Penerapan Metode Regresi Linear Berganda Untuk Memprediksi Angka Buta Huruf Di Provinsi Jawa Barat

Kiki Mustaqim[1)*], Noviana Riza[2)], Nita Sawalia Fitri[3)], Nasywa Nur Handayani[4)]
[1,2,3,4)]Universitas Logistik dan Bisnis Internasional, Fakultas Logistik, Teknologi dan Bisnis, Program Studi Sains Data
kiki@ulbiac.id1)*, novianariza@ulbi.ac.id[2)], nitasawalia71@gmail.com[3)], nasywanurr1047@gmail.com[4)]

Received: 21 January 2025 Accepted: 4 February 2025 Published: 6 February 2025

* kiki@ulbiac.id

Kata Kunci: Prediksi, Regresi Linear Berganda, Angka Buta Huruf, Jawa Barat

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Angka buta huruf di Indonesia yang diartikan sebagai persentase penduduk berusia 15 tahun ke atas yang tidak memiliki kemampuan membaca dan menulis kalimat sederhana, menjadi isu yang sangat penting, terutama di provinsi dengan jumlah penduduk besar seperti Jawa Barat. Meskipun telah dilakukan berbagai upaya untuk meningkatkan tingkat melek huruf, tantangan tetap ada, dan data menunjukkan bahwa angka buta huruf masih cukup tinggi di beberapa daerah. Penelitian ini mengidentifikasi sejumlah variabel yang diduga berkontribusi terhadap angka buta huruf, termasuk jumlah penduduk dan kondisi ekonomi. ini bertujuan untuk memprediksi angka buta huruf di Provinsi Jawa Barat menggunakan metode regresi linear berganda. Variabel independen yang digunakan dalam penelitian ini adalah jumlah penduduk (X1), jumlah sekolah dasar (X2), dan jumlah penduduk miskin (X3), dan variabel dependennya adalah angka buta huruf. Data yang digunakan merupakan data sekunder dari instansi terkait di Provinsi Jawa Barat. Analisis regresi linear berganda diterapkan untuk mengetahui pengaruh ketiga variabel independen terhadap angka buta huruf serta memprediksi angka buta huruf di masa yang akan datang. Hasil penelitian menunjukkan bahwa ketiga variabel independen memiliki pengaruh terhadap angka buta huruf di Jawa Barat. Model regresi yang dihasilkan dapat digunakan untuk memprediksi angka buta huruf berdasarkan perubahan pada jumlah penduduk, jumlah sekolah dasar, dan jumlah penduduk miskin.

PENDAHULUAN

Persentase penduduk yang berumur 15 tahun ke atas dan tidak mampu membaca serta menulis kalimat sederhana dalam sebuah bahasa disebut sebagai angka buta huruf, yang merupakan kebalikan dari angka melek huruf. pendidikan berperan penting sebagai faktor utama dalam meningkatkan kualitas sumber daya manusia di suatu daerah. UNESCO (2019) [1], mendefinisikan "Angka melek huruf mencerminkan persentase penduduk usia 15 tahun ke atas yang memiliki kemampuan membaca dan menulis kalimat sederhana dalam bahasa tertentu." Jawa Barat, sebagai salah satu provinsi dengan jumlah penduduk terbesar di Indonesia, memiliki tantangan dalam meningkatkan angka melek huruf masyarakatnya yang masih berada di bawah angka nasional." [2]. Terlepas dari upaya yang telah dilakukan untuk mengurangi buta huruf di Indonesia, angka ini masih cukup tinggi di beberapa wilayah, seperti Jawa Barat, dan memerlukan perhatian khusus. Penelitian mendalam tentang faktor-faktor yang mempengaruhi angka buta huruf di Jawa Barat diperlukan untuk membuat kebijakan dan strategi yang efektif untuk mengurangi angka tersebut. Untuk menyelidiki hubungan antara variabel terikat dan berbagai variabel bebas yang mempengaruhinya, dapat menggunakan metode analisis regresi linear berganda. Sebagaimana dijelaskan oleh Wibisono et al. (2019) [13], "Analisis regresi linier berganda merupakan analisis statistik yang digunakan untuk memprediksi nilai suatu variabel berdasarkan dua atau lebih variabel bebas."

Faktor-faktor yang dapat mempengaruhi angka buta huruf dapat ditemukan dengan menggunakan analisis regresi linear berganda. Selain itu, kontribusi masing-masing faktor tersebut dapat dihitung [13]. Informasi

ini sangat penting untuk menetapkan prioritas faktor-faktor yang harus ditangani secara khusus dalam upaya mengurangi buta huruf di Jawa Barat. Penelitian ini bertujuan untuk memodelkan hubungan antara angka buta huruf di Jawa Barat dengan berbagai faktor yang mempengaruhinya. Untuk mencapai tujuan ini, akan digunakan metode analisis regresi linear berganda. Dengan memanfaatkan data dari sumber yang valid dan dapat dipercaya, model regresi yang akan dikembangkan diharapkan dapat menggambarkan secara akurat pengaruh dari faktor-faktor tersebut terhadap angka buta huruf. Hasil penelitian ini diharapkan dapat membantu pemerintah daerah dan para pemangku kepentingan dalam merancang kebijakan, program, dan strategi yang efektif untuk mengurangi angka buta huruf di Jawa Barat.

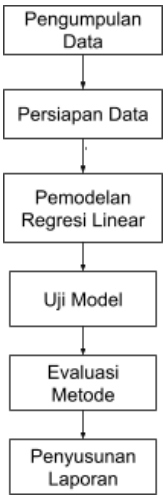
.

