

# Tugas 3: - Tugas Praktikum Mandiri

## ROHMATUL HIDAYAT - 0110224015

Teknik Informatika, STT Terpadu Nurul Fikri, Depok

\*E-mail: 0110224015@student.nurulfikri.ac.id -

### 1. Menghubungkan Google Colab ke Google Drive

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

Kode	Penjelasan
from google.colab import drive	Mengimpor modul khusus Colab untuk akses Drive
drive.mount('/content/drive')	<ul style="list-style-type: none"><li>Memasang Google Drive ke path /content/drive di Colab</li><li>Akan minta izin akses ke akun Google (hanya pertama kali)</li></ul>

### 2. Load Dataset

```
import pandas as pd
file_path= '/content/drive/MyDrive/ML3/praktikum/TUGAS PRAKTIKUM/tugas praktikum 3/Data/day.csv'
df = pd.read_csv(file_path)
df.head()
```

Kode	Penjelasan
import pandas as pd	Mengimpor library pandas untuk manipulasi data
file_path = '/content/drive/MyDrive/ML3/praktikum/TUGAS PRAKTIKUM/tugas praktikum 3/data/day.csv'	Menentukan lokasi file CSV di Google Drive
df = pd.read_csv(file_path)	Membaca file CSV dan mengubahnya menjadi DataFrame pandas
df.head()	Menampilkan 5 baris pertama data untuk inspeksi cepat

Tabel dengan 15 kolom yang berisi:

- Kolom Waktu:** dteday, season, yr, mnth
- Kolom Cuaca:** temp, hum, windspeed, weathersit
- Kolom Target:** cnt (total sepeda yang disewa)

### 3. Import Library

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
```

Kode	Penjelasan
import pandas as pd	Mengimpor library pandas untuk manipulasi data
from sklearn.model_selection import train_test_split	Berfungsi untuk membagi dataset menjadi data training & testing
from sklearn.linear_model import LinearRegression	Algoritma machine learning untuk prediksi nilai kontinu, kegunaannya untuk Menguji performa model pada data yang belum dilihat

from sklearn.metrics import mean_squared_error, r2_score	<ul style="list-style-type: none"> <li>• <b>mean_squared_error</b>: Mengukur rata-rata kuadrat error (semakin kecil semakin baik)</li> <li>• <b>r2_score</b>: Mengukur seberapa baik model menjelaskan variasi data (0-1, semakin mendekati 1 semakin baik)</li> </ul>
import matplotlib.pyplot as plt	Membuat grafik, plot, chart untuk analisis visual

#### 4. Persiapan Data

```
# X: variabel independent, y: variabel dependent (target)
X = df[['temp', 'season', 'weathersit', 'hum', 'windspeed']]
y = df['cnt']
```

Penjelasan	
X	Variabel input (suhu, musim, cuaca, kelembaban, angin)
y	Variabel target (jumlah sepeda yang disewa)

#### 5. Split Data

```
# Split data: 80% training, 20% testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Penjelasan	
80%	data untuk training (latih model)
20%	data untuk testing (uji model)

#### 6. Buat model regresi linear

```
model = LinearRegression()
model.fit(X_train, y_train)
```

Penjelasan	
<ul style="list-style-type: none"> <li>• Menggunakan algoritma <b>Linear Regression</b></li> <li>• Model belajar pola dari data training</li> </ul>	

#### 7. Prediksi & Evaluasi Model

```
# Prediksi
y_pred = model.predict(X_test)

# Evaluasi model
mse = mean_squared_error(y_test, y_pred)
rmse = mse ** 0.5
r2 = r2_score(y_test, y_pred)
```

Penjelasan	
<ul style="list-style-type: none"> <li>• Model membuat prediksi pada data testing</li> <li>• Hitung RMSE (error rata-rata) dan R<sup>2</sup> (akurasi model)</li> </ul>	

# OUTPUT

## 1. HASIL MODEL PREDIKSI



=== HASIL MODEL PREDIKSI ===  
RMSE: 1356.79  
R<sup>2</sup> Score: 0.5409  
Akurasi: 54.09%

### Penjelasan

- RMSE 1356: Rata-rata selisih prediksi dengan kenyataan = 1356 sepeda
- Akurasi 54%: Model bisa menjelaskan 54% pola data, 46% sisanya masih error

