

Analysis of Factors Influencing Literacy and Reading Activity Index Using Regression Models with Variable Selection

Analisis Faktor yang Memengaruhi Indeks Aktivitas Literasi dan Membaca Menggunakan Regresi Berganda dengan Seleksi Peubah

Lutfi S Lubis^{1‡}, Reynd H Pasaribu¹, Fedora Ilahi¹, Febri D Cahyanto¹, Reyza S Dewi¹, Yenni Anggraini¹, and Muhlis Ardiansyah¹

¹Department of Statistics, IPB University, Indonesia

[‡]corresponding author: lutfisyahrezalubis@apps.ipb.ac.id

Abstract

The Indonesian Literacy and Reading Activity Index (Alibaca) in 2019 is classified in the low category. This category contradicts the fact that Indonesia has the second-largest number of libraries in the world. This study aims to select factors that significantly influence literacy and reading activities in Indonesia. The method used is multiple linear regression through variable selection, assumption testing, and handling assumption violations. The data for this study were obtained from publications by the Center for Research on Education and Culture Policy, the Agency for Research and Development, and the Ministry of Education and Culture in 2019. The data comprises one response variable and 16 explanatory variables from 34 provinces in Indonesia. The results of the study show that there are only 3 out of 16 variables that significantly influenced The Literacy and Reading Activity Index in Indonesia in 2019, namely: (1) the percentage of households buying newspapers/newspapers; (2) the percentage of the population using computers; and the percentage of people visiting the library. These three factors can explain the diversity of the Alibaca index by 95.8 percent.

Keywords: Literacy index, multiple regression, variable selection

Abstrak

Indeks Aktivitas Literasi dan Membaca (Alibaca) di Indonesia berada dalam kategori rendah. Hal ini berkontradiksi dengan fakta bahwa Indonesia sebagai negara dengan jumlah perpustakaan terbanyak kedua di dunia. Penelitian ini bertujuan memilih faktor-faktor yang secara signifikan memengaruhi aktivitas literasi dan membaca di Indonesia. Metode yang digunakan adalah regresi linier berganda melalui seleksi peubah, pengujian asumsi, dan penanganan pelanggaran asumsi. Data pada penelitian ini diperoleh dari publikasi Pusat Penelitian Kebijakan Pendidikan dan Kebudayaan, Badan Penelitian dan Pengembangan, serta Kementerian Pendidikan dan Kebudayaan pada tahun 2019. Data tersusun atas satu peubah respon dan 16 peubah penjelas yang berasal dari 34 provinsi di Indonesia. Hasil penelitian menunjukkan bahwa hanya terdapat 3 dari 16 peubah yang signifikan memengaruhi Indeks Alibaca di Indonesia tahun 2019, yaitu: (1) persentase rumah tangga membeli surat kabar/koran; (2) persentase penduduk menggunakan komputer; dan (3) persentase penduduk mengunjungi perpustakaan. Ketiga faktor tersebut mampu menjelaskan keragaman indeks Alibaca sebesar sebesar 95.8 persen.

Kata Kunci: Indeks literasi, regresi berganda, seleksi peubah

1. Pendahuluan

1.1 Latar Belakang

Kondisi literasi masyarakat Indonesia kian memprihatinkan. Unesco menyatakan hanya satu dari seribu orang masyarakat Indonesia yang memiliki minat membaca (Jatmiko *et al.*, 2021). Pemerintah melalui Gerakan Literasi Nasional (GLN) yang dicanangkan oleh Kemendikbud sejak tahun 2016 berupaya merangsang minat literasi di lingkungan sekolah dan masyarakat. Namun, upaya pemetaan literasi ini belum memadai karena disebabkan oleh berbagai kendala (Puslitjakdikbud, 2019). Secara umum, ada tiga parameter yang menunjukkan permasalahan yang melatarbelakangi kendala literasi masyarakat Indonesia: [1] akses di sekolah yang minim, [2] akses di masyarakat yang minim, dan [3] perilaku membaca yang rendah. Parameter-parameter tersebut hanya beberapa dari banyaknya faktor yang memengaruhi aktivitas membaca masyarakat. Oleh karena itu, penelitian ini bertujuan mengetahui faktor-faktor yang secara signifikan menggambarkan indeks Alibaca di Indonesia.

Pemilihan peubah didasarkan oleh faktor-faktor yang secara umum berpengaruh terhadap indeks Alibaca. Faktor-faktor tersebut meliputi faktor kecakapan, akses, alternatif, dan faktor kebudayaan (McKenna, 2016). Perihal indeks Alibaca menjadi tanggungjawab semua pihak. Pemerintah menyediakan akses kepada masyarakat sebanyak 160 ribu perpustakaan beserta 22 juta koleksi buku yang menyebar di penjuru Indonesia (Puspresnas, 2021 dan Duniaperpustakaan.com, 2023). Di sisi lain, calon guru perlu memenuhi standar atau kualifikasi dalam membantu memotivasi masyarakat untuk membaca di lingkungan sekolah (Ningsih dan Alpusari, 2019). Terlebih, pemerintah dan pihak swasta berusaha menyediakan pendidikan kepada masyarakat hingga tingkat Perguruan Tinggi yang hingga saat ini terdapat lebih dari tiga ribu universitas di seluruh Indonesia (BPS, 2022).

