



Analisis Principal Component Regression pada Kapasitas Pembangkit dan Produksi Listrik terhadap Distribusi Listrik Antarprovinsi di Indonesia Tahun 2024

Iqva Junita, Aqil Afritzal Saldi, Nadia Rikza Ananda, Feyza Zaki Nandana

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

iqva.4243260005@mhs.unimed.ac.id, aqilsaldi.4243260030@mhs.unimed.ac.id, nadiarikza.4241260023@mhs.unimed.ac.id, feyzazaki.4243260023@mhs.unimed.ac.id

Abstrak

Distribusi listrik antarprovinsi di Indonesia memiliki perbedaan yang cukup besar, sehingga diperlukan analisis untuk melihat keterkaitannya dengan kondisi pembangkitan listrik di setiap wilayah. Penelitian ini bertujuan menganalisis hubungan kapasitas pembangkit dan produksi listrik terhadap distribusi listrik antarprovinsi di Indonesia tahun 2024 menggunakan metode Principal Component Regression (PCR). Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS) yang mencakup 34 provinsi. Distribusi listrik digunakan sebagai variabel respons, sedangkan kapasitas pembangkit dan produksi listrik digunakan sebagai variabel prediktor. Analisis dilakukan melalui pemeriksaan korelasi dan multikolinearitas, pembentukan Principal Component Analysis (PCA), pemilihan komponen berdasarkan beberapa kriteria, kemudian dilanjutkan dengan regresi menggunakan komponen utama. Hasil analisis menunjukkan bahwa kapasitas pembangkit dan produksi listrik memiliki hubungan yang kuat dengan distribusi listrik. Berdasarkan validasi silang Leave-One-Out Cross-Validation (LOOCV), dua komponen utama dipertahankan dalam model PCR. Kedua komponen tersebut menunjukkan hubungan yang signifikan dengan distribusi listrik. Model juga memiliki kemampuan prediksi yang lebih baik dibandingkan model nol. Namun, hasil pengujian asumsi menunjukkan adanya heteroskedastisitas dan beberapa provinsi teridentifikasi sebagai observasi yang berpengaruh. Karena seluruh komponen utama tetap digunakan, PCR dalam penelitian ini secara numerik setara dengan regresi linear berganda, sehingga manfaat utamanya lebih terlihat dalam menangani multikolinearitas dan menggambarkan struktur hubungan antarvariabel. Hasil penelitian menunjukkan bahwa distribusi listrik tidak hanya berkaitan dengan kapasitas dan produksi listrik lokal, tetapi juga perlu mempertimbangkan keterhubungan jaringan listrik antarwilayah.

Kata kunci: PCA, PCR, Produksi Listrik, Distribusi Listrik, Regresi

Abstract

Inter-provincial electricity distribution in Indonesia exhibits considerable differences across various regions, thus requiring comprehensive analysis to examine its correlation with power generation conditions in each respective area. This study aims to analyze the relationship between generation capacity and electricity production on inter-provincial electricity distribution in Indonesia in 2024 using the Principal Component Regression (PCR) method. The data used in this research are secondary data obtained from Statistics Indonesia covering 34 provinces. Electricity distribution is utilized as the response variable, while generation capacity and electricity production serve as the predictor variables. The analytical procedure was conducted through correlation and multicollinearity examinations, the formation of Principal Component Analysis, component selection based on several established criteria, and subsequently followed by regression modeling using principal components. The analytical results demonstrate that generation capacity and electricity production possess a strong relationship with electricity distribution nationwide. Based on Leave-One-Out Cross-Validation, two principal components were retained in the final PCR model. Both components exhibit a statistically significant relationship with electricity distribution. Furthermore, the model possesses superior predictive capability compared to the null model. However, assumption testing results indicate the presence of heteroscedasticity and several provinces identified as influential observations.

Keywords: PCA, PCR, Electricity Production, Electricity Distribution, Regression

1. Pendahuluan

Listrik merupakan salah satu infrastruktur penting untuk mendukung kegiatan rumah tangga, industri, layanan publik, dan pertumbuhan ekonomi. Ketersediaan listrik tidak hanya berkaitan dengan kemampuan pembangkit

Analisis Principal Component Regression pada Kapasitas Pembangkit dan Produksi Listrik terhadap Distribusi Listrik Antarprovinsi di Indonesia Tahun 2024

menghasilkan energi, tetapi juga dengan bagaimana energi tersebut disalurkan dan dimanfaatkan di berbagai wilayah. Dalam sistem kelistrikan, kapasitas terpasang menggambarkan kemampuan maksimum pembangkit, sedangkan produksi listrik menunjukkan energi yang benar-benar dihasilkan selama periode tertentu. Distribusi listrik menggambarkan jumlah energi yang disalurkan kepada pengguna. Ketiga indikator tersebut saling berhubungan, tetapi memiliki makna yang berbeda sehingga perlu dianalisis secara cermat ketika digunakan dalam pemodelan statistik.

Perbedaan kondisi kelistrikan antardaerah menjadi salah satu hal yang dapat diamati dari data provinsi di Indonesia. Badan Pusat Statistik (BPS) menerbitkan data kapasitas pembangkit, tenaga listrik yang dibangkitkan, dan listrik yang didistribusikan menurut provinsi. Pada data distribusi listrik tahun 2024, nilai yang tercantum berbeda cukup besar antarprovinsi; misalnya, Jawa Barat tercatat sebesar 61.637,47 GWh, sedangkan Kalimantan Utara sebesar 266,98 GWh. Angka tersebut menggambarkan variasi distribusi yang tercatat, bukan bukti bahwa kapasitas pembangkit atau produksi listrik menjadi satu-satunya penyebab perbedaan distribusi (BPS, 2026a). Selain itu, tabel BPS memberi tanda “-” pada beberapa provinsi baru, sehingga kelengkapan observasi perlu diperiksa sebelum pemodelan dilakukan.