TINJAUAN LITERATUR

Penelitian sebelumnya oleh Sukri (2021) [6] menganalisis pengaruh angka partisipasi sekolah dan penduduk buta huruf terhadap jumlah masyarakat miskin di Kota Padang sidimpuan menggunakan regresi linear berganda. Hasil penelitian menunjukkan adanya pengaruh negatif yang signifikan dari angka partisipasi sekolah (koefisien -0,042, p=0,005) dan pengaruh positif signifikan dari jumlah penduduk buta huruf (koefisien 2,973, p=0,005) terhadap jumlah masyarakat miskin. Penelitian lain mengenai variabel-variabel yang mempengaruhi Angka Buta Huruf (ABH) di Provinsi Papua tahun 2020 [17] menemukan bahwa beberapa faktor seperti Angka Kesiapan Sekolah (AKS), Angka Putus Sekolah (APTS) SD, jumlah sekolah, dan jumlah penduduk memiliki pengaruh signifikan terhadap ABH. Penelitian ini mengungkap adanya ketimpangan pembangunan pendidikan dan sistem pendidikan dasar yang belum efektif di wilayah tersebut. Sanusi et al. (2023) [5] mengambil pendekatan berbeda dengan menggunakan metode Geographically Weighted Regression (GWR) untuk menganalisis angka buta huruf di Provinsi Sulawesi Selatan. Penelitian ini menghasilkan temuan bahwa model GWR memberikan tingkat akurasi yang lebih tinggi (86,1%) dibandingkan model regresi linear global (69,2%). Penelitian ini berkontribusi pada literatur dengan mendemonstrasikan variasi pengaruh variabel- variabel terhadap angka buta huruf di berbagai lokasi geografis, meskipun beberapa daerah seperti Palopo, Tana Toraja, dan Toraja Utara tidak menunjukkan efek yang signifikan.

METODE PENELITIAN

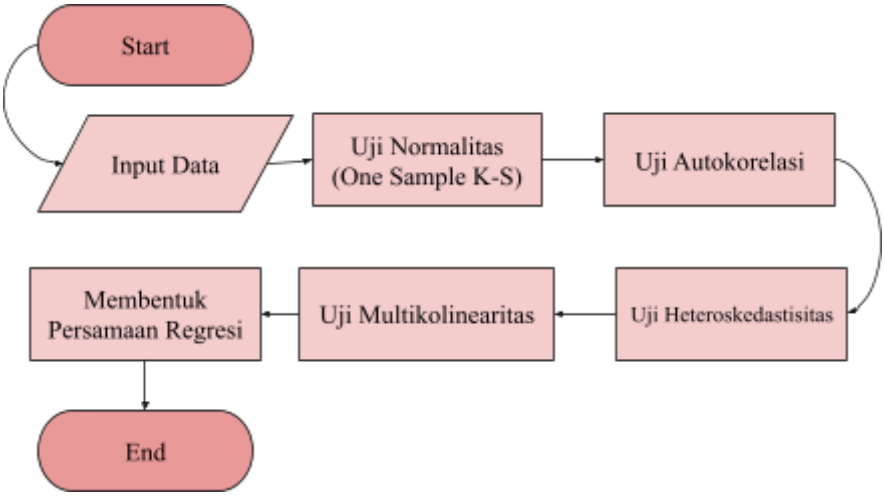
Metode penelitian merupakan sebuah ilmu yang berisikan berbagai metode penelitian. Metodologi penelitian membahas tentang konsep teoritis dari bermacam metode baik kelebihan maupun kekurangan pada suatu penelitian [11]. Pada penelitian ini, dibuat sebuah diagram alur yang memaparkan secara singkat mengenai tahapan penelitian yang akan dilakukan. Berikut merupakan diagram alur metodologi penelitian.



Gambar 1 Alur Penelitian

Tahapan-Tahapan Diagram Alur Penelitian

1. Pengumpulan Data
   Data yang digunakan dalam penelitian ini diperoleh dengan metode dokumentasi. Metode dokumentasi merupakan metode pengumpulan data sekunder dari sumber-sumber yang ada dengan mencatat dokumen yang berkaitan dengan penelitian [11]. Maka metode ini digunakan untuk mengumpulkan data dalam bentuk CSV. Data ini dapat diunduh dari situs Badan Pusat Statistik Indonesia (bps.go.id) dan Data Jabar
2. Preprocessing Data
   Dalam tahapan ini dilakukan penyamaan satuan dengan konversi semua data kedalam satuan yang sama untuk setiap variabel. Data angka buta huruf dan data penduduk miskin awalnya dalam bentuk persen, dikonversi ke dalam bilangan real. Dataset jumlah penduduk dalam bentuk jutaan ribu dan ratusan ribu, jumlah SD dalam bentuk ratusan dan ribuan, jumlah penduduk miskin dalam bentuk ribuan, dan angka buta huruf dalam bentuk ribuan
3. Penerapan Regresi Linear Berganda
   Alur tahapan dari penerapan regresi linear berganda dapat dilihat dari gambar berikut ini:



Gambar 2 Tahapan Penerapan Regresi Linear

Pada dataset yang digunakan dalam penelitian ini, terdapat satu variabel terikat (Y) dan 3 variabel bebas (X). Variabel terikat tersebut adalah Angka buta huruf, sedangkan variabel bebasnya adalah Jumlah Penduduk, Jumlah SD, dan Persentase penduduk Miskin.

