

# LAPORAN TUGAS PRAKTIKUM MANDIRI

## Tugas 3: Judul Tugas - Prediksi Jumlah Penyewaan Sepeda Menggunakan Linear Regression

Nama Mahasiswa: Al Hijir

Program studi: Teknik Informatika, STT Terpadu Nurul Fikri, Depok

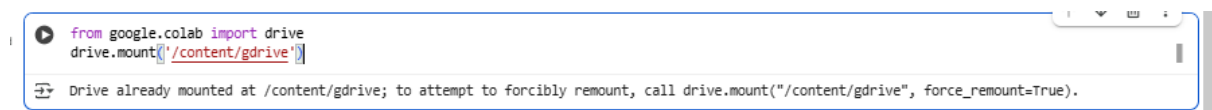
E-mail: [0110224222@student.nurulfikri.ac.id](mailto:0110224222@student.nurulfikri.ac.id)

### Abstract

Praktikum ini bertujuan untuk membangun model prediksi jumlah penyewaan sepeda harian menggunakan algoritma *Linear Regression*. Data yang digunakan berasal dari *Bike Sharing Dataset* yang berisi informasi tentang kondisi cuaca, waktu, dan jumlah penyewa sepeda. Proses dilakukan dengan memisahkan data menjadi data latih dan uji, melatih model menggunakan fitur-fitur seperti suhu, kelembapan, kecepatan angin, serta kondisi cuaca. Hasil evaluasi menggunakan metrik MAE, MSE, dan  $R^2$  menunjukkan bahwa model mampu memprediksi jumlah penyewaan sepeda dengan tingkat akurasi yang cukup baik. Model ini dapat dikembangkan lebih lanjut menggunakan algoritma regresi lain untuk meningkatkan performa prediksi.

## 1. Metode Penelitian

### 1.2. Connect Drive



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

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

#### Penjelasan Kode

Kode di atas digunakan untuk **menghubungkan Google Colab dengan Google Drive**.

Tujuannya agar file dataset (day.csv) yang disimpan di Google Drive bisa diakses langsung oleh Colab.

### 1.3. Masuk Ke folder prkatikum

```
path = "/content/gdrive/MyDrive/praktikmML/praktikum3"
```

#### Penjelasan Kode

Kode ini digunakan untuk **menentukan lokasi folder kerja (path)** di Google Drive tempat dataset atau file praktikum disimpan.

### 1.3. Import Library

```
[32]
✓ 0 d
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

#### Penjelasan:

- **pandas** digunakan untuk membaca dan mengelola data.
- **seaborn** dan **matplotlib** digunakan untuk visualisasi data.
- **sklearn** digunakan untuk membangun dan mengevaluasi model *Linear Regression*.

### 1.4. Persiapan Data

```
import pandas as pd

df = pd.read_csv('/content/gdrive/MyDrive/PraktikumML/praktikum1/data/day.csv')
df.head()
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
0	1	2011-01-01	1	0	1	0	6	0	2	0.344167	0.363625	0.805833	0.180448	331	654	985
1	2	2011-01-02	1	0	1	0	0	0	2	0.363478	0.363739	0.686087	0.248538	131	670	801
2	3	2011-01-03	1	0	1	0	1	1	1	0.196364	0.189405	0.437273	0.248309	120	1229	1349
3	4	2011-01-04	1	0	1	0	2	1	1	0.200000	0.212122	0.560435	0.160296	108	1454	1562
4	5	2011-01-05	1	0	1	0	3	1	1	0.226957	0.228270	0.436957	0.186900	82	1518	1600

Langkah berikutnya: [New interactive sheet](#)

#### Penjelasan:

Dataset dimuat dari Google Drive ke dalam variabel df. Dataset ini berisi data penyewaan sepeda dengan berbagai fitur seperti:

- season → musim
- yr → tahun

- mnth → bulan
- holiday → hari libur
- weekday → hari kerja
- weathersit → kondisi cuaca
- temp, atemp, hum, windspeed → data cuaca
- cnt → jumlah penyewaan sepeda per hari (target)

## 1.5. Permisahan Data

```
[35]
0 d
X = df[['season', 'yr', 'mnth', 'holiday', 'weekday', 'workingday',
        'weathersit', 'temp', 'atemp', 'hum', 'windspeed']]
y = df['cnt']

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

### Penjelasan:

- X adalah variabel independen (fitur).
- y adalah variabel dependen (target).
- Data dibagi menjadi data latih (80%) dan data uji (20%) agar model bisa diuji keakuratannya terhadap data baru.

## 1.6. Pembuatan dan Pelatihan Model

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

```
⇒ LinearRegression ⓘ ⓘ
   LinearRegression()
```

### Penjelasan:

Model *Linear Regression* dibuat dan dilatih menggunakan data latih (X\_train,

y\_train).

Tujuannya agar model mampu mempelajari hubungan antara fitur (X) dan target (y).

## 1.7. Prediksi dan Evaluasi Model