Kapasitas pembangkit dan produksi listrik secara konseptual berpotensi memiliki hubungan yang kuat karena kapasitas merupakan salah satu kondisi yang mendukung kemampuan produksi. Ketika prediktor dalam regresi saling berkorelasi tinggi, model dapat mengalami multikolinearitas. Multikolinearitas dapat memperbesar varians estimasi koefisien dan membuat koefisien regresi lebih sulit ditafsirkan secara stabil. Davino et al. (2022) membahas persoalan kolinearitas pada pemodelan regresi dan menunjukkan bahwa regresi komponen utama dapat digunakan sebagai salah satu pendekatan untuk menangani kondisi tersebut. Nasution dan Pane (2024) juga membandingkan PCR dengan LASSO dalam berbagai kondisi korelasi prediktor, sedangkan Ergişi et al. (2024) membandingkan PCR, regresi ridge, LASSO, dan kuadrat terkecil pada data dengan multikolinearitas. Temuan dari studi-studi tersebut memperlihatkan bahwa pemilihan metode perlu disesuaikan dengan karakteristik data dan ukuran evaluasi yang digunakan.

Principal Component Regression (PCR) merupakan metode yang menggabungkan Principal Component Analysis (PCA) dan regresi. PCA mentransformasikan prediktor asli menjadi komponen utama yang saling tidak berkorelasi, kemudian komponen terpilih digunakan sebagai prediktor dalam model regresi. Pendekatan ini telah diterapkan pada berbagai konteks, termasuk prediksi Nilai Tukar Petani di Indonesia (Effendi et al., 2023), pemodelan Indeks Pembangunan Manusia (Indhova et al., 2023; Khoiro et al., 2024), dan analisis faktor penanganan stunting (Qur’ani, 2023). Balqis et al. (2022) mengembangkan pemodelan PCR dengan pendekatan Bayesian untuk menghadapi multikolinearitas pada proporsi data hilang, sementara Gu (2023) membahas pemilihan komponen dan estimasi galat baku dalam PCR multivariat. Rong dan Liu (2024) juga mengkaji pengembangan estimator PCR dan membandingkannya dengan pendekatan klasik. Berbagai penelitian tersebut menunjukkan bahwa PCR masih menjadi topik metodologis yang relevan, tetapi pemilihan komponen dan evaluasi model tetap merupakan bagian penting dari analisis.

Penggunaan PCA dan PCR juga ditemukan dalam kajian energi dan kelistrikan. Veeramsetty et al. (2022) menggabungkan PCA dengan jaringan saraf untuk peramalan beban listrik jangka pendek. Yang dan Yin (2022) menggunakan PCA bersama metode pengelompokan untuk menganalisis pola perilaku konsumsi listrik, sedangkan penelitian mengenai pemilihan variabel berbasis PCA mengkaji kontribusi variabel untuk peramalan beban energi listrik (International Journal of Energy Sector Management, 2022). Iswan et al. (2021) mengaplikasikan optimasi variabel berbasis PCA dan pendekatan PCR pada peramalan penjualan energi listrik berdasarkan kelompok wilayah. Parhizkar et al. (2021) turut menguji reduksi fitur berbasis PCA untuk meningkatkan model prediksi konsumsi energi. Literatur tersebut memberi konteks bahwa teknik komponen utama dapat dimanfaatkan dalam pengolahan data energi, meskipun tujuan, jenis data, dan model yang digunakan pada penelitian-penelitian tersebut tidak sama dengan penelitian ini.

Meskipun penelitian mengenai PCR dan penerapan PCA pada data energi telah dilakukan, kajian yang secara khusus menggunakan kapasitas pembangkit dan produksi listrik untuk menganalisis distribusi listrik antarprovinsi di Indonesia pada tahun 2024 masih perlu ditelaah melalui pemodelan yang terarah. Penelitian ini menempatkan distribusi listrik sebagai variabel respons, sedangkan kapasitas pembangkit dan produksi listrik sebagai variabel prediktor. Dengan hanya dua prediktor, jumlah komponen utama yang mungkin terbentuk paling banyak dua. Oleh sebab itu, penggunaan PCR tidak langsung diasumsikan lebih baik daripada regresi biasa; kebutuhan reduksi komponen perlu didukung oleh hasil korelasi, proporsi keragaman yang dijelaskan, dan evaluasi kinerja model.

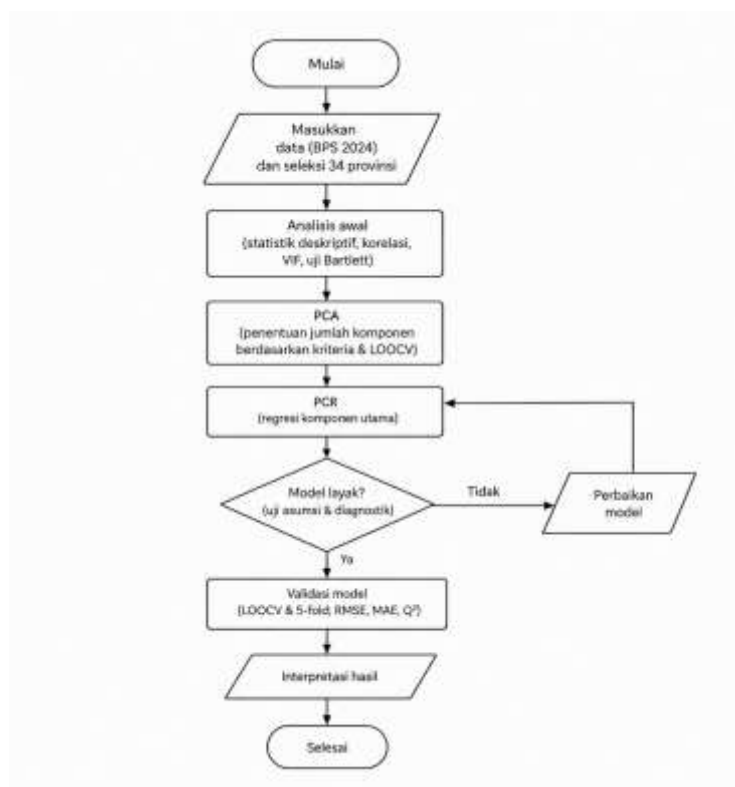
Pendekatan ini sejalan dengan pembahasan pemilihan komponen pada Gu (2023) serta kajian perbandingan model regresi oleh Shah et al. (2022).

Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan Principal Component Regression untuk menganalisis hubungan kapasitas pembangkit dan produksi listrik dengan distribusi listrik antarprovinsi di Indonesia tahun 2024. Rencana pemecahan masalah meliputi pengumpulan data sekunder dari BPS, pemeriksaan konsistensi satuan dan kelengkapan data, analisis korelasi antarvariabel, pembentukan serta pemilihan komponen utama, penyusunan model PCR, dan evaluasi model menggunakan ukuran yang sesuai, seperti koefisien determinasi dan galat prediksi. Penelitian ini diharapkan memberikan gambaran statistik mengenai hubungan indikator pembangkitan dengan distribusi listrik serta menjelaskan kegunaan dan keterbatasan PCR pada data kelistrikan antarprovinsi. Hasil penelitian akan ditafsirkan sebagai hubungan statistik, bukan sebagai bukti kausalitas..

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder tahun 2024 yang diperoleh dari Badan Pusat Statistik (BPS). Data meliputi kapasitas terpasang pembangkit listrik (MW), tenaga listrik yang dibangkitkan (GWh), dan listrik yang didistribusikan (GWh) menurut provinsi. Ketiga data digabungkan berdasarkan nama provinsi. Dari 38 provinsi, empat provinsi tidak disertakan karena tidak tersedia data distribusi listrik, sehingga diperoleh 34 provinsi sebagai unit analisis. Variabel respons adalah distribusi listrik, sedangkan variabel prediktor terdiri atas kapasitas pembangkit dan produksi listrik.

Analisis diawali dengan statistik deskriptif dan korelasi Pearson serta Spearman untuk menggambarkan karakteristik dan hubungan antarvariabel. Multikolinearitas diperiksa menggunakan Variance Inflation Factor (VIF), sedangkan kelayakan PCA diuji menggunakan Bartlett. Selanjutnya, prediktor distandardisasi dan dianalisis menggunakan Principal Component Analysis (PCA). Jumlah komponen utama ditentukan berdasarkan eigenvalue, proporsi varians kumulatif, parallel analysis, broken-stick, serta evaluasi prediksi menggunakan Leave-One-Out Cross-Validation (LOOCV). Berikut diagram alur analisis untuk lebih mudah dipahami:



Gambar 1. Diagram Alur Analisis Principal Component Regression

Berdasarkan komponen terpilih, dilakukan Principal Component Regression (PCR) dengan meregresikan distribusi listrik terhadap skor komponen utama. Signifikansi model diuji menggunakan uji F dan uji t, sedangkan kemampuan penjelasan model dievaluasi melalui koefisien determinasi terkoreksi dan galat baku residual. Koefisien regresi kemudian ditransformasikan kembali ke skala variabel asli untuk memudahkan interpretasi. Evaluasi model mencakup uji normalitas residual Shapiro-Wilk, heteroskedastisitas Breusch-Pagan, autokorelasi Durbin-Watson, dan Ramsey RESET, serta diagnostik observasi berpengaruh. Kemampuan prediksi dievaluasi menggunakan LOOCV dan 5-fold cross-validation dengan ukuran RMSE, MAE, dan Q^2 . Kinerja PCR juga dibandingkan dengan model nol dan Ordinary Least Squares (OLS). Seluruh analisis dilakukan menggunakan perangkat lunak R dengan tingkat signifikansi 5%.

3. Hasil dan Pembahasan

3.1 Hasil

Data penelitian mencakup 38 provinsi, tetapi empat provinsi hasil pemekaran (Papua Barat Daya, Papua Selatan, Papua Tengah, dan Papua Pegunungan) tidak memiliki data distribusi listrik sehingga dikeluarkan dari analisis. Analisis akhir menggunakan 34 provinsi. Statistik deskriptif disajikan pada Tabel 1.

Tabel 1. Statistik deskriptif variabel penelitian ($n = 34$)

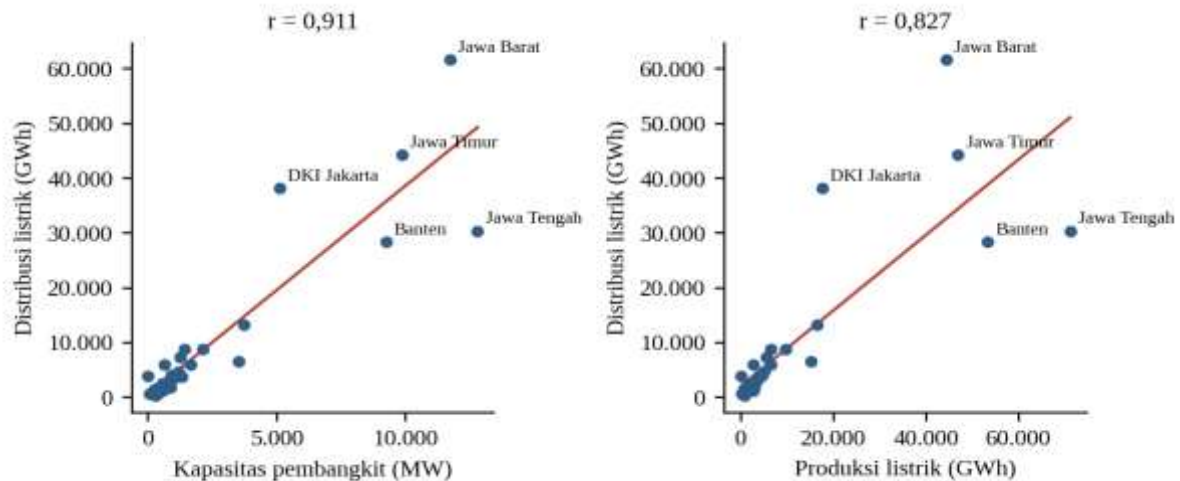
Variabel	Rata-rata	Simpangan baku	Min.	Median	Maks.	CV (%)	Skewness
Distribusi listrik (GWh)	9.006,45	14.381,88	266,98	3.628,01	61.637,47	159,7	2,32
Kapasitas pembangkit (MW)	2.223,36	3.440,71	0,78	808,71	12.814,08	154,8	2,08
Produksi listrik (GWh)	10.079,10	17.217,22	3,64	3.104,39	71.068,16	170,8	2,33