HASIL PENELITIAN

1. Data Yang Digunakan Pada Analisis Data
   Dataset yang digunakan pada penelitian ini adalah dataset berisi 189 kolom dan 6 baris yang didapatkan dari situs web bps.go.id dan opendata.jabarprov.go.id. Berikut merupakan data yang digunakan pada analisis data :

Tabel 1 Cuplikan Data

Kota/Kabupaten	Tahun	Jumlah Penduduk	Jumlah SD	Jumlah Penduduk Miskin	Angka Buta Huruf
KABUPATEN BOGOR	2017	4358	1411	373	22
KABUPATEN SUKABUMI	2017	2537	947	204	4
KABUPATEN CIANJUR	2017	225	1086	257	6
KABUPATEN BANDUNG	2017	3525	1801	260	12

KABUPATEN GARUT	2017	2222	651	250	10
KABUPATEN CIAMIS	2017	123	1577	101	0.6
KABUPATEN KUNINGAN	2017	1137	599	151	4
KABUPATEN CIREBON	2017	2100	744	272	21
KABUPATEN MAJALENGKA	2017	1269	892	160	8
...	...	...	...	...	...
KABUPATEN PANGANDARAN	2023	442	658	40	4
KOTA BOGOR	2023	1127	671	75	10
KOTA SUKABUMI	2023	365	421	27	2
KOTA BANDUNG	2023	2569	283	102	6
KOTA CIREBON	2023	352	226	32	6
KOTA BEKASI	2023	2514	287	103	21
KOTA DEPOK	2023	1941	118	46	22
KOTA CIMAHI	2023	575	164	27	2
KOTA TASIKMALAYA	2023	758	127	87	2
KOTA BANJAR	2023	208	88	13	3
KABUPATEN TASIKMALAYA	2017	1733	284	188	2

1. Preprocessing Data
   a. Memerika Jumlah Baris dan Kolom
   Fungsi "shape" digunakan untuk mengetahui jumlah baris dan kolom pada dataset yang ada. Dataset yang digunakan dalam penelitian ini memiliki 189 baris dan 6 kolom.

```
data.shape
(189, 6)
```

Gambar 3 Jumlah Baris dan Kolom

   b. Memeriksa Kolom dengan Nilai Kosong
   Fungsi `isnull().sum()` digunakan untuk memeriksa apakah terdapat nilai kosong dalam dataset yang digunakan.

```
data.isnull().sum()
```

	0
Kota/Kabupaten	0
Tahun	0
Jumlah Penduduk	0
Jumlah SD	0
Jumlah Penduduk Miskin	0
Angka Buta Huruf	0

dtype: int64

Gambar 4 Jumlah Data Kosong

Dari output yang diperoleh, tidak ada data yang memiliki nilai kosong.

c. Statistika Deskriptif

Fungsi `describe` digunakan untuk mengetahui statistika deskriptif yang digunakan pada dataset, seperti rata-rata, standar deviasi, minimal, maksimal, kuartil 1, kuartil 2, dan kuartil 3.

```
[ ] data.describe()
```

	Tahun	Jumlah Penduduk	Jumlah SD	Jumlah Penduduk Miskin	Angka Buta Huruf
count	189.000000	189.000000	189.000000	189.000000	189.000000
mean	2020.000000	1712.142857	726.645503	144.230688	23.610582
std	2.005312	1112.140980	458.348906	97.752164	34.562898
min	2017.000000	101.000000	86.000000	10.100000	0.000000
25%	2018.000000	936.000000	284.000000	75.900000	3.000000
50%	2020.000000	1650.000000	673.000000	136.000000	9.000000
75%	2022.000000	2438.000000	974.000000	199.900000	25.000000
max	2023.000000	5559.000000	1903.000000	491.200000	173.000000

Gambar 3 Jumlah Data Kosong

d. Uji Normalitas

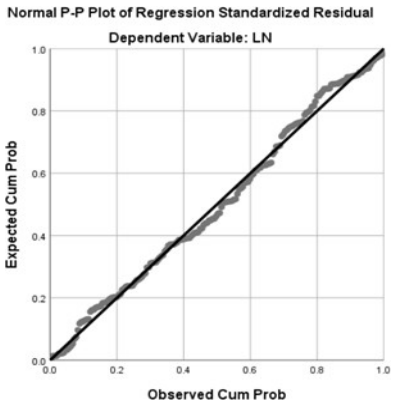
One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		188
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	1.18207189
Most Extreme Differences	Absolute	.053
	Positive	.042
	Negative	-.053
Test Statistic		.053
Asymp. Sig. (2-tailed)		.200[c,d]

a. Test distribution is Normal.
b. Calculated from data.
c. Lilliefors Significance Correction.
d. This is a lower bound of the true significance.

Gambar 4 Uji Normalitas

Kriteria Pengujian : Jika Nilai Asymp.sig. (2-tailed) lebih besar dari 0,05 maka berkesimpulan Data Berdistribusi Normal. Analisis Data, Nilai Asymp.sig. (2-tailed) yang didapat sebesar 0,200 (>0,05) maka berkesimpulan Data Berdistribusi secara Normal. Karena nilai yang didapat lebih besar dari 0,05.



Gambar 5 Grafik Uji Normalitas

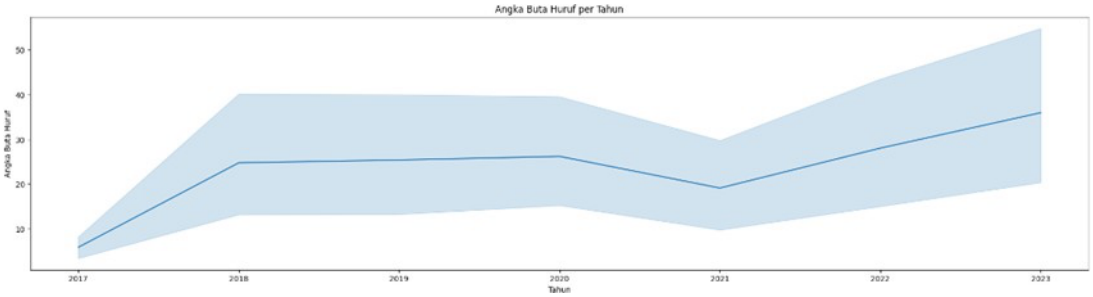
Pada grafik gambar 5, data tersebar mendekati garis, kesimpulan yang dapat diambil adalah asumsi normalitas terpenuhi.

