

Tugas 1: Judul tugas – Regresi dan evaluasi model

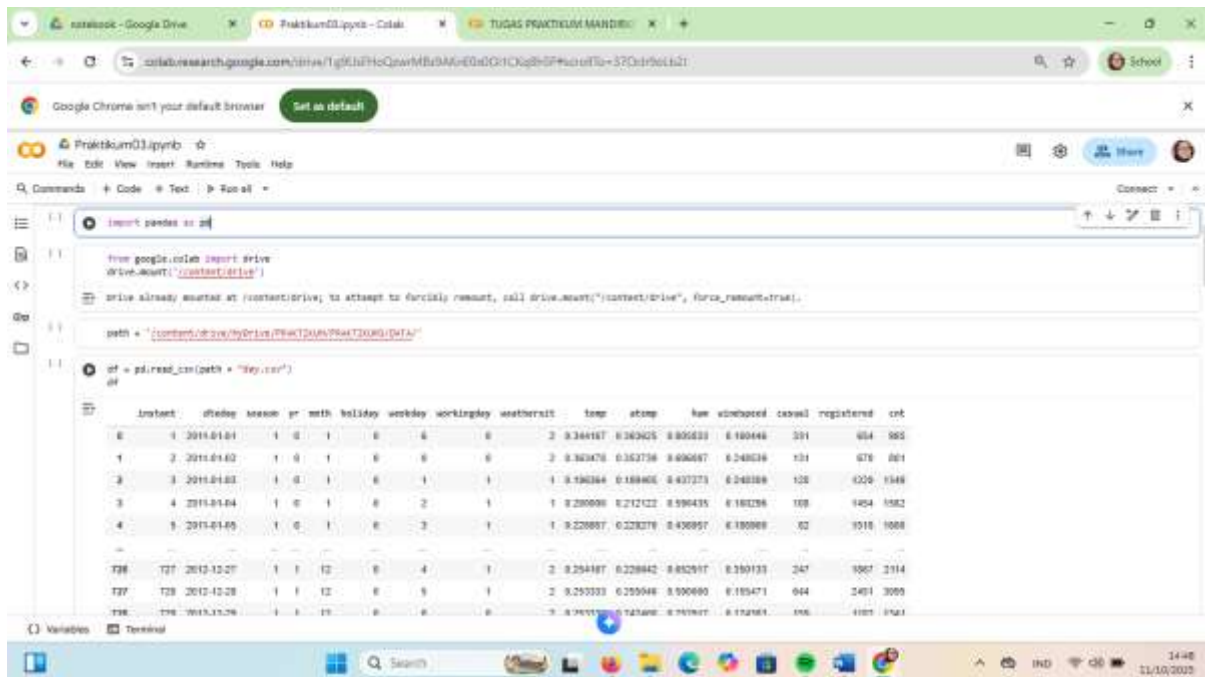
Siti aisah - 0110222129¹

¹ Teknik Informatika, STT Terpadu Nurul Fikri, Depok

*E-mail: siti22129ti@student.nurulfikri.ac.id

Abstract. Pembelajaran *Machine Learning* merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma dan model statistik untuk memungkinkan sistem komputer belajar dari data dan membuat prediksi atau keputusan secara otomatis tanpa pemrograman eksplisit..

