

Prediksi Tinggi Badan Berdasarkan Berat Badan dengan Regresi Linear Sederhana

Edman Tangjong¹, Erfan Hasmin²

Program Studi Teknik Informatika, Universitas Dipa Makassar, Indonesia

Info Artikel

Riwayat Artikel:

Disusun Jun 19, 2025

Direvisi Jun 19, 2025

Diterima Jun 19, 2025

Kata Kunci:

Berat badan

Prediksi

Regresi linear sederhana

Tinggi badan

ABSTRAK

Penelitian ini bertujuan memprediksi tinggi badan berdasarkan berat badan menggunakan regresi linear sederhana. Dataset yang digunakan telah melalui proses pembersihan dan transformasi. Analisis korelasi menunjukkan hubungan positif antara berat dan tinggi badan dengan koefisien Pearson sebesar 0.48. Model yang dibangun menghasilkan persamaan: Tinggi badan (cm) = $0.44 \times \text{Berat badan (kg)} + 147.62$. Evaluasi dilakukan menggunakan MAE, validasi silang K-Fold, dan standar deviasi. Hasilnya menunjukkan nilai MAE sebesar 3.25 cm, rata-rata MAE 3.28 cm, dan standar deviasi 0.04 cm. Model ini terbukti akurat dan konsisten, serta relevan untuk estimasi tinggi badan dalam konteks kesehatan dan gizi.

This is an open access article under the [CC BY-SA](#) license.



Korespondensi:

Edman Tangjong, Erfan Hasmin

Program Studi Teknik Informatika,

Universitas Dipa Makassar,

Perintis Kemerdekaan No. KM.9, Tamalanrea Indah, Tamalanrea, Makassar, Sulawesi Selatan, Indonesia.

Email: mhs.232007@undipa.ac.id

<https://doi.org/10.00000/joscex.0000.00.00.000>

1. PENDAHULUAN

Prediksi data merupakan elemen penting dalam pengambilan keputusan berbasis data, terutama dalam bidang kesehatan, pendidikan, dan ilmu sosial. Salah satu pendekatan statistik yang banyak digunakan untuk prediksi adalah regresi linear sederhana (Simple Linear Regression), yang memungkinkan pemodelan hubungan linear antara dua variabel kuantitatif. Berbagai penelitian sebelumnya telah menerapkan regresi linear untuk menganalisis hubungan antara berat badan dan tinggi badan. [5] menyatakan bahwa regresi linear sederhana efektif dalam menjelaskan hubungan dua variabel dengan pendekatan matematis yang mudah dipahami. [11] menekankan pentingnya prediksi kuantitatif dalam sistem pengambilan keputusan, khususnya dalam konteks data terstruktur seperti antropometri. [9] menggunakan tinggi dan berat badan sebagai indikator status gizi global, yang menunjukkan relevansi metode regresi dalam kajian kesehatan. [3] memperkenalkan K-Fold Cross Validation sebagai pendekatan yang efektif dalam memastikan kestabilan performa model. Namun, sebagian besar penelitian sebelumnya hanya fokus pada pembangunan model prediktif tanpa evaluasi mendalam terhadap stabilitas dan konsistensinya. Selain itu, visualisasi distribusi data sebelum dan sesudah transformasi masih jarang disajikan sebagai bagian integral dari proses pengembangan model. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memodelkan hubungan antara berat badan dan tinggi badan menggunakan regresi linear sederhana, serta mengevaluasi performa model secara menyeluruh.

menggunakan MAE, validasi silang K-Fold, dan standar deviasi. Penelitian ini juga menyertakan proses visualisasi sebelum dan sesudah pembersihan data untuk menunjukkan pentingnya transformasi dalam meningkatkan kualitas prediksi. Dengan demikian, penelitian ini menghadirkan pendekatan yang tidak hanya fokus pada akurasi model, tetapi juga konsistensi dan reliabilitasnya.

2. METODE

Penelitian ini dilakukan secara daring dengan supervisi dari dosen pembimbing di Universitas Dipa Makassar selama lima bulan, dari September 2024 hingga Februari 2025. Tujuan utama penelitian adalah membangun model prediksi tinggi badan berdasarkan berat badan menggunakan regresi linear sederhana, serta mengevaluasi akurasi melalui pendekatan statistik.

Dataset dan Pengumpulan Data

Tabel 1. Dataset (SOCR-HeightWeight.csv), dataset berat (Pon) dan tinggi badan (Inci) yang bersifat publik

Tinggi badan (Inci)	Berat badan (Pon)
65.78	112.99
71.51	136.48
69.39	153.02
68.21	142.33
.....