e. Visualisasi Dataset

1. Visualisasi Angka Buta Huruf per Tahun

```
# Visualisasi Angka Buta Huruf per Tahun
plt.figure(figsize=(25,6))
sns.lineplot(x = data['Tahun'], y = data['Angka Buta Huruf'],data=data)
plt.title('Angka Buta Huruf per Tahun')
plt.show()
```

Gambar 6 Sintaks Hasil Visualisasi Angka Buta Huruf per Tahun



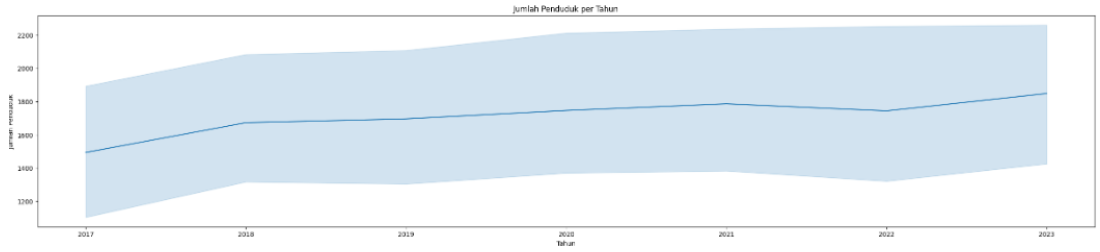
Gambar 7 Hasil Visualisasi Angka Buta Huruf per Tahun

Visualisasi pada gambar 7, sempat ada penurunan dari tahun 2020 sampai tahun 2021, kemudian terdapat kenaikan yang cukup signifikan dari tahun 2021.

2. Visualisasi Jumlah Penduduk per Tahun

```
# Visualisasi Jumlah Penduduk per Tahun
plt.figure(figsize=(30,6))
sns.lineplot(x = data['Tahun'], y = data['Jumlah Penduduk'],data=data)
plt.title('Jumlah Penduduk per Tahun')
plt.show()
```

Gambar 8 Sintaks Visualisasi Jumlah Penduduk per Tahun



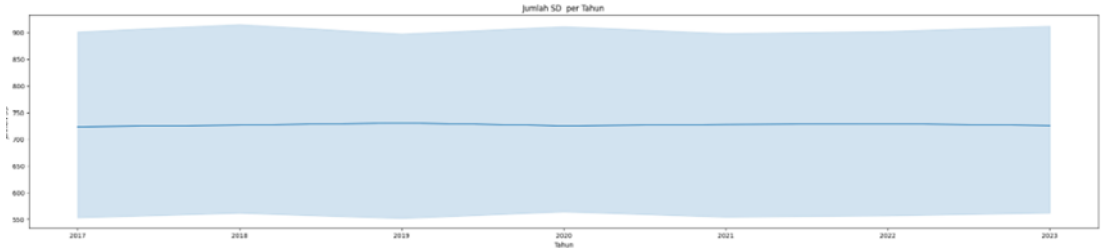
Gambar 9 Hasil Visualisasi Jumlah Penduduk per Tahun

Visualisasi pada gambar 9, terdapat kenaikan yang cukup signifikan dari tahun 2017.

3. Visualisasi Jumlah SD

```
# Visualisasi Jumlah SD  per Tahun
plt.figure(figsize=(30,6))
sns.lineplot(x = data['Tahun'], y = data['Jumlah SD '],data=data)
plt.title('Jumlah SD  per Tahun')
plt.show()
```

Gambar 10 Sintaks Hasil Visualisasi Jumlah SD per Tahun



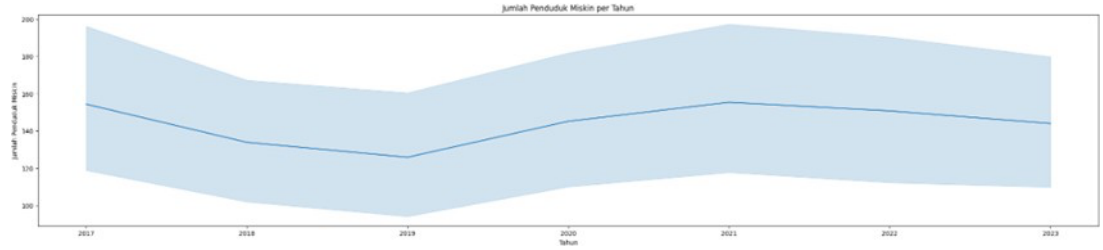
Gambar 11 Hasil Visualisasi Jumlah SD per Tahun

Visualisasi pada gambar 11, Jumlah SD stagnan dari tahun 2017 hingga 2023

4. Visualisasi Jumlah Penduduk Miskin

```
# Visualisasi Jumlah Penduduk Miskin per Tahun
plt.figure(figsize=(30,6))
sns.lineplot(x = data['Tahun'], y = data['Jumlah Penduduk Miskin'])
plt.title('Jumlah Penduduk Miskin per Tahun')
plt.show()
```

Gambar 12 Sintaks Hasil Visualisasi Jumlah Penduduk Miskin per Tahun



Gambar 13 Hasil Visualisasi Jumlah Penduduk Miskin per Tahun

Visualisasi pada gambar 13, terdapat penurunan pada tahun 2018 lalu mengalami kenaikan pada tahun 2020.

