MASALIQ
Jurnal Pendidikan dan Sains

e-ISSN : 2808-8115
p-ISSN : 2809-1051

Terindeks : Sinta 5, Copernicus, Dimensions, Scilit, Lens, Crossref, Garuda, Google Scholar, Base, etc

https://doi.org/10.58578/masaliq.v5i1.4335

IMPLEMENTASI REGRESI LINEAR BERGANDA PREDIKSI FAKTOR-FAKTOR INDEKS PEMBANGUNAN MANUSIA DI PROVINSI JAWA BARAT

Implementation of Multiple Linear Regression to Predict Factors Affecting the Human Development Index in West Java Province

Noviana Riza[1], Fatia Amalia Maresti[2], Siti Salwa Azzahra[3], Salsa Paringga Pangestu Ningsih[4]
Universitas Logistik dan Bisnis Internasional
novianariza@ulbi.ac.id

Article Info:

Submitted:	Revised:	Accepted:	Published:
Nov 15, 2024	Dec 1, 2024	Dec 9, 2024	Dec 14, 2024

Abstract

The Human Development Index (HDI) is a key indicator for measuring the welfare and prosperity of a region, including West Java Province. This study aims to analyze the factors influencing HDI and predict its future trends. The analysis was conducted using a multiple linear regression method implemented with the Python programming language, with independent variables including Life Expectancy, Expected Years of Schooling, Mean Years of Schooling, and Adjusted Per Capita Expenditure. The results show that Expected Years of Schooling (X3) and Adjusted Per Capita Expenditure (X4) are the most significant factors influencing HDI in West Java, particularly due to the declining trends in these variables. Based on the model, the predicted HDI values for 2024, 2025, and 2026 are 73.19, 72.59, and 71.62, respectively, which fall under the medium HDI category. These findings provide valuable insights for strategic planning to improve HDI in West Java, particularly through interventions targeting the significant variables.

Masaliq Journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

Keywords: HDI; Prediction; Multiple Linear Regression; Python

Abstrak: Indeks Pembangunan Manusia (IPM) merupakan indikator utama untuk mengukur tingkat kesejahteraan masyarakat di suatu wilayah, termasuk Provinsi Jawa Barat. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi IPM serta memprediksi perkembangannya pada tahun-tahun mendatang. Analisis dilakukan menggunakan metode regresi linear berganda yang diimplementasikan dengan bahasa pemrograman Python, dengan variabel bebas meliputi Umur Harapan Hidup, Harapan Lama Sekolah, Rata-Rata Lama Sekolah, dan Pengeluaran per Kapita. Hasil penelitian menunjukkan bahwa Harapan Lama Sekolah (X3) dan Pengeluaran per Kapita (X4) merupakan variabel yang paling signifikan dalam memengaruhi nilai IPM di Jawa Barat, terutama akibat tren penurunan pada kedua variabel tersebut. Berdasarkan model yang dihasilkan, prediksi nilai IPM untuk tahun 2024, 2025, dan 2026 masing-masing adalah 73,19; 72,59; dan 71,62, yang tergolong dalam kategori IPM sedang. Temuan ini memberikan wawasan penting bagi perencanaan strategis dalam meningkatkan IPM di Jawa Barat, terutama melalui intervensi pada variabel yang signifikan.

Kata Kunci: Indeks Pembangunan Manusia; Prediksi; Regresi Linear Berganda; Python

PENDAHULUAN

Pembangunan merupakan suatu proses yang bertujuan untuk mengubah kondisi saat ini menjadi lebih baik di masa depan. Konsep ini dapat diterapkan dalam pembangunan nasional, dimana perubahan terus berlanjut hingga semua aspek kehidupan masyarakat dapat meningkat pada kesejahteraan dan kemakmuran masyarakat. Negara Indonesia sebagai negara berkembang, pembangunan merupakan salah satu indikator yang penting. Seperti halnya pembangunan manusia yang dapat diupayakan oleh pemerintah agar dapat mencapai kesejahteraan dan kemakmuran pada masyarakatnya (Khotimah et al., 2024). Kesejahteraan dan kemakmuran masyarakat dapat diukur dari Indeks Pembangunan Manusia (IPM) di daerah tersebut (Anastashya et al., 2023).

Menurut data dari BPS Jawa Barat menunjukkan variasi tingkat Indeks Pembangunan Manusia (IPM) antar kota atau kabupaten di Jawa Barat. Misalnya, pada tahun 2019 di Sukabumi dengan IPM sebesar 66,87 sementara Bogor memiliki IPM sebesar 70,65. Kota-kota lain seperti Bandung, Cianjur dan Tasikmalaya juga memiliki variasi tingkat IPM yang signifikan (BPS, 2019).

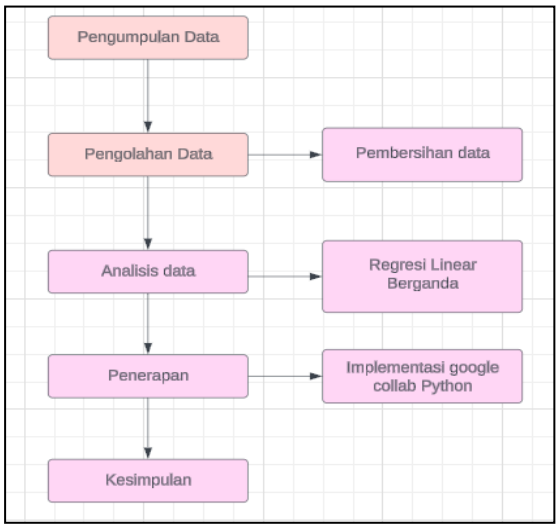
Menurut Badan Pusat Statistik 2023, Indeks Pembangunan Manusia (IPM) dapat dikatakan berhasil jika penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya (Sihite et al., 2024). Indeks Pembangunan Manusia (IPM) dapat dikatakan indeks yang fundamental dengan komponen seperti umur

panjang dan kesehatan dapat dilihat dari indikator Umur Harapan Hidup (UHH), pengetahuan dapat dilihat dari Harapan Lama Sekolah (HLS) dan rata-rata Lama Sekolah (RLS) dan taraf hidup yang layak dapat dilihat dari indikator Pengeluaran per Kapita (PPK) yang disesuaikan (Ul Hasanah & Yollanda, 2022).