Pengukuran faktor-faktor yang secara signifikan memengaruhi indeks Alibaca menerapkan prinsip analisis regresi dengan seleksi peubah. Kekuatan hubungan indeks Alibaca dengan keempat faktor yang memengaruhinya dilihat dengan mengukur hubungan antara variabel respon dan variabel bebas. Model regresi dapat digunakan untuk menduga pengaruh suatu variabel atau beberapa variabel bebas dengan variabel responnya (Iriawan dan Astuti, 2006). Analisis regresi yang

digunakan untuk melihat hubungan indeks Alibaca dengan berbagai faktor internal dan eksternalnya sehingga terbentuk model regresi terbaik menggunakan analisis regresi linear berganda.

1.2 Tujuan Penelitian

- a. Mengetahui faktor-faktor apa saja yang secara signifikan berpengaruh terhadap Indeks Aktivitas Literasi dan Membaca (Alibaca) di Indonesia.
- b. Menentukan model linear terbaik dari peubah yang berpengaruh secara signifikan.

2. Tinjauan Pustaka

2.1 Analisis Regresi

Analisis regresi merupakan salah satu metode pengolahan data yang menggunakan perhitungan statistika dengan memprediksi pengaruh suatu variabel bebas terhadap variabel lain (variabel tak bebas/ terikat). Variabel terikat dinotasikan sebagai Y dan variabel bebasnya dinotasikan dengan X_1, X_2, \dots, X_p , huruf p pada notasi tersebut melambangkan nomor dari variabel-variabel bebas yang ada.

Hubungan antara Y dengan X_1, X_2, \dots, X_p bisa didekati dengan pemodelan $Y = f(X_1, X_2, \dots, X_p) + \varepsilon$, ε diasumsikan sebagai galat acak yang menunjukkan perbedaan pada nilai taksiran. Fungsi $f(X_1, X_2, \dots, X_p)$ menunjukkan hubungan antara Y dan X_1, X_2, \dots, X_p yang dapat dimodelkan secara regresi linier, yakni $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$, dengan $\beta_0, \beta_1, \dots, \beta_p$ yang menunjukkan parameter regresi atau koefisien yang diketahui nilainya dari data yang digunakan (Chatterjee dan Hadi, 2012).

2.2 Indeks Aktivitas Literasi dan Membaca (Alibaca)

Literasi diartikan sebagai melek huruf, kemampuan baca tulis, kemelekwancahan atau kecakapan dalam membaca dan menulis (Padamadewi dan Artini, 2018). Secara nasional, Indeks Aktivitas Literasi Membaca (Alibaca) pada 34 provinsi dikategorikan dalam “aktivitas literasi rendah” (Puslitjakdikbud, 2019). Indeks Alibaca tidak semata dipengaruhi oleh kemampuan internal oleh masyarakat (Ningsih dan Alpusari, 2019). Minat, bakat, kecerdasan, budaya, dan lainnya turut berperan aktif dalam merespon indeks Alibaca.

2.3 Uji Asumsi

Setelah memilih model, langkah berikutnya adalah mendeteksi asumsi Gauss-Markov untuk mengidentifikasi masalah yang mungkin ada dalam data dan model penelitian. Jika asumsi klasik dipenuhi, maka secara teoritis penggunaan analisis regresi linier berganda akan menghasilkan estimasi parameter yang sah (Ningsih dan Dukalang, 2019).

2.3.1 Uji Normalitas

Uji Normalitas dilakukan untuk memeriksa apakah variabel terikat dan variabel bebas dalam model regresi memiliki distribusi normal atau tidak, serta apakah nilai residualnya memiliki distribusi normal atau tidak. Dalam penelitian ini, digunakan uji normalitas *Kolmogorov Smirnov*. Residual data berdistribusi normal ketika nilai signifikan > 0.05 .

2.3.2 Uji Multikolinearitas

Uji Multikolinearitas digunakan untuk memeriksa adanya korelasi antara variabel bebas dalam model regresi. Model regresi yang baik seharusnya tidak memiliki korelasi antara variabel bebas. Untuk mendeteksi adanya Multikolinearitas dalam regresi, dapat dilihat dari nilai *Variance Inflation Factor* (VIF) dan nilai *Tolerance*. Multikolinearitas terjadi jika nilai VIF lebih dari 10.

2.3.3 Uji Heteroskedastisitas

Uji heteroskedastisitas digunakan untuk menentukan ketidaksamaan varian residual antara pengamatan dalam model regresi. Jika residual antara pengamatan memiliki varian yang sama, maka model tersebut disebut homoskedastis, sedangkan jika memiliki varian yang berbeda-beda, maka disebut heteroskedastis. Model regresi yang baik adalah yang tidak memiliki heteroskedastisitas. Pendeteksian heteroskedastisitas dapat menggunakan uji *Glejser*. Heteroskedastisitas dideteksi ketika nilai signifikansi < 0.05 .