Data yang digunakan dalam penelitian ini berasal dari open dataset publik (SOCR-HeightWeight.csv) yang memuat berat badan (dalam pound) dan tinggi badan (dalam inci).

```
import pandas as pd
df=pd.read_csv(r"C:\Users\user\Desktop\SOCR-HeightWeight.csv")
df['tinggi_badan']=df['tinggi_badan']*2.54
df['berat_badan']=df['berat_badan']*0.453592
df.to_csv(r"C:\Users\user\Desktop\uji.csv",index=False)
```

Figur 1. Mengubah satuan Inci dan Pon ke Centimeter dan Kilogram

Dataset tersebut dikonversi ke satuan metrik (Kilogram dan Centimeter) sebelum digunakan. Setelah itu, dilakukan dua tahap pembersihan data:

Tabel 2. Dataset (uji.csv), hasil konversi dari satuan Inci dan Pon ke Centimeter dan Kilogram

Baris	tinggi badan (Centimeter)	berat badan (Kilogram)
1	167.08	51.25
2	181.64	61.9
3	176.27	69.41
4	173.27	64.56
.... 25000	174.94	56.64

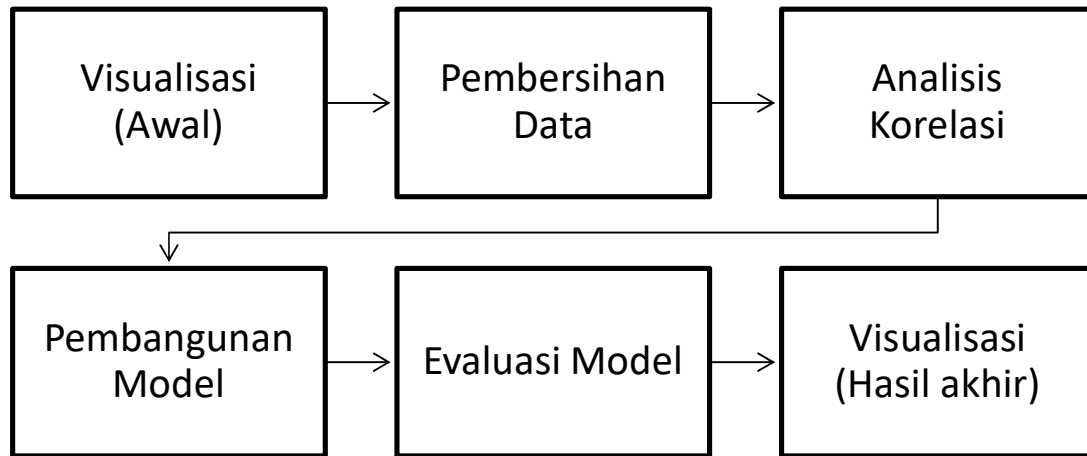
1) Visualisasi dan deteksi outlier awal (uji.csv)

Tabel 3. Dataset (latih.csv), hasil tanpa outlier dari dataset (uji.csv)

Baris	tinggi badan (Centimeter)	berat badan (Kilogram)
1	167.08	51.25
2	181.64	61.9
3	176.27	69.41
4	173.27	64.56
.... 24657	174.94	56.64

2) Pembuangan outlier dan penyimpanan data bersih sebagai latih.csv untuk keperluan pelatihan dan evaluasi model.

Prosedur dan Tahapan Penelitian



Figur 2. Tahapan penelitian

Langkah-langkah dalam penelitian ini dirancang agar dapat direproduksi sebagai berikut:

```

import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:\Users\user\Desktop\uji.csv")
height='tinggi_badan'
weight='berat_badan'
Q1_height=df[height].quantile(0.25)
Q3_height=df[height].quantile(0.75)
IQR_height=Q3_height-Q1_height
Q1_weight=df[weight].quantile(0.25)
Q3_weight=df[weight].quantile(0.75)
IQR_weight=Q3_weight-Q1_weight
plt.figure(figsize=(10,6),dpi=200)
plt.scatter(df[weight],df[height],color='red',alpha=0.6)
plt.xlabel('Berat badan (Kilogram)')
plt.ylabel('Tinggi badan (Centimeter)')
plt.title('Distribusi data tinggi badan berdasarkan berat badan')
plt.grid(False)
plt.axhline(Q1_height,color='blue',linestyle='--',label='Q1 Tinggi badan')
plt.axhline(Q3_height,color='green',linestyle='--',label='Q3 Tinggi badan')
plt.axvline(Q1_weight,color='purple',linestyle='--',label='Q1 Berat badan')
plt.axvline(Q3_weight,color='orange',linestyle='--',label='Q3 Berat badan')
plt.legend()
plt.show()
  