Data BPS tahun 2020 menunjukkan bahwa wabah covid-19 yang menyebabkan beberapa wilayah provinsi terbesar di Indonesia, salah satunya adalah Provinsi Jawa Barat mengalami perubahan dari nilai faktor-faktor IPM yang mengakibatkan kenaikan maupun penurunan nilai IPM secara signifikan. Berdasarkan hal tersebut, oleh karena itu diperlukannya pengembangan suatu sistem lebih lanjut untuk mengetahui dalam memprediksi IPM Jawa Barat yang bisa saja terjadi kenaikan maupun penurunan di tahun mendatang menggunakan metode regresi linear berganda dengan bahasa pemrograman Python. Ada empat faktor dari Indeks Pembangunan Manusia (IPM) yang akan dianalisa dan prediksi yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), rata-rata Lama Sekolah (RLS) dan Pengeluaran per Kapita (PPK) (Aprila et al., 2023). Dalam menganalisa dan memprediksi Indeks Pembangunan Manusia (IPM) untuk tahun mendatang kami menggunakan metode statistika yaitu analisis regresi linear berganda dan bahasa pemrograman python (Khaerunisa et al., 2024).

Penelitian ini dapat digunakan sebagai dasar tentang penerapan metode regresi linear berganda dan bahasa pemrograman *python* untuk menganalisa dan memprediksi faktor yang mempengaruhi Indeks Pembangunan Manusia di Jawa Barat termasuk dan manfaatnya untuk masa mendatang.

METODE



Gambar 1. Alur Pemrosesan Analisis untuk Memprediksi IPM Jawa Barat

Metode penelitian dalam studi ini menggunakan pendekatan kuantitatif dengan analisis data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik (bps.go.id). Data yang digunakan bersifat historis dan mencakup indikator-indikator yang relevan dengan Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Barat. Variabel-variabel yang dianalisis terdiri dari variabel independen, yaitu Umur Harapan Hidup (UHH), Rata-Rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Pengeluaran Perkapita (PPK), serta variabel dependen, yaitu IPM. Penelitian ini bertujuan untuk mengevaluasi hubungan antara variabel-variabel tersebut serta memprediksi IPM di masa mendatang.

Tahapan penelitian dimulai dengan proses pengumpulan data yang kemudian diolah menggunakan perangkat lunak IBM SPSS Statistics versi 25 untuk analisis statistik. Data mentah dalam format Excel diolah melalui tahapan pembersihan untuk memastikan akurasi dan reliabilitas. Selanjutnya, dilakukan analisis deskriptif untuk menggambarkan karakteristik data. Tahapan berikutnya adalah uji asumsi klasik, meliputi uji normalitas, multikolinearitas, autokorelasi, dan heteroskedastisitas. Uji normalitas dilakukan dengan menggunakan metode *One-Sample Kolmogorov-Smirnov* Test untuk memastikan distribusi residual bersifat normal. Uji multikolinearitas digunakan untuk mendeteksi adanya hubungan linear yang kuat antarvariabel independen. Uji autokorelasi dilakukan untuk memastikan tidak adanya hubungan antara residual-residual dalam data berurutan, sedangkan uji heteroskedastisitas digunakan untuk memeriksa kesamaan varians residual. Setelah tahap uji asumsi, analisis dilanjutkan dengan penerapan regresi linier berganda untuk menentukan pengaruh variabel independen terhadap variabel dependen (Syahza, 2021).

Selain itu, analisis prediktif dilakukan dengan menggunakan Python di Google Collab. Menggunakan pustaka seperti *pandas* untuk pengolahan data, serta *statsmodels* untuk analisis regresi linier berganda, prediksi IPM untuk tahun mendatang dihitung berdasarkan model yang dihasilkan. Proses ini mencakup eksplorasi data, penghitungan rata-rata variabel, serta pembuatan model prediksi dengan validasi kinerja model. Penelitian ini dilakukan dalam

HASIL

Penelitian ini melewati beberapa tahapan. Bagian pertama adalah melakukan tahapan analisis menggunakan *tools spss*, tahapan yang kedua menggunakan *excel* untuk mencari model persamaan regresi linear berganda dan yang ketiga menggunakan bahasa pemrograman *google collab python*. Data yang digunakan bersumber dari situs *online* yaitu *bps.go.id*. Data yang

digunakan dalam penelitian ini bersifat sekunder dan berasal dari kumpulan data yang diambil dari situs *bps.go.id*. Data tersebut akan di lanjut untuk dijadikan sebagai bahan penelitian untuk memprediksikan Indeks Pembangunan Manusia (IPM) pada Provinsi Jawa Barat.

Uji Normalitas

Tabel 1. Uji Normalitas (SPPS)

		Unstandardized Residual
N		135
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	.45121880
Most Extreme Differences	Absolute	.055
	Positive	.040
	Negative	-.055
Test Statistic		.055
Asymp. Sig. (2-tailed)		.200[c,d]

Sumber: Data Diolah, 2024

Dari hasil *one sample kolmogorov-smirnov test* menunjukan hasil signifikannya adalah 0,200 yang artinya nilai residualnya lebih dari 0,05 dapat menunjukan data tersebut berdistribusi normal.