2.3.4 Uji Kebebasan Sisaan

Uji kebebasan sisaan berguna untuk memeriksa apakah terdapat ketidakbebasan pada sisaan model regresi. Uji ini dilakukan dengan menggunakan *run test*, dan apabila hasil signifikansi $> 0,05$ disimpulkan bahwa sisaan model tersebut saling bebas.

2.4 Pendeteksian Data Pencilan, Leverage, dan Amatan Berpengaruh

2.4.1 Pencilan

Data yang memiliki nilai sangat jauh dari pusat data dan tidak mengikuti sebaran dari sebagian besar dari pola disebut sebagai data pencilan atau *outlier* (Sari et al. 2018). Pencilan dapat memengaruhi nilai-nilai statistik lainnya dan

mengakibatkan kekeliruan dalam penentuan model yang tepat serta efisiensinya.

2.4.2 Leverage

Leverage terjadi saat nilai suatu amatan yang jauh berbeda, namun masih mengikuti pola amatan lainnya. Oleh karena itu, data leverage harus tetap dipertimbangkan karena dapat berpengaruh terhadap *R-Square*.

2.4.3 Amatan Berpengaruh

Amatan berpengaruh adalah amatan yang mampu menarik model regresi yang ada menuju amatan itu dan penghapusannya akan mengubah regresi yang dipasang sehingga akan berpengaruh terhadap koefisien regresi yang didapat (Prayogo dan Sukim 2020). Dalam menganalisis amatan berpengaruh pada suatu model regresi, metode yang sering digunakan adalah pengukuran nilai *DFBETAS*, *DFFITS*, dan *COVRATIO* atau *Cook's Distance*.

2.5 Pendugaan Model Terbaik

2.5.1 Metode Forward

Metode *forward* adalah metode pendugaan model terbaik di mana peubah bebas dimasukkan satu per satu diurutkan berdasarkan nilai mutlak koefisien korelasi yang terbesar terhadap peubah penjelas, dan berhenti bila semua syarat telah terpenuhi (Samsosir *et al.* 2014). Pada metode ini, model awalnya adalah model tanpa peubah penjelas.

2.5.2 Metode Backward

Metode *backward* adalah metode pendugaan model terbaik di mana peubah dieliminasi satu per satu berdasarkan nilai mutlak koefisien korelasi yang terkecil terhadap peubah penjelas, dan berhenti bila semua syarat telah terpenuhi (Samsosir *et al.* 2014). Pada metode ini, model awalnya adalah model dengan semua peubah penjelas.

2.5.3 Metode Stepwise

Metode *stepwise* adalah metode pemilihan peubah penjelas dengan langkah memasukkan satu per satu peubah bebas berdasarkan koefisien korelasi parsial yang terbesar terhadap peubah respons (Pujilestari *et al.* 2014). Pada metode ini, digunakan Akaike Information Criterion (AIC) untuk mengetahui persamaan mana yang terbaik.

2.5.4 Adjusted R-Square

Model terbaik adalah model yang memiliki nilai *adjusted r-square* yang paling tinggi. Nilai kuadrat tengah sisaan yang kecil akan menaikkan nilai *adjusted r-squared*. Nilai dari *adjusted R-square* dinilai lebih efektif daripada nilai *R-square* pada model dengan lebih dari satu peubah karena kecenderungan *R-square* yang akan terus menaik jika peubah bertambah, sehingga hasilnya akan berbias. *Adjusted R-square* hadir untuk mengatasi hal tersebut.

2.5.5 CP Mallow

CP Mallow digunakan untuk menilai kecocokan model regresi. Kita dapat mengidentifikasi model regresi “terbaik” dengan mengidentifikasi model dengan nilai CP terendah yang mendekati $k+1$, di mana k adalah jumlah

peubah dalam model. Jika dalam model tersedia nilai CP yang rendah, pilihlah nilai yang paling kecil di antaranya.

3. Data dan Metodologi

3.1 Data

Data pada penelitian ini diperoleh dari publikasi Pusat Penelitian Kebijakan Pendidikan dan Kebudayaan, Badan Penelitian dan Pengembangan, serta Kementerian Pendidikan dan Kebudayaan pada tahun 2019. Data tersusun atas satu peubah respon dan 16 peubah penjelas yang berasal dari 34 provinsi di Indonesia. Peubah-peubah yang dipilih pada penelitian ini berdasarkan penelitian terdahulu yang telah dilakukan. Peubah-peubah yang dipilih dijelaskan pada tabel di bawah ini.

Tabel 1: Daftar peubah yang digunakan dalam model regresi

Kode	Peubah	Satuan
Y	Indeks Alibaca	Poin
X1	Melek Huruf Latin	Persen
X2	Rata-rata Lama Sekolah	Tahun
X3	Jumlah Perpustakaan Kondisi Baik	Persen
X4	Jumlah Petugas Pengelola Perpustakaan Sekolah	Persen
X5	Jumlah Perpustakaan Umum	Persen
X6	Jumlah TBM dan Pustaka Bergerak	Poin
X7	Persentase Rumah Tangga Membeli Surat kabar/koran	Persen
X8	Persentase Rumah Tangga Membeli Majalah/Tabloid	Persen
X9	Sekolah Memiliki Akses Internet	Persen
X10	Persentase penduduk mengakses internet	Persen
X11	Persentase penduduk menggunakan komputer	Persen
X12	Persentase penduduk membaca surat kabar/ koran atau majalah cetak	Persen
X13	Persentase penduduk membaca buku cetak kitab suci	Persen
X14	Persentase penduduk membaca berita/artikel dari media elektronik, internet	Persen
X15	Persentase penduduk mengunjungi perpustakaan	Persen
X16	Persentase penduduk memanfaatkan taman baca	Persen

3.2 Metode Penelitian

Penelitian ini dilaksanakan dengan prosedur sebagai berikut (terlampir).