Distribusi listrik rata-rata sebesar 9.006,45 GWh dengan median 3.628,01 GWh. Nilai tertinggi berada di Jawa Barat (61.637,47 GWh) dan terendah di Kalimantan Utara (266,98 GWh). Kapasitas pembangkit tertinggi terdapat di Jawa Tengah (12.814,08 MW) dan terendah di DI Yogyakarta (0,78 MW), sedangkan produksi listrik tertinggi juga di Jawa Tengah (71.068,16 GWh) dan terendah di DI Yogyakarta (3,64 GWh). Ketiga variabel menjulur kuat ke kanan (skewness > 2; CV > 150%) dan menyimpang dari distribusi normal (Shapiro-Wilk, $p < 0,001$). Lima provinsi, yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten, menyumbang 66,3% total distribusi listrik.

Distribusi listrik berkorelasi positif kuat dengan kapasitas pembangkit ($r = 0,911$) dan produksi listrik ($r = 0,827$) (Gambar 2; Tabel 2). Kedua prediktor sendiri berkorelasi sangat kuat ($r = 0,981$; $p < 0,001$) dengan nilai VIF 26,53 pada masing-masing prediktor. Uji Bartlett menolak hipotesis bahwa matriks korelasi adalah matriks identitas ($\chi^2 = 103,26$; $db = 1$; $p < 0,001$).

Tabel 2. Koefisien korelasi antarvariabel

Pasangan variabel	Pearson (r)	Spearman (ρ)
Kapasitas – Produksi	0,981	0,982
Kapasitas – Distribusi	0,911	0,885
Produksi – Distribusi	0,827	0,895

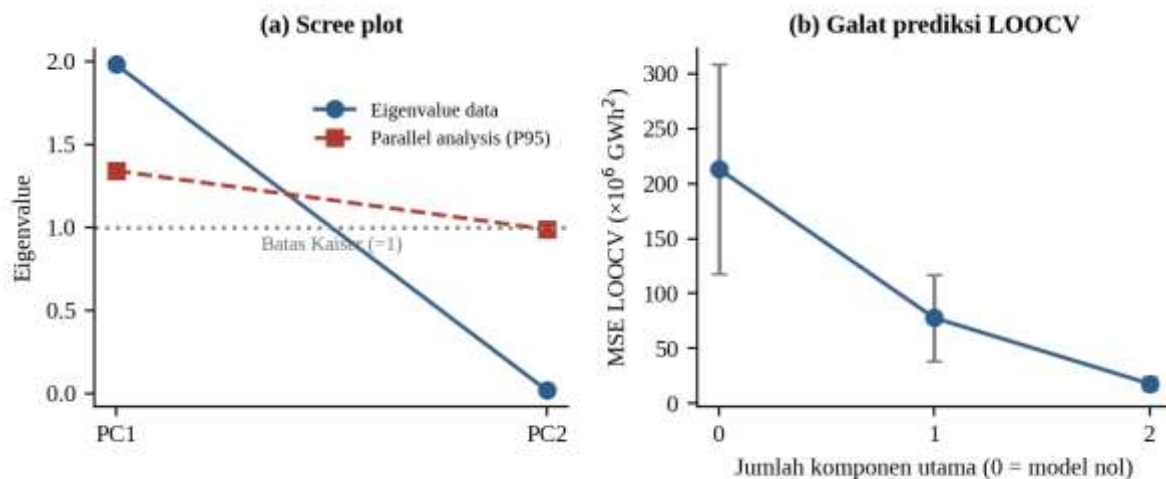


Gambar 2. Diagram pencar distribusi listrik terhadap kapasitas pembangkit dan produksi listrik

