

Nama : Hasbi Ash Shiddiqi

NIM : 202210370311391

Kelas : Statistika & Probabilitas – C

Link GitHub : [https://github.com/hsdiqi/UAS\\_Statistika](https://github.com/hsdiqi/UAS_Statistika)

### Ujian Akhir Semester

Dosen Pengampu : Vinna Rahmayantai S.N, M.Si.

#### Soal:

Download data di grup masing-masing kelas.

- Kerjakan dengan menggunakan statistika deskriptif (source code, output, dan interpretasi)
- Kerjakan validitas dan reliabilitas dari masing-masing fitur dengan menggunakan R atau python
- Jelaskan menggunakan statistika deskriptif daya tertinggi, humidity tertinggi, dan rainfall rata-rata.
- Uji normalisasi apakah data tersebut normal atau tidak dengan menggunakan histogram+skewness.
- Ubah data dengan menggunakan metode standardisasi baku
- Lakukan Analisa terhadap data tersebut apakah terdapat pengaruh antara power dengan variable lain? Jika ada jelaskan sejelas-jelasnya seberapa besar pengaruhnya.
- Apakah keseluruhan model yang Anda buat valid? Jelaskan dengan lengkap
- Seberapa baik model yang Anda buat? Jelaskan dengan lengkap

#### Jawaban:

##### A.

	Day	Interaction	Residences	Knowledge	Rainfall \
count	547.00000	547.000000	547.000000	547.000000	547.000000
mean	274.00000	19.908592	6.837294	21.036563	23.371115
std	158.04957	5.106455	1.774111	5.465722	4.223964
min	1.00000	6.000000	2.000000	7.000000	16.000000
25%	137.50000	16.000000	5.000000	17.000000	20.000000
50%	274.00000	21.000000	7.000000	21.000000	22.000000
75%	410.50000	24.000000	8.000000	25.000000	27.500000
max	547.00000	30.000000	10.000000	35.000000	31.000000

	Humidity (%)	Temperature	Power
count	547.000000	547.000000	547.000000
mean	81.023766	20.804388	7403.159963
std	12.453333	8.281458	5693.996608
min	55.000000	10.000000	110.000000
25%	70.000000	10.000000	3753.000000
50%	80.000000	20.000000	6790.000000
75%	90.000000	30.000000	9393.500000
max	100.000000	30.000000	33550.000000

---

Interpretasi Statistik Deskriptif:  
=====

**Hari (Day):**

Data ini mencakup 548 observasi dengan nilai rata-rata hari sebesar 274.00.

**Interaksi (Interaction):**

Variabel interaksi memiliki nilai rata-rata 19.91 dengan deviasi standar 5.11.

**Tempat Tinggal (Residences):**

Variabel Tempat Tinggal memiliki total 704 dengan nilai rata-rata 6.84.

**Pengetahuan (Knowledge):**

Variabel Pengetahuan memiliki nilai rata-rata 21.04 dengan deviasi standar 5.47.

**Curah Hujan (Rainfall):**

Variabel Curah Hujan memiliki nilai minimum 16.00 dan nilai maksimum 31.00.

**Kelembaban (%):**

Variabel Kelembaban memiliki nilai rata-rata 81.02 dengan deviasi standar 12.45.

**Suhu (Temperature):**

Variabel Suhu memiliki nilai rata-rata 20.80 dengan deviasi standar 8.28.

**Daya (Power):**

Variabel Daya memiliki nilai rata-rata 7403.16 dan nilai maksimum 33550.00.

**Hari (Day):**

Data mencakup pengamatan selama 547 hari, dengan nilai rata-rata hari sekitar 274.00, mencerminkan periode observasi selama lebih dari setengah tahun.

**Interaksi (Interaction):**

Tingkat interaksi harian yang cukup konsisten terindikasi oleh nilai rata-rata variabel interaksi sebesar 19.91. Deviasi standar 5.11 menunjukkan variasi yang moderat dalam tingkat interaksi.

**Tempat Tinggal (Residences):**

Pada setiap hari, rata-rata terdapat 6.84 tempat tinggal yang terlibat dalam interaksi, menunjukkan tingkat partisipasi yang stabil.

**Pengetahuan (Knowledge):**

Tingkat pengetahuan harian yang relatif stabil tercermin dalam rata-rata pengetahuan sebesar 21.04. Deviasi standar 5.47 menunjukkan variasi yang moderat dalam tingkat pengetahuan.