- a. Eksplorasi data

Eksplorasi pada data indeks Alibaca dilakukan menggunakan Histogram, Boxplot, dan Statistik lima serangkai. Selain itu, eksplorasi juga dilakukan dengan matriks korelasi untuk melihat korelasi antar peubah.

b. Pendugaan Model Regresi Linear Awal

Pada tahap ini ingin diketahui model regresi linear dari model awal tanpa adanya perlakuan.

c. Pendeteksian Multikolinearitas

Pengujian multikolinearitas menggunakan batas nilai VIF. Peubah yang melewati batas VIF akan direduksi secara bertahap, kemudian dilihat kembali nilai VIF-nya.

d. Pengujian Asumsi

Pengujian asumsi menerapkan kondisi Gauss-Markov dan uji normalitas.

e. Pendeteksian Data Pencilan, Leverage, dan Amatan Berpengaruh

f. Pendugaan Model Terbaik

g. Menentukan Peubah Penjelas

h. Menentukan Model Terbaik

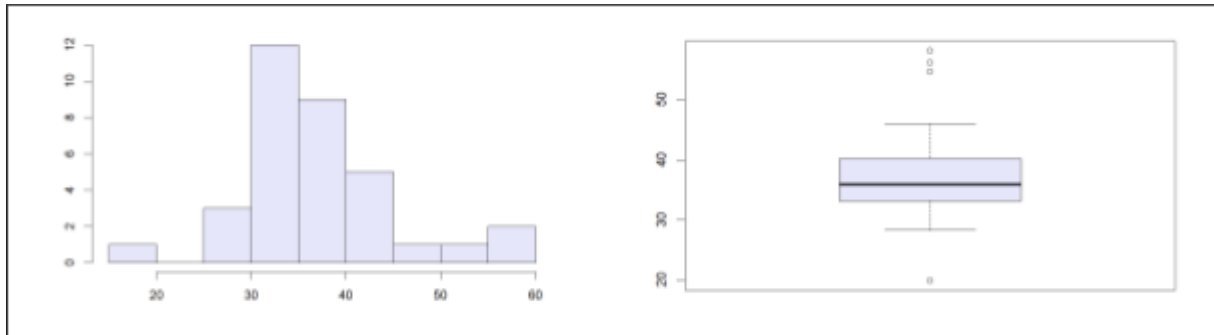
4. Hasil dan Pembahasan

4.1 Eksplorasi Data

Tabel 2: Deskriptif statistik Indeks Alibaca Indonesia tahun 2019

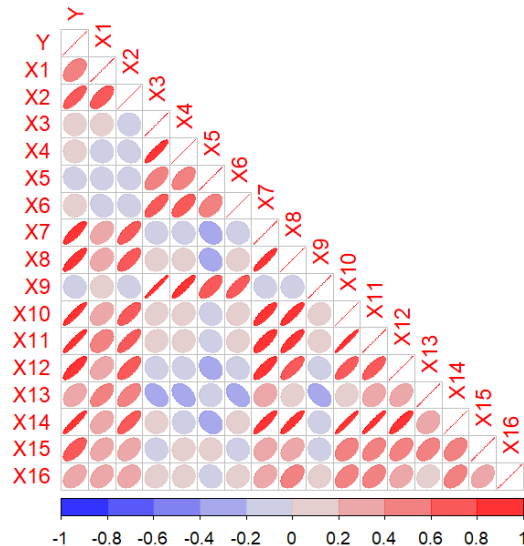
Minimum	19.90
Maksimum	58.16
Kuartil Pertama	33.22
Median	35.90
Rata-Rata	37.32
Kuartil Ketiga	40.02
Ragam	62.35
Total Provinsi	34

Tabel di atas menunjukkan deskripsi Indeks Alibaca di Indonesia pada tahun 2019. Indeks Alibaca pada seluruh provinsi di Indonesia merentang dari 19.90 hingga maksimum 58.16 poin. Rata-rata Indeks Alibaca sendiri berada pada nilai 37.32 poin yang hampir sama dengan mediannya, yakni 35.90 poin. Angka ini menunjukkan bahwa aktivitas membaca di Indonesia tergolong “rendah” (Puslitjakdikbud, 2019).