Uji Asumsi Multikolinearitas

Tabel 2. Uji Asumsi Multikolinearitas

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	25.543	2.240		11.405	.000		
	Jumlah_UHH	.163	.029	.054	5.612	.000	.762	1.312
	Jumlah_RLS	1.470	.059	.451	25.036	.000	.218	4.594
	Jumlah_HLS	.927	.091	.151	10.199	.000	.322	3.101
	Jumlah_PPK	.001	.000	.444	32.538	.000	.380	2.629

Sumber: Data Diolah, 2024

Jika nilai toleransi > 0,0100 dan VIF < 10,00 maka disimpulkan tidak terjadi fenomena *multikolinearitas*, dan jika nilai toleransi < 0,100 dan VIF > 10,00 maka disimpulkan terjadi fenomena *multikolinearitas*. Dengan nilai toleransi didapat sebesar 0.161 dan VIF sebesar 6.220, maka tidak termasuk *multikolinearitas* karena nilai toleransi tidak rendah secara signifikan dan nilai VIF tidak tinggi melebihi batas.

Uji Autokorelasi

Tabel 3. Uji Autokorelasi

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.995[a]	.991	.991	.45811	1.891

Nilai 1.891 tidak termasuk dalam hasil statistik *Durbin-Watson* karena nilainya di luar rentang yang dimungkinkan.

Uji Heteroskedastisitas

Tabel 4. Uji Heteroskedastisitas

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-159.533	41.484		-3.846	.000		
	LN_X1	36.009	9.604	.357	3.749	.000	.756	1.323
	LN_X2	-4.971	2.268	-.373	-2.192	.030	.236	4.244
	LN_X3	4.960	5.265	.136	.942	.348	.330	3.034
	LN_X4	-.013	1.456	-.001	-.009	.993	.408	2.449

Dari hasil di atas nilai probabilitas (*Sig*) untuk koefisien regresi dari variabel bebas tersebut lebih besar dari tingkat signifikansi maka disimpulkan bahwa nilai probabilitas (*Sig*) koefisien regresi variabel independen lebih besar dari tingkat signifikansi $\alpha = 0{,}05$ sehingga disimpulkan tidak terdapat variansi heterogen. Dengan kata lain asumsi heterogenitas terpenuhi.

Model Regresi Linear Berganda

Tabel 5. Model Regresi Linear Berganda SPSS

Model		Unstandardized B
1	(Constant)	25.543
	Jumlah_UHH	.163
	Jumlah_RLS	1.470
	Jumlah_HLS	.927
	Jumlah_PPK	.001

Dapat dilihat dari kolom B merupakan hasil persamaan regresi linear bergandanya, setelah itu dapat diimplementasikan ke dalam model regresi linear berganda yaitu Y = 25,543 + 0,163X1 + 1,470X2 + 0,927X3 + 0,001X4.

Google Collab Python

Penelitian ini juga menggunakan *Google Collab Phython* untuk melakukan pengolah data dalam penelitian ini, sehingga hasil yang diharapkan bisa lebih memiliki tingkat akurasi yang tinggi. Berikut merupakan gambar proses olah data menggunakan *Google Collab Phython:*

```
# Import yang diperlukan
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Import model
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

Gambar 2. ***Import library***

```
df = pd.read_excel('DATA IPM FIX BGT.xlsx')

# Tampilkan data frame
print(df)
```

Gambar 3. Sintak Input Data

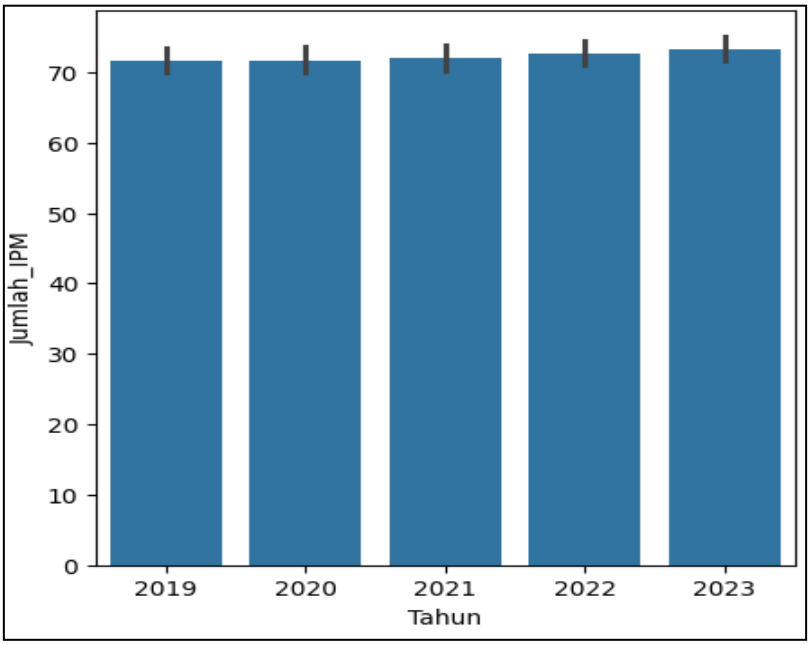
```
      Provinsi  Tahun             Kota  Jumlah_UHH     UHH  Jumlah_RLS  \
0    Jawa Barat   2019            Bogor       71.01  Tinggi        8.29
1    Jawa Barat   2019         Sukabumi       70.73  Tinggi        7.02
2    Jawa Barat   2019          Cianjur       69.91  Sedang        6.97
3    Jawa Barat   2019          Bandung       73.40  Tinggi        8.79
4    Jawa Barat   2019            Garut       71.22  Tinggi        7.51
..          ...    ...              ...         ...     ...         ...
130  Jawa Barat   2023     Kota Bekasi        75.86  Tinggi       11.66
131  Jawa Barat   2023      Kota Depok        75.53  Tinggi       11.58
132  Jawa Barat   2023      Kota Cimahi       75.27  Tinggi       11.39
133  Jawa Barat   2023  Kota Tasikmalaya      75.08  Tinggi        9.54
134  Jawa Barat   2023     Kota Banjar        74.77  Tinggi        8.79

