasis data. Penelitian ini bertujuan untuk memprediksi dan membandingkan performa model regresi linear pada tiga dataset berbeda, yaitu: (1) Peringkat e-Government Indonesia tahun 2010–2020, (2) Produksi Gas Indonesia tahun 1995–1998, dan (3) Produk Domestik Bruto (GDP) Indonesia tahun 1960–1973. Masing-masing dataset dianalisis menggunakan algoritma regresi linear, dengan evaluasi performa model berdasarkan metrik R-squared (R^2), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE). Hasil analisis menunjukkan bahwa model memberikan prediksi terbaik pada dataset e-Government dengan R^2 sebesar 0,9506 dan RMSE sebesar 1,78. Dataset GDP menunjukkan hasil cukup akurat dengan R^2 sebesar 0,8583, sedangkan dataset Produksi Gas memiliki R^2 sebesar 0,6251 dan RMSE sebesar 18.141,70 yang menunjukkan bahwa model masih mampu mengenali pola meskipun terdapat fluktuasi. Temuan ini mengindikasikan bahwa regresi linear tetap relevan untuk prediksi tren jangka pendek pada data time-series sederhana di berbagai sektor nasional, serta berpotensi digunakan dalam sistem pendukung keputusan kebijakan publik.

Kata kunci: *Regresi Linear, Prediksi, e-Government, GDP, Produksi Gas, R-squared*

ABSTRACT

The application of data analysis in strategic national sectors is crucial to support data-driven decision-making. This study aims to predict and compare the performance of linear regression models on three different datasets: (1) Indonesia's e-Government rankings (2010–2020), (2) Indonesia's gas production (1995–1998), and (3) Indonesia's Gross Domestic Product (GDP) (1960–1973). Each dataset was analyzed using the linear regression algorithm, with performance evaluated using R-squared (R^2), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Results show that the model performs best on the e-Government dataset with an R^2 of 0.9506 and an RMSE of 1.78. The GDP dataset also demonstrated strong accuracy with an R^2 of 0.8583, while the gas production dataset achieved an R^2 of 0.6251 and RMSE of 18,141.70, indicating the model was still able to capture general patterns despite high fluctuations. These findings suggest that linear regression remains a relevant method for short-term trend prediction on simple time-series data across various national sectors and may contribute to more effective policy-making.

Keywords: Linear Regression, Prediction, e-Government, GDP, Gas Production, R-squared

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0>

PENDAHULUAN

Pemanfaatan analisis data menjadi kunci utama dalam mendukung pengambilan keputusan berbasis bukti, terutama dalam konteks pembangunan sosial dan ekonomi

nasional. Di era digital saat ini, berbagai sektor strategis seperti pemerintahan, energi, dan ekonomi makro menghasilkan data yang dapat diolah untuk mengidentifikasi tren, merumuskan kebijakan, dan merancang perencanaan jangka panjang. Oleh karena itu, penerapan metode prediksi berbasis data menjadi langkah penting untuk menjawab kebutuhan tersebut.

Penelitian ini secara khusus bertujuan untuk menerapkan algoritma regresi linear pada tiga sektor penting Indonesia: peringkat e-Government, produksi gas, dan Produk Domestik Bruto (GDP). Ketiga indikator tersebut mencerminkan performa sektor digital, energi, dan ekonomi secara menyeluruh. Pemilihan regresi linear sebagai metode utama didasarkan pada kesederhanaannya, kemampuannya untuk menangkap pola linier dalam data time-series, dan relevansinya untuk prediksi jangka pendek. Penelitian ini akan membandingkan performa model pada masing-masing dataset dengan mengevaluasi nilai R-squared (R^2), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE).

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder dari sumber resmi, yaitu: (1) Indeks e-Government Development Index (EGDI) Indonesia dari situs UNDESA (2020), (2) data produksi gas nasional dari laporan Kementerian ESDM periode 1995–1998, dan (3) data GDP Indonesia tahun 1960–1973 dari publikasi Badan Pusat Statistik (BPS). Ketiganya diolah dan dianalisis secara terpisah, lalu dibandingkan hasilnya untuk melihat keandalan model pada berbagai karakteristik data.