f. Menghapus Kolom Kota/Kabupaten dan Tahun

```
df = data.copy()
df = df.drop(columns=['Kota/Kabupaten','Tahun'])
df.head()
```

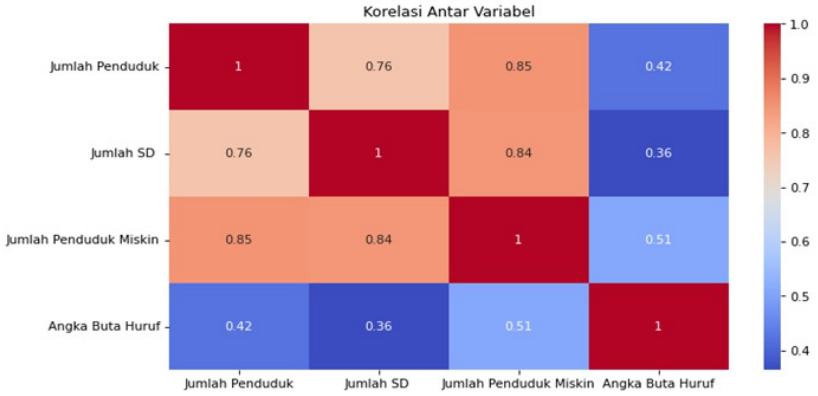
	Jumlah Penduduk	Jumlah SD	Jumlah Penduduk Miskin	Angka Buta Huruf
0	4358	1411	487.3	22.0
1	2537	947	197.1	4.0
2	225	1086	257.4	6.0
3	3525	1801	268.0	12.0
4	2222	651	291.2	10.0

Gambar 14 Penghapusan Kolom Kota/Kabupaten dan Tahun

Fungsi drop digunakan untuk menghapus kolom ataupun baris. Karena kolom kota/kabupaten dan tahun tidak termasuk variabel yang digunakan dalam pemodelan maka dilakukan penghapusan kolom.

g. Matriks Korelasi

```
fig, ax = plt.subplots(figsize=(10,5), dpi=80)
sns.heatmap(df.loc[:, 'Jumlah Penduduk':'Angka Buta Huruf'].corr(), cmap='coolwarm', annot=True
plt.title('Korelasi Antar Variabel')
plt.show()
```



Gambar 15 Matriks Korelasi

Heatmap pada gambar 15, nilai statistik korelasi jumlah penduduk miskin terhadap angka buta huruf adalah 0,53. Nilai tersebut lebih besar dari 0,5 sehingga jumlah penduduk miskin berpengaruh secara signifikan terhadap angka buta huruf, sedangkan variabel bebas lainnya bernilai di bawah 0,5 sehingga variabel-variabel tersebut tidak berpengaruh signifikan terhadap angka buta huruf.

h. Mencari Nilai R2 *Score* Terbaik dengan *Random State*

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score

x = df.drop(columns='Angka Buta Huruf')
y = df['Angka Buta Huruf']

# Inisialisasi variabel untuk menyimpan hasil terbaik
best_random_state = None
best_r2_score = -float('inf')  # Inisialisasi dengan nilai R^2 yang sangat rendah agar dapat diperbarui

# Coba beberapa nilai random_state dan evaluasi performa model
for random_state in range(5000):  # Lakukan percobaan untuk beberapa nilai random_state
    # Bagi data menggunakan nilai random_state saat ini
    X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=random_state)

    # Bangun model
    model = LinearRegression()
    model.fit(X_train, y_train)

    # Lakukan prediksi
    y_pred = model.predict(X_test)

    # Hitung R^2 score
    r2 = r2_score(y_test, y_pred)

    # Perbarui nilai random_state terbaik jika R^2 score saat ini lebih baik
    if r2 > best_r2_score:
        best_random_state = random_state
        best_r2_score = r2

# Cetak nilai random_state terbaik
print("Best random_state:", best_random_state)
print(best_r2_score)
```

Gambar 16 Sintaks Nilai Score R2

```
Best random_state: 3980
0.6443514214294263
```

Gambar 17 Nilai Score R2

Sintaks pada gambar 16 merupakan sebuah proses untuk mendefinisikan variabel X dan Y dimana variabel X berisikan kolom-kolom pada *dataframe* df selain kolom angka buta huruf, sedangkan variabel Y hanya berisikan kolom angka buta huruf saja. kemudian melakukan proses untuk mencari *random state* dengan nilai R^2 *score* terbaik dari 5000 kondisi acak, menggunakan fungsi *looping*.

i. *Train Test Split*

```
x = df.drop(columns='Angka Buta Huruf')
y = df[['Angka Buta Huruf']]
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = best_random_state)
model_reg = LinearRegression()
model_reg.fit(x_train, y_train)
```

Gambar 18 *Train* and *Test Split*

Sintaks pada gambar 18 digunakan untuk membagi data menjadi data latih dan data uji dengan menggunakan random state yang telah diperoleh sebelumnya. Data uji digunakan untuk mengevaluasi keakuratan model, sedangkan data latih digunakan untuk membangun model tersebut.

j. Evaluasi

```
y_pred = model_reg.predict(x_test)
print(y_pred[:5])
print(y_test.head())

#R Squared
print(f'R Squared : {r2_score(y_test, y_pred)}')
#MAE
print(f'MAE : {mean_absolute_error(y_test, y_pred)}')
#MSE
print(f'MSE : {mean_squared_error(y_test, y_pred)}')
```