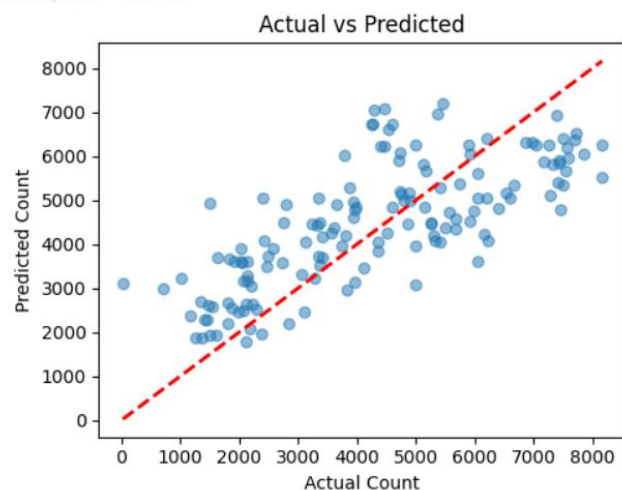
## 2. PENGARUH VARIABLE

=== PENGARUH VARIABLE ===  
temp: 5564.45  
season: 421.35  
weathersit: -470.91  
hum: -2231.35  
windspeed: -2717.20

### Penjelasan

TEMP: +5564.45	Setiap kenaikan 1 unit suhu (dalam skala ternormalisasi 0-1), jumlah sepeda yang disewa meningkat sekitar <b>5564 unit</b>
SEASON: +421.35	Setiap perubahan musim (1->2->3->4) mempengaruhi sekitar <b>421 sepeda</b>
WEATHERSIT: -470.91	Cuaca buruk mengurangi sekitar <b>471 sepeda</b>
HUM: -2231.35	Kelembaban tinggi sangat mengurangi minat bersepeda
WINDSPEED: -2717.20	Angin kencang paling mengurangi minat bersepeda

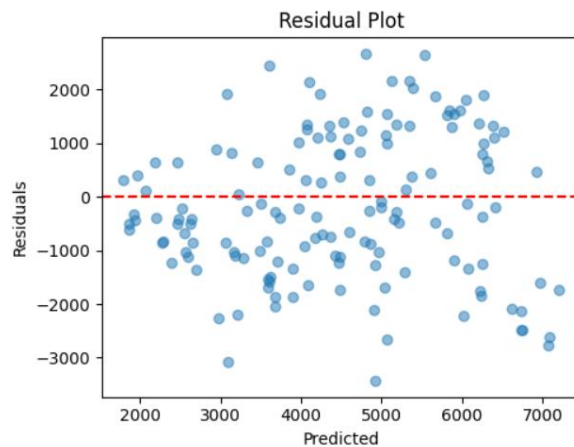
## 3. GRAFIK 1: Actual vs Predicted



### Penjelasan

- Titik-titik: Hasil prediksi vs kenyataan
- Garis merah: Garis sempurna (prediksi = aktual)
- Semakin dekat titik ke garis = prediksi semakin akurat

#### 4. GRAFIK 2: Residual Plot



##### Penjelasan

- Titik-titik: Error/selisih prediksi
- Garis merah di 0: Ideal (tidak ada error)
- Pola acak = model cukup baik
- Beberapa error besar = ada kondisi ekstrem yang sulit diprediksi

#### 5. Contoh Prediksi

```
=== CONTOH PREDIKSI ===
Data testing pertama:
temp      0.475833
season    4.000000
weathersit 1.000000
hum       0.733750
windspeed 0.174129
Name: 703, dtype: float64
Actual count: 6606
Predicted count: 5063
```

##### Penjelasan

Data testing pertama:

- temp 0.475833 (Cukup Hangat)
- season 4.000000 (Musim Dingin)
- weathersit 1.000000 (Cuaca Cerah)
- hum 0.733750 (Cukup Lembab)
- windspeed 0.174129 (Angin Pelan)
- Name: 703, dtype: float64
- Actual count: 6606
- Predicted count: 5063

**Selisih 1543 sepeda:** Model under-predict (prediksi lebih rendah dari kenyataan)

**Kemungkinan sebab:** Ada faktor lain seperti event khusus/promosi

#### KESIMPULAN IMPLEMENTASI

Model Linear Regression ini sudah memadai untuk prediksi dasar penyewaan sepeda berdasarkan kondisi cuaca, dengan suhu sebagai faktor penentu utama dan angin sebagai penghambat terbesar. Cocok untuk perencanaan stok harian, namun perlu dikembangkan lebih lanjut untuk akurasi yang lebih tinggi.