Gambar 1: Histogram Indeks Alibaca di Indonesia tahun 2019; Gambar 2: Box plot Indeks Alibaca di Indonesia tahun 2019

Pada Gambar 2 ditunjukkan bahwa Indeks Alibaca di Indonesia pada tahun 2019 menyebar secara merata di rentang 20 hingga 45 poin. Histogram tersebut juga menunjukkan bahwa terdapat satu provinsi yang memiliki angka indeks kurang dari 20 dan terdapat beberapa provinsi dengan angka indeks yang lebih dari angka 45. Gambar 3 mendeteksi bahwa provinsi-provinsi tersebut yang berada di luar rentang 20 sampai 45 poin merupakan pencilan. Provinsi tersebut adalah Provinsi Papua, Kepulauan Riau, DI Yogyakarta, dan DKI Jakarta dengan angka Indeks Alibaca 19.9, 54.76, 56.2, dan 58,16 secara berurutan. Oleh karena itu, Provinsi Papua dinyatakan sebagai provinsi dengan kategori Indeks Alibaca “sangat rendah” serta Provinsi Kepulauan Riau, DI Yogyakarta, dan DKI Jakarta berada dalam “kategori sedang” di Indonesia pada tahun 2019.



Gambar 3: Matriks Korelasi Antar Peubah

Matriks korelasi di atas menjelaskan hubungan antar peubah yang digunakan dalam penelitian. Semakin merah hasil visualisasi dari hubungannya menandakan korelasi yang tinggi atau mendekati satu. Selain itu, warna yang semakin biru menjelaskan hubungan yang tinggi juga namun bernilai negatif atau mendekati nilai negatif satu. Visualisasi yang digambarkan dengan warna yang kurang jelas juga menggambarkan hubungan yang rendah di antara peubahnya. Hasil dari matriks korelasi di atas menunjukkan peubah bebas X1, X2, X7, X8, X10, X11, X12, X14,

dan X15 berkorelasi tinggi terhadap peubah respon Y. Di sisi lain, peubah bebas X13 dan X16 memiliki korelasi terhadap peubah respon Y, namun hubungannya tergolong sedang. Sedangkan peubah bebas X3, X4, X5, X6, dan X9 menunjukkan korelasi yang rendah terhadap peubah respon Y.

4.2 Pendugaan Model Regresi Awal

Model regresi dugaan awal yang didapatkan sebagai berikut.

Tabel 3: Daftar koefisien model regresi awal

Kode	Koefisien	Kode	Koefisien
X1	0.2091	X9	-0.0003
X2	-0.5296	X10	0.2104
X3	-0.0002	X11	0.0804
X4	0.0009	X12	-0.0531
X5	0.0016	X13	0.0405
X6	0.0034	X14	0.4044
X7	0.2348	X15	0.7422
X8	0.2021	X16	-2.4973

4.3 Pendeteksian Multikolinearitas

Tabel 4: Nilai VIF seluruh peubah

Kode	VIF	Kode	VIF
X1	4.014	X9	43.304
X2	8.024	X10	54.949
X3	30.182	X11	41.763
X4	8.023	X12	6.586
X5	3.984	X13	5.591
X6	3.371	X14	33.396
X7	11.934	X15	2.182
X8	9.446	X16	1.965

Tabel di atas menunjukkan bahwa terdapat multikolinearitas pada beberapa peubah. Multikolinearitas diindikasikan berdasarkan nilai VIF yang dihasilkan oleh masing-masing peubah. Peubah-peubah tersebut adalah X3, X7, X9, X10, X11, dan X14 yang memiliki nilai VIF lebih dari sepuluh. Selanjutnya, dilakukan penghapusan peubah terdeteksi multikolinearitas dengan berbagai ketentuan. Ketentuan pertama,

menghapus peubah-peubah tersebut satu per satu dan dilihat nilai VIF-nya. Ketentuan kedua, menghilangkan peubah-peubah tersebut berdasarkan nilai VIF tertinggi hingga terendah. Ketiga, melakukan penghapusan pada beberapa peubah secara simultan. Pada tahap eksplorasi ini, didapatkan model yang lebih baik, yakni dengan menghilangkan peubah X9, X10, dan X14 secara bersamaan. Model ini memberikan nilai VIF kurang dari sepuluh untuk semua peubah.

Model regresi linear dugaan setelah peubah direduksi:

$$\hat{y} = -6.3725 + 0.2367X_1 - 0.4282X_2 - 0.0008X_3 + 0.0004X_4 + 0.0015X_5 + \\ + 0.0036X_6 + 0.3432X_7 + 0.1854X_8 + 0.8378X_{11} + 0.0952X_{12} - 0.1412X_{13} \\ + 0.7185X_{15} - 1.7491X_{16}$$

4.4 Pengujian Asumsi

Pengujian asumsi dilakukan pada model regresi terbaru yang memiliki tiga belas peubah. Pada tabel di bawah dapat terlihat bahwa dilakukan empat macam pengujian asumsi pada model regresi terbaru. Uji asumsi yang dilakukan terdiri dari pengujian normalitas dan Gauss-Markov. Uji normalitas dilakukan dengan dua macam uji formal, *Kolmogorov-Smirnov* dan *Shapiro-Wilk*. Berdasarkan hasil uji pada taraf nyata 5%, didapatkan nilai masing-masing *p-value* sebesar 0.7288 dan 0.6481. Kedua nilai tersebut lebih besar dari $\alpha = 0,05$ maka dapat dikatakan terima H_0 sehingga pada model regresi galat berdistribusi normal.