```
[[ 5.93459305]
 [13.5854298 ]
 [32.70570655]
 [48.10780341]
 [18.88568114]]
     Angka Buta Huruf
125               3.0
74               10.0
97                6.0
14               12.0
171              45.0
R Squared : 0.6443514214294263
MAE : 11.02327323015212
MSE : 268.47457030102254
```

Gambar 19 Evaluasi Model Regresi

Langkah berikutnya setelah model dibangun adalah mengevaluasi model dengan data uji, lalu dihitung nilai R2, MAE, dan MSE. Nilai R2, MAE, dan MSE yang diperoleh secara berurutan adalah 0,639, 12.885227, dan 334.5167031046139

k. Persamaan Regresi

```
# Print the coefficients
print('Coefficients:')
print('Intercept :', model_reg.intercept_)
print('Koefisien :', model_reg.coef_)
```

```
Coefficients:
Intercept : [1.83733452]
Koefisien : [[ 9.55991635e-06 -2.03812876e-02  2.52721349e-01]]
```

Gambar 20 Persamaan Regresi

Sintaks pada gambar 4.47, digunakan untuk mengembalikan nilai konstanta dan koefisien persamaan regresi. Dari hasil di atas persamaan yang didapatkan adalah:

$y = 1.83733452 + 9.55991635\,x1 - 2.03812876\,x2 + 2.52721349\,x3.$

l. Model Prediksi Faktor (Variabel Bebas)

MODEL PREDIKSI FAKTOR (VARIABEL BEBAS)

```
[ ] #import data
    data2 = pd.read_excel("Data_Rata-Rata_Pertahun.xlsx")
    data2.reset_index(drop=True, inplace=True)
    df1 = data2.copy()
    df1 = df1.drop(columns=['Angka Buta Huruf'])
    df1.head()
```

	Tahun	Jumlah Penduduk	Jumlah SD	Jumlah Penduduk Miskin
0	2017	1493.222222	723.444444	154.396296
1	2018	1672.518519	1672.518519	133.914815
2	2019	1695.555556	730.296296	125.907407
3	2020	1745.962963	725.074074	145.192593
4	2021	1786.037037	727.370370	155.396296

Gambar 21 Model Prediksi Faktor (Variabel Bebas)

Data yang digunakan dalam Sintaks pada gambar 21 adalah data rata-rata variabel bebas secara keseluruhan di Provinsi Jawa Barat setiap tahunnya. Sintaks tersebut juga digunakan untuk mengimpor data yang akan digunakan untuk memprediksi variabel bebas. Setelah memprediksi, lalu dilakukan evaluasi prediksi dengan data uji, lalu dihitung nilai R2, MAE, dan MSE. Nilai R2, MAE, dan MSE yang diperoleh secara berurutan adalah 0,639 (63,9%), 12.885227, dan 334.5167031046139. Dari hasil tersebut dapat diketahui bahwa model memiliki kemampuan prediksi yang cukup baik dengan R2 sebesar 63,9% yang menunjukkan variasi variabel terikat dapat dijelaskan oleh variabel bebas. Nilai MAE sebesar 12.885227 menunjukkan rata-rata kesalahan absolut antara nilai prediksi dan aktual, sementara MSE sebesar 334.5167031046139 mengindikasikan rata-rata kuadrat error dari model.

m. Import Library *Prophet*

Library Prophet digunakan untuk memprediksi nilai variabel bebas hingga tahun 2026. Setiap kolom diprediksi menggunakan looping, dan hasilnya disimpan dalam dataframe forecasted_df.

Prediksi Faktor hingga 2026

```
[ ] # Import library
    from prophet import Prophet
```

Gambar 22 Import Library Prophet

n. Prediksi Nilai – Nilai Faktor

```
print(forecasted_df)
```

	Jumlah Penduduk	Jumlah SD	Jumlah Penduduk Miskin
0	1978.341444	502.448016	148.008266
1	2034.471578	413.235875	154.442182
2	2087.845289	317.100631	164.236647

Gambar 23 Prediksi Nilai – Nilai Faktor

Setelah diprediksi, Jumlah SD semakin menurun setiap tahunnya. Setelah dilakukan berbagai macam riset, maka ditemukan penjelasan dari beberapa artikel dan jurnal yang menjelaskan tentang penurunan Jumlah SD ini. Beberapa faktor alasan yang dijelaskan, diantaranya :

1. Variasi antar kabupaten/kota dimana terjadi situasi yang berbeda-beda di setiap kabupaten/kota di Jawa Barat. Beberapa daerah mungkin mengalami penurunan, sementara yang lain mungkin stabil atau bahkan meningkat [15].

2. Faktor-faktor yang mempengaruhi jika memang terjadi penurunan Jumlah SD ini bisa disebabkan oleh berbagai faktor seperti penurunan angka kelahiran, perpindahan penduduk, atau kebijakan konsolidasi sekolah [16].
3. Kualitas vs kuantitas. Penurunan jumlah sekolah tidak selalu berarti penurunan kualitas pendidikan. Terkadang, penggabungan sekolah dilakukan untuk meningkatkan efisiensi dan kualitas pendidikan [16].

o. Prediksi 2024-2026

Prediksi 2024-2026