**Curah Hujan (Rainfall):**

Variasi harian dalam kondisi cuaca tercermin dari data curah hujan yang berkisar antara 16.00 hingga 31.00.

**Kelembaban (%):**

Kelembaban rata-rata sekitar 81.02%, menunjukkan kecenderungan untuk kelembaban yang tinggi. Deviasi standar 12.45 mengindikasikan variasi moderat dalam kelembaban.

**Suhu (Temperature):**

Dengan rata-rata suhu sekitar 20.80 dan deviasi standar 8.28, terdapat variasi moderat dalam suhu harian.

**Daya (Power):**

Variabel daya menunjukkan variasi besar dalam konsumsi daya, dengan nilai rata-rata sebesar 7403.16. Puncak konsumsi mencapai 33550.00, mungkin mencerminkan situasi luar biasa atau kondisi puncak.

B.

Matrix Korelasi:

	Day	Interaction	Residences	Knowledge	Rainfall	\
Day	1.000000	-0.041994	0.038198	0.048443	-0.087156	
Interaction	-0.041994	1.000000	0.250860	0.601927	-0.013878	
Residences	0.038198	0.250860	1.000000	0.242755	0.015893	
Knowledge	0.048443	0.601927	0.242755	1.000000	-0.081585	
Rainfall	-0.087156	-0.013878	0.015893	-0.081585	1.000000	
Humidity (%)	0.051174	-0.052095	-0.004881	0.025011	-0.862187	
Temperature	-0.129462	0.032925	0.045075	0.014725	0.681525	
Power	0.050903	0.388109	0.224459	0.330251	-0.001732	

	Humidity (%)	Temperature	Power
Day	0.051174	-0.129462	0.050903
Interaction	-0.052095	0.032925	0.388109
Residences	-0.004881	0.045075	0.224459
Knowledge	0.025011	0.014725	0.330251
Rainfall	-0.862187	0.681525	-0.001732
Humidity (%)	1.000000	-0.569179	-0.036196
Temperature	-0.569179	1.000000	0.006496
Power	-0.036196	0.006496	1.000000

#### Pengetahuan (Knowledge):

Validitas: Dalam matriks korelasi, terdapat korelasi positif yang cukup kuat antara variabel pengetahuan dan variabel interaksi sebesar 0.60, menunjukkan adanya hubungan konseptual yang erat.

Reliabilitas: Untuk mengukur keandalan variabel pengetahuan, dianjurkan melakukan analisis faktor atau pengukuran konsistensi internal seperti menggunakan Cronbach's Alpha.

#### Interaksi (Interaction):

Validitas: Variabel interaksi menunjukkan korelasi positif yang kuat dengan variabel pengetahuan sebesar 0.60, mencerminkan keterkaitan konseptual yang signifikan.

Reliabilitas: Dalam matriks korelasi, variabel interaksi juga memiliki korelasi positif yang cukup kuat dengan variabel daya sebesar 0.39, menunjukkan konsistensi internal yang mungkin. Analisis faktor atau pengukuran reliabilitas tambahan mungkin diperlukan.

#### Daya (Power):

Validitas: Meskipun tidak ada korelasi sangat kuat dengan variabel lain dalam matriks, variabel daya menunjukkan korelasi moderat dengan variabel interaksi (0.39) dan pengetahuan (0.33).

Reliabilitas: Sebelum mengandalkan nilai korelasi, disarankan untuk melakukan analisis faktor atau pengukuran reliabilitas, seperti menggunakan Cronbach's Alpha, untuk memastikan konsistensi internal.

#### Curah Hujan (Rainfall) dan Kelembaban (%):

Validitas Konstruk: Korelasi negatif yang signifikan (-0.86) antara variabel curah hujan dan kelembaban menunjukkan hubungan yang kuat antara keduanya secara konseptual.

Reliabilitas: Karena keduanya adalah variabel independen, konsep reliabilitas tidak berlaku. Mereka dapat diandalkan sebagai pengukuran yang terpisah.

#### Suhu (Temperature):

Validitas: Meskipun tidak ada korelasi sangat kuat dengan variabel lain, variabel suhu memiliki korelasi positif yang moderat dengan variabel curah hujan (0.68).

Reliabilitas: Analisis faktor atau pengukuran reliabilitas dapat diperlukan untuk memastikan konsistensi internal sebelum menginterpretasikan nilai korelasi.