1. Praktikum mandiri



The screenshot shows a Jupyter Notebook interface with the following code and output:

```
import pandas as pd

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

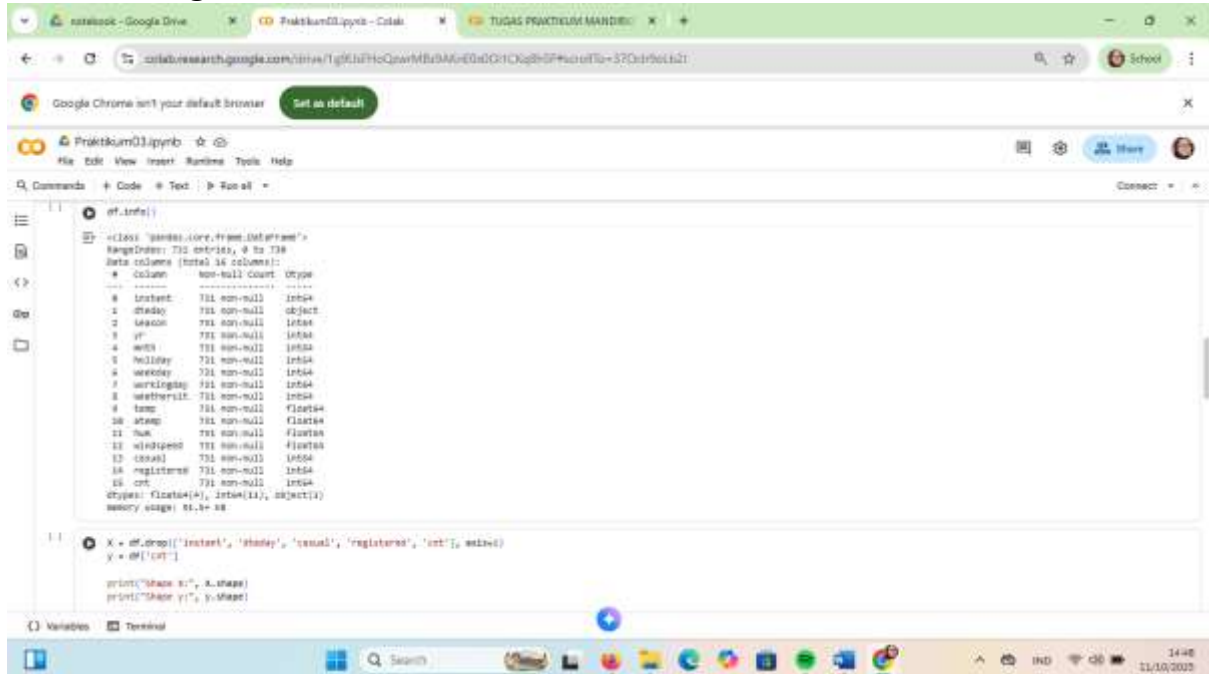
path = '/content/drive/MyDrive/Praktikum/Praktikum1/Data/'

df = pd.read_csv(path + 'day.csv')
df
```

	instant	date	day	season	yr	month	holiday	weekday	workingday	weather	temp	atemp	cas	registered	cnt		
0	1	2011-01-01	1	0	1	0	0	0	0	2	5.344167	5.282425	0.805633	0.180448	331	684	885
1	2	2011-01-02	1	0	1	0	0	0	0	2	5.363478	5.352738	0.806687	0.248038	131	678	881
2	3	2011-01-03	1	0	1	0	0	1	1	1	5.196384	5.188405	0.837273	0.248038	128	1220	1348
3	4	2011-01-04	1	0	1	0	0	2	1	1	5.280808	5.271222	0.896435	0.181238	108	1454	1562
4	5	2011-01-05	1	0	1	0	0	3	1	1	5.228857	5.228378	0.836857	0.180808	82	1516	1608
...
726	727	2012-12-27	1	1	12	0	0	4	1	2	8.254187	8.228842	0.852917	0.350133	247	1067	2114
727	728	2012-12-28	1	1	12	0	0	5	1	2	8.253333	8.255048	0.500680	0.185471	844	2481	3000
728	729	2013-01-06	1	1	1	0	0	0	0	3	8.191110	8.174808	0.761817	0.172861	106	1187	1293

1.1 . mengimport librsry psndas dan membaca dataset.