```

Figur 3. Visualisasi (awal) untuk mengetahui sebaran dan outlier pada Dataset (uji.csv)

- 1) Visualisasi Awal: Data divisualisasikan menggunakan pustaka matplotlib untuk mengidentifikasi sebaran awal dan potensi outlier

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:\Users\user\Desktop\uji.csv")
height='tinggi_badan'
weight='berat_badan'
Q1_height=df[height].quantile(0.25)
Q3_height=df[height].quantile(0.75)
IQR_height=Q3_height-Q1_height
Q1_weight=df[weight].quantile(0.25)
Q3_weight=df[weight].quantile(0.75)
IQR_weight=Q3_weight-Q1_weight
lower_bound_height=Q1_height-1.5*IQR_height
upper_bound_height=Q3_height+1.5*IQR_height
lower_bound_weight=Q1_weight-1.5*IQR_weight
upper_bound_weight=Q3_weight+1.5*IQR_weight
df_cleaned=df[
    (df[height]>=lower_bound_height)&(df[height]<=upper_bound_height)&
    (df[weight]>=lower_bound_weight)&(df[weight]<=upper_bound_weight)
]
plt.figure(figsize=(10,6),dpi=200)
plt.scatter(df_cleaned[weight],df_cleaned[height],color='green',alpha=0.6)
plt.xlabel('Berat badan (Kilogram)')
plt.ylabel('Tinggi badan (Centimeter)')
plt.title('Distribusi data tinggi badan berdasarkan berat badan (Tanpa outlier)')
plt.grid(True)
plt.show()
df_cleaned.to_csv(r"C:\Users\user\Desktop\latih.csv",index=False)
```

Figur 4. Menghapus outlier dalam batas interkuartil

- 2) Pembersihan Data: Outlier dihapus berdasarkan batas interkuartil (IQR)

$$IQR = Q3 - Q1$$
 Digunakan untuk deteksi dan penghapusan outlier (nilai ekstrem di luar jangkauan normal).

```
import pandas as pd
from scipy.stats import pearsonr
df=pd.read_csv(r"C:\Users\user\Desktop\latih.csv")
x=df['berat_badan']
y=df['tinggi_badan']
correlation,p_value=pearsonr(x,y)
print(f"Koefisien Korelasi Pearson: {correlation}")
print(f"Nilai-p : {p_value}")
```

Figur 5. Menganalisis keterkaitan terhadap variabel tinggi_badan dan berat_badan

- 3) Analisis Korelasi: Menggunakan koefisien korelasi Pearson untuk mengukur keterkaitan dan arah hubungan linear antara variabel [6].
 $r > 0$: Korelasi positif.
 $r < 0$: Korelasi negative.
 $r = 0$: Tidak ada hubungan lienar.

Nilai antara -1 hingga 1 yang menunjukkan kekuatan dan arah hubungan linear antara dua variabel.

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import linregress
data=pd.read_csv(r"C:\Users\user\Desktop\latih.csv")
x=data['berat_badan']
y=data['tinggi_badan']
slope,intercept,r_value,p_value,std_err=linregress(x,y)
plt.figure(figsize=(10,6),dpi=200)
sns.scatterplot(x=x,y=y,color='green',label='Titik data')
plt.plot(x,intercept+slope*x,color='red',label=f'Garis SLR\Ny= {slope:.2f}x + {intercept:.2f}')
plt.title('Korelasi antara tinggi badan berdasarkan berat badan (Simple linear regression)')
plt.xlabel('Berat badan (Kilogram)')
plt.ylabel('Tinggi badan (Centimeter)')
plt.legend()
plt.show()

```

Figur 6. Membangun model awal regresi linear sederhana

- 4) Pembangunan Model: Regresi linear sederhana diterapkan dengan persamaan $Y = b * X + a$, di mana Y adalah tinggi badan dan X adalah berat badan [2], [3]
- 5) Evaluasi Model:

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:\Users\user\Desktop\latih.csv")
X=df[['berat_badan']]
y=df['tinggi_badan']
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=42)
model=LinearRegression()
model.fit(X_train,y_train)
y_pred=model.predict(X_test)
mae=mean_absolute_error(y_test,y_pred)
print(f'Nilai MAE (Mean Absolute Error): {mae} Centimeter')
print(f'Hasil Simple Linear Regression: Tinggi badan (Centimeter) = {model.coef_[0]:.2f} * Berat badan (Kilogram) + {model.intercept_:.2f}')
plt.figure(figsize=(10,6),dpi=200)
plt.scatter(X_test,y_test,color='lightgreen',label='Titik data')
plt.plot(X_test,y_pred,color='red',linewidth=2,label=f'Tinggi badan (Centimeter) = {model.coef_[0]:.2f} * Berat badan (Kilogram) + {model.intercept_:.2f}')
plt.xlabel('Berat Badan (Kilogram)')
plt.ylabel('Tinggi Badan (Centimeter)')
plt.title('Korelasi antara tinggi badan berdasarkan berat badan (Simple Linear Regression)')
plt.legend()
plt.show()

```

Figur 7. Mengukur nilai MAE (Mean absolute error) pada Dataset (latih.csv)

Mean Absolute Error (MAE): Untuk mengukur rata-rata kesalahan absolut antara nilai prediksi dan aktual [10]

$$MAE = \frac{1}{n} \sum_{i=1}^n |\text{Nilai aktual}_i - \text{Hasil prediksi}_i|$$

Digunakan untuk mengukur tingkat kesalahan prediksi rata-rata dalam satuan sebenarnya (cm).