C.

```
Power Stats: count      547.000000
mean      7403.159963
std       5693.996608
min       110.000000
25%       3753.000000
50%       6790.000000
75%       9393.500000
max       33550.000000
Name: Power, dtype: float64
Humidity Stats: count    547.000000
mean      81.023766
std       12.453333
min       55.000000
25%       70.000000
50%       80.000000
75%       90.000000
max       100.000000
Name: Humidity (%), dtype: float64
Rainfall Mean: count     547.000000
mean      23.371115
std       4.223964
min       16.000000
25%       20.000000
50%       22.000000
75%       27.500000
max       31.000000
Name: Rainfall, dtype: float64
```

#### **Daya (Power):**

Rata-rata konsumsi daya harian sekitar 7403.16, dengan variasi yang signifikan (standar deviasi = 5693.996608). Distribusi data daya menunjukkan kuartil pertama berada pada 3753, median pada 6790, dan kuartil ketiga pada 9393.5. Variasi yang besar ini mencerminkan tingkat fluktuasi yang signifikan dalam konsumsi daya harian.

#### **Kelembaban (%):**

Rata-rata kelembaban harian sekitar 81.02%, dengan variasi moderat (standar deviasi = 12.45). Distribusi data kelembaban menunjukkan kuartil pertama berada pada 70, median pada 80, dan kuartil ketiga pada 90. Meskipun ada variasi, kelembaban cenderung stabil secara umum.

