

Madani: Jurnal Ilmiah Multidisiplin
Volume 3, Nomor 2, Maret 2025, P. 519-528
 E-ISSN: 2986-6340
 Licenced by CC BY-SA 4.0
 DOI: <https://doi.org/10.5281/zenodo.15073386>

Analisis Faktor yang Mempengaruhi Angka Melek Huruf di Indonesia Dengan Regresi Robust

Indy Pratiwi^{1*}, Wira Triono², Rizky Rafiza Panjaitan³, Risa Rozzaqi Rambe⁴, Syairal Fahmy Dalimunthe⁵

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan
 Email : ^{1*}indykisan2323@gmail.com, ²wiratriono3366@gmail.com, ³rizkyraviza84@gmail.com,
⁴risarozzaqi@gmail.com

Abstract

This study aims to analyze the factors that affect literacy rates in Indonesia using a robust regression approach. The data used included data from 34 provinces, focusing on the dependent variables, namely the literacy rate and the independent variables, namely the average length of schooling, the number of teachers, out-of-school children, PDRB per capita, and the number of poor people. The results of the analysis show the equation $\hat{Y} = 97.7766 - 0.41220PCA1 - 0.2886PCA2 - 0.000002x_4$, where $PCA1$ = a combination of $X1$ and $X2$ and $PCA2$ = a combination of $X2$ and $X5$. Every 1 unit increase in $PCA1$ will reduce the literacy rate by about 0.4120 or 41.2%, an increase of 1 unit of $PCA2$ will reduce the literacy rate by about 0.2886 or 28.86%, while in x_4 it only affects about 0.00022%. The number of children who do not attend school, children who drop out of school and the number of poor people can significantly increase the illiteracy rate. Efforts that can be made by the government are to facilitate education so that all groups can access education and educate the public that education is very important and will be the foundation of the nation and state in the future..

Keywords: robust regression, literacy rate, factors, outliers, ols

Abstrak

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi angka melek huruf di Indonesia dengan menggunakan pendekatan regresi robust. Data yang digunakan mencakup data dari 34 provinsi, dengan berfokus pada variabel dependen yaitu angka melek huruf dan variabel independen yaitu rata-rata lama sekolah, jumlah guru, anak tidak sekolah, PDRB perkapita, dan jumlah penduduk miskin. Hasil analisis menunjukkan persamaan $\hat{Y} = 97.7766 - 0.41220PCA1 - 0.2886PCA2 - 0.000002X4$, dimana $PCA1$ = kombinasi $X1$ dan $X2$ dan $PCA2$ = kombinasi $X2$ dan $X5$. Setiap peningkatan 1 unit pada $PCA1$ akan mengurangi angka melek huruf sekitar 0.4120 atau 41,2%, peningkatan 1 unit $PCA2$ akan mengurangi angka melek huruf sekitar 0.2886 atau 28,86%, sedangkan pada x_4 hanya memengaruhi sekitar 0.00022%. Banyaknya anak yang tidak bersekolah, anak putus sekolah dan jumlah penduduk miskin dapat meningkatkan angka buta huruf secara signifikan. Upaya yang dapat dilakukan oleh pemerintah adalah memfasilitasi pendidikan sehingga semua golongan dapat mengakses pendidikan serta mengedukasi masyarakat bahwa pendidikan itu sangat penting dan akan menjadi pondasi bangsa dan negara dimasa yang akan datang.

Kata Kunci: regresi robust, angka melek huruf, faktor, pencilan, metode ols

Article Info

Received date: 12 March 2025

Revised date: 17 March 2025

Accepted date: 22 March 2025

PENDAHULUAN

Pendidikan adalah hal yang penting bagi kehidupan, pendidikan tidak hanya ditunjukkan agar seseorang menjadi pandai dan ahli dalam suatu hal tertentu. Pendidikan juga dipakai dalam proses pembentukan karakter peserta didik, pembelajaran dalam penerapan budi pekerti, akhlak dan pengembangan bakat peserta didik dan juga untuk mengembangkan keterampilan yang belum atau sudah dimiliki peserta didik (Lestari .A.Y.B. et all, 2020).

Persoalan buta huruf masih menjadi isu sentral di Indonesia. Tak hanya menggambarkan ketidakmampuan seseorang dalam membaca dan menulis, tetapi juga kemanusiaan yang berdampak pada perkembangan pendidikan. Tingginya angka buta huruf di Indonesia dipengaruhi oleh 2 masalah utama, yaitu masalah dalam sistem pendidikan dan angka kemiskinan. Meski pemerintah sudah melakukan banyak upaya, penuntasan buta huruf selalu terhalang oleh kedua faktor tersebut. Pembangunan yang tidak merata, hal ini menyebabkan masyarakat yang tinggal di daerah terpencil menjadi lebih sulit dalam mengakses pendidikan. Angka putus sekolah dan menurunnya perhatian

pemerintah terhadap pendidikan juga menjadi masalah utama dalam kepekaan masyarakat dalam membaca dan menulis. Seperti yang dikatakan oleh Anis Baswedan “anak-anak yang tidak terdidik di republik ini merupakan ‘dosa’ dari setiap orang terdidik yang dimiliki oleh republik ini. Anak-anak nusantara tidak berbeda. Mereka semua berpotensi. Mereka hanya dibedakan oleh keadaan.”(parsande. P. et all, 2020).