       RLS  Jumlah_HLS     HLS  Pengeluaran     PPK  Jumlah_IPM     IPM
0    Sedang       12.47  Tinggi        10683  Tinggi       70.65  Sedang
1    Sedang       12.22  Tinggi         8973  Sedang       66.87  Sedang
2    Rendah       11.98  Sedang         8290  Sedang       65.38  Sedang
3    Sedang       12.68  Tinggi        10502  Tinggi       72.41  Sedang
4    Sedang       11.82  Sedang         8099  Sedang       66.22  Sedang
..      ...         ...     ...          ...     ...         ...     ...
130  Tinggi       14.12  Tinggi        16479  Tinggi       83.03  Tinggi
131  Tinggi       13.96  Tinggi        16279  Tinggi       82.38  Tinggi
132  Tinggi       13.84  Tinggi        12883  Tinggi       79.46  Sedang
133  Sedang       13.49  Tinggi        11063  Tinggi       74.47  Sedang
134  Sedang       13.27  Tinggi        11356  Tinggi       73.08  Sedang
```

Gambar 4. Hasil *Input* Data

```
plt.figure(figsize=(5,5))
sns.barplot(x=df['Tahun'], y=df['Jumlah_IPM'],data=df)
plt.title('Indeks pembangunan manusia pertahun')
plt.show()
```

Gambar 5. Sintaks visualisasi IPM per Tahun



Gambar 6. Visualisasi IPM per Tahun

Pada gambar di atas ditunjukkan bahwa pada tahun 2019 sampai 2023 adanya peningkatan terhadap indeks pembangunan manusia.

```
df.describe()   #statistika deskriptif
```

	Tahun	Jumlah_UHH	Jumlah_RLS	Jumlah_HLS	Pengeluaran	Jumlah_IPM
count	135.000000	135.000000	135.000000	135.000000	135.000000	135.000000
mean	2021.000000	73.653778	8.654593	12.819037	11029.029630	72.216296
std	1.419481	1.560100	1.444490	0.766472	2290.558568	4.705129
min	2019.000000	69.210000	5.990000	11.690000	7829.000000	65.360000
25%	2020.000000	72.315000	7.520000	12.195000	9470.500000	68.625000
50%	2021.000000	74.370000	8.200000	12.590000	10670.000000	71.200000
75%	2022.000000	74.730000	9.585000	13.450000	11805.000000	74.745000
max	2023.000000	75.860000	11.660000	14.290000	18236.000000	83.040000

Gambar 7. Visualisasi IPM per Tahun

Fungsi *describe* untuk mencari statistika deskripsi seperti banyaknya data, rata-rata, standar deviasi, minimal dan maksimal.

```
#Pilih UHH, RLS, HLS,PPK untuk variabel X dan IPM sebagai variabel y:
feature_names=['Jumlah_UHH', 'Jumlah_RLS', 'Jumlah_HLS', 'Pengeluaran']
X=data[feature_names]
X
y=data.Jumlah_IPM
```

Gambar 8. Variabel X dan Y

Tahap ini menentukan variabel X dan Y. X adalah jumlah IPM dan Y adalah jumlah UHH, jumlah HLS, jumlah RLS dan pengeluaran.

```
#Kemudian memisahkan X dan y ke dalam data latih (train) dan data pengujian (test):
X_train, X_test, y_train, y_test = train_test_split(X,y,random_state=1)

# Linear Regression Model
Linreg=LinearRegression()

#Membuat model dengan data latih
Linreg.fit(X_train,y_train)

#Membuat prediksi pada data pengujian
y_pred=Linreg.predict(X_test)
```

Gambar 9. Train dan Split

Memisahkan data menjadi data latih (*train*) dan data pengujian (*test*) membuat model yang efektif dan menghindari masalah pada data.

```
[ ]  #menambahkan variabel konstan
     X=sm.add_constant(X)
     model=sm.OLS(y,X).fit()
     model.summary()
```

Gambar 10. Sintaks Uji Normalitas

	coef	std err	t	P>|t|	[0.025	0.975]
const	25.5432	2.240	11.405	0.000	21.112	29.974
Jumlah_UHH	0.1630	0.029	5.612	0.000	0.106	0.220
Jumlah_RLS	1.4702	0.059	25.036	0.000	1.354	1.586
Jumlah_HLS	0.9274	0.091	10.199	0.000	0.747	1.107
Pengeluaran	0.0009	2.8e-05	32.538	0.000	0.001	0.001

Omnibus:	5.669	Durbin-Watson:	1.891
Prob(Omnibus):	0.059	Jarque-Bera (JB):	5.546
Skew:	-0.496	Prob(JB):	0.0625
Kurtosis:	3.023	Cond. No.	6.40e+05

Gambar 11. Hasil Uji Normalitas

Dengan uji normalitas yang didapat 0.0625 hasil tersebut mendekati distribusi normal. Meskipun nilai probabilitas untuk Omnibus dan JB sedikit di atas 0.05, indikasi ini menunjukkan bahwa asumsi normal dikarenakan distribusi residu cenderung normal, tetapi dengan sedikit skewness ke kiri.