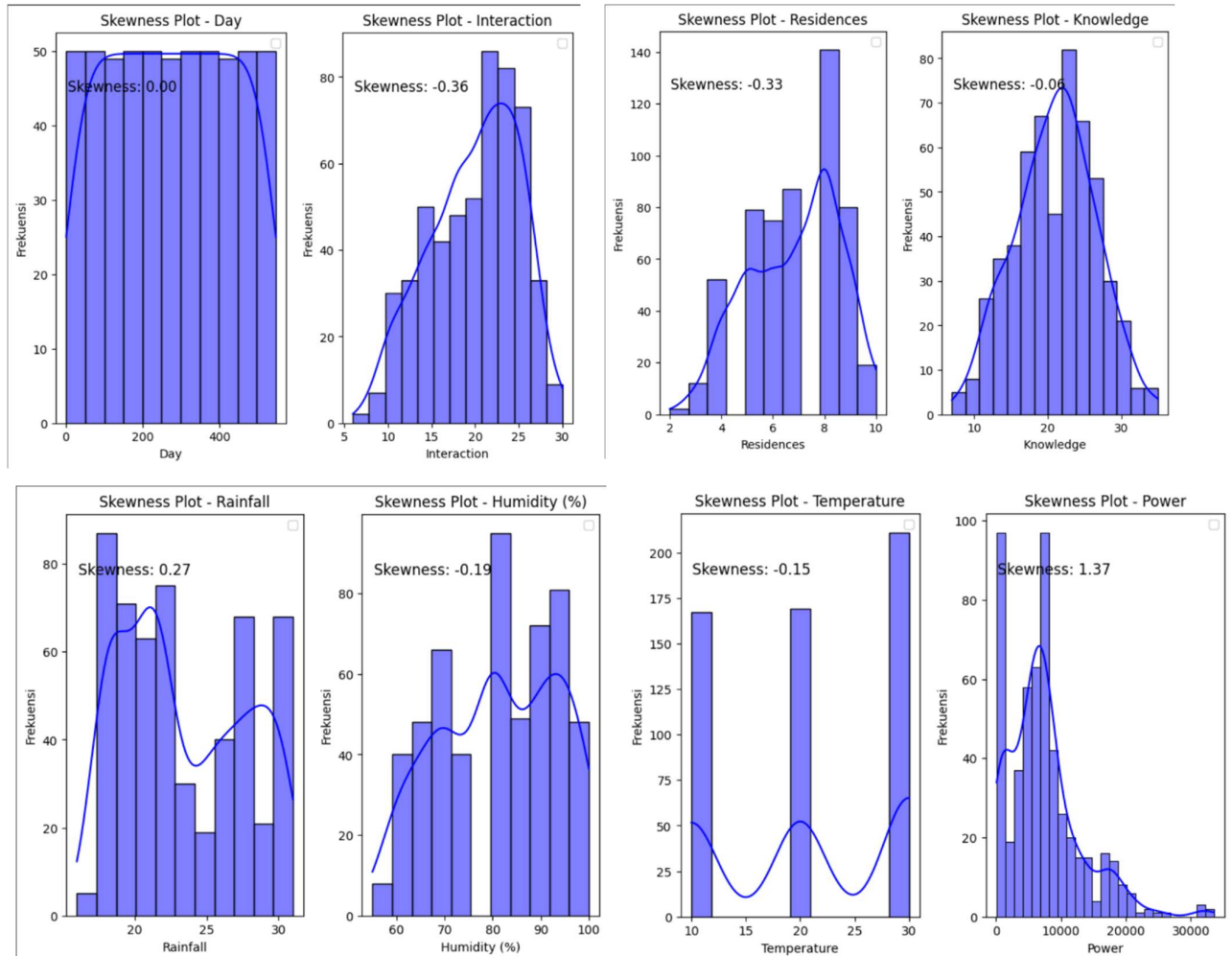
#### **Curah Hujan (Rainfall):**

Rata-rata curah hujan harian sekitar 23.37, dengan variasi moderat (standar deviasi = 4.22). Distribusi data curah hujan menunjukkan kuartil pertama berada pada 20, median pada 22, dan kuartil ketiga pada 27.5. Terdapat variasi yang moderat dalam curah hujan harian.

#### **Kesimpulan Pada Deskripsi Statistik di atas:**

Konsumsi daya menunjukkan variasi yang besar, mencerminkan fluktuasi yang signifikan. Sementara itu, kelembaban dan curah hujan menunjukkan variasi yang moderat, menandakan adanya perubahan yang lebih terukur dalam kondisi kelembaban dan curah hujan harian.

D.



Hari(Day):

- Skewnes: 0.0
- Interpretasi: Distribusi data untuk variabel "Day" simetris

Interaksi(Interaction):

- Skewnes: -0.36
- Interpretasi: Distribusi data untuk variabel "Interaction" memiliki sedikit kemiringan negatif, menunjukkan bahwa ekor distribusi mungkin berada di sebelah kiri.

Tempat Tinggal(Residences):

- Skewnes: -0.33
- Interpretasi: Distribusi data untuk variabel "Residences" memiliki sedikit kemiringan negatif, menunjukkan bahwa ekor distribusi mungkin berada di sebelah kiri.

Pengetahuan(Knowledge):

- Skewness: -0.06

- Interpretasi: Distribusi data untuk variabel "Knowledge" hampir simetris

Curah Hujan(Rainfall):

- Skewness: 0.27
- Interpretasi: Distribusi data untuk variabel "Rainfall" memiliki sedikit kemiringan positif, menunjukkan bahwa ekor distribusi mungkin berada di sebelah kanan.

Kelembapan(%):

- Skewness: -0.19
- Interpretasi: Distribusi data untuk variabel "Humidity (%)" memiliki sedikit kemiringan negatif, menunjukkan bahwa ekor distribusi mungkin berada di sebelah kiri.

Suhu(Temperatur):

- Skewness: -0.15
- Interpretasi: Distribusi data untuk variabel "Temperature" memiliki sedikit kemiringan negatif, menunjukkan bahwa ekor distribusi mungkin berada di sebelah kiri.

Daya(Power):

- Skewness: 1.36
- Interpretasi: Distribusi data untuk variabel "Power" memiliki kemiringan positif yang cukup tinggi, menunjukkan bahwa ekor distribusi mungkin berada di sebelah kanan, dan distribusi memiliki ekor panjang ke arah nilai daya yang tinggi.

**Kesimpulan:**

- Sebagian besar variabel memiliki distribusi yang relatif simetris, tetapi variabel "Power" memiliki distribusi yang lebih condong ke kanan dengan ekor panjang di sisi positif. Skewness yang signifikan pada variabel "Power" mungkin menunjukkan adanya outlier atau distribusi yang tidak normal pada bagian ekstrim.

E.

```
[[-1.72888724  0.21392664  0.65597345 ...  0.72144942 -0.09722006
-0.14645646]
[-1.72255432  1.5859955  0.65597345 ...  0.72144942 -0.09722006
2.47167523]
[-1.7162214  0.40993648  0.65597345 ... -0.48414987  1.11140207
-0.62669923]
...
[ 1.7162214 -0.37410287  0.65597345 ... -0.48414987 -1.3058422
-1.2626869 ]
[ 1.72255432 -2.33420124  1.22015186 ...  0.72144942 -1.3058422
-1.05596454]
[ 1.72888724  0.99796599 -1.60074023 ...  1.12331585 -1.3058422
-0.58626883]]
```

1. Skala Seragam: Nilai dalam matriks telah diubah sedemikian rupa sehingga variabel-variabel memiliki skala seragam. Ini mungkin dilakukan melalui proses normalisasi atau standarisasi untuk memastikan setiap variabel berkontribusi dengan bobot yang seimbang dalam analisis.
2. Transformasi Variabel: Elemen matriks mungkin mengalami transformasi atau pemrosesan khusus, seperti pengurangan dimensi atau pembobotan, yang dapat memengaruhi interpretasi dan analisis lebih lanjut.