Sejumlah studi terdahulu menunjukkan bahwa regresi linear banyak digunakan untuk prediksi indikator ekonomi dan sosial. Fitria & Fitriana (2021) menggunakan regresi linear untuk memprediksi penjualan UMKM dengan hasil cukup akurat. Sulantari et al. (2020) menerapkan regresi linier berganda untuk memodelkan utang negara Indonesia. Sementara itu, UNDESA (2020) menjadikan EGDI sebagai tolok ukur digitalisasi pemerintahan secara global. Ayuni et al. (2022) dan Alfatar & Soebagiyo (2023) juga menunjukkan bahwa regresi linear dapat digunakan untuk memodelkan berbagai fenomena sosial ekonomi.

METODE PENELITIAN

1. Sumber Data

Penelitian ini menggunakan tiga dataset berbeda yang mewakili sektor-sektor strategis nasional:

- a. **Dataset e-Government:** Merupakan data *e-Government Development Index (EGDI)* Indonesia dari tahun 2010 hingga 2020 yang diperoleh dari situs resmi *United Nations Department of Economic and Social Affairs (UNDESA)*. Dataset ini mencakup empat indikator digital, yaitu EGDI, Online Service Index (OSI), Telecommunication Infrastructure Index (TII), dan Human Capital Index (HCI), serta peringkat global Indonesia.
- b. **Dataset GDP:** Berisi nilai Produk Domestik Bruto (GDP) Indonesia dari tahun 1960 hingga 1973, diperoleh dari publikasi statistik resmi oleh Badan Pusat Statistik (BPS).
- c. **Dataset Produksi Gas:** Mencakup data bulanan produksi gas nasional selama tahun 1995–1998, yang diperoleh dari laporan energi nasional Kementerian ESDM dan buku statistik industri

2. Pra-pemrosesan Data

Sebelum dilakukan pemodelan, seluruh dataset melalui tahapan pra-pemrosesan, yaitu:

- a. Konversi tipe data ke numerik jika terdapat kolom dalam format non-numerik.
- b. Penghapusan baris dengan nilai kosong atau tidak valid (*missing values*).
- c. Penentuan variabel input (X) dan target (y):

- d. **EGDI**: X = EGDI, OSI, TII, HCI; y = Peringkat
- e. **GDP**: X = Tahun; y = GDP
- f. **Produksi Gas**: X = Tahun 1995–1997; y = Produksi 1998

3. Algoritma Regresi Linear

Penelitian ini menggunakan algoritma **regresi linear** sebagai metode utama. Regresi linear merupakan pendekatan statistik untuk memodelkan hubungan linier antara variabel independen dan dependen. Metode ini efektif digunakan pada data time-series yang cenderung memiliki pola linier dan dapat diaplikasikan dengan kompleksitas rendah namun interpretasi yang jelas

4. Evaluasi Model

Evaluasi performa model dilakukan berdasarkan jenis data target:

- a. **EGDI**: Menggunakan **R-squared (R^2)** karena target berupa peringkat ordinal.
- b. **GDP dan Produksi Gas**: Menggunakan **Mean Squared Error (MSE)** dan **Root Mean Squared Error (RMSE)** karena target berupa angka kontinu.