Pertumbuhan ekonomi dinyatakan oleh Pendapatan Regional Domestik Bruto (PDRB) per kapita karena menunjukkan tingkat pendapatan individu secara rata-rata. Memasukkan variabel investasi fisik, belanja pemerintah, serta indeks pembangunan manusia sebagai input persamaan pertumbuhan ekonomi. Pengeluaran pemerintah meningkatkan akses masyarakat terhadap fasilitas pembangunan sehingga proses kegiatan ekonomi berjalan semakin efisien dan produktif. Pertumbuhan ekonomi, yang diukur dengan Peningkatan Produk Domestik Regional Bruto atau PDRB. Sering digunakan sebagai tolak ukur utama keberhasilan pembangunan dalam suatu wilayah. Meningkatnya PDRB secara terus-menerus diharapkan dapat mengurangi angka pengangguran dengan pembukaan lapangan pekerjaan dan pada akhirnya akan menurunkan tingkat kemiskinan. Kemiskinan selalu menjadi masalah yang kompleks bagi negara terutama negara berkembang, contohnya Indonesia. Kemiskinan tidak hanya terdapat pada wilayah pedesaan tapi juga perkotaan, karena sumber daya manusia yang rendah serta urbanisasi yang kurang terkendali hal ini menyebabkan pertumbuhan ekonomi melambat dan meningkatnya angka kemiskinan.

Menurut BPS (Badan Pusat Statistik Indonesia) kemiskinan itu adalah ketika seseorang tidak mampu mencukupi keperluan dasar untuk kehidupan yang seimbang (termasuk pangan dan non-pangan). Sedangkan menurut *world bank* (2000) kemiskinan merupakan kehilangan kesejahteraan (*deprivation of well being*). Kemiskinan yang diabaikan secara terus menerus bakal berakibat buruk dalam berbagai aspek kehidupan, seperti rendahnya tingkatan penghasilan penduduk yang akan mengakibatkan rendahnya tingkat kesehatan serta Pendidikan, tingginya tingkat pengangguran serta kriminalitas, dan juga buruknya keadaan lingkungan hidup (Hasibuan R.R.A, et all, 2022).

Rata rata lama sekolah merupakan rata rata jumlah tahun yang telah diselesaikan oleh penduduk pada seluruh jenjang pendidikan formal yang pernah dijalani, perhitungan dilakukan pada usia 25 tahun keatas. Rata rata lama sekolah mengidentifikasikan semakin tinggi pendidikan yang dicapai oleh masyarakat suatu daerah maka indeks pembangunan manusia pun semakin baik (Manurung. E.N & Hutabarat. F, 2021).

Melek huruf merupakan kemampuan individu dalam membaca, menulis, dan memahami informasi tertulis yang menjadi dasar untuk berpartisipasi dalam kehidupan sosial dan ekonomi (UNESCO, 2006). Angka Melek Huruf (AMH) menurut Badan Pusat Statistik (BPS, 2021) merujuk pada persentase penduduk berusia 15 tahun ke atas yang memiliki kemampuan membaca dan menulis dalam huruf Latin maupun aksara lainnya. Kemampuan ini tidak mensyaratkan pemahaman terhadap isi tulisan, tetapi hanya mencerminkan keterampilan dasar dalam membaca dan menulis di kalangan penduduk usia tersebut. Beberapa faktor utama yang memengaruhi angka melek huruf meliputi :

- Tingkat Kemiskinan: Kemiskinan menjadi faktor utama rendahnya angka melek huruf. Keterbatasan ekonomi membuat masyarakat sulit memenuhi kebutuhan dasar, sehingga pendidikan seringkali tidak menjadi prioritas (*Lubis et al., 2022*).
- Akses terhadap Teknologi Informasi dan Komunikasi (TIK): Ketersediaan dan penggunaan perangkat teknologi, seperti telepon seluler dan akses internet, berperan dalam meningkatkan literasi. Rumah tangga yang memiliki akses ke TIK cenderung memiliki tingkat melek huruf yang lebih tinggi (*Lake & Utami, 2020*).
- Fasilitas Pendidikan: Ketersediaan dan kualitas lembaga pendidikan, terutama di daerah terpencil, mempengaruhi tingkat melek huruf. Daerah dengan akses terbatas ke sekolah cenderung memiliki angka buta huruf yang lebih tinggi (*Lubis et al., 2022*).
- Tingkat Partisipasi Sekolah Dasar dan Angka Putus Sekolah: Daerah dengan partisipasi sekolah dasar yang rendah dan angka putus sekolah yang tinggi cenderung memiliki tingkat buta huruf yang lebih tinggi (*Lake & Utami, 2020*).

Sejak tahun 2010, Badan Pusat Statistik (BPS) Indonesia telah mengadopsi metodologi baru dalam perhitungan Indeks Pembangunan Manusia (IPM). Sebagai gantinya, digunakan dua indikator baru, yaitu:

1. **Harapan Lama Sekolah (HLS):** Menggambarkan jumlah tahun sekolah yang diharapkan akan ditempuh oleh anak-anak usia 7 tahun ke atas.
2. **Rata-rata Lama Sekolah (RLS):** Menunjukkan rata-rata jumlah tahun sekolah yang telah diselesaikan oleh penduduk usia 25 tahun ke atas.

Perubahan ini dilakukan karena AMH dianggap kurang mampu mencerminkan kualitas pendidikan secara menyeluruh, terutama karena angka melek huruf di banyak daerah sudah mencapai tingkat yang sangat tinggi, sehingga tidak lagi efektif dalam membedakan tingkat pendidikan antar wilayah. Dengan menggunakan HLS dan RLS, diharapkan dapat diperoleh gambaran yang lebih komprehensif mengenai capaian pendidikan Masyarakat (Humaira, 2022)

Tujuan penelitian ini adalah untuk mengidentifikasi dan menganalisis hubungan antara angka melek huruf dengan variabel-variabel independen, menganalisis variasi tingkat melek huruf di berbagai daerah, membandingkan efektivitas model regresi OLS dan regresi robust, serta memberikan rekomendasi kebijakan untuk meningkatkan angka melek huruf melalui pengembangan pendidikan dan pengurangan faktor-faktor sosial ekonomi yang berpengaruh.

METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah data sekunder dengan 34 provinsi di Indonesia. Data pada penelitian ini bersumber dari laman <https://www.bps.go.id/id> dan <https://dapo.dikdasmen.go.id/> yaitu data Angka Melek Huruf (15 tahun keatas) sebagai variabel Y, Rata-rata lama sekolah (15 tahun keatas) sebagai variabel X1, jumlah guru sebagai variabel X2, Angka anak tidak sekolah sebagai variabel X3, PDRB per kapita sebagai variabel X4 dan jumlah penduduk miskin sebagai variabel X5. Seluruh data yang digunakan merupakan data pada tahun 2023.

Sebelum melanjutkan ke regresi robust, data diuji terlebih dahulu dengan estimasi *ordinary least square* jika salah satu dari beberapa asumsi klasik seperti normalitas, *homoskedastisitas*, *autokorelasi*, dan *multikolinearitas* tidak terpenuhi maka lanjut ke regresi robust. Berikut adalah langkah-langkah yang dilakukan pada penelitian ini:

1. Uji korelasi antar variabel

Uji korelasi yang digunakan pada penelitian ini adalah *spearman rank correlation* yang mana tidak memerlukan asumsi variabel harus terdistribusi normal dan berasal dari data ordinal (Nelvidawati & Kasman, 2023).

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

keterangan :

ρ = Nilai korelasi spearman

d = Selisih antara X dan Y

n = Jumlah pasangan (data)

metode korelasi “(1)” menghasilkan nilai ρ mulai dari -1 yang berarti korelasi sempurna berbanding terbalik dan +1 berarti korelasi sempurna berbanding lurus. Jika antar variabel independen terdapat korelasi yang kuat maka dilakukan PCA (*principal component analysis*) untuk mengatasi *multikolinearitas*.

2. Model OLS (*Ordinary Least Square*)

Ordinary least square adalah metode untuk mendapatkan parameter regresi linier dengan meminimalkan jumlah kuadrat kesalahan antara nilai observasi dan nilai prediksi (Putra & Muhammad, 2024). Model OLS digunakan sebagai model *null* yang kemudian akan dilakukan uji asumsi untuk memastikan apakah model OLS sudah cukup untuk dijadikan sebagai penilaian faktor dibanding model *robust*. Dengan “(2)” adalah bentuk dari perhitungan estimasi dari OLS :

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2)$$

Dengan X adalah bentuk matrix dari variabel independen (prediktor) dan Y bentuk vektor dari variabel dependen (respons).

3. Uji Normalitas

Asumsi yang pertama yang harus dipenuhi adalah *residual* dari model OLS harus berdistribusi normal, karena data yang digunakan pada penelitian ini berjumlah 34 data

maka digunakan uji Shapiro-Wilk karena lebih akurat untuk sampel kecil atau dibawah 40 sampel (Ahadi & Zain, 2023).

4. Uji Homoskedastisitas

Salah satu uji Homoskedastisitas adalah Uji Breusch-Pagan, dalam uji Breusch-Pagan nilai statistik dihitung dengan kuadrat dari *residual* yang terstandarisasi terhadap variabel-variabel independen (Anggraeni & Putri, 2021). Dimana Jika P-value < 0.05, maka tolak $H_0 \rightarrow$ terdapat heteroskedastisitas (tidak homogen).

5. Uji Multikolinearitas

Salah satu cara untuk mendeteksi *multikolinearitas* adalah dengan *Variance Inflation Factor* (VIF). Multikolinearitas terjadi ketika variabel independen dalam regresi memiliki korelasi tinggi satu sama lain.

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3)$$

Persamaan “(3)” menunjukkan di mana R_i^2 adalah koefisien determinasi dari regresi antara variabel independen X_i dengan variabel *independen* lainnya. Jika nilai $VIF > 10$ maka terdapat bukti kuat adanya *multikolinearitas* antar variabel *independen* (Azizah, Arum, & Wasono, 2021).

6. Uji Autokorelasi

Uji Breusch-Godfrey digunakan untuk menguji *autokorelasi residual* dalam regresi linier, terutama ketika ada *autokorelasi* orde lebih tinggi (lag lebih dari 1) yang tidak dapat dideteksi oleh Durbin-Watson. Jika p-value < 0.05, maka tolak $H_0 \rightarrow$ ada *autokorelasi* pada *residual*.

7. Deteksi Pencilan

Salah satu cara untuk mendeteksi pencilan adalah dengan metode Z-Score yaitu mengukur seberapa jauh suatu nilai dari rata-rata dalam satuan standar deviasi. Jika nilai dari $Z_i \geq 3$ maka data tersebut adalah pencilan.

8. Model Regresi Robust

Regresi robust digunakan ketika data memiliki pencilan atau pelanggaran asumsi OLS (seperti heteroskedastisitas atau distribusi error yang tidak normal). Model robust akan diestimasi dengan metode *MM-estimator*. merupakan pengestimasian parameter regresi menggunakan S-estimator yang meminimumkan scale dari galat pada M-estimator kemudian dilanjutkan dengan M-estimator (putri, yozza, & yanuar, 2021).

$$\begin{aligned} \hat{\beta} &= \arg \min_{\beta} \sum_{i=1}^n \rho(u_i) \\ &= \arg \min_{\beta} \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{s} \right) \end{aligned} \quad (4)$$