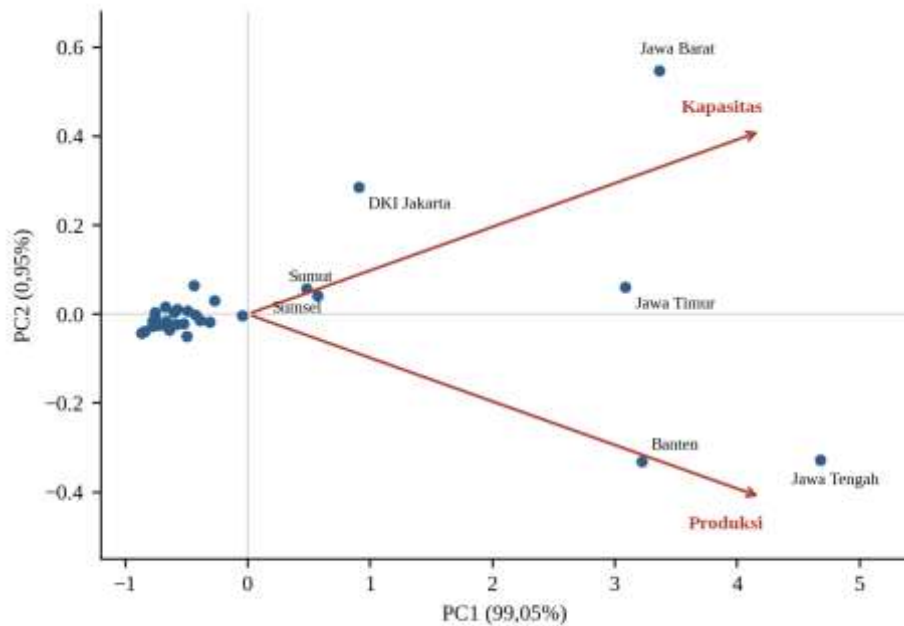
Analisis komponen utama pada prediktor terstandarisasi menghasilkan dua komponen (Tabel 3). PC1 menjelaskan 99,05% varians dengan loading kedua variabel sama besar dan bertanda positif (0,995), sedangkan PC2 menjelaskan 0,95% varians dengan loading berlawanan tanda (0,098 dan $-0,098$). Kriteria Kaiser, varians kumulatif $\geq 80\%$, *parallel analysis*, dan *broken-stick* masing-masing menyarankan satu komponen, sedangkan validasi silang *leave-one-out* (LOOCV) memilih dua komponen (Gambar 3). Model satu komponen menghasilkan RMSE LOOCV 8.821,26 GWh ($Q^2 = 0,612$; $R^2 = 0,762$), sedangkan model dua komponen menghasilkan RMSE 4.240,20 GWh ($Q^2 = 0,910$). Dua komponen karenanya dipertahankan. Skor provinsi pada kedua komponen disajikan pada Gambar 4.

Tabel 3. Eigenvalue, proporsi varians, dan loading komponen utama

Komponen	Eigenvalue	Varians (%)	Kumulatif (%)	Loading kapasitas	Loading produksi
PC1	1,981	99,05	99,05	0,995	0,995
PC2	0,019	0,95	100,00	0,098	$-0,098$



Gambar 3. Scree plot dan galat prediksi LOOCV menurut jumlah komponen utama



Gambar 4. Biplot skor provinsi dan arah variabel pada PC1 dan PC2

Hipotesis uji simultan adalah $H_0: \gamma_1 = \gamma_2 = 0$ melawan H_1 : minimal satu $\gamma_j \neq 0$. Uji F menolak H_0 ($F(2; 31) = 281,36; p < 0,001$), sehingga model PCR signifikan. Uji parsial menunjukkan PC1 dan PC2 sama-sama berpengaruh signifikan terhadap distribusi listrik ($p < 0,001$) (Tabel 4). Model memiliki $R^2 = 0,9478$, R^2 terkoreksi = 0,9444, dan galat baku residual 3.390,67 GWh. Setelah ditransformasi balik ke variabel asli, persamaan model adalah $\hat{Y} = -692,87 + 11,057 \text{ Kapasitas} - 1,477 \text{ Produksi}$.

Tabel 4. Hasil regresi komponen utama dan koefisien pada variabel asli

Parameter	Koefisien	Galat baku	<i>t</i>	<i>p</i>
<i>Model komponen utama</i>				
Intersep	9.006,45	581,49	15,49	< 0,001
PC1	8.922,64	419,36	21,28	< 0,001
PC2	44.878,03	4.278,71	10,49	< 0,001
<i>Variabel asli (transformasi balik)</i>				
Intersep	-692,87	—	—	—
Kapasitas (MW)	11,057	0,884	12,51	< 0,001
Produksi (GWh)	-1,477	0,177	-8,36	< 0,001

Pengujian asumsi residual (Tabel 5) menunjukkan residual berdistribusi normal dan tidak terdapat autokorelasi, tetapi terdapat heteroskedastisitas; uji RESET berada tepat pada batas taraf signifikansi. Diagnostik pengaruh menandai Jawa Tengah, Banten, Jawa Barat, DKI Jakarta, Jawa Timur, dan Sumatera Selatan sebagai pencilan atau observasi berpengaruh (Gambar 4a). Setelah lima provinsi berpengaruh (DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten) dikeluarkan ($n = 29$), koefisien PC1 tetap signifikan (9.056,87; $p < 0,001$), tetapi R^2 turun menjadi 0,752.

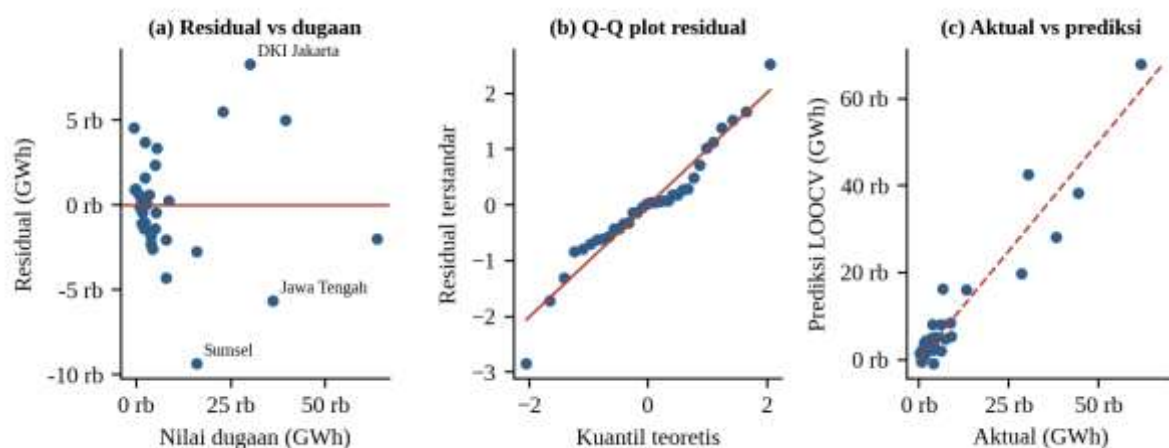
Tabel 5. Hasil uji asumsi residual

Uji	Statistik	<i>p</i>	Keputusan
Shapiro-Wilk (normalitas)	0,956	0,188	Residual normal
Breusch-Pagan (homoskedastisitas)	6,483	0,039	Heteroskedastis
Durbin-Watson (autokorelasi)	1,684	0,156	Tidak ada autokorelasi
RESET (bentuk fungsi)	3,327	0,050	Memadai (batas)