5. Tools dan Bahasa Pemrograman

Seluruh proses dilakukan menggunakan bahasa pemrograman **Python**, dengan pustaka sebagai berikut:

- a. pandas, numpy untuk manipulasi dan analisis data
- b. scikit-learn untuk membangun model regresi linear dan evaluasi metrik
- c. matplotlib, seaborn untuk visualisasi grafik tren dan korelasi

6. Implementasi Kode

- a). Untuk Dataset EDGI:

```
from sklearn.linear_model import
LinearRegression
from sklearn.metrics import r2_score

X = df[['EGDI', 'OSI', 'TII', 'HCI']]
y = df['Peringkat']

model = LinearRegression()
model.fit(X, y)
y_pred = model.predict(X)

print(f"R2 Score: {r2_score(y, y_pred):.4f}")
```

- b). Untuk Dataset GDP / Produksi Gas:

```
from sklearn.linear_model import
LinearRegression
from sklearn.metrics import
mean_squared_error

X = df[['Tahun']] # atau ['1995', '1996', '1997']
y = df['GDP']     # atau df['1998']

model = LinearRegression()
model.fit(X, y)
y_pred = model.predict(X)
```

```
print(f"MSE: {mean_squared_error(y,
y_pred):.2f}")
```

7. Alur Proses Umum

Langkah-langkah umum yang diterapkan pada setiap dataset adalah sebagai berikut:

- Tentukan variabel input (X) dan target (y)
- Latih model regresi linear menggunakan dataset
- Lakukan prediksi terhadap data target
- Evaluasi model menggunakan R^2 atau MSE sesuai karakteristik data

8. Keterbatasan

Penelitian ini belum menerapkan teknik validasi silang (cross-validation) dan menggunakan dataset berukuran kecil, sehingga generalisasi model pada data baru mungkin terbatas. Pengembangan lebih lanjut dapat melibatkan algoritma regresi non-linier serta penggunaan data yang lebih besar dan bervariasi.

9. Hasil dan Pembahasan

Penelitian ini menerapkan algoritma regresi linear untuk menganalisis dan memprediksi tren dari tiga dataset berbeda yang merepresentasikan indikator sosial ekonomi nasional, yaitu peringkat e-Government, produksi gas, dan Produk Domestik Bruto (GDP) Indonesia. Setiap dataset dievaluasi secara individual menggunakan metrik yang sesuai untuk mengukur performa model.

Hasil Analisis Regresi Linear

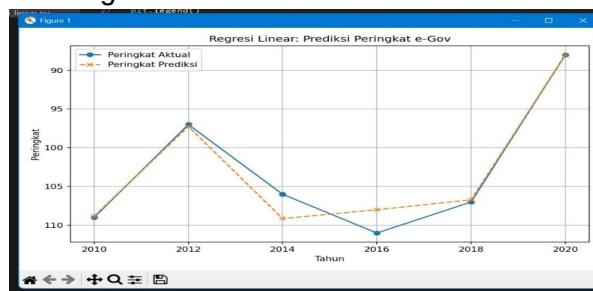
Dataset e-Government Indonesia

Dataset ini mencakup tahun 2010 hingga 2020, dengan variabel input berupa EGDI, OSI, TII, dan HCI, serta target berupa peringkat e-Government. Evaluasi performa model ditampilkan dalam Tabel 1.

Tabel 1. Evaluasi Model Dataset e-Government

Matrik Evaluasi	Nilai
R-squared (R^2)	0.9506
MSE	3.18
RMSE	1.78

Prediksi model untuk tahun 2022, 2024, dan 2026 menghasilkan peringkat masing-masing sebesar **88.15**, **85.48**, dan **79.20**, menunjukkan tren perbaikan peringkat nasional. Grafik prediksi dan aktual menunjukkan bahwa model mampu mengikuti pola data dengan sangat baik, mendekati garis identitas.