```
[ ] future_df = model_reg.predict(forecasted_df)
    tahun = [2024,2025,2026]
    df_tahun = pd.DataFrame(tahun, columns=['Tahun'])
    df_future = pd.DataFrame(future_df, columns=['Angka Buta Huruf'])
    df_3yr = pd.concat([df_tahun, df_future], axis=1)
    df_3yr
```

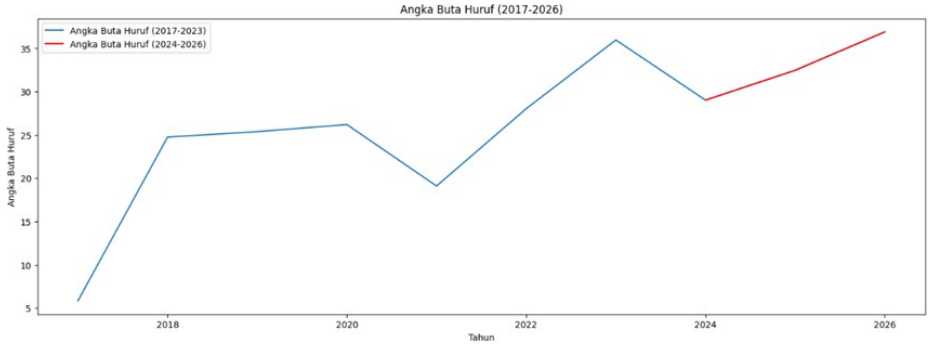
	Tahun	Angka Buta Huruf
0	2024	29.020558
1	2025	32.465341
2	2026	36.900482

Gambar 24 Prediksi 2024-2026

Nilai variabel bebas yang diprediksi digunakan dalam model regresi untuk memperkirakan angka buta huruf tiga tahun ke depan. Hasil prediksi menunjukkan angka buta huruf sebesar 29,02 pada tahun 2024, 32,46 pada tahun 2025, dan 36,90 pada tahun 2026.

p. Visualisasi Angka Buta Huruf dari Tahun ke Tahun

Data hasil Angka Buta Huruf digabungkan dengan data awal untuk membuat visualisasi prediksi Angka Buta Huruf setiap tahunnya.



Gambar 24 Visualisasi Angka Buta Huruf dari Tahun ke Tahun

Visualisasi pada gambar 4.31, prediksi Angka Buta Huruf menunjukkan kenaikan pada tahun 2024 hingga 2026..

PEMBAHASAN

Penelitian ini menganalisis faktor-faktor yang mempengaruhi angka buta huruf di Provinsi Jawa Barat menggunakan metode regresi linear berganda. Hasil analisis menunjukkan bahwa model regresi yang dikembangkan memiliki nilai R^2 sebesar 0,639, yang mengindikasikan bahwa 63,9% variasi dalam angka buta huruf dapat dijelaskan oleh variabel-variabel independen yang digunakan dalam model. Berdasarkan matriks korelasi, ditemukan bahwa jumlah penduduk miskin memiliki pengaruh yang signifikan terhadap angka buta huruf dengan nilai korelasi 0,53. Hal ini sejalan dengan penelitian sebelumnya yang menunjukkan adanya hubungan kuat antara kemiskinan dan tingkat literasi. Sementara itu, variabel independen lainnya menunjukkan korelasi yang lebih lemah (di bawah 0,5), mengindikasikan pengaruh yang tidak terlalu signifikan terhadap angka buta huruf. Model regresi yang dihasilkan memiliki persamaan: $y = 1{,}83733452 + 9{,}55991635x_1 - 2{,}03812876x_2 + 2{,}52721349x_3$. Evaluasi kinerja model menggunakan metrics MAE dan MSE menghasilkan nilai masing-masing sebesar 12,885227 dan 334,5167031046139. Nilai MAE yang relatif kecil menunjukkan bahwa model memiliki tingkat akurasi yang

cukup baik dalam memprediksi angka buta huruf.

Hasil prediksi untuk tahun 2024-2026 menunjukkan tren kenaikan angka buta huruf, dengan prediksi sebesar 29,02 pada tahun 2024, 32,46 pada tahun 2025, dan 36,90 pada tahun 2026. Tren kenaikan ini perlu mendapat perhatian khusus dari pemangku kebijakan untuk mengambil langkah-langkah preventif dalam menangani masalah buta huruf di Provinsi Jawa Barat. Analisis temporal menunjukkan adanya penurunan seiring waktu. Fenomena ini dapat dijelaskan oleh beberapa faktor, termasuk variasi antar kabupaten/kota, perubahan demografis, dan kebijakan konsolidasi sekolah. Meskipun terjadi penurunan kuantitas, hal ini tidak selalu mengindikasikan penurunan kualitas pendidikan, karena dapat merupakan bagian dari strategi efisiensi dan peningkatan mutu pendidikan. Keterbatasan penelitian ini terletak pada jumlah variabel independen yang digunakan. Penelitian selanjutnya dapat mempertimbangkan variabel tambahan seperti aksesibilitas pendidikan, kualitas pengajaran, dan faktor sosio-ekonomi lainnya untuk meningkatkan akurasi prediksi. Selain itu, penggunaan metode machine learning yang lebih advanced seperti ensemble learning atau deep learning mungkin dapat menghasilkan prediksi yang lebih akurat.

Hasil penerapan dari penelitian ini adalah pentingnya fokus pada pengentasan kemiskinan sebagai strategi untuk menurunkan angka buta huruf, mengingat kuatnya korelasi antara kedua variabel tersebut. Selain itu, hasil prediksi yang menunjukkan tren kenaikan dapat digunakan sebagai dasar untuk merumuskan kebijakan preventif dalam menangani masalah buta huruf di masa mendatang.

KESIMPULAN