Pada validasi silang (Tabel 6; Gambar 5c), model PCR menghasilkan RMSE LOOCV 4.240,20 GWh, MAE 2.825,96 GWh, dan $Q^2 = 0,910$, jauh lebih baik daripada model nol (RMSE 14.598,16 GWh; $Q^2 = -0,062$). Hasil LOOCV PCR identik dengan OLS karena seluruh komponen digunakan; selisih kecil pada 5-fold berasal dari pengacakan lipatan.

Tabel 6. Perbandingan kinerja prediksi berdasarkan validasi silang

Model	RMSE LOOCV	MAE LOOCV	Q^2 LOOCV	RMSE 5-fold	Q^2 5-fold
Model nol (rata-rata)	14.598,16	9.826,13	-0,062	14.834,08	-0,098
PCR (2 PC)	4.240,20	2.825,96	0,910	4.838,75	0,872
OLS	4.240,20	2.825,96	0,910	4.777,21	0,875



Gambar 5. Diagnostik model PCR: (a) residual terhadap nilai dugaan, (b) Q-Q plot residual, dan (c) nilai aktual terhadap prediksi LOOCV

3.2 Pembahasan

Hasil penelitian menjawab rumusan masalah bahwa kapasitas pembangkit dan produksi listrik secara bersama-sama berkaitan sangat erat dengan distribusi listrik antarprovinsi di Indonesia tahun 2024. Dua komponen utama menjelaskan 94,78% variasi distribusi dan memprediksi provinsi yang tidak dilibatkan dalam pemodelan dengan $Q^2 = 0,910$. Temuan ini sejalan dengan logika sistem kelistrikan bahwa ketersediaan pembangkit menjadi dasar penyediaan listrik. Ketimpangan yang besar, dengan lima provinsi menguasai dua pertiga distribusi nasional, juga mencerminkan konsentrasi penduduk dan kegiatan ekonomi di Pulau Jawa (Sekretariat Kabinet Republik Indonesia, 2021). Ketimpangan spasial kelistrikan antarwilayah di Indonesia juga dikaitkan dengan kepadatan penduduk dan PDB per kapita (Rosidin et al., 2026).

Korelasi kapasitas dan produksi yang sangat tinggi ($r = 0,981$; $VIF = 26,53$) menjadi alasan pemakaian PCR (Montgomery et al., 2021; Ningrum et al., 2025). Komponen utama bersifat saling ortogonal sehingga inferensi pada tingkat komponen stabil. PC1, yang menjelaskan 99,05% varians prediktor, dapat dimaknai sebagai indeks skala sistem pembangkitan: Jawa Tengah, Jawa Barat, dan Banten memiliki skor PC1 tertinggi (Gambar 3), dan koefisien PC1 yang positif berarti semakin besar skala sistem pembangkitan suatu provinsi, semakin besar listrik yang didistribusikan. PC2 hanya memuat 0,95% varians, tetapi signifikan dan menaikkan Q^2 dari 0,612 menjadi 0,910. PCR konvensional mengandalkan anggapan bahwa arah varians terbesar pada prediktor paling berkaitan dengan respons (James et al., 2021), sedangkan pemilihan komponen yang hanya memakai informasi prediktor dapat kalah akurat dibanding yang memanfaatkan informasi respons (Ekvall, 2022). Karena itu, pemilihan komponen berdasarkan varians saja kurang memadai dan kriteria prediktif seperti LOOCV lebih tepat.

PC2 merupakan kontras antara kapasitas dan produksi relatif terhadap skala provinsi. Provinsi dengan produksi relatif besar terhadap kapasitasnya, seperti Jawa Tengah dan Banten, bernilai negatif pada PC2 sehingga dugaan distribusinya lebih rendah. Pola residual mendukung hal ini. Distribusi DKI Jakarta lebih tinggi dari dugaan model (residual +8.284,57 GWh), sedangkan Sumatera Selatan (-9.350,84 GWh) dan Jawa Tengah (-5.651,93 GWh) lebih rendah. Hal ini diduga karena listrik tidak selalu didistribusikan di provinsi tempat ia dibangkitkan. Provinsi pusat konsumsi seperti DKI Jakarta dapat memperoleh pasokan dari sistem interkoneksi, sedangkan provinsi dengan pembangkit besar dapat memasok wilayah lain. Kasus paling ekstrem adalah DI Yogyakarta yang hanya memiliki kapasitas 0,78 MW dan produksi 3,64 GWh tetapi mendistribusikan 3.891,18 GWh. Dugaan ini perlu dikonfirmasi dengan data aliran listrik antarwilayah.

Koefisien produksi bernilai negatif (-1,477) meskipun korelasi sederhananya dengan distribusi positif ($r = 0,827$). Hal ini merupakan efek multikolinearitas (*suppression*): koefisien tersebut adalah pengaruh bersyarat produksi ketika kapasitas dijaga tetap, sehingga kedua koefisien paling tepat dibaca sebagai satu kontras dan bersifat asosiatif, bukan kausal. Perlu dicatat pula bahwa karena kedua komponen dipertahankan, PCR secara numerik ekuivalen dengan OLS, sebagaimana terlihat dari LOOCV yang identik. Dengan dua prediktor, manfaat PCR terletak pada penanganan multikolinearitas dan diagnosis struktur korelasi, bukan pada reduksi dimensi.