Gambar 1. Grafik Prediksi Peringkat e-Gov

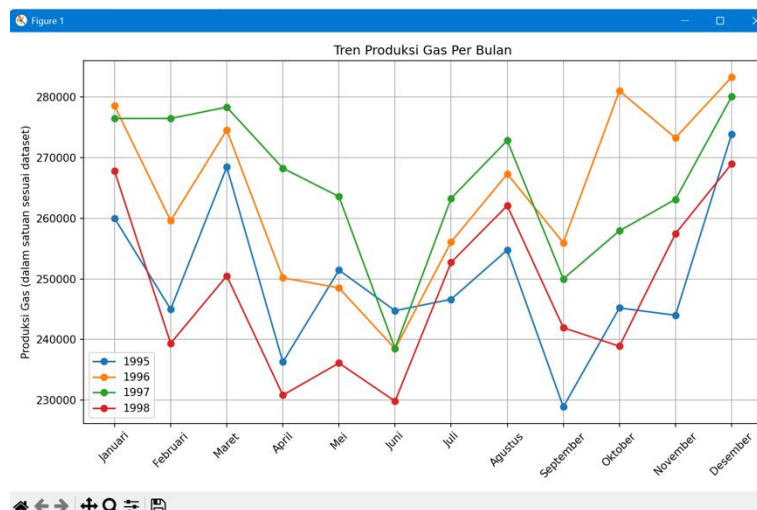
Dataset Produksi Gas Indonesia

Dataset ini memuat data produksi gas bulanan dari Januari hingga Desember pada tahun 1995–1998. Model dilatih menggunakan data tahun 1995–1997 untuk memprediksi tahun 1998. Hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Evaluasi Model Dataset Produksi Gas

Matrik Evaluasi	Nilai
R-squared (R^2)	0.6251
MSE	329,121,312.93
RMSE	18,141.70

Nilai R^2 menunjukkan bahwa sekitar 62,5% variasi dapat dijelaskan oleh model. Performa yang moderat ini diduga dipengaruhi oleh **jumlah data yang terbatas** (12 bulan) dan **fluktuasi musiman**, yang tidak dapat sepenuhnya ditangkap oleh model linier.



Gambar 2. Grafik Tren Produksi Gas

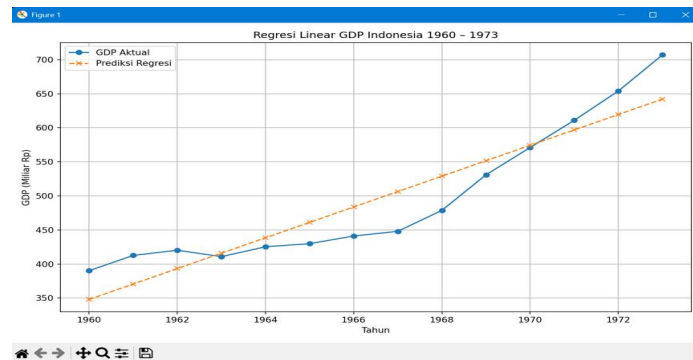
Dataset GDP Indonesia

Dataset ini mencakup nilai GDP dari tahun 1960 hingga 1973, dengan input berupa tahun dan target berupa GDP. Hasil evaluasi model ditampilkan pada Tabel 3.

Tabel 3. Evaluasi Model Dataset GDP Indonesia

Matrik Evaluasi	Nilai
R-squared (R^2)	0.8583
MSE	1373.23
RMSE	37.05

Model menunjukkan hubungan linier yang cukup kuat, dengan R^2 sebesar 0.8583. Grafik regresi menunjukkan bahwa model mengikuti tren pertumbuhan GDP secara konsisten, meskipun terdapat sedikit deviasi pada tahun-tahun awal.