Pengujian asumsi Gauss-Markov dilakukan pada taraf nyata 5%. Berdasarkan hasil uji *runs*, didapatkan nilai *p-value* sebesar 1. Nilai tersebut lebih besar dari $\alpha = 0,05$ maka dapat dikatakan terima H_0 sehingga cukup bukti untuk menyatakan bahwa sisaan saling bebas. Uji asumsi ketiga yang digunakan adalah uji homoskedastisitas. Berdasarkan hasil uji *Breusch-Pagan* pada taraf nyata 5%, didapatkan nilai *p-value* sebesar 0.5176. Nilai tersebut lebih besar dari $\alpha = 0,05$ maka dapat dikatakan terima H_0 sehingga pada model regresi tidak terjadi homoskedastisitas. Uji asumsi keempat pada kondisi Gauss-Markov adalah asumsi nilai harapan sisaan sama dengan nol. Berdasarkan hasil uji *t* pada taraf nyata 5%, didapatkan nilai *p-value* sebesar 1. Nilai tersebut lebih besar dari $\alpha = 0,05$ maka dapat dikatakan terima H_0 sehingga cukup bukti untuk menyatakan bahwa nilai harapan galat sama dengan nol.

Tabel 5: Hasil uji asumsi sisaan

Asumsi Sisaan	Uji Formal	P-Value
Normalitas	<i>Kolmogorov-Smirnov</i>	0.7288
	<i>Shapiro-Wilk</i>	0.6481
Kebebasan Sisaan	<i>Runs test</i>	1
Homoskedastisitas	<i>Breush-Pagan</i>	0.5176
$E[\epsilon_i] = 0$	<i>t-test</i>	1

4.5 Pendeteksian Data Pencilan, Leverage, dan Amatan Berpengaruh

4.5.1 Pendeteksian data pencilan

Tabel 6: Nilai ri

Pengamatan	ri
3	2.0639
20	-2.2126
22	2.1823

Apabila nilai mutlak ri melebihi 2 atau 3, maka data tersebut dapat dianggap sebagai outlier. Berdasarkan tabel tersebut didapatkan dua pencilan yaitu pada amatan ke-3, 20, dan 22.

4.5.2 Pendeteksian data leverage

Berikut ini merupakan rumus yang digunakan untuk mengidentifikasi adanya data leverage:

$$H = X(X'X)^{-1}X'$$

Jika nilai $h_{ii} > 2p/n$, amatan tersebut dapat dikatakan sebagai titik *leverage*. Dengan $n = 34$; $p = 14$, maka $2p/n = 0.8235294$.

h_{ii} : unsur diagonal dari matriks H

p : banyaknya parameter

n : banyaknya amatan

Berdasarkan formula di atas didapatkan nilai h_{ii} pada amatan ke-24 (0.8435887) $> 2p/n$, sehingga amatan 24 termasuk data *leverage*.

4.5.3 Pendeteksian amatan berpengaruh

Dengan $n = 34$; $p = 14$, maka; $F_{(14,20,0.95)} = 2.225$. Jika suatu amatan terdeteksi sebagai amatan berpengaruh jika nilai $D_i > F_{(p,n-p,1-\alpha)}$. Berdasarkan nilai D_i tidak ditemukan amatan yang berpengaruh atau nilai D_i lebih dari 2.24.

4.6 Menduga Model Terbaik

Tabel 7: Nilai *adjusted R-Square* tanpa amatan

Model Regresi	<i>Adjusted R-Square</i>
Tanpa amatan 24	0.9403
Tanpa amatan 3	0.9505

Tanpa amatan 20	0.9524
Tanpa amatan 22	0.9520
Tanpa amatan 3, 24	0.9541
Tanpa amatan 24, 20	0.9497
Tanpa amatan 24, 22	0.9462
Tanpa amatan 3, 20	0.9585
Tanpa amatan 3, 22	0.9647
Tanpa amatan 20, 22	0.9666
Tanpa amatan 24, 3, 20	0.9583
Tanpa amatan 24, 3, 22	0.9611
Tanpa amatan 3, 20, 22	0.9734
Tanpa amatan 24, 20, 22	0.9698
Tanpa amatan 24, 3, 20, 22	0.9684

Berdasarkan Tabel didapatkan model regresi dengan nilai Adjusted R-squared tertinggi adalah model tanpa amatan ke 3, 20, dan 22 sehingga model terbaik yang akan digunakan adalah model dengan data tanpa pencilan.

4.7 Memilih Model Terbaik

Dalam menentukan model terbaik, dilakukan dua analisa hubungan, yaitu hubungan banyak peubah terhadap nilai R-square dan hubungan banyak peubah terhadap nilai CP Mallow. Berdasarkan analisis tersebut, diketahui bahwa model terbaik yang terpilih adalah model dengan 3 peubah. Hal ini dikarenakan model dengan 3 peubah memiliki nilai R-Square tertinggi dan nilai CP Mallow terendah.