```

import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_score, KFold
from sklearn.linear_model import LinearRegression
from sklearn.metrics import make_scorer, mean_absolute_error
import matplotlib.pyplot as plt
import numpy as np
df = pd.read_csv(r"C:\Users\user\Desktop\latih.csv")
X = df[['berat_badan']]
y = df['tinggi_badan']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mae = mean_absolute_error(y_test, y_pred)
print(f'Simple Linear Regression: Tinggi badan (Centimeter) = {model.coef_[0]:.2f} * Berat badan (Kilogram) + {model.intercept_:.2f}')
print(f'Nilai MAE (Mean Absolute Error): {mae} Centimeter')
k = 10
kf = KFold(n_splits=k, shuffle=True, random_state=42)
mae_scorer = make_scorer(mean_absolute_error, greater_is_better=False)
neg_mae_scores = cross_val_score(model, X, y, cv=kf, scoring=mae_scorer)
mae_scores = -neg_mae_scores
average_mae = mae_scores.mean()
std_mae = mae_scores.std()
for i, score in enumerate(mae_scores, 1):
    print(f'MAE pada fold ke- {i}: {score:.2f} Centimeter')
print(f'Rata-rata MAE dari {k} fold: {average_mae} Centimeter')
print(f'Standar deviasi MAE dari {k} fold: {std_mae:.2f} Centimeter')
plt.figure(figsize=(10, 6), dpi=200)
plt.scatter(X_test, y_test, color='lightgreen', label='Titik data')
plt.plot(X_test, y_pred, color='red', linewidth=2, label=f'Tinggi badan (Centimeter) = {model.coef_[0]:.2f} * Berat badan (Kilogram) + {model.intercept_:.2f}')
plt.fill_between(X_test.values.flatten(), y_pred - average_mae, y_pred + mae, color='blue', alpha=0.175, label=f'Mean absolute error ± {mae}')
plt.xlabel('Berat Badan (Kilogram)')
plt.ylabel('Tinggi Badan (Centimeter)')
plt.title(f'Korelasi antara rata-rata tinggi badan berdasarkan berat badan')
plt.legend()
plt.show()

```

Figur 8. Uji validitas dan reabilitas terhadap Dataset (latih.csv)

Validasi Silang K-Fold ($K = 10$): Untuk mengukur kestabilan dan reliabilitas model [5], [8]
 Standar Deviasi MAE: Untuk mengevaluasi variasi performa model di seluruh fold

Tabel 4. Pengujian terhadap nilai-nilai Latih-Uji

Latih (Persen)	Uji (Persen)
55%	45%
60%	40%
65%	35%
70%	30%
75%	25%
80%	20%
85%	15%
90%	10%
95%	5%

- 6) Visualisasi Hasil Akhir: Dihasilkan grafik regresi linear akhir dengan daerah kesalahan $\pm MAE$ menggunakan pustaka seaborn dan matplotlib.

Catatan:

X : Variabel independen (bebas), yaitu berat badan dalam satuan Kilogram (kg).