Gambar 3. Grafik GDP Indonesia

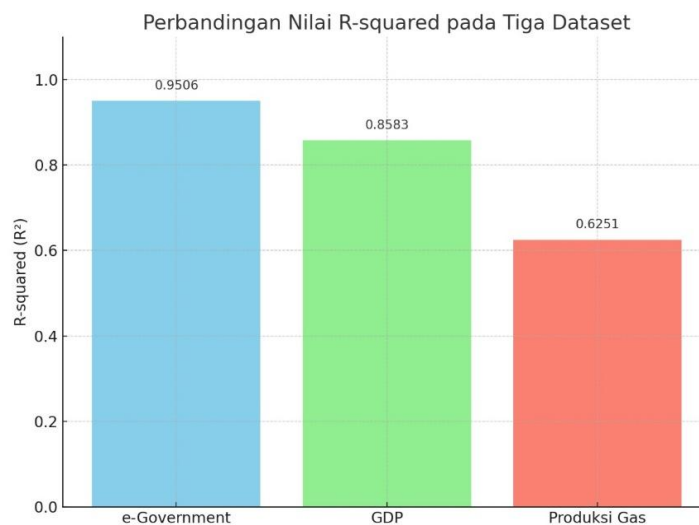
Perbandingan Hasil Antar Dataset

Tabel berikut menyajikan perbandingan performa model regresi linear pada ketiga dataset.

Tabel 4. Perbandingan Evaluasi Regresi Linear

Dataset	R^2	MSE	RMSE
e-Government	0.9506	3.18	1.78
Produksi Gas	0.6251	329,121,312.93	18,141.70
GDP Indonesia	0.8583	1373.23	-

Hasil ini menunjukkan bahwa model paling cocok digunakan pada dataset e-Government, diikuti oleh GDP, dan yang paling rendah pada dataset Produksi Gas. Hal ini memperkuat bahwa model regresi linear lebih optimal digunakan pada data dengan pola tren linier yang kuat dan stabil.



Gambar 4. Grafik Perbandingan

Interprestasi Umum

Hasil analisis menunjukkan bahwa algoritma regresi linear sangat efektif diterapkan pada data yang memiliki pola tren linier yang kuat dan konsisten, seperti pada dataset e-

Government dan GDP. Pada kedua dataset tersebut, nilai R-squared (R^2) yang tinggi mencerminkan bahwa sebagian besar variasi data dapat dijelaskan oleh model.

Sebaliknya, performa model pada dataset Produksi Gas cenderung moderat, dengan nilai R^2 yang lebih rendah dan error yang cukup tinggi. Hal ini kemungkinan besar disebabkan oleh sifat data yang bersifat musiman dan fluktuatif, serta jumlah data yang relatif sedikit (hanya 12 baris). Regresi linear sebagai model linier sederhana memiliki keterbatasan dalam menangkap pola musiman atau hubungan non-linier yang kompleks.

Dengan demikian, untuk kasus dataset yang menunjukkan pola musiman atau ketidakteraturan yang tinggi, metode prediksi alternatif seperti regresi non-linier, polynomial regression, atau model time-series seperti ARIMA, SARIMA, maupun Prophet, dapat menjadi solusi yang lebih tepat. Hal ini membuka peluang untuk pengembangan metode lanjutan pada penelitian selanjutnya guna memperoleh prediksi yang lebih akurat.

KESIMPULAN

Penelitian ini berhasil menerapkan algoritma regresi linear pada tiga dataset berbeda yang merepresentasikan sektor strategis nasional: e-Government, Produksi Gas, dan GDP Indonesia. Hasil evaluasi menunjukkan bahwa model memberikan hasil prediksi terbaik pada dataset e-Government dengan nilai R^2 sebesar 0,9506. Dataset GDP juga menunjukkan performa kuat dengan R^2 sebesar 0,8583, sedangkan dataset Produksi Gas memiliki akurasi lebih rendah karena fluktuasi data dan jumlah data yang terbatas. Temuan ini mengindikasikan bahwa regresi linear cukup efektif digunakan untuk prediksi tren jangka pendek pada data dengan pola linier yang stabil, dan dapat menjadi dasar dalam pengambilan keputusan berbasis data. Untuk dataset dengan pola musiman atau fluktuatif, metode non-linier dapat dijadikan alternatif.