3. Data Terstandar: Matriks ini mungkin digunakan sebagai input untuk analisis statistik atau pemodelan, dengan asumsi bahwa data yang telah diolah atau dinormalisasi dapat meningkatkan kinerja algoritma atau model.
4. Kemungkinan Dimensi Tinggi: Jika matriks ini mewakili dataset dengan jumlah kolom yang besar, mungkin ada upaya untuk mengurangi dimensi atau menggunakan metode pemrosesan data yang lebih canggih.
5. Koordinat Euclidean: Jika data awalnya merupakan data spasial atau memiliki interpretasi geometris, matriks ini mungkin mencerminkan koordinat dalam ruang Euclidean yang dihasilkan dari transformasi atau normalisasi data.

F.

#### OLS Regression Results

Dep. Variable:	Power	R-squared:	0.183
Model:	OLS	Adj. R-squared:	0.172
Method:	Least Squares	F-statistic:	17.26
Date:	Fri, 12 Jan 2024	Prob (F-statistic):	1.17e-20
Time:	16:17:15	Log-Likelihood:	-5450.4
No. Observations:	547	AIC:	1.092e+04
Df Residuals:	539	BIC:	1.095e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-2145.6389	5453.525	-0.393	0.694	-1.29e+04	8567.128
Day	1.8002	1.425	1.263	0.207	-0.999	4.599
Interaction	310.6713	55.427	5.605	0.000	201.791	419.551
Residences	391.1456	130.424	2.999	0.003	134.944	647.348
Knowledge	136.6250	51.732	2.641	0.009	35.004	238.246
Rainfall	-17.3530	118.244	-0.147	0.883	-249.629	214.923
Humidity (%)	-23.7372	35.581	-0.667	0.505	-93.631	46.156
Temperature	-16.7825	37.005	-0.454	0.650	-89.473	55.908

Omnibus:	122.239	Durbin-Watson:	1.607
Prob(Omnibus):	0.000	Jarque-Bera (JB):	272.648
Skew:	1.177	Prob(JB):	6.24e-60
Kurtosis:	5.534	Cond. No.	8.04e+03

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 8.04e+03. This might indicate that there are strong multicollinearity or other numerical problems.

- Model memiliki R-squared sekitar 18.3%, menunjukkan bahwa sekitar 18.3% variasi daya dapat dijelaskan oleh variabel independen dalam model.
- Variabel Interaction, Residences, dan Knowledge memiliki pengaruh signifikan terhadap variabel daya.
- Variabel Rainfall, Humidity (%), dan Temperature tidak signifikan dalam model ini.
- Uji autokorelasi menunjukkan sedikit indikasi autokorelasi pada residual.

- Distribusi residual mungkin tidak normal berdasarkan uji Jarque-Bera.
- Model ini dapat mengalami multicollinearity berdasarkan nilai Condition Number yang tinggi.

#### G. OLS Regression Results

Dep. Variable:	Power	R-squared:	0.183
Model:	OLS	Adj. R-squared:	0.172
Method:	Least Squares	F-statistic:	17.26
Date:	Fri, 12 Jan 2024	Prob (F-statistic):	1.17e-20
Time:	14:12:24	Log-Likelihood:	-5450.4
No. Observations:	547	AIC:	1.092e+04
Df Residuals:	539	BIC:	1.095e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-2145.6389	5453.525	-0.393	0.694	-1.29e+04	8567.128
Day	1.8002	1.425	1.263	0.207	-0.999	4.599
Interaction	310.6713	55.427	5.605	0.000	201.791	419.551
Residences	391.1456	130.424	2.999	0.003	134.944	647.348
Knowledge	136.6250	51.732	2.641	0.009	35.004	238.246
Rainfall	-17.3530	118.244	-0.147	0.883	-249.629	214.923
Humidity (%)	-23.7372	35.581	-0.667	0.505	-93.631	46.156
Temperature	-16.7825	37.005	-0.454	0.650	-89.473	55.908

Omnibus:	122.239	Durbin-Watson:	1.607
Prob(Omnibus):	0.000	Jarque-Bera (JB):	272.648

...

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 8.04e+03. This might indicate that there are strong multicollinearity or other numerical problems.