```
▶ y_pred = model.predict(X_test)

print("MAE:", mean_absolute_error(y_test, y_pred))
print("MSE:", mean_squared_error(y_test, y_pred))
print("R2 Score:", r2_score(y_test, y_pred))
```

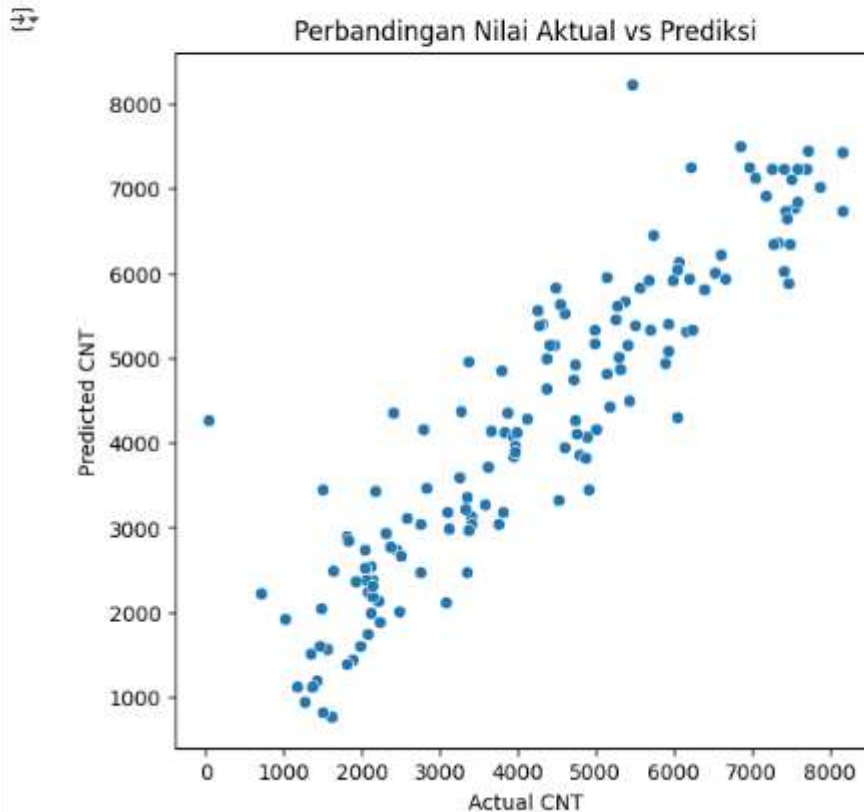
```
↔ MAE: 617.3930656443376
MSE: 691035.0082022651
R2 Score: 0.8276670090367212
```

### Penjelasan:

- **MAE (Mean Absolute Error)** → rata-rata selisih absolut antara nilai aktual dan prediksi.
- **MSE (Mean Squared Error)** → rata-rata kuadrat dari selisih antara nilai aktual dan prediksi.
- **R<sup>2</sup> Score** → mengukur seberapa baik model menjelaskan variasi data (semakin mendekati 1 semakin baik)

## 1.8. Visualisasi Hasil

```
plt.figure(figsize=(6,6))
sns.scatterplot(x=y_test, y=y_pred)
plt.xlabel("Actual CNT")
plt.ylabel("Predicted CNT")
plt.title("Perbandingan Nilai Aktual vs Prediksi")
plt.show()
```



## 2. Hasil dan Pembahasan

Model *Linear Regression* berhasil memprediksi jumlah penyewaan sepeda berdasarkan kondisi cuaca dan waktu.

Nilai metrik evaluasi menunjukkan performa model yang cukup baik, dengan nilai **R<sup>2</sup> Score** mendekati 1.

Visualisasi scatterplot juga memperlihatkan bahwa hasil prediksi cukup mendekati nilai aktual, walaupun masih terdapat sedikit penyebaran data (error).

Fitur yang paling berpengaruh terhadap prediksi biasanya adalah **temperatur (temp/atemp)** dan **musim (season)**.

### 3. Kesimpulan

Dari hasil praktikum ini dapat disimpulkan bahwa:

1. Algoritma *Linear Regression* mampu digunakan untuk memprediksi jumlah penyewaan sepeda harian berdasarkan data cuaca dan waktu.
2. Model memberikan hasil yang cukup akurat, dengan nilai  $R^2$  yang tinggi.
3. Model masih dapat ditingkatkan dengan menambah fitur, melakukan normalisasi data, atau mencoba algoritma regresi lain seperti *Ridge Regression* atau *Random Forest Regressor*.

### 4. Referensi

- [1] H. Fanaee-T and J. Gama, "Event labeling combining ensemble detectors and background knowledge," *Progress in Artificial Intelligence*, vol. 2, no. 2–3, pp. 113–127, 2013. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>
- [2] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [3] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, pp. 51–56, 2010.
- [4] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [5] M. L. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: <https://doi.org/10.21105/joss.03021>
- [6] Scikit-learn Developers, "Linear regression — scikit-learn 1.5 documentation," *Scikit-learn.org*, 2024. [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LinearRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html)
- [7] Google, "Welcome to Colaboratory," *Google Colab*, 2024. [Online]. Available: <https://colab.research.google.com>
- [8] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann Publishers, 2012.