Persamaan “(4)” menunjukkan dimana:

x_{ij} = data ke-i pada variabel bebas ke-j

s = standar deviasi yang diperoleh dari galat dari S-estimator

ρ = fungsi objektif Huber

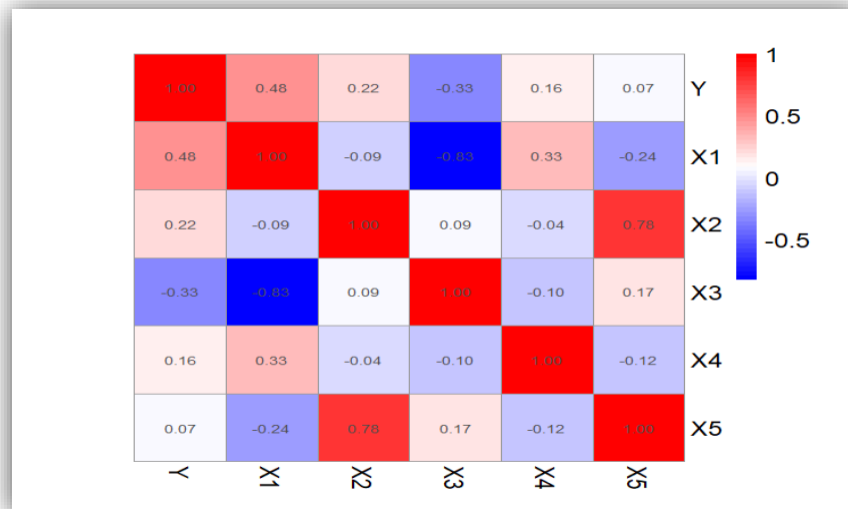
HASIL DAN PEMBAHASAN

Analisis Korelasi Antar Variabel

Analisis korelasi menggunakan metode Spearman bertujuan untuk mengukur kekuatan dan arah hubungan antara variabel-variabel dalam dataset yang bersifat ordinal atau tidak terdistribusi normal. Hasil korelasi akan ditampilkan dalam bentuk matriks dimana setiap nilai dalam matriks menunjukkan koefisien korelasi antara dua variabel, dengan kriteria nilai :

- Nilai korelasi yang mendekati 1 menunjukkan adanya hubungan positif yang kuat antara dua variabel (jika satu variabel meningkat, maka variabel lainnya cenderung meningkat).
- Nilai korelasi yang mendekati -1 menunjukkan adanya hubungan negatif antara dua variabel (jika satu variabel meningkat, maka variabel lainnya cenderung menurun). N
- Nilai korelasi mendekati 0 menunjukkan tidak adanya hubungan yang jelas antara kedua variabel (Perubahan pada satu variabel tidak diikuti oleh perubahan pada variabel lain).

Berikut adalah hasil analisis korelasi dengan menggunakan metode Spearman



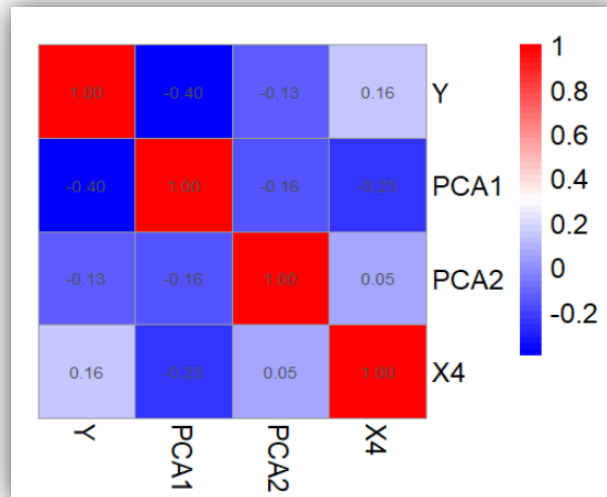
Gambar 1. Matriks Korelasi Antar Variabel

- Dari matriks korelasi dapat dilihat bahwa hubungan antara variabel Y dan variabel X adalah sebagai berikut :
 1. **X1 (rata-rata lama sekolah)** : Hasil analisis menunjukkan nilai korelasi positif yang cukup signifikan antara angka melek huruf dan rata-rata lama sekolah, dengan koefisien korelasi sebesar 0.48. Ini mengindikasikan bahwa peningkatan rata-rata lama sekolah dapat berpotensi meningkatkan angka melek huruf di masyarakat.
 2. **X2 (jumlah guru)** : Hasil analisis menunjukkan nilai korelasi yang rendah antara angka melek huruf dan jumlah guru, dengan koefisien korelasi sebesar 0.22. Ini mengindikasikan bahwa Jumlah guru tidak memiliki pengaruh signifikan terhadap angka melek huruf, meskipun Jumlah guru penting, faktor lain mungkin lebih berpengaruh.
 3. **X3 (angka anak tidak sekolah)** : Hasil analisis menunjukkan nilai korelasi negatif yang kecil antara angka melek huruf dan angka anak tidak sekolah, dengan koefisien korelasi sebesar -0.33. Hal ini mengindikasikan bahwa peningkatan jumlah anak tidak sekolah berhubungan dengan penurunan angka melek huruf.
 4. **X4 (PDRB per kapita)** : Hasil analisis menunjukkan nilai korelasi yang cukup rendah antara angka melek huruf dan PDRB per kapita dengan koefisien korelasi sebesar 0.16. Ini mengindikasikan bahwa PDRB per kapita tidak menjadi faktor utama dalam meningkatkan angka melek huruf.
 5. **X5 (jumlah penduduk miskin)** : Hasil analisis menunjukkan nilai korelasi yang sangat rendah antara angka melek huruf dan jumlah penduduk miskin, dengan koefisien korelasi sebesar 0.07 (mendekati 0). Ini mengindikasikan bahwa tidak ada hubungan yang signifikan antara Jumlah penduduk miskin dan angka melek huruf.
- Dari matriks korelasi juga dapat dilihat hubungan antar variabel X itu sendiri, khususnya korelasi antar variabel X1 dan X3 serta korelasi antara variabel X2 dan X5 yang sangat signifikan, yaitu sebagai berikut :
 1. **X1 dan X3** : Terdapat hubungan negatif yang cukup besar antara rata-rata lama sekolah dan angka anak tidak sekolah, dengan koefisien korelasi sebesar -0.83. Ini mengindikasikan bahwa peningkatan rata-rata lama sekolah sangat berhubungan dengan penurunan angka anak tidak sekolah.
 2. **X2 dan X5** : Terdapat hubungan positif yang cukup besar antara jumlah guru dan Jumlah penduduk miskin, dengan koefisien korelasi sebesar 0.78. Ini mengindikasikan bahwa meningkatkan jumlah guru di wilayah dengan Tingkat kemiskinan tinggi dapat membantu dalam meningkatkan akses Pendidikan dan kualitas Pendidikan di wilayah tersebut.