 Y : Variabel dependen (tak bebas), yaitu tinggi badan dalam satuan Centimeter (cm).
 a : intersep (titik potong sumbu- Y), nilai Y saat $X = 0$.
 b : adalah koefisien regresi (kemiringan garis), menunjukkan perubahan rata-rata Y terhadap perubahan X .
 r : Koefisien korelasi pearson.
 IQR : Jangkauan antara kuartil ke-1 dan kuartil ke-3.
 $Q1$: Kuartil ke-1.
 $Q3$: Kuartil ke-3.
 $K = 10$: K-Fold Cross Validation, Teknik validasi model di mana data dibagi menjadi 10 bagian (fold), dan model dilatih serta diuji secara bergilir. Memberikan ukuran kestabilan performa model.
 σ : Mengukur variasi atau penyebaran nilai MAE antar fold pada validasi silang. Semakin kecil nilai standar deviasi, semakin konsisten model.
 $\pm MAE$: Menunjukkan rentang kesalahan prediksi sekitar garis regresi; digunakan dalam visualisasi untuk menampilkan area ketidakpastian atau batas prediksi.

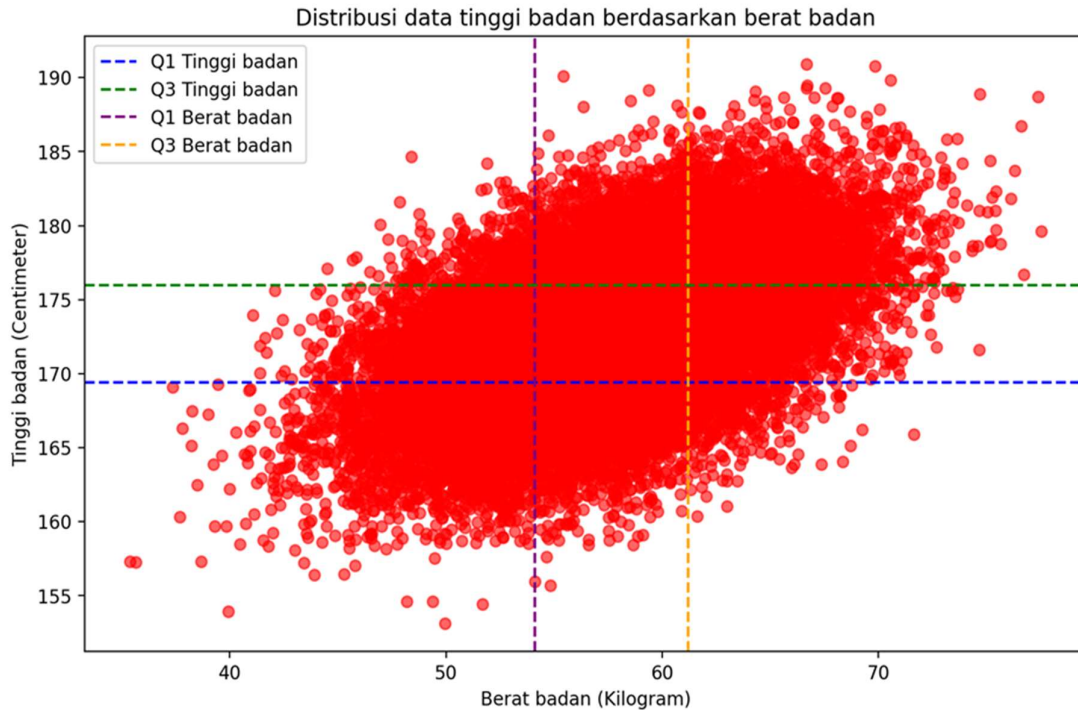
Perangkat dan Alat Analisis

Perangkat Keras: Laptop minimum Intel Core i3, RAM 6GB.

Perangkat Lunak: Python (JupyterLab & VS Code), pustaka pandas, numpy, matplotlib, scikit-learn, dan seaborn.

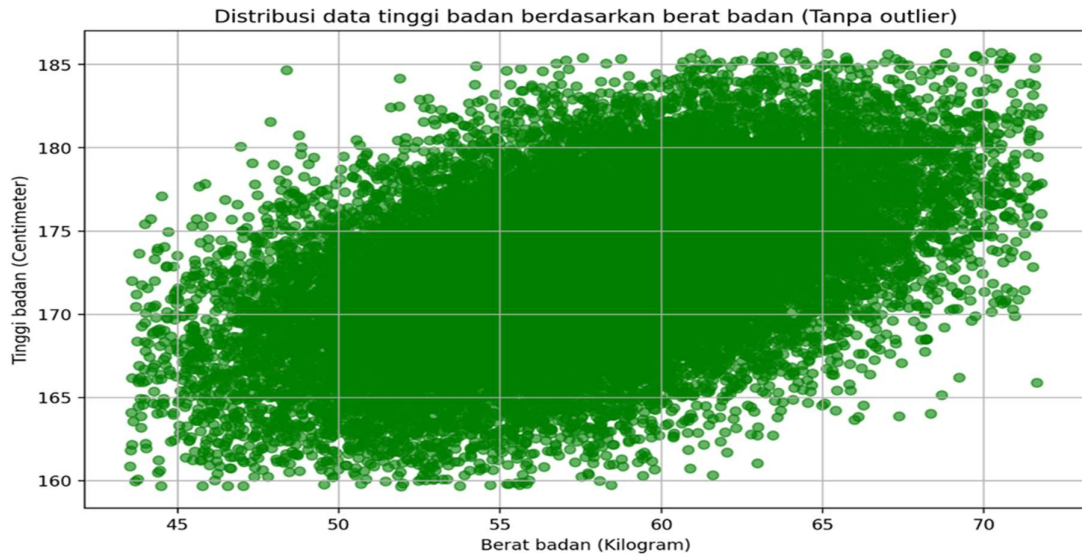
3. DISKUSI DAN HASIL

Penelitian ini dimulai dengan eksplorasi data melalui visualisasi awal untuk memahami sebaran data berat badan dan tinggi badan individu. Berdasarkan visualisasi pada *Gambar 1*, terlihat adanya data pencilon (outlier) yang berpotensi memengaruhi akurasi model.



Gambar 1. Distribusi data tinggi badan berdasarkan berat badan pada dataset (uji.csv)

Setelah proses pembersihan data menggunakan metode batas interkuartil (IQR), data divisualisasikan ulang seperti ditunjukkan pada *Gambar 2*, menunjukkan distribusi yang lebih terpusat dan bebas dari outlier ekstrem.

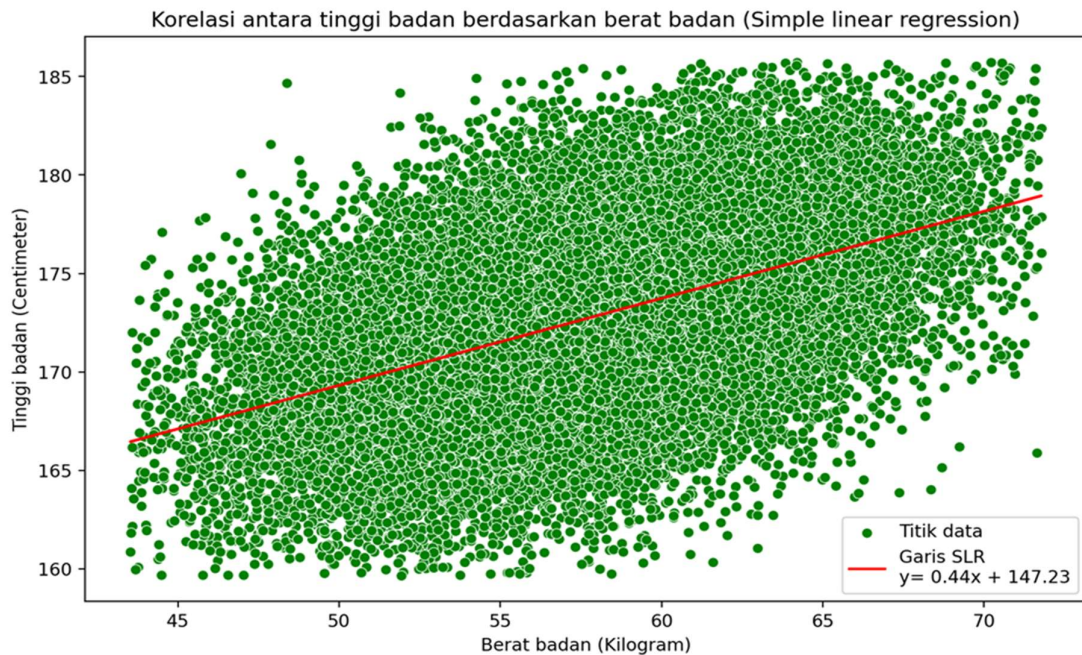


Gambar 2. Distribusi data tinggi badan berdasarkan berat badan (Tanpa outlier) dari dataset (latih.csv)

Koefisien Korelasi Pearson: 0.48205747725983883
Nilai-p : 0.0

Figur 9. Hasil analisis koefisien korelasi Pearson

Analisis korelasi menggunakan koefisien Pearson menghasilkan nilai sebesar $r = 0.48$, yang menunjukkan hubungan positif sedang antara berat dan tinggi badan. Nilai p sebesar 0 mengindikasikan bahwa hubungan tersebut signifikan secara statistik.



Gambar 3. Hasil awal model regresi linear sederhana

Model regresi linear sederhana awal dibangun berdasarkan data bersih dengan nilai Latih-Uji (Default) dan menghasilkan persamaan regresi sebagai berikut:

$$Y = 0.44 * X + 147.23$$

Dimana Y adalah tinggi badan dalam Centimeter dan X adalah berat badan dalam Kilogram.

Evaluasi model awal dilakukan menggunakan MAE (Mean Absolute Error) pada dataset (latih.csv) dengan nilai Latih-Uji (80: 20), dan diperoleh hasil:

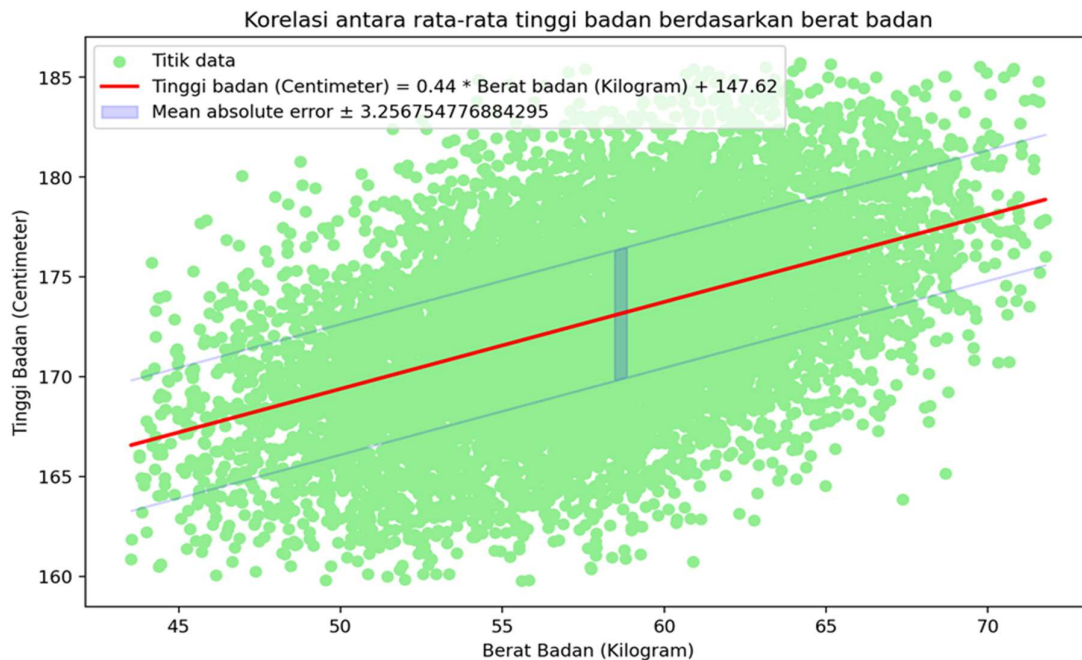
Nilai MAE (Mean Absolute Error): 3.2780453770837963 Centimeter
 Hasil Simple Linear Regression: Tinggi badan (Centimeter) = 0.44 * Berat badan (Kilogram) + 147.48

Figur 10. Hasil evaluasi MAE dengan nilai Latih-Uji (80:20)

Tabel 5. Hasil validasi silang k-fold pada tiap nilai Latih-Uji

Data latih	Data uji	Regresi linear	MAE	Rata-rata MAE 10-fold	Standar deviasi MAE 10-fold
55%	45%	$Y=0.44X+147.43$	3.26 Cm	3.28 Cm	0.04 Cm
60%	40%	$Y=0.44X+147.62$	3.25 Cm	3.28 Cm	0.04 Cm
65%	35%	$Y=0.44X+147.71$	3.27 Cm	3.28 Cm	0.04 Cm
70%	30%	$Y=0.44X+147.56$	3.27 Cm	3.28 Cm	0.04 Cm
75%	25%	$Y=0.44X+147.58$	3.28 Cm	3.28 Cm	0.04 Cm
80%	20%	$Y=0.44X+147.48$	3.27 Cm	3.28 Cm	0.04 Cm
85%	15%	$Y=0.44X+147.42$	3.27 Cm	3.28 Cm	0.04 Cm
90%	10%	$Y=0.44X+147.24$	3.27 Cm	3.28 Cm	0.04 Cm
95%	5%	$Y=0.44X+147.17$	3.31 Cm	3.28 Cm	0.04 Cm

Untuk menguji konsistensi model, digunakan validasi silang K-Fold dengan $K = 10$. Hasil dari tiap fold diperlihatkan pada Tabel 5. Karena hasil validasi pada tiap nilai Latih-Uji bersifat konsisten dan stabil, maka dipilih model regresi linear sederhana dengan nilai MAE (Mean Absolute Error) terkecil dengan nilai Latih-Uji (60: 40).



Gambar 4. Model regresi linear akhir dan batas kesalahan \pm MAE

Visualisasi garis regresi dan batas kesalahan MAE ditunjukkan pada Gambar 4, yang menggambarkan kesesuaian antara data aktual dan hasil prediksi:

$$Y = 0.44 * X + 147.62$$

$$MAE \pm 3.25$$

Dimana Y adalah tinggi badan (Centimeter), X adalah berat badan (Kilogram), dengan nilai error (MAE) hasil prediksi yang kurang atau lebih 3.25 Centimeter.

Secara umum, model mampu memberikan prediksi tinggi badan yang baik. Evaluasi dengan MAE yang rendah serta standar deviasi kecil membuktikan bahwa model ini tidak mengalami overfitting dan dapat diaplikasikan secara praktis.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi tinggi badan berdasarkan berat badan menggunakan pendekatan regresi linear sederhana. Seperti yang diharapkan pada *Bagian 1*, penelitian ini mampu menunjukkan adanya hubungan linear positif antara kedua variabel, dengan koefisien korelasi Pearson sebesar 0.48 dan nilai $p = 0$ yang signifikan secara statistik.