Pada tahap sebelumnya, telah ditentukan bahwa peubah X_7 , X_{11} , dan X_{15} merupakan peubah untuk model terbaik. Hal itu telah dibuktikan dengan metode *Forward*, *Backward*, dan *Stepwise*. Maka model terbaik yang didapat adalah :

$$Y = \beta_0 + \beta_1 X_7 + \beta_2 X_{11} + \beta_3 X_{15}$$

$$Y = 9.5690 + 0.4568X_7 + 0.7772X_{11} + 0.7632X_{15}$$

Dengan X_7 adalah **persentase rumah tangga membeli surat kabar**, X_{11} adalah **persentase penduduk menggunakan komputer**, dan X_{15} adalah **persentase penduduk mengunjungi perpustakaan**. Melalui model tersebut, didapat nilai R-square yang sudah cukup besar, yaitu 0.958

4.8 Pengujian Asumsi Sisaan Kembali Terhadap Model Terbaik

Setelah ditetapkan model regresi terbaik, uji asumsi normalitas dan Gauss-Markov kembali dilakukan pada taraf nyata 5%. Berdasarkan tabel di bawah, nilai p -value pada masing-masing uji asumsi lebih besar dari $\alpha = 0,05$, maka dapat dikatakan terima H_0 . Oleh karena itu, uji asumsi normalitas dan *Gauss-Markov* terpenuhi.

Tabel 8: Hasil uji asumsi sisaan model terbaik

Asumsi Sisaan	Uji Formal	P-Value
Normalitas	<i>Kolmogorov-Smirnov</i>	0.6431
	<i>Shapiro-Wilk</i>	0.3480
Kebebasan Sisaan	<i>Runs test</i>	1
Homoskedastisitas	<i>Breush-Pagan</i>	0.9816
$E[\varepsilon_i] = 0$	<i>t-test</i>	1

4.9 Interpretasi Model Terbaik

Berdasarkan hasil analisis pada model regresi terbaik, didapat uji-T dan uji-F sebagai berikut:

Tabel 9: Pengujian model terbaik

Kode	Peubah	Koefisien	Sig. t
X7	Persentase rumah tangga membeli surat kabar/koran	0.4568	5.890
X11	Persentase penduduk menggunakan komputer	0.7772	6.754
X15	Persentase penduduk mengunjungi perpustakaan	0.7632	6.901

$$R^2 = 0.9622; R_{adj}^2 = 0.9580; RSE = 1.68$$

Hipotesis uji-t:

$$H_0: \beta_1 = 0 \text{ (Peubah tak berpengaruh)}$$

$$H_1: \beta_1 \neq 0 \text{ (intersep berpengaruh)}$$

Berdasarkan tabel di atas, setiap peubah akan signifikan pada taraf nyata 5%. Oleh karena itu, model terbaik yang didapat adalah :

$$Y = 9.5690 + 0.4568X_7 + 0.7772X_{11} + 0.7632X_{15} + e$$

Dengan keragaman yang dapat dijelaskan oleh model (R^2) sebesar 96.22% yang dapat dikatakan hampir sempurna. Hal ini menandakan bahwa model tersebut sudah sangat baik untuk menggambarkan peubah-peubah yang mempengaruhi Indeks Aktivitas Literasi Membaca (Alibaca) di Indonesia.

Secara umum, terdapat tiga faktor yang paling memengaruhi Indeks Alibaca di

Indonesia pada tahun 2019. Faktor-faktor tersebut meliputi persentase rumah tangga membeli surat kabar/ koran, persentase penduduk menggunakan komputer, dan persentase penduduk mengunjungi perpustakaan. Berdasarkan ketiga faktor tersebut, dua dari tiga faktor berdimensikan akses atau berkaitan dengan bagaimana masyarakat mendapatkan sumber daya literasi. Faktor yang merepresentasikan persentase rumah tangga membeli surat kabar/ koran menggambarkan bagaimana sumber daya literasi didapatkan di lingkungan masyarakat. Semakin tinggi tingkat rumah tangga yang menyediakan surat kabar/ koran mengimplikasikan tingginya akses bahan bacaan cetak bagi rumah tangga tersebut. Hal ini sesuai dengan penelitian Saputri *et al.* (2017) bahwa pengenalan huruf kepada anak usia prasekolah bersifat penting.

Selain akses terhadap surat kabar, persentase penduduk yang mengunjungi perpustakaan secara signifikan juga memengaruhi indeks Alibaca. Indonesia sebagai negara ke dua dengan jumlah perpustakaan terbanyak di dunia seharusnya mampu menyokong kebutuhan masyarakat terhadap akses bahan bacaan. Namun, hal ini menjadi kontradiksi sebagaimana indeks Alibaca di Indonesia yang masih tergolong rendah. Hasil ini menandakan bahwa minat masyarakat untuk mengunjungi perpustakaan juga masih rendah.