2. Mengecek informasi dataset



```
df.info()

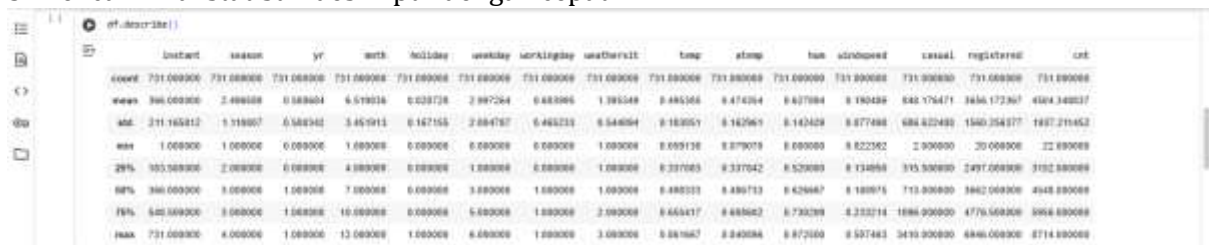
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 16 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   instant             731 non-null    int64  
 1   season              731 non-null    object  
 2   yr                  731 non-null    int64  
 3   mo                  731 non-null    int64  
 4   aed                 731 non-null    int64  
 5   holiday             731 non-null    int64  
 6   weekday             731 non-null    int64  
 7   workingday          731 non-null    int64  
 8   weatherid           731 non-null    int64  
 9   temp               731 non-null    float64 
10   atemp              731 non-null    float64 
11   hum                731 non-null    float64 
12   windspeed          731 non-null    float64 
13   casual             731 non-null    int64  
14   registered          731 non-null    int64  
15   cnt                731 non-null    int64  
dtypes: float64(4), int64(11), object(1)
memory usage: 51.5+ KB

X = df.drop(['instant', 'holiday', 'casual', 'registered', 'cnt'], axis=1)
y = df['cnt']

print("Shape X:", X.shape)
print("Shape y:", y.shape)
```

1.2 Hasil mengecek informasi dataset

3. mencari nilai statistik deskriptif dengan cepat



	instant	season	yr	mo	holiday	weekday	workingday	weatherid	temp	atemp	hum	windspeed	casual	registered	cnt
count	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000
mean	361.000000	2.000000	0.000000	6.519018	0.020728	2.997264	0.683965	1.395548	9.485366	9.474304	9.437084	9.190489	842.176471	3654.172367	4504.348837
std	211.165812	1.118007	0.000000	3.451913	0.167155	2.084787	0.465233	0.548084	9.183051	9.142428	9.877488	684.622480	1540.354377	1897.211652	1937.211652
min	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	0.001136	0.079078	0.000000	0.000000	0.000000	20.000000	22.000000
25%	105.500000	2.000000	0.000000	4.000000	0.000000	1.000000	0.000000	1.000000	0.337085	0.337042	0.520000	0.134956	315.500000	2497.000000	3102.000000
50%	361.000000	3.000000	1.000000	7.000000	0.000000	3.000000	1.000000	1.000000	8.880333	8.880733	9.426667	9.180976	713.000000	3642.000000	4504.000000
75%	540.500000	3.000000	1.000000	10.000000	0.000000	5.000000	1.000000	3.000000	9.654417	9.655682	9.730289	9.233214	1084.000000	4776.500000	5954.000000
max	731.000000	4.000000	1.000000	12.000000	1.000000	6.000000	1.000000	3.000000	9.581667	9.583086	9.877000	9.527483	3410.000000	6846.000000	8714.000000

1.2 hasil dari mencari nilai statistic deskriptif dengan cepat

4. menentukan variable independent dan dependent

```
1 X = df.drop(['instant', 'date', 'casual', 'registered', 'cnt'], axis=1)
2 y = df['cnt']
3
4 print("Shape X:", X.shape)
5 print("Shape y:", y.shape)
6
7 Shape X: (731, 11)
8 Shape y: (731,)
```

1.4 menentukan variable independent dan dependent

5. membagi data testing dan training

```
1 from sklearn.metrics import r2_score, mean_squared_error
2 import numpy as np
3
4 r2 = r2_score(y_test, y_pred)
5 rmse = np.sqrt(mean_squared_error(y_test, y_pred))
6
7 print("R² Score:", r2)
8 print("RMSE:", rmse)
9
10 R² Score: 0.8276670808367213
11 RMSE: 0.31285154662688
```