Berdasarkan uji korelasi dapat dilihat bahwa korelasi antara variabel X1 dan X3 serta korelasi antara variabel X2 dan X4 itu cukup besar yang artinya bahwa kedua variabel tersebut memiliki informasi yang serupa. Maka dalam hal ini dilakukan analisis PCA (*Principal Component Analysis*) yang

digunakan untuk mengurangi dimensi data dengan cara mengubah sekumpulan variabel yang berkorelasi menjadi sekumpulan variabel baru yang tidak berkorelasi, yang disebut dengan komponen utama.

Peneliti berfokus pada dua komponen utama yang dihasilkan dari PCA, yaitu PCA1 (komponen yang dihasilkan dari kombinasi variabel X1 (rata-rata lama sekolah) dan X3 (angka anak tidak sekolah)) dan PCA2 (komponen yang dihasilkan dari kombinasi variabel X2 (jumlah guru) dan X5 (Jumlah penduduk miskin)). Berikut adalah analisis korelasi yang dihasilkan setelah melakukan analisis PCA (Principal Component Analysis)



Gambar 2. Matriks Korelasi Setelah Analisis PCA

Sehingga secara keseluruhan dapat disimpulkan, bahwa tidak ada korelasi yang sangat kuat antara variabel independen dalam data setelah dilakukan analisis PCA. Hal ini menunjukkan bahwa meskipun terdapat hubungan antara variabel-variabel, dampaknya tidak signifikan. Hasil PCA ini kemudian akan dilakukan untuk analisis yang lebih lanjut

Membuat Model OLS (Ordinary Least Squares)

OLS (*Ordinary Least Squares*) bertujuan untuk menemukan garis regresi yang meminimalkan jumlah kuadrat dari selisih antara nilai yang diamati (data sebenarnya) dan nilai yang diprediksi oleh model. Dengan kata lain, OLS mencari parameter model yang menghasilkan prediksi paling mendekati data asli. Berikut adalah model OLS yang dihasilkan:

```

Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)  9.777663e+01  7.989015e-01  122.38885 < 2.22e-16 ***
PCA1         -1.576656e+00  3.757157e-01  -4.19641  0.00022199 ***
PCA2         -6.723894e-01  3.382874e-01  -1.98763  0.05604782 .
X4           -8.049009e-06  1.370363e-05  -0.58736  0.56135479
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.622703 on 30 degrees of freedom
Multiple R-squared:  0.3943417, Adjusted R-squared:  0.3337758
F-statistic:  6.51096 on 3 and 30 DF,  p-value: 0.001592323

```

Gambar 3. Model OLS

Berdasarkan koefisien dan hasil analisis yang ditampilkan, model OLS dapat dirumuskan sebagai berikut:

$$\hat{Y} = \beta_0 + \beta_1 PCA1 + \beta_2 PCA2 + \beta_3 x_4$$

$$\hat{Y} = 97.7766 + (-1.5766) PCA1 + (-0.6723) PCA2 + (-0.000008) X_4$$

$$\hat{Y} = 97.7766 - 1.5766 PCA1 - 0.6723 PCA2 - 0.000008 X_4$$

R-squared adalah ukuran yang menunjukkan proporsi variabilitas dalam variabel dependen (dalam hal ini, angka melek huruf) yang dapat dijelaskan oleh variabel independen dalam model. Nilai R-Square sebesar 0.3943417 berarti bahwa sekitar 39,43% variasi angka melek huruf dapat dijelaskan

oleh model regresi ini. Sebaliknya, sekitar 60.66% variasi dalam data tidak dapat dijelaskan oleh model dan mungkin disebabkan oleh faktor lain atau variabel yang tidak termasuk dalam model.

Uji Normalitas

Dari model regresi yang telah dibangun menggunakan metode Ordinary Least Squares (OLS), langkah selanjutnya adalah melakukan uji normalitas terhadap residual model. Uji ini bertujuan untuk memastikan bahwa residual terdistribusi normal, yang merupakan salah satu asumsi dasar dalam analisis regresi. Berikut adalah hasil uji normalitas dengan menggunakan uji *Shapiro-Wilk*, dengan hipotesis :

- H_0 : Residual mengikuti distribusi normal.
- H_1 : Residual tidak mengikuti distribusi normal.

```
Shapiro-Wilk normality test

data: residuals_ols
W = 0.88512497, p-value = 0.001899074
```

Gambar 4. Hasil Uji Normalitas dengan Kolmogorov-Smirnov

Dari hasil uji normalitas residual model dengan *Shapiro-Wilk* menunjukkan bahwa p-value $0.001899074 < 0.05$. Karena p-value < 0.05 , maka terdapat cukup bukti untuk menolak hipotesis nol yang menyatakan bahwa residual terdistribusi normal. Yang artinya bahwa residual dari model OLS yang diuji menunjukkan penyimpangan yang signifikan dari distribusi normal (asumsi normalitas tidak terpenuhi).