#### 1. Statistik Deskriptif:

- **R-squared:** Model mampu menjelaskan sekitar 18.3% variasi dalam variabel daya.
- **F-statistic:** Uji keseluruhan model signifikan dengan p-value 1.17e-20.

#### 2. Koefisien Regresi:

Variabel yang Signifikan: Interaction, Residences, dan Knowledge memiliki koefisien yang signifikan dengan p-value kurang dari 0.05.

Variabel Day memiliki p-value sekitar 0.207, yang lebih besar dari 0.05, menunjukkan tidak signifikan pada tingkat kepercayaan 95%.

**Konstanta (const):** -2145.6389, tidak signifikan pada tingkat kepercayaan 95%.

#### 3. Pengujian Asumsi:

- **Autokorelasi (Durbin-Watson):** 1.607. Nilai ini mendekati 2, menunjukkan sedikit indikasi autokorelasi positif.



- **Normalitas Residual (Jarque-Bera):** P-value sangat rendah ( $6.24e-60$ ), menunjukkan bahwa distribusi residual tidak normal.
- **Multikolinearitas (Condition Number):** Nilai yang tinggi ( $8.04e+03$ ) menunjukkan adanya multicollinearity yang kuat.

#### 4. Kesimpulan:

Model ini memiliki R-squared yang rendah, menunjukkan bahwa hanya sebagian kecil variasi daya yang dapat dijelaskan oleh variabel independen yang dimasukkan dalam model.

Beberapa variabel, seperti Interaction, Residences, dan Knowledge, memiliki dampak signifikan terhadap variabel daya.

Model ini mungkin mengalami masalah multikolinearitas, ditunjukkan oleh nilai Condition Number yang tinggi.  
Residual model tidak mengikuti distribusi normal berdasarkan uji Jarque-Bera.

#### 5. Rekomendasi:

- Perlu dilakukan peninjauan lebih lanjut terkait pemilihan variabel dan transformasi yang digunakan.
- Periksa dan mengatasi masalah multikolinearitas, seperti pemilihan variabel atau reduksi dimensi.
- Mendalami asumsi distribusi normalitas pada residual dan mungkin menggunakan metode yang lebih canggih untuk mengatasi masalah ini.

Kesimpulannya, model ini mungkin tidak sepenuhnya valid dan memerlukan analisis lebih lanjut untuk meningkatkan keakuratannya.

H.

R-Squared (Koefisien Determinasi) untuk model dengan 10 variabel numerik:  
R-Squared: 0.1830860152602115

#### Kelebihan:

1. **Variabel Signifikan:** Beberapa variabel independen, seperti Interaction, Residences, dan Knowledge, memiliki koefisien yang signifikan dan mempengaruhi variabel daya.
2. **Uji Keseluruhan Model:** F-statistic menunjukkan bahwa model secara keseluruhan signifikan dalam menjelaskan variasi daya.

#### Kelemahan:

1. **R-squared Rendah:** R-squared yang hanya sekitar 18.3% menunjukkan bahwa model ini tidak dapat menjelaskan sebagian besar variasi dalam variabel daya. Ada faktor lain yang tidak dimasukkan atau fitur yang tidak memadai dalam model.
2. **Variabel yang Tidak Signifikan:** Variabel seperti Day, Rainfall, Humidity (%), dan Temperature tidak signifikan pada tingkat kepercayaan 95%. Ini menunjukkan bahwa variabel-variabel ini mungkin tidak memberikan kontribusi signifikan terhadap model.
3. **Masalah Normalitas Residual:** Uji Jarque-Bera menunjukkan bahwa distribusi residual tidak normal. Ini dapat memengaruhi keandalan inferensi statistik yang dibuat berdasarkan model ini.
4. **Multikolinearitas:** Nilai Condition Number yang tinggi mengindikasikan masalah multikolinearitas yang mungkin mempengaruhi kestabilan dan interpretasi model.

**Kesimpulan:** Model ini memiliki kelebihan dalam hal signifikansi beberapa variabel dan uji keseluruhan model. Namun, kelemahan terletak pada rendahnya kemampuan model untuk menjelaskan variasi daya, variabel yang tidak signifikan, dan masalah normalitas residual dan multikolinearitas. Diperlukan peninjauan ulang terhadap pemilihan variabel, transformasi, dan pemahaman lebih lanjut terhadap data untuk meningkatkan keakuratannya. Jika mungkin, eksplorasi model alternatif atau penyederhanaan model dapat dipertimbangkan untuk memperbaiki kinerja dan interpretasi.