Hasil ini perlu ditafsirkan dengan mempertimbangkan beberapa keterbatasan. Terdapat heteroskedastisitas dan sejumlah provinsi berpengaruh, dan R^2 turun dari 0,948 menjadi 0,752 ketika provinsi-provinsi Jawa tersebut dikeluarkan, sehingga tingginya daya jelas model sangat dipengaruhi provinsi tersebut meskipun arah dan signifikansi hubungan tetap. PC2 hanya memuat 0,95% varians, tetapi signifikan dan menaikkan Q^2 dari 0,612 menjadi 0,910. PCR konvensional mengandalkan anggapan bahwa arah varians terbesar pada prediktor paling berkaitan dengan respons (James et al., 2021), sedangkan pemilihan komponen yang hanya memakai informasi prediktor dapat kalah akurat dibanding yang memanfaatkan informasi respons (Ekvall, 2022). Karena itu, pemilihan komponen berdasarkan varians saja kurang memadai dan kriteria prediktif seperti LOOCV lebih tepat. Penelitian selanjutnya dapat menambahkan variabel tersebut, memakai transformasi logaritma atau galat baku *robust*, serta membandingkan PCR dengan regresi *ridge*. Secara praktis, temuan ini menunjukkan bahwa perencanaan distribusi listrik antarprovinsi perlu mempertimbangkan interkoneksi dan bukan hanya kapasitas pembangkit lokal.

4. Kesimpulan

Berdasarkan hasil analisis Principal Component Regression (PCR) terhadap kapasitas pembangkit dan produksi listrik dalam menjelaskan distribusi listrik antarprovinsi di Indonesia tahun 2024, dapat disimpulkan bahwa kedua variabel tersebut secara bersama-sama memiliki hubungan yang signifikan dengan distribusi listrik. Analisis

komponen utama menghasilkan dua komponen yang dipertahankan berdasarkan hasil validasi silang leave-one-out (LOOCV). Komponen utama pertama (PC1) merepresentasikan skala sistem pembangkitan, dengan koefisien positif yang menunjukkan bahwa peningkatan skala pembangkitan berkaitan dengan peningkatan distribusi listrik. Sementara itu, komponen utama kedua (PC2) menggambarkan kontras relatif antara kapasitas pembangkit dan produksi listrik serta turut memberikan informasi dalam menjelaskan variasi distribusi listrik antarprovinsi. Hasil penelitian juga menunjukkan adanya ketimpangan distribusi listrik yang cukup besar antarprovinsi, dengan lima provinsi, yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten, menyumbang sekitar dua pertiga distribusi listrik nasional. Meskipun model PCR menunjukkan kemampuan prediksi yang lebih baik dibandingkan model nol, hasil diagnostik mengindikasikan adanya heteroskedastisitas dan beberapa observasi berpengaruh yang perlu diperhatikan dalam interpretasi model. Selain itu, karena seluruh komponen utama dipertahankan, model PCR secara numerik ekuivalen dengan regresi linear berganda (OLS), sehingga manfaat utamanya dalam penelitian ini terletak pada penanganan multikolinearitas dan pemahaman struktur hubungan antarvariabel, bukan pada reduksi dimensi. Temuan ini menegaskan bahwa distribusi listrik antarprovinsi tidak cukup dijelaskan hanya berdasarkan kapasitas pembangkit dan produksi listrik di wilayah masing-masing, tetapi juga perlu mempertimbangkan interkoneksi jaringan serta kemungkinan penyaluran listrik antarwilayah. Penelitian selanjutnya disarankan untuk melibatkan variabel tambahan, seperti jumlah penduduk, aktivitas industri, dan aliran listrik antarwilayah, serta mempertimbangkan transformasi logaritma, galat baku robust, atau metode regresi alternatif untuk mengatasi keterbatasan model. Dengan demikian, analisis yang lebih komprehensif diharapkan dapat memberikan pemahaman yang lebih baik mengenai faktor-faktor yang berkaitan dengan distribusi listrik dan mendukung perencanaan sistem kelistrikan yang lebih terintegrasi di Indonesia.