Model regresi yang dikembangkan menghasilkan persamaan:

$$\text{Tinggi badan (Centimeter)} = 0.44 * \text{Berat badan (Kilogram)} + 147.62$$

dan menunjukkan performa yang baik, dengan nilai MAE sebesar 3.25 Centimeter dan standar deviasi MAE sebesar 0.04 Centimeter berdasarkan validasi silang 10-fold.

Evaluasi menyeluruh menunjukkan bahwa model ini lumayan baik juga stabil dan konsisten, serta bebas dari overfitting. Temuan ini membuktikan bahwa regresi linear sederhana dapat digunakan secara efektif dalam memprediksi tinggi badan dari data berat badan, khususnya dalam konteks data antropometri.

Model ini dapat dikembangkan lebih lanjut dengan menambahkan variabel tambahan seperti usia, jenis kelamin, dan indeks massa tubuh (BMI) untuk meningkatkan akurasi. Selain itu, penerapan model ini dalam bentuk aplikasi berpotensi mendukung pemeriksaan awal status gizi, terutama di bidang kesehatan anak dan remaja.

REFERENSI

- [1] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.
- [2] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, New York, NY, USA: Springer, 2013.
- [3] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, New York, NY, USA: Springer, 2013.
- [4] D. S. Moore, G. P. McCabe, and B. A. Craig, *Introduction to the Practice of Statistics*, New York, NY, USA: W. H. Freeman, 2012.
- [5] D. S. Moore, G. P. McCabe, and B. A. Craig, *Introduction to the Practice of Statistics*, 9th ed., New York, NY, USA: W. H. Freeman and Company, 2017.
- [6] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, 2012.
- [7] R. A. Puspasari and T. Hidayat, "Evaluasi model prediksi berat badan anak dengan regresi linier dan MAE," *J. Informatika*, 2020.
- [8] L. Rahmawati, S. E. Nugroho, and D. Susanti, "K-Fold Cross Validation pada regresi linier sederhana," *J. Teknol. dan Sist. Komput.*, 2021.
- [9] World Health Organization, *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and Development*, Geneva, Switzerland: WHO Press, 2006.
- [10] C. J. Willmott and K. Matsuura, "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005, doi: 10.3354/cr030079.
- [11] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Burlington, MA, USA: Morgan Kaufmann, 2016.
- [12] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 5th ed., Mason, OH, USA: Cengage Learning, 2013.