Berdasarkan hasil analisis menggunakan regresi linear berganda terhadap angka buta huruf di Provinsi Jawa Barat, ditemukan bahwa ketiga variabel independen (jumlah penduduk, jumlah SD, dan jumlah penduduk miskin) secara bersama-sama mempengaruhi angka buta huruf dengan kontribusi sebesar 63,9%. Model regresi menghasilkan persamaan y = -1,20187617 + 0,00263x1 - 0,022716x2 + 0,262265x3, di mana jumlah penduduk miskin memiliki pengaruh positif paling signifikan, sementara jumlah SD menunjukkan pengaruh negatif yang mengindikasikan potensi penurunan angka buta huruf. Hasil prediksi angka buta huruf menunjukkan tren peningkatan dari 29 jiwa (2024) menjadi 36 jiwa (2026), dengan tingkat akurasi oleh MAE sebesar 12,885227 dan MSE sebesar 334,5167. Dari ketiga faktor tersebut, jumlah penduduk miskin memberikan pengaruh terbesar terhadap meningkatnya angka buta huruf, sedangkan bertambahnya jumlah SD dapat membantu menurunkan angka buta huruf. Oleh karena itu, pemerintah perlu melakukan beberapa upaya seperti menambah akses pendidikan dasar terutama di daerah miskin, memperkuat program membaca dan menulis, serta mengoptimalkan fasilitas sekolah yang sudah ada. Untuk penelitian selanjutnya, disarankan untuk menambahkan faktor-faktor lain seperti ketersediaan internet, jumlah guru, dan program literasi yang sedang berjalan agar hasil prediksi lebih akurat.

REFERENCES

[1] UNESCO, *Global education monitoring report 2022: Gender Report*. 2022.

[2] Asiva Noor Rachmayani, "Statistik Pendidikan Provinsi Jawa Barat 2022," p. 6, 2015.

[3] K. P. dan K. R. Indonesia, "Kementerian Pendidikan Dan Kebudayaan Republik Indonesia," 2020. [Online]. Available: https://www.kemdikbud.go.id/main/blog/2020/09/pemerintah-terus-berkomitmen-dalam-mengentaskan-buta-aksara.

[4] Sudariana and Yoedani, "Analisis Statistik Regresi Linier Berganda," *Seniman Trans.*, vol. 2, no. 2, pp. 1– 11, 2022.

[5] W. Sanusi, A. N. Arkas, and & Rusli, "Pemodelan Angka Buta Huruf di Provinsi Sulawesi Selatan Menggunakan Geographically Weighted Regression," *ARRUS J. Math. Appl. Sci.*, vol. 3, no. 2, pp. 55–64, 2023, [Online]. Available: https://doi.org/10.35877/mathscience2144

[6] S. R, "PENGARUH ANGKA PARTISIPASI SEKOLAH DAN PENDUDUK BUTA HURUF TERHADAP JUMLAH MASYARAKAT MISKIN DI KOTA PADANGSIDIMPUAN," *J. LPPM UGN*, vol. Vol. 12 No, no. Vol. 12 No. 1A (2021): Jurnal LPPM UGN, 2021, [Online]. Available: https://jurnal.ugn.ac.id/index.php/jurnalLPPM/article/view/1157

[7] A. P. Satvika, N. L. P. Suciptawati, and M. Susilawati, "Memodelkan Angka Buta Huruf di Provinsi Nusa Tenggara Barat," *J. Educ.*, vol. 06, no. 01, pp. 1499–1507, 2023.

[8] N. Afriansyah,H.,Permatasari,D, "Penerapan Sistem Manajemen Mutu untuk Meningkatkan Kualitas Pendidikan di Sekolah Dasar," *Manaj. Pendidik.*, pp. 3–4, 2019, [Online]. Available: https://www.google.com/url?sa=t&source=web&rct=j&url=https://osf.io/qg6u3/download/%3Fformat%3Dpdf&ved=2ahUKEwiTgumQsOf0AhU04nMBHdhnAkkQFnoECAUQAQ&usg=AOvVaw0NV2CvsP0ZuhitRWy7fEdT

[9] J. R. Ritonga, S. Ginting, E. Naibaho, A. Situmorang, and K. Kunci, "Pengaruh Angka Melek Huruf Dan Angka Partisipasi Sekolah Terhadap Jumlah Kemiskinan Di Provinsi Sumatera Utara Tahun 2017-2023," vol. 10, pp. 676–681, 2024.

[10] A. M. Dawis *et al.*, *Pengantar Metodologi Penelitian*. 2023.

[11] M. Asqia, *Video Penjelasan Regresi Linier Berganda pada Python*, (2021). [Online]. Available: https://youtu.be/VmiLme-Eb_Q?si=_70Ca7_H6TNzPnVE

[12] A. Wibisono, M. Rofik, and E. Purwanto, "Penerapan Analisis Regresi Linier Berganda dalam Penyelesaian Skripsi Mahasiswa," *J. ABDINUS J. Pengabdi. Nusant.*, vol. 3, no. 1, p. 30, 2019, doi: 10.29407/ja.v3i1.13512.

[13] Badan Pusat Statistik Indonesia, "Profil Kemiskinan di Indonesia Maret 2023," bps.go.id. Accessed: Mar. 04, 2024. [Online]. Available: https://www.bps.go.id/id/pressrelease/2023/07/17/2016/profil-kemiskinan-di-indonesia-maret-2023.html

[14] D. N. Megasari and L. Purnastuti, "Disparitas Gender Dalam Tingkat Pengembalian Investasi Pendidikan Di Jawa Barat," *J. Econ.*, vol. 12, no. 1, p. 23, 2016, doi: 10.21831/economia.v12i1.8463.

[15] I. D. Nauraida, "Efektivitas Sekolah Rujukan Terhadap Peningkatan Mutu Pada Jenjang Pendidikan Menengah," *JDMP (Jurnal Din. Manaj. Pendidikan)*, vol. 7, no. 2, pp. 104–113, 2023, doi: 10.26740/jdmp.v7n2.p104-113.

[16] B. S. Lake and E. D. Utami, "Variabel-Variabel yang Memengaruhi Angka Buta Huruf (ABH) di Provinsi Papua Tahun 2020," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 113–122, 2022, doi: 10.34123/semnasoffstat.v2022i1.1216.