Reference

1. Akbar, L., Mafruhah, I., & Gravitiani, E. (2024). Determinants of variables that affect electrical energy consumption in Indonesia 2011-2020. *International Journal of Energy Economics and Policy*, 14(1), 165–171. <https://doi.org/10.32479/ijeep.11069>
2. Badan Pusat Statistik. (2025). *Statistik listrik 2020-2024*. Badan Pusat Statistik. <https://www.bps.go.id/id/publication/2025/12/31/c4f7aade5d4801236ec79b81/statistik-listrik-2020-2024.html>
3. Badan Pusat Statistik. (2025). *Statistik listrik 2020-2024*. <https://www.bps.go.id/id/publication/2025/12/31/c4f7aade5d4801236ec79b81/electric-statistics-2020-2024.html>
4. Badan Pusat Statistik. (2026a). *Electricity distributed by province (GWh), 2024*. <https://www.bps.go.id/en/statistics-table/2/ODU5IzI%3D/electricity-distributed-by-province--gwh-.html>
5. Balqis, N. A., Astutik, S., & Solimun. (2022). Principal component regression modelling with variational Bayesian approach to overcome multicollinearity at various levels of missing data proportion. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 6(4), 1013–1026. <https://journal.ummat.ac.id/index.php/jtam/article/download/10223/pdf>
6. Davino, C., Romano, R., & Vistocco, D. (2022). Handling multicollinearity in quantile regression through the use of principal component regression. *METRON*, 80, 153–174. <https://doi.org/10.1007/s40300-022-00230-3>
7. Effendi, M., Ardhyatirta, R., Angelina, S. G., & Ohyyer, M. (2023). Predict farmer exchange rate in the food crop sector using principal component regression. *Enthusiastic: International Journal of Applied Statistics and Data Science*, 3(1). <https://doi.org/10.20885/enthusiastic.vol3.iss1.art7>
8. Ekvall, K. O. (2022). Targeted principal components regression. *Journal of Multivariate Analysis*, 190, Article 104995. <https://doi.org/10.1016/j.jmva.2022.104995>
9. Ergisi, S., Doğanay, B., & Yavuz, Y. (2024). Performance comparison of least squares, ridge, lasso and principal component regression for addressing multicollinearity in regression analysis. *İstatistik Araştırma Dergisi*, 14(2). <https://dergipark.org.tr/en/pub/jsstr/article/1549414>
10. Gu, F. (2023). A model-based approach to multivariate principal component regression: Selecting principal components and estimating standard errors for unstandardized regression coefficients. *British Journal of Mathematical and Statistical Psychology*, 76(3), 605–622. <https://doi.org/10.1111/bmsp.12301>
11. Indhova, R. A., Suparti, & Hakim, A. R. (2023). Perbandingan Gulud Regression dan Principal Component Regression (PCR) terhadap pemodelan Indeks Pembangunan Manusia Provinsi Jawa Timur. *Jurnal Gaussian*, 12(2), 199–208. <https://doi.org/10.14710/j.gauss.12.2.199-208>
12. International Journal of Energy Sector Management. (2022). A PCA-based variable ranking and selection approach for electric energy load forecasting. *International Journal of Energy Sector Management*, 16(6), 1172–1191. <https://doi.org/10.1108/IJESM-12-2019-0009>
13. Iswan, I., Garniwa, I., & Surjandari, I. (2021). Data optimization on the accuracy of forecasting electricity energy sales using principal component analysis based on spatial. *International Journal of Energy Economics and Policy*. <https://scholar.ui.ac.id/en/publications/data-optimization-on-the-accuracy-of-forecasting-electricity-ener/>
14. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
15. Khoiro, I., Asfa'ani, E., & Musthofa, S. (2024). Perbandingan Principal Component Regression dan regresi ridge pada analisis faktor-faktor Indeks Pembangunan Manusia di Provinsi Sumatera Barat. *MApp (Mathematics and Applications) Journal*. <https://doi.org/10.15548/map.v6i2.10168>
16. Menteri Energi dan Sumber Daya Mineral Republik Indonesia. (2007). *Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 03 Tahun 2007 tentang Aturan Jaringan Sistem Tenaga Listrik Jawa-Madura-Bali*. https://gatrik.esdm.go.id/assets/uploads/download_index/files/41992-permen-esdm-03-2007.pdf
17. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). John Wiley & Sons. <https://www.wiley-vch.de/de/fachgebiete/mathematik-und-statistik/introduction-to-linear-regression-analysis-978-1-119-57872-7>

18. Nasution, E. F., & Pane, R. (2024). Perbandingan regresi LASSO dan Principal Component Regression dalam mengatasi masalah multikolinearitas. *MES: Journal of Mathematics Education and Science*, 10(1). <https://jurnal.uisu.ac.id/index.php/mesuisu/article/view/9656>
19. Ningrum, S. H. P., Hisan, K., Ramdhani, T. P., Luzianawati, L., Zindawi, M. D. R., & Harsyiah, L. (2025). Regresi komponen utama dalam mengatasi multikolinieritas pada faktor-faktor yang mempengaruhi inflasi di Indonesia. *Indonesian Journal of Applied Statistics and Data Science*, 2(1), 34–43. <https://doi.org/10.29303/ijasds.v2i1.5827>
20. Parhizkar, T., Rafieipour, E., & Parhizkar, A. (2021). Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction. *Journal of Cleaner Production*, 279, 123866. <https://doi.org/10.1016/j.jclepro.2020.123866>
21. Qur'ani, A. Y. (2023). Pemodelan Principal Component Regression Analysis dari faktor penanganan stunting saat pandemi Covid-19 di Indonesia. *ULIL ALBAB: Jurnal Ilmiah Multidisiplin*, 2(8), 3922–3931. <https://doi.org/10.56799/jim.v2i8.2053>
22. Republik Indonesia. (2009). *Undang-Undang Nomor 30 Tahun 2009 tentang Ketenagalistrikan*. https://gatrik.esdm.go.id/assets/uploads/download_index/files/9ef73-03.uu-30-2009-tentang-ketenagalistrikan.pdf
23. Rong, J.-Y., & Liu, X.-Q. (2024). Hybrid principal component regression estimation in linear regression. *Electronic Research Archive*, 32(6), 3758–3776. <https://doi.org/10.3934/era.2024171>
24. Rosidin, I. S., de Groot, H. L. F., & Mulder, P. (2026). Power islands: The geography of electricity diffusion in Indonesia, 1975–2018. *Energy Strategy Reviews*, 64, Article 102077. <https://doi.org/10.1016/j.esr.2026.102077>
25. Sekretariat Kabinet Republik Indonesia. (2021, 23 Januari). *Hasil Sensus Penduduk 2020; BPS: Meski lambat, ada pergeseran penduduk antarpulau*. <https://setkab.go.id/hasil-sensus-penduduk-2020-bps-meski-lambat-ada-pergeseran-penduduk-antarpulau/>
26. Shah, H. I., Iftikhar, H., & Ali, S. (2022). Modeling and forecasting electricity demand and prices: A comparison of alternative approaches. *Journal of Mathematics*, 2022, 3581037. <https://doi.org/10.1155/2022/3581037>
27. Veeramsetty, V., Chandra, D. R., Grimaccia, F., & Mussetta, M. (2022). Short term electric power load forecasting using principal component analysis and recurrent neural networks. *Forecasting*, 4(1), 149–164. <https://doi.org/10.3390/forecast4010008>
28. Yang, Q., & Yin, S. (2022). Analysis of electricity consumption behaviors based on principal component analysis and density peak clustering. *Concurrency and Computation: Practice and Experience*. <https://doi.org/10.1002/cpe.7126>