1.5 hasil dari menentukan variable independent dan dependent

6. menginstall model linear regresi

```
1 from sklearn.linear_model import LinearRegression
2
3 model = LinearRegression()
4 model.fit(X_train, y_train)
5
6 LinearRegression
7 ~LinearRegression()
```

1.6 hasil dari menentukan variable independent dan dependent

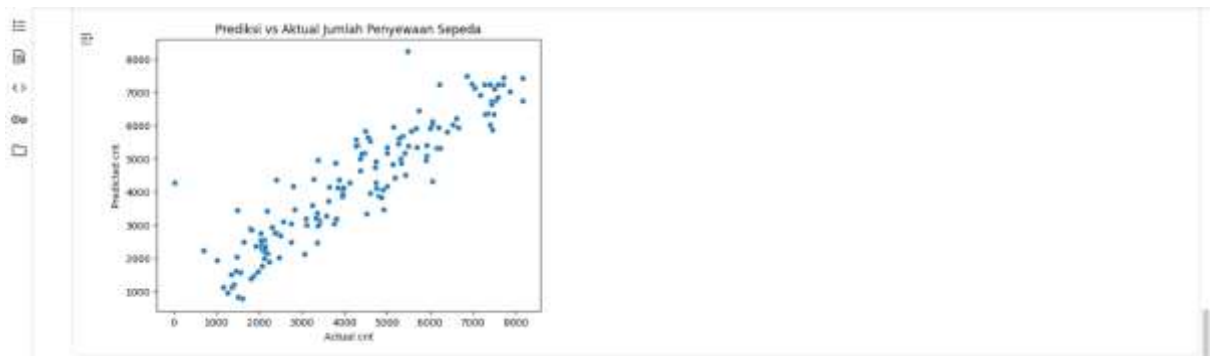
7. menyiapkan model prediksi

```
1 y_pred = model.predict(X_test)
2
3 from sklearn.metrics import r2_score, mean_squared_error
4 import numpy as np
5
6 r2 = r2_score(y_test, y_pred)
7 rmse = np.sqrt(mean_squared_error(y_test, y_pred))
8
9 print("R² score:", r2)
10 print("RMSE:", rmse)
11
12 R² score: 0.8276670808367213
13 RMSE: 0.31285154662688
```

1.7 hasil dari menyiapkan model prediksi

8. visualisasi data

```
1 import matplotlib.pyplot as plt
2 import seaborn as sns
3
4 plt.figure(figsize=(7,3))
5 sns.scatterplot(x=X_test, y=y_pred)
6 plt.xlabel('actual cnt')
7 plt.ylabel('predicted cnt')
8 plt.title('Predicted vs actual jumlah penyewaan sepeda')
9 plt.show()
```



1.8 hasil dari . visualisasi data

Referensi:

- Munir, S., Seminar, K. B., Sudradjat, Sukoco, H., & Buono, A. (2022). The Use of Random Forest Regression for Estimating Leaf Nitrogen Content of Oil Palm Based on Sentinel 1-A Imagery. *Information*, 14(1), 10. <https://doi.org/10.3390/info14010010>
- Seminar, K. B., Imantho, H., Sudradjat, Yahya, S., Munir, S., Kaliana, I., Mei Haryadi, F., Noor Baroroh, A., Supriyanto, Handoyo, G. C., Kurnia Wijayanto, A., Ijang Wahyudin, C., Liyantono, Budiman, R., Bakir Pasaman, A., Rusiawan, D., & Sulastri. (2024). PreciPalm: An Intelligent System for Calculating Macronutrient Status and Fertilizer Recommendations for Oil Palm on Mineral Soils Based on a Precision Agriculture Approach. *Scientific World Journal*, 2024(1). <https://doi.org/10.1155/2024/1788726>

LINK GITHUB: <https://github.com/Sitiaisah1604/machine-learning>