Uji Homoskedastisitas

Uji homoskedastisitas dilakukan untuk memeriksa apakah asumsi kunci dalam analisis regresi terpenuhi, yaitu bahwa varians dari residual (kesalahan prediksi) adalah konstan di seluruh rentang nilai variabel independen. Uji Homoskedastisitas yang digunakan dalam hal ini adalah dengan metode uji *Breusch-Pagan*. Hipotesis dalam uji ini dinyatakan sebagai berikut:

- H_0 : Varians residual adalah konstan (homoskedastis)
- H_1 : Varians residual tidak konstan (heteroskedastis)

Hasil dari uji Homoskedastisitas yang telah dilakukan adalah sebagai berikut :

```
studentized Breusch-Pagan test

data: model_ols
BP = 6.9444288, df = 3, p-value = 0.07369027
```

Gambar 5. Hasil Uji Homoskedastisitas dengan Breusch-Pagan

Dari hasil uji homoskedastisitas didapat nilai p-value sebesar $0.07369027 > 0.05$. Karena p-value > 0.05 , maka tidak terdapat cukup bukti untuk menolak hipotesis nol, yang menyatakan bahwa varians residual dari model adalah konstan. Sehingga asumsi homoskedastisitas terpenuhi dan dapat diandalkan untuk analisis lebih lanjut.

Uji Multikolinearitas

Variance Inflation Factor (VIF) digunakan untuk mengukur tingkat multikolinearitas di antara variabel independen dalam model regresi. VIF menunjukkan seberapa banyak varians dari estimasi koefisien regresi meningkat karena adanya multikolinearitas. Semakin tinggi nilai VIF, semakin besar kemungkinan adanya multikolinearitas. Berdasarkan pengujian multikolinearitas dengan melihat nilai VIF diperoleh hasil sebagai berikut :

```
> print(vvif)
      PCA1      PCA2      X4
1.229605627 1.069491615 1.169816621
```

Gambar 6. Nilai vvif untuk Uji Multikolinearitas

Secara keseluruhan, nilai VIF untuk variabel independen dalam model dibawah 10. Hal ini menunjukkan tidak adanya multikolinearitas yang serius dalam variable-variable yang diteliti.

Uji Autokorelasi

Uji Autokorelasi berfungsi untuk menganalisis adanya hubungan antara residual (kesalahan) model regresi pada waktu yang berbeda. Autokorelasi terjadi Ketika nilai residual pada satu waktu berkorelasi dengan nilai residual pada waktu lainnya. Dalam uji autokorelasi ini memakai metode *Brusch-Godfrey* yang dengan hipotesis sebagai berikut:

- H_0 : Tidak ada autokorelasi dalam residual model. Ini berarti bahwa residual bersifat independen satu sama lain.
- H_1 : Ada autokorelasi dalam residual model. Ini berarti bahwa residual tidak independen.

```
> bgtest(model_ols)

Breusch-Godfrey test for serial correlation of order up to 1

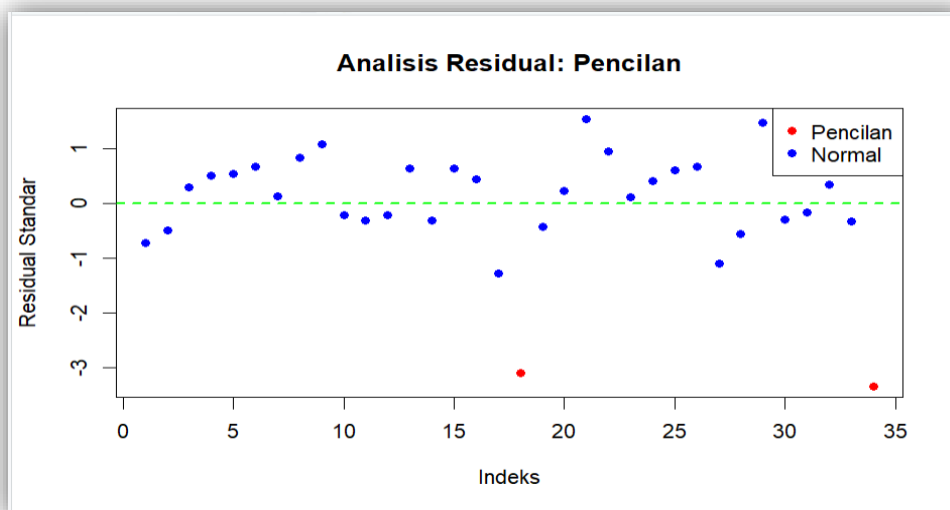
data: model_ols
LM test = 3.4056642, df = 1, p-value = 0.06497295
```

Gambar 7. Hasil Uji Autokorelasi dengan Metode *Brusch-Godfrey*

Dari hasil uji autokorelasi didapat nilai p-value sebesar $0.06497295 > 0.05$. Karena p-value > 0.05 , maka tidak terdapat cukup bukti untuk menolak hipotesis nol, yang menyatakan bahwa tidak ada autokorelasi dalam residual model. Hal ini menjadi indikasi positif bahwa model regresi berfungsi dengan baik dan hasil analisis dapat diandalkan.