DAFTAR PUSTAKA

- Ayuni, T., & Ardiansyah, R. (2022). Analisis regresi linier pada indeks digitalisasi pemerintahan daerah. *Jurnal Sains dan Informatika*, 6(2), 44–51.
- Badan Pusat Statistik. (1974). *Produk Domestik Bruto Indonesia Tahun 1960–1973*. BPS.
- Budiarto, M., & Sari, I. A. (2024). Analisis regresi linier berganda dalam prediksi kinerja ekonomi daerah di Indonesia. *Jurnal Informatika dan Komputasi Sosial*, 4(2), 123–131.
- Fadillah, R., Hidayat, S. A., & Firmansyah, I. (2024). Prediksi IPM menggunakan regresi linier dan korelasi variabel sosial ekonomi. *Jurnal Ilmiah Teknologi dan Komputer*, 10(1), 55–63.
- Fitria, F., & Fitriana, L. (2021). Prediksi penjualan UMKM menggunakan regresi linear. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(1), 15–23.
- Herlina, L., & Rahmi, P. P. (2022). The gross domestic product analysis in Indonesia for 2008–2021. *Journal of Business and Management Inaba*, 1(2), 64–74.
- Kementerian Energi dan Sumber Daya Mineral. (1999). *Statistik Produksi Gas Indonesia 1995–1998*. Kementerian ESDM.
- Kusumawati, D., & Saifudin. (2023). Adopsi e-Government di era kebiasaan baru. *Jurnal Ilmiah Komunikasi Makna*, 11(2), 175–186.

- Lestari, A., & Firmansyah, K. (2023). Pengaruh belanja modal TIK terhadap efektivitas e-Government. *Jurnal Sistem Informasi*, 9(1), 77–85.
- Maisarah, U., Farisha, M., Rizal, Y., & Safuridar, S. (2024). Pengaruh subsidi energi dan konsumsi listrik terhadap pertumbuhan ekonomi di Indonesia. *APKE*, 1(4), 100–107.
- Maulidiyah, L., & Ramadhan, A. (2023). Model regresi untuk memprediksi Produk Domestik Bruto Indonesia berdasarkan sektor energi. *Jurnal Ekonomi dan Teknologi*, 5(3), 210–217.
- Mulyadi, A., & Wibowo, A. (2021). Penerapan metode regresi linier dalam peramalan angka pengangguran terbuka di Provinsi Jawa Barat. *Jurnal Ilmiah Informatika Global*, 12(2), 145–152.
- Nugroho, S., & Cahyani, Y. (2022). Pengaruh infrastruktur digital terhadap perkembangan e-Government di Indonesia menggunakan regresi. *Jurnal Teknologi Informasi dan Komputer*, 8(1), 91–97.
- Prahutama, A., & Rahmawati, R. (2024). Analisis regresi linier berganda pada data survei untuk pemodelan total pengeluaran di Jawa Tengah, Indonesia. *Jurnal Gaussian*, 13(2), 394–404.
- Puspasari, N. K., & Darmawan, A. (2017). Pengaruh ekspor dan nilai tukar perdagangan terhadap produk domestik bruto Indonesia. *Jurnal Administrasi Bisnis*, 50(1), 190–199.
- Putri, Y., & Suryadi, D. (2023). Evaluasi performa regresi linier dalam memodelkan pertumbuhan ekonomi berbasis data gas alam. *Jurnal Sains Terapan*, 7(4), 370–378.
- Sari, H., & Gunawan, R. (2022). Analisis regresi linier untuk konsumsi gas per kapita di Indonesia. *Jurnal Energi dan Informatika*, 6(2), 132–140.
- Sulantari, N., Hariyanto, B., & Budi, A. P. (2020). Pemodelan utang negara Indonesia dengan regresi linier berganda. *Jurnal Ekonomi dan Bisnis Indonesia*, 12(1), 88–95.
- Susanto, D., & Raharjo, R. F. (2023). Implementasi regresi linear menggunakan Python untuk prediksi ekonomi. *Jurnal Teknologi dan Informatika*, 9(1), 67–75.
- United Nations Department of Economic and Social Affairs. (2020). E-Government Development Index (EGDI) 2020.