```
[ ]  #menambahkan variabel konstan
     X=sm.add_constant(X)
     model=sm.OLS(y,X).fit()
     model.summary()
```

Gambar 12. Sintaks Uji Autokorelasi

	coef	std err	t	P>|t|	[0.025	0.975]
const	25.5432	2.240	11.405	0.000	21.112	29.974
Jumlah_UHH	0.1630	0.029	5.612	0.000	0.106	0.220
Jumlah_RLS	1.4702	0.059	25.036	0.000	1.354	1.586
Jumlah_HLS	0.9274	0.091	10.199	0.000	0.747	1.107
Pengeluaran	0.0009	2.8e-05	32.538	0.000	0.001	0.001

Omnibus:	5.669	Durbin-Watson:	1.891
Prob(Omnibus):	0.059	Jarque-Bera (JB):	5.546
Skew:	-0.496	Prob(JB):	0.0625
Kurtosis:	3.023	Cond. No.	6.40e+05

Gambar 13. Hasil Uji Autokorelasi

Dari *output* nilai 1.891 mendekati 2, yang menunjukkan bahwa tidak ada autokorelasi yang signifikan dalam residual model. Ini berarti asumsi bahwa residual model tidak berkorelasi satu sama lain terpenuhi.

```
import statsmodels.formula.api as smf
import matplotlib.pyplot as plt

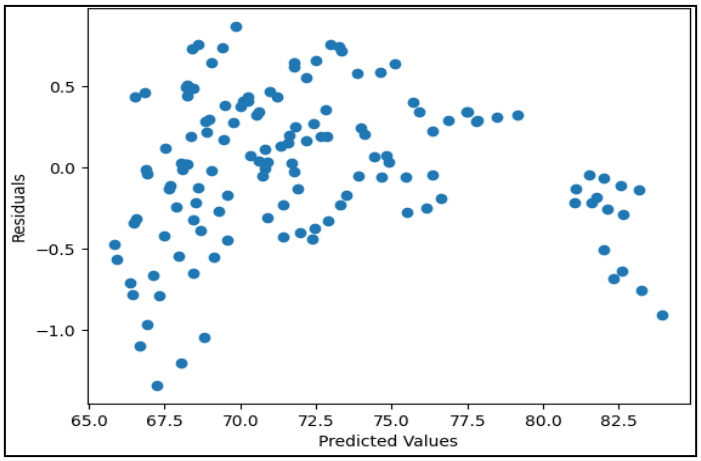
# Membuat model regresi dengan formula yang benar
lm = smf.ols(formula='Jumlah_IPM ~ Jumlah_UHH + Jumlah_RLS + Jumlah_HLS + Pengeluaran', data=data).fit()

# Menampilkan hasil model
print(lm.summary())

# Menghitung residual
resid = lm.resid

# Plot scatter plot antara nilai prediksi dan residual
plt.scatter(lm.predict(), resid)
plt.xlabel('Predicted Values')
plt.ylabel('Residuals')
plt.title('Scatter plot between Predicted Values and Residuals')
plt.show()
```

Gambar 14. Sintaks Uji Heteroskedastisitas



Gambar 15. Plot Uji Heteroskedastisitas

Dari hasil di atas terlihat bahwa titik-titik tersebut tidak mempunyai bentuk yang jelas. Dan titik sebar berada di atas dan di bawah nol pada sumbu Y sehingga dapat disimpulkan tidak terjadi masalah heteroskedastisitas pada model regresi.

```
[ ]  from patsy import dmatrices
     from statsmodels.stats.outliers_influence import variance_inflation_factor
     import statsmodels.formula.api as smf
     import pandas as pd

     # Contoh data
     # Pastikan 'data' adalah DataFrame yang sudah terdefinisi sebelumnya dengan kolom yang sesuai
     # Anda harus menggantikan 'Jumlah_IPM', 'Jumlah_UHH', 'Jumlah_RLS', 'Jumlah_HLS', 'Pengeluaran'
     # dengan nama kolom yang benar sesuai dengan DataFrame Anda.

     # Membuat model regresi dengan formula yang benar
     lm = smf.ols(formula='Jumlah_IPM ~ Jumlah_UHH + Jumlah_RLS + Jumlah_HLS + Pengeluaran', data=data).fit()

     # Membuat matriks desain (design matrices) menggunakan dmatrices
     y, X = dmatrices("Jumlah_IPM ~ Jumlah_UHH + Jumlah_RLS + Jumlah_HLS + Pengeluaran", data=data, return_type="dataframe")

     # Menghitung Variance Inflation Factor (VIF)
     vif = pd.DataFrame()
     vif["Variable"] = X.columns
     vif["VIF"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

     # Menampilkan hasil VIF
     print(vif)
```

Gambar 16. Sintaks Uji Multikolinearitas

	Variable	VIF
0	Intercept	3226.481133
1	Jumlah_UHH	1.311611
2	Jumlah_RLS	4.594175
3	Jumlah_HLS	3.101343
4	Pengeluaran	2.629477

Gambar 17. Hasil Uji Multikolinearitas

Dari *output* diatas didapat UHH dengan VIF 1.311611, RLS dengan VIF 4.594175, HLS dengan VIF 3.101343, Pengeluaran dengan VIF 2.629477. Semua nilai VIF (Jumlah UHH, Jumlah RLS, Jumlah HLS, dan Pengeluaran) lebih kecil dari 10. Maka dapat disimpulkan variabel tersebut tidak termasuk kedalam uji multikolinearitas.