Faktor terakhir yang secara signifikan mampu mendorong tingkat indeks Alibaca di Indonesia adalah persentase penduduk menggunakan komputer. Faktor ini berkaitan dengan alternatif yang dimiliki masyarakat Indonesia dalam mengakses informasi menggunakan perangkat elektronik atau digital. Persentase ini hanya mencakup penggunaan terhadap komputer, laptop, dan tablet. Hal ini sangat menguntungkan bagi masyarakat, terutama golongan pelajar yang sering berinteraksi dengan komputer yang tanpa sadar mampu meningkatkan kemampuan literasi serta aktivitas membaca. Namun, sangat disayangkan bahwa faktor mengakses internet yang juga merupakan faktor alternatif tidak mampu untuk mendorong minat dan aktivitas membaca masyarakat. Terlebih hampir seluruh masyarakat Indonesia menghabiskan waktunya untuk mengakses internet.

5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa Indeks Aktivitas Literasi Membaca (Alibaca) dapat dipengaruhi oleh persentase rumah tangga membeli surat kabar/koran, persentase penduduk menggunakan komputer, dan persentase penduduk mengunjungi perpustakaan. Ketiga faktor tersebut hanya mencakup dua dari empat dimensi yang memengaruhi indeks Alibaca, yaitu dimensi akses dan alternatif. Sementara dua faktor lainnya, dimensi kecakapan dan budaya, belum mampu mendorong minat dan aktivitas membaca masyarakat Indonesia.

DAFTAR PUSTAKA

- Chatterjee S., Hadi A. S. 2012. Regression Analysis by Example Fifth Edition. New York: WILEY.
- Iriawan N., Astuti S. P. 2006. Mengolah Data Statistik dengan menggunakan Minitab 14. Yogyakarta: Andi.
- Jatmiko, E. S., Aini, Q., & Puspitasari, D. (2021). Penataan Taman Baca Grio Mukti Untuk Meningkatkan Literasi Membaca di Desa Pringombo Kcamatan Tempuran. *Khidmatan*, 1(2), 141-149.
- Kurnia, A., Kusumaningrum, D., Soleh, A. M., Handayani, D., & Anisa, R. (2015). Small area estimation with winsorization method for poverty alleviation at a sub-district level. *International Journal of Applied Mathematics and Statistics™*, 53(6): 77–84.
- Miller, John W. & Micahel M. McKenna. (2016). World Literacy: How Countries Rank and Why It Matters. New York: Routledge.
- Ningsih, H.S., & Alpusari, M. (2019). Identifikasi Faktor-Faktor yang Mempengaruhi Literasi Siswa Sekolah Dasar Kelas Rendah. Prossiding Seminar Nasional Pendidikan Guru Sekolah Dasar, halaman 122-126. ISBN: 978-623-91681-0-0, DOI: <http://dx.doi.org/10.33578/psn.v1i1.7787>.
- Ningsih S, Dukalang H. 2019. Penerapan metode suksesif interval pada analisis regresi linier berganda. *Jambura Journal of Mathematics*. 1(1) : 43-53.
- Padmadewi, N. N., Artini, L. P., Nitiasih, P. K., & Swandana, I. W. (2018). Memberdayakan keterlibatan orang tua dalam pembelajaran literasi di sekolah dasar. *Jurnal ilmu sosial dan humaniora*, 7(1), 64-76.
- Prayogo, D., & Sukim, S. (2021). Determinan daya beli masyarakat Indonesia selama pandemi Covid-19 tahun 2020. *Prosiding Seminar Nasional Official Statistics*, 1(1): 631-640.
- Pujilestari, S. Dwidayati, N. Sugiman, S. 2017. Pemilihan Model Regresi Linier Berganda Terbaik pada Kasus Multikolinieritas berdasarkan Metode Principal Component Analysis (PCA) dan Metode Stepwise. *UNNES Journal of Mathematics*. 6(1) : 70-81.
- Puslitjakdikbud. 2019. Indeks Aktivitas Literasi Membaca 34 Provinsi. https://repositori.kemdikbud.go.id/13033/1/Puslitjakdikbud_Indeks%20Aktivitas%20Literasi%20Membaca%2034%20Provinsi diakses pada 11 April 2023.
- Saefuddin, A., Notodiputro, K. A., Alamudi, A., & Sadik, K. (2009). *Statistika Dasar*. Jakarta(ID): Grasindo.
- Samosir, N., Siagian, P., & Bangun, P. (2014). Analisa Metode Backward dan Metode Forward untuk Menentukan Persamaan Regresi Linier Berganda (Kasus Jumlah Kecelakaan Lalu Lintas di Kotamadya). *Saintia Matematika*, 2(4), 345-360.
- Sari, Y., Dwidayanti, N. K., & Hendikawati, P. (2018). Estimasi parameter pada regresi Spatial Error Model (SEM) yang memuat outlier menggunakan Iterative Z Algorithm. *Prosiding Seminar Nasional Matematika*, 1(1): 456-463.
- Soleh, A. M., Darusman, L. K., & Rafi, M. (2008). Model otentikasi komposisi obat bahan alam berdasarkan spektra inframerah dan komponen utama studi kasus: obat bahan alam/fitofarmaka penurun tekanan darah. *Forum Statistika dan Komputasi*, 13(1): 77-84.
- Soleh, A. M, Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2015). Statistical downscaling to predict monthly rainfall using linear regression with L1 regularization (LASSO). *Applied Mathematical Sciences*, 9(108): 5361–5369.