Uji Pencilan

Uji pencilan dilakukan untuk mengidentifikasi data yang tidak biasa atau outlier dalam dataset. Outlier merupakan pengamatan yang berada jauh dari pola umum data dan dapat mempengaruhi hasil analisis regresi secara signifikan. Ketika data tidak memenuhi asumsi normalitas, maka keberadaan pencilan menjadi signifikan karena dapat menyebabkan pelanggaran asumsi dasar model regresi. Oleh karena itu, penting untuk mengenali dan menangani pencilan. Plot berikut menunjukkan pencilan yang terdapat dalam data :



Gambar 8. Uji Pencilan

Plot ini berguna untuk mendeteksi dan menganalisis pencilan dalam data yang digunakan untuk model regresi. Dalam grafik analisis residual yang ditampilkan, terdapat dua pencilan yang teridentifikasi yaitu pada indeks ke-18 (Bali) dan indeks ke-34 (Papua Barat). Ini menunjukkan bahwa nilai residual untuk Kedua wilayah ini berbeda signifikan dari nilai residual wilayah lainnya. Hal ini mungkin disebabkan oleh karakteristik unik dari Kedua wilayah, seperti perbedaan ekonomi, social, atau lingkungan yang tidak tercatat dalam model. Sehingga perlu dilakukan analisis regresi robust untuk mengurangi pengaruh pencilan.

Membuat Model Regresi Robust

Dengan hasil uji normalitas yang menunjukkan bahwa residual tidak terdistribusi normal, terutama karena adanya pencilan dalam data maka dalam hal ini dilakukan penerapan model regresi robust. Model ini dapat mengurangi pengaruh pencilan dan memberikan estimasi parameter yang lebih stabil dan akurat meskipun data mengalami penyimpangan dari distribusi normal. Berikut adalah hasil analisis dengan regresi robust :

```
Call: rlm(formula = Y ~ PCA1 + PCA2 + X4, data = PENDIDIKAN, method = "MM")
Residuals:
    Min       1Q   Median       3Q      Max
-12.58419107  -0.77638247  -0.05846273   0.48603056   1.33660279

Coefficients:
            Value      Std. Error t value
(Intercept) 98.2951206    0.3758747 261.5103260
PCA1        -0.4120086    0.1767703  -2.3307575
PCA2        -0.2886534    0.1591607  -1.8135977
X4           0.0000022    0.0000064   0.3473624

Residual standard error: 0.9665702 on 30 degrees of freedom
```

Gambar 9. Model dengan Regresi Robust

Berdasarkan koefisien dan hasil analisis yang detail, model regresi robust dapat dirumuskan sebagai berikut:

$$\hat{Y} = \beta_0 + \beta_1 PCA1 + \beta_2 PCA2 + \beta_3 x_4$$

$$\hat{Y} = 98.2951 + (-0.4120) PCA1 + (-0.2886) PCA2 + 0.000002X_4$$

$$\hat{Y} = 97.7766 - 0.4120 PCA1 - 0.2886 PCA2 - 0.000002X_4$$

Residual Standard Error (RSE) adalah ukuran yang menunjukkan seberapa baik model regresi dapat memprediksi nilai variabel dependen. RSE mengukur rata-rata deviasi residual, yaitu selisih antara nilai yang diamati dan nilai yang diprediksi oleh model. Berikut adalah perbandingan antara nilai Residual Standar Error (RSE) untuk model OLS dan model regresi robust :

- Dari model OLS diperoleh nilai RSE sebesar 2.622703. RSE ini menunjukkan bahwa, secara rata-rata prediksi model OLS menyimpang dari nilai actual sekitar 2.62 unit. Nilai RSE yang lebih tinggi ini mengindikasikan bahwa model OLS mungkin dipengaruhi oleh pencilan yang signifikan, sehingga akurasi prediksinya berkurang.
- Sedangkan dari model Regersi Robust diperoleh nilai RSE sebesar 0.9665702. RSE ini menunjukkan bahwa secara rata-rata prediksi model regresi robust menyimpang dari nilai actual sekitar 0.97 unit. Nilai RSE yang lebih rendah ini menunjukkan bahwa modle robust mampu memberikan estimasi yang lebih akurat, terutama dalam mengatasi pengaruh pencilan yang adad alam data.

Interpretasi Koefisien

Dari hasil analisis regresi robust yang diperoleh dapat dilakukan interpretasi terhadap masing-masing koefisien regresi, sehingga dapat menjelaskan bagaimana setiap variabel independen mempengaruhi variabel dependen (angka melek huruf) :

- Nilai intercept sebesar 98.2951 yang berarti ketika semua variabel independen (X1, X2, X3, X4, dan X5) bernilai nol, angka melek huruf diprediksi sebesar 98.2951.
- Nilai PCA1 yaitu komponen utama yang dihasilkan dari kombinasi dua variabel yaitu : X1 (rata-rata lama sekolah) yang menunjukkan tingkat pendidikan atau lama belajar yang dicapai oleh populasi, dan X3 (angka anak tidak sekolah) yang menunjukkan jumlah anak yang tidak mendapatkan pendidikan formal. Nilai PCA1 sebesar -0.4120 menunjukkan bahwa ketika nilai PCA1 meningkat, maka angka melek huruf cenderung menurun. Setiap peningkatan 1 unit pada PCA1 akan mengurangi angka melek huruf sekitar 0.4120.
- Nilai PCA2 yaitu komponen utama yang dihasilkan dari kombinasi dua variabel yaitu : X2 (Jumlah guru) yang menunjukkan jumlah guru yang tersedia dalam suatu provinsi, dan X5 (jumlah penduduk miskin) yang menunjukkan jumlah penduduk yang hidup di bawah garis kemiskinan. Nilai PCA2 sebesar -0.2886 menunjukkan bahwa nilai PCA2 meningkat, maka angka melek huruf cenderung menurun. Setiap peningkatan 1 unit pada PCA2 akan mengurangi angka melek huruf sekitar 0.2886.