```
#Model Prediksi IPM
from sklearn.model_selection import train_test_split    #misahkan data latih sama data uji
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error         #semakin kecil semakin bagus
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score                    #semakin mendekati 1 semakin bagus

#Pisahkan variabel bebas (X) dan variabel terkait (Y)
x = data.drop(columns='Jumlah_IPM')
y = data[['Jumlah_IPM']]
```

Gambar 18. Input Model Prediksi

```
# 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
```

Gambar 19. Sintaks Random State

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

```
Best random_state: 1632
0.9979314201776063
```

Gambar 20. Nilai Random State

Random state adalah bilangan bulat non-negatif atau *none*, yang digunakan untuk menginisialisasi generator bilangan acak yang digunakan dalam algoritma seperti pembagian data training/testing, inisialisasi model, atau pemilihan nilai awal yang berpengaruh dalam hasil algoritma. Hasil yang didapat adalah 0.99793141776063.

```
[ ]  #Regresi Linear Berganda
     #Bagi data menjadi set pelatihan dan pengujian
     x_train, x_test, y_train, y_test = train_test_split(x,y, test_size = 0.1, random_state = best_random_state)
     model = LinearRegression()
     model.fit(x_train, y_train)

     ▾ LinearRegression
     LinearRegression()
```

Gambar 21. Sintaks *train and test split*

Pada fungsi diatas untuk membagi data menjadi data latih dan data uji dengan *random state*.

```
[ ]  #EVALUASI IPM
     y_pred = model.predict(x_test)
     print(y_pred[:5])
     print(y_test.head())

     #R squared
     print(f'R squared : {r2_score(y_test, y_pred)}')
     #MAE (menghitung rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual)
     print(f'MAE : {mean_absolute_error(y_test, y_pred)}')     #rata" mutlak error     #absolute nya dinyatakan positif
     #MSE (menghitung rata-rata dari selisih kuadrat antara nilai prediksi dan nilai aktual)
     print(f'MSE : {mean_squared_error(y_test, y_pred)}')      #rata" error kuadrat    #kuadrat/pangkat 2 biar positif

     [[68.04265063]
      [77.46690006]
      [71.76687531]
      [74.88907044]
      [82.66442344]]
          Jumlah_IPM
     44        68.06
     51        77.83
     26        71.75
     21        74.92
     131       82.38
     R squared : 0.9979314201776063
     MAE : 0.15317074439048842
     MSE : 0.04031274531748009
```

Gambar 22. Sintaks Evaluasi Model Regresi

Evaluasi model untuk mencari nilai R^2, MAE, dan MSE. Nilai R^2, MAE, dan MSE yang diperoleh adalah 0.997, 0.153 dan 0.040.

```
[ ]  #PERSAMAAN REGRESI
     print('Intercept : ', model.intercept_)
     print('Koefisien : ', model.coef_)

     Intercept :  [25.89732355]
     Koefisien :  [[1.60997490e-01 1.47141595e+00 9.06037904e-01 9.16498165e-04]]
```

Gambar 23. Persamaan Regresi

Output di atas digunakan untuk mengembalikan nilai konstanta dan koefisien persamaan regresi. Dari hasil persamaan yang didapat Y = 1.609 + 1.471 + 9.0603 + 9.164.

```
#MODEL PREDIKSI FAKTOR (VARIABEL BEBAS)
#import data
data1 = pd.read_excel('IPM_Modelfix (1).xlsx')
df1 = data1.copy()

#Menghapus kolom IPM
df1 = df1.drop(columns='IPM')
df1.head()
```

	Tahun	Tahun.1	Jumlah_UHH	Jumlah_RLS	Jumlah_HLS	Pengeluaran
0	2019	2019	72.050741	8.419630	12.763333	10996.07407
1	2020	2020	72.241481	8.540741	12.790000	10717.18519
2	2021	2021	74.380370	8.657778	12.816296	10762.96296
3	2022	2022	74.664815	8.784815	12.840741	11143.11111
4	2023	2023	74.931481	8.870000	12.884815	11525.81481

Gambar 24. Impor Data Variabel Bebas

Pada *input* di atas digunakan untuk memprediksi variabel bebas. Data yang digunakan adalah rata rata dari seluruh variabel bebas di Provinsi Jawa Barat.

```
# Import library
from prophet import Prophet

# Definisikan kolom yang akan diprediksi
columns_to_forecast = ['Jumlah_UHH', 'Jumlah_RLS', 'Jumlah_HLS', 'Pengeluaran']

# Buat DataFrame kosong untuk menampung hasil prediksi
forecasted_df = pd.DataFrame()

# Prediksi setiap kolom dan masukkan ke dalam DataFrame
for column in columns_to_forecast:
  # Membangun model forecasting
  modelp = Prophet(interval_width=0.95)

  # Mengganti nama kolom untuk prediksi
  df_prophet = df1[['Tahun', column]].rename(columns={'Tahun': 'ds', column: 'y'})

  # Melatih model
  modelp.fit(df_prophet)

  # Menentukan tanggal prediksi
  future_dates = pd.date_range(start='2024-01-01', periods=3, freq='y')
  future_df = pd.DataFrame({'ds': future_dates})

  # Prediksi nilai faktor-faktor
  forecast = modelp.predict(future_df)
```

Gambar 25. Sintaks Prediksi Variabel Bebas hingga Tahun 2026

```
# Mengubah kembali nama kolom
forecasted_values.rename(columns={'yhat': column}, inplace=True)

# Memasukkan kolom prediksi ke dalam DataFrame
forecasted_df = pd.concat([forecasted_df, forecasted_values], axis=1)
```

Gambar 26. Library Prophet

Library Prophet adalah sebuah algoritma prediksi yang dikembangkan oleh Meta untuk memprediksi data seri waktu. Memprediksi dari variabel bebas dialam data tersebut.

