

Nama : Muhammad Mahda

Kelompok : A\_Infotech

## Homework Machine Learning II

1. Dari data 'SALES.txt', saya merubah format data yang ada dengan menambahkan colum 'Sales' dan column 'Advertising'. Sehingga data tersebut dapat diolah, contohnya sebagai berikut :

```
import pandas as pd

# Data yang diberikan
data_values = [
    [12.0, 15.0],
    [20.5, 16.0],
    [21.0, 18.0],
    [15.5, 27.0],
    [15.3, 21.0],
    [22.5, 49.0],
    [24.5, 21.0],
    [21.3, 22.0],
    [23.5, 20.0],
    [28.0, 36.0],
    [24.0, 40.0],
    [15.5, 3.0],
    [17.3, 21.0],
    [25.3, 20.0],
    [25.0, 62.0],
    [36.5, 65.0],
    [36.5, 46.0],
    [29.6, 44.0],
    [30.5, 33.0],
    [28.0, 62.0],
    [26.0, 22.0],
    [21.5, 12.0],
    [19.7, 34.0],
    [19.0, 3.0],
    [16.0, 5.0],
    [20.7, 14.0],
    [26.5, 36.0],
    [30.6, 40.0],
    [32.3, 49.0],
    [29.5, 7.0],
    [28.3, 52.0],
    [31.3, 65.0],
    [32.3, 47.0],
    [26.4, 5.0],
    [23.4, 17.0],
    [16.4, 1.0]
]

# Membuat DataFrame
df = pd.DataFrame(data_values, columns=['Sales', 'Advertising'])

# Simpan DataFrame ke dalam file CSV
df.to_csv("Sales_Advertising_data.csv", index=False)

# Menampilkan pesan setelah menyimpan
print("Data telah disimpan ke dalam file Sales_Advertising_data.csv")
```

2. Lakukan prediksi menggunakan simple regression, Contoh Code nya :

```
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 numpy as np

# Memuat data dari file CSV
df = pd.read_csv("Sales_Advertising_data.csv")

# Memberikan nama kolom saat memisahkan fitur (X) dan target (y)
X = df[['Sales']] # fitur
X.columns = ['Sales'] # Menambahkan nama kolom 'Sales'
y = df['Advertising'] # target

# Membagi data menjadi data training dan data testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Membuat model regresi linear
model = LinearRegression()

# Melatih model menggunakan data training
model.fit(X_train, y_train)
```

3. Klien ingin mengetahui berapa cost Advertising yang harus dikeluarkan jika :
  - a. 50 sales

- b. 100 sales
  - c. 150 sales
4. Hitung prediksi cost advertising untuk masing-masing sales tersebut, contoh codenya sebagai berikut:

```
# Prediksi biaya advertising untuk masing-masing sales
predicted_ad_costs = model.predict(sales_values)
print("Predicted Advertising Costs:")
for sales, cost in zip([50, 100, 150], predicted_ad_costs):
    print(f"Sales: {sales}, Predicted Advertising Cost: {cost}")
```

Predicted Advertising Costs:  
Sales: 50, Predicted Advertising Cost: 70.64773341965265  
Sales: 100, Predicted Advertising Cost: 153.75066092686419  
Sales: 150, Predicted Advertising Cost: 236.85358843407573

5. Lakukan Evaluasi menggunakan RMSE dan R2 score, contohnya sebagai berikut:

```
# Prediksi menggunakan testing set
y_pred = model.predict(x_test)

# Hitung RMSE
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
print(f"RMSE: {rmse}")

# Hitung R2 score
r2 = r2_score(y_test, y_pred)
print(f"R2 Score: {r2}")
```

RMSE: 12.8197563902471  
R2 Score: 0.5875233784338036

6. Lakukan interpretasi,
- a. Prediksi Biaya Advertising  
Dari model regresi linear sederhana yang telah kita latih, kita telah melakukan prediksi biaya advertising untuk nilai sales tertentu. Misalnya, untuk sales sebesar