- Nilai X_4 (PDRB Per kapita) sebesar 0.000002 menunjukkan bahwa ada hubungan positif antara PDRB per kapita dan angka melek huruf. Ini berarti bahwa dengan setiap peningkatan satu unit dalam PDRB per kapita, angka melek huruf diprediksi meningkat sekitar 0.0000022. Hal ini menunjukkan bahwa provinsi dengan PDRB per kapita yang lebih tinggi cenderung memiliki angka melek huruf yang lebih baik.

Hubungan negatif antara PCA1 dan PCA2 dengan angka melek huruf menunjukkan bahwa hanya meningkatkan lama sekolah atau jumlah guru tidak cukup untuk meningkatkan melek huruf. Pendidikan yang efektif membutuhkan perhatian pada kualitas pendidikan, dukungan sosial, dan faktor ekonomi.

SIMPULAN

Dalam penelitian ini ditemukan bahwa model regresi robust lebih mewakili dengan eror yang lebih kecil dibandingkan dengan metode OLS. Menggunakan model Regresi Robust diperoleh nilai RSE sebesar 0.9665702. RSE ini menunjukkan bahwa secara rata-rata prediksi model regresi robust menyimpang dari nilai actual sekitar 0.97 unit. Nilai RSE yang lebih rendah ini menunjukkan bahwa model robust mampu memberikan estimasi yang lebih akurat, terutama dalam mengatasi pengaruh pencila yang ada dalam data. Sedangkan dengan metode OLS didapatkan nilai R-Square sebesar 0.3943417 berarti bahwa sekitar 39,43% variasi angka melek huruf dapat dijelaskan oleh model regresi ini. Sebaliknya, sekitar 60.66% variasi dalam data tidak dapat dijelaskan oleh model dan mungkin disebabkan oleh faktor lain atau variabel yang tidak termasuk dalam model.

Pada model robust ditemukan persamaan $\hat{Y} = 97.7766 - 0.4120 PCA1 - 0.2886 PCA2 - 0.000002 X_4$ artinya Setiap peningkatan 1 unit pada PCA1 (x_1 = rata-rata lama sekolah ; x_2 = angka anak tidak sekolah) akan mengurangi angka melek huruf sekitar 0.4120 atau 41,2% dan nilai PCA2 (x_3 = Jumlah guru ; x_5 = jumlah penduduk miskin) sebesar -0.2886 menunjukkan bahwa nilai PCA2 meningkat, maka angka melek huruf cenderung menurun. Setiap peningkatan 1 unit pada PCA2 akan mengurangi angka melek huruf sekitar 0.2886 atau 28,86%. Sedangkan untuk PDRB memiliki hubungan positif yang sangat kecil yaitu 0.000002, artinya setiap kenaikan 1 unit PDRB meningkatkan angka melek huruf 0.00022%. Banyaknya anak yang tidak bersekolah, anak putus sekolah, dan penduduk miskin. Meningkatkan angka buta huruf meningkat secara signifikan, upaya yang dapat dilakukan oleh pemerintah adalah memfasilitasi pendidikan sehingga semua golongan dapat mengakses pendidikan serta mengedukasi kepada masyarakat bahwa pendidikan itu sangat penting yang akan menjadi pondasi bangsa maupun negara di masa yang akan datang.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyelesaian penelitian ini. Secara khusus peneliti ingin menyampaikan penghargaan dan rasa terimakasih kepada Bapak Dr, Syairal Fahmy Dalimunthe, S.Sos., M.I.Kom. atas bimbingan, saran dan dukungan selama proses penelitian ini. Ucapan terimakasih juga tercurahkan kepada keluarga dan teman-teman atas dukungan dan motivasi serta arahan selama proses penulisan penelitian ini. Tanpa dukungan dari pihak-pihak tersebut, penelitian ini tidak akan diselesaikan dengan baik. Peneliti sangat menghargai segala bentuk kontribusi dan bantuan yang telah diberikan.

REFERENSI

- Ahadi, G. D., & Zain, N. N. L. E. (2023). Pemeriksaan Uji Kenormalan dengan Kolmogorov-Smirnov, Anderson-Darling dan Shapiro-Wilk. *Eigen Mathematics Journal*, 11-19.
- Anggraeni, D. P., & Putri, D. N. (2021). The Model of Coals Production in Indonesia. *EVOLUSI: JOURNAL OF MATHEMATICS AND SCIENCES*, 5(2), 58-69.
- Azizah, I. N., Arum, P. R., & Wasono, R. (2021). Model terbaik uji multikolinearitas untuk analisis faktor-faktor yang mempengaruhi produksi padi di Kabupaten Blora tahun 2020. In *Prosiding Seminar Nasional UNIMUS* (Vol. 4).
- Lestari, A. Y., Kurniawan, F., & Ardi, R. B. (2020). Penyebab Tingginya Angka Anak Putus Sekolah Jenjang Sekolah Dasar (SD). *Jurnal Ilmiah Sekolah Dasar*, 4(1) 299-308.

- Manurung, E. N., & Hutabarat, F. (2021). Pengaruh Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Pengeluaran per Kapita Terhadap Indeks Pembangunan Manusia. *Jurnal Ilmiah Akuntansi Manajemen*, 4(2), 121-129.
- Nelvidawati, N., & Kasman, M. (2023). Penggunaan Korelasi Spearman untuk Menguji Hubungan Suhu dan Besarnya Curah Hujan Bulanan di Kota Padang. *Jurnal Daur Lingkungan*, 6(1), 34-39.
- Pasande, P., Katelu, M., & Tari, E. (2020). PERAN GURU DALAM MENGATASI BUTA HURUF DI SUKU TAA WANA DESA TARONGGO. *EDUMASPUL: Jurnal Pendidikan*, 4(2),
- Putra, B., & Muhammad, A. H. (2024). Prediksi Prevalensi Stunting di Indonesia dengan Ordinary Least Square (OLS). *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1890-1900.