```
print(forecasted_df)
```

	Jumlah_UHH	Jumlah_RLS	Jumlah_HLS	Pengeluaran
0	75.439345	8.960001	12.579582	11536.250856
1	75.686598	8.970476	12.231350	11165.332954
2	75.932123	8.961766	11.753534	10545.059384

Gambar 27. Hasil Prediksi Variabel Bebas 2024-2026

Hasil *output* diatas dimasukkan ke dalam model regresi untuk memprediksi indeks pembangunan manusia tiga tahun kedepan.

```
# Menghitung penurunan persen dari baris pertama ke baris terakhir
def penurunan_persen(nilai_awal, nilai_akhir):
    return ((nilai_awal - nilai_akhir) / nilai_awal) * 100

penurunan = {}
for col in data.columns:
    penurunan[col] = penurunan_persen(data[col].iloc[0], data[col].iloc[-1])

penurunan_df = pd.DataFrame.from_dict(penurunan, orient='index', columns=['Penurunan Persen'])
penurunan_df
```

	Penurunan Persen
Jumlah_UHH	-0.653211
Jumlah_RLS	-0.019699
Jumlah_HLS	6.566577
Pengeluaran	8.591972

Gambar 28. Sintaks Menghitung Persentase Faktor IPM

Untuk tahun 2024-2025 untuk variabel umur harapan hidup dengan -0.6%, variabel rata lama sekolah -0.01%, variabel harapan lama sekolah 6.5% dan variabel pengeluaran perkapita 8.5%.

```
[ ] #PREDIKSI FAKTOR 3 TAHUN KEDEPAN
    future_pred = model.predict(forecasted_df)
    df1 = pd.DataFrame(future_pred, index=[2024,2025,2026], columns=['IPM'])
    df1
```

	IPM
2024	73.197288
2025	72.597051
2026	71.622365

Gambar 29. Hasil Prediksi Variabel Bebas 2024-2026

Hasil menunjukkan pada tahun 2024 adalah sebesar 73.19, pada tahun 2025 adalah sebesar 72.59, pada tahun 2026 adalah sebesar 71.62.

Untuk dapat mengakses *Google Collab Phython* versi lengkap dari penlitian ini, pembaca dapat mengakses pada link berikut ini:

https://colab.research.google.com/drive/10xwQGgCYtUdOJnKiDRTwK-VUSMgzjW9p#scrollTo=XZp3u39NPF8X

PEMBAHASAN

Indeks Pembangunan Manusia (IPM) merupakan tolak ukur kesejahteraan dan kemakmuran, khususnya masyarakat di Provinsi Jawa Barat. Faktor-faktor seperti Umur Harapan Hidup, Harapan Lama Sekolah, rata-rata rama rekolah, dan Pengeluaran per Kapita yang disesuaikan memiliki peran penting dalam menentukan kualitas hidup masyarakat (Kadri et al., 2020). Penurunan IPM yang signifikan pada tahun 2020 akibat pandemi COVID-19 menunjukkan bagaimana faktor eksternal dapat memengaruhi pembangunan manusia secara drastis. Oleh karena itu, penelitian ini menawarkan pendekatan berbasis data untuk memprediksi tren IPM di masa depan dan memberikan rekomendasi strategis bagi para pemangku kepentingan (Sulistianingsih et al., 2023).

Hasil penelitian ini menunjukkan bahwa variabel Harapan Lama Sekolah (HLS) dan Pengeluaran Perkapita (PPK) memiliki pengaruh yang sangat signifikan terhadap IPM. Penurunan kedua variabel ini dapat mengindikasikan tantangan dalam akses pendidikan yang berkualitas dan ketersediaan sumber daya ekonomi bagi masyarakat (Alwi et al., 2021). Temuan ini memberikan dampak penting, khususnya dalam merancang kebijakan pembangunan. Pemerintah dapat menggunakan informasi ini untuk memprioritaskan investasi di sektor pendidikan dan ekonomi guna memperbaiki indikator yang berdampak langsung pada kesejahteraan masyarakat (Sari et al., 2023).

Selain itu, prediksi nilai IPM yang masuk ke dalam kategori "sedang" pada tahun 2024 hingga 2026 menegaskan perlunya strategi yang lebih terarah untuk meningkatkan kualitas hidup masyarakat Jawa Barat. Misalnya, pengembangan program pendidikan yang lebih inklusif dan upaya peningkatan daya beli masyarakat dapat menjadi solusi untuk mengatasi tantangan tersebut. Penelitian ini juga memiliki dampak jangka panjang dalam memandu

pengalokasian sumber daya secara lebih efisien, sehingga dapat meningkatkan kapasitas pembangunan yang berkelanjutan di masa depan.

Lebih jauh, pendekatan berbasis data yang digunakan dalam penelitian ini, seperti penerapan regresi linier berganda dengan Python, memberikan manfaat praktis bagi institusi pendidikan, pemerintah daerah, dan peneliti lainnya. Dengan metodologi yang dapat direplikasi, penelitian ini dapat menjadi dasar untuk studi serupa di wilayah lain. Dampak positif dari penelitian ini tidak hanya terbatas pada pengambilan kebijakan, tetapi juga berkontribusi pada pengembangan wawasan ilmiah tentang hubungan antara pendidikan, ekonomi, dan pembangunan manusia di Indonesia.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan dari hasil penelitian mengenai implementasi regresi linear berganda untuk mempengaruhi indeks pembangunan manusia di jawa barat maka didapatkan kesimpulan sebagai berikut.

1. Variabel harapan lama sekolah (X3) dan variabel pengeluaran perkapita (X4) sangat berpengaruh untuk variabel indeks pembangunan manusia (Y) karena kedua variabel tersebut menurun. Melihat faktor variabel harapan lama sekolah (X3) dan variabel pengeluaran perkapita (X4) menurun, maka sangat berpengaruh terhadap penurunan nilai IPM Jawa Barat.
2. Dari hasil analisis menunjukkan prediksi untuk tahun 2024 sebesar 73.19, pada tahun 2025 sebesar 72.59, pada tahun 2026 sebesar 71.62 nilai tersebut masuk ke dalam kategori IPM sedang.

Berdasarkan hasil analisis yang dilakukan, saran yang dapat diberikan yaitu sebagai berikut.

1. Penelitian selanjutnya diharapkan dapat menggunakan variabel dan metode yang berbeda untuk memberikan referensi yang lebih baik bagi para peneliti lain yang tertarik pada topik ini.
2. Untuk pemerintah perlu mempersiapkan tindakan terkait kenaikan maupun penurunan secara signifikan dalam prediksi Indeks Pembangunan Manusia (IPM) di tahun yang akan datang.

DAFTAR PUSTAKA

Alwi, W., Irwan, M., & Musfirah, M. (2021). Penerapan Regresi Nonparametrik Spline Dalam Memodelkan Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia (Ipm) Di Indonesia Tahun 2018. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, *9*(2). https://doi.org/10.24252/msa.v9i2.23055

Anastashya, M., Putri, A. E., Pinaringan, G. A., & Hendrawati, T. (2023). Faktor-faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi-Provinsi Indonesia. *Seminar Nasional Statistika Aktuaria II*, *2*(2), 520–530. http://prosiding.snsa.statistics.unpad.ac.id

Aprila, D., Andriani, W., & Ananto, R. P. (2023). Financial Management of Nagari Owned Enterprises (BUMNAG) and Its Impact on Community Welfare. *Jurnal Akuntansi Bisnis*, *16*(2), 210–225. https://doi.org/10.30813/jab.v16 i2.4461

Badan Pusat Statistik, "Indeks Pembangunan Manusia(IPM) Jawa Barat 2019," Badan Pusat Statistik, 2019. Diakses pada 05 Maret 2024, dari https://jabar.bps.go.id/pressrelease/2020/02/17/771/indeks-pembangunan-manusia--ipm--provinsi-jawa-barat-tahun-2019-mencapai-72-03-.html.

Badan Pusat Statistik, "Indeks Pembangunan Manusia(IPM) Jawa Barat 2023," Badan Pusat Statistik, 2023. Diakses pada 05 Maret 2024, dari https://jabar.bps.go.id

Kadri, I. A., Susilawati, M., & Sari, K. (2020). Faktor–Faktor Yang Berpengaruh Signifikan Terhadap Indeks Pembangunan Manusia Di Provinsi Papua. *E-Jurnal Matematika*, *9*(1), 31. https://doi.org/10.24843/mtk.2020.v09.i01.p275

Khaerunisa, S., Nur Padilah, T., & Haerul Jaman, J. (2024). Implementasi Data Mining Menggunakan Metode Regresi Data Panel Untuk Memprediksi Capaian Indeks Pembangunan Manusia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(5), 3399–3406. https://doi.org/10.36040/jati.v7i5.7260

Khotimah, A. K., Rahman, A. A., Alam, M. Z., Nur, Y. H., & Aufi, T. R. (2024). Analisis Regresi Linier Berganda Dalam Estimasi Indeks Pembangunan Manusia di Indonesia Multiple Linear Regression Analysis In Estimating The Human Development Index In Indonesia. *Jurnal Eksponensial*, *15*(November), 90–99. https://doi.org/10.30872/eksponensial.v15i2.1318

Sari, D. T., Khusna, N. I., & Wulandari, F. (2023). Analisis Tingkat Kemiskinan Di Provinsi Jawa Tengah: Suatu Kajian Berdasarkan Faktor Pendidikan, Sosial, Ekonomi, Lokasi Dan Indeks Pembangunan Manusia. *Jurnal PIPSI (Jurnal Pendidikan IPS Indonesia)*, *8*(1), 37. https://doi.org/10.26737/jpipsi.v8i1.3978

Sihite, K., Fatimah, F., Sagala, S. M., Asnidar, A., & Ridha, A. (2024). Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia (IPM) Di Provinsi Jawa Tengah. *Santri: Jurnal Ekonomi Dan Keuangan Islam*, *2 (1)*(1), 6. https://doi.org/10.61132/santri.v2i1.188

Sulistianingsih, E., Suparti, S., & Ispriyanti, D. (2023). Pemodelan Indeks Pembangunan Manusia Di Jawa Tengah Menggunakan Metode Regresi Ridge Dan Regresi Stepwise. *Jurnal Gaussian*, *11*(3), 468–477. https://doi.org/10.14710/j.gauss.11.3.468-477

Syahza, A. (2021). *Metodologi Penelitian (Edisi Revisi Tahun 2021)* (Issue September). Unri Press. https://www.researchgate.net

Ul Hasanah, F. R., & Yollanda, M. (2022). Penerapan Model Regresi Logistik Terhadap Indeks Pembangunan Manusia (IPM) di Provinsi Sumatera Barat Tahun 2019 – 2021. *JOSTECH: Journal of Science and Technology*, *2*(2), 199–208. https://doi.org/10.15548/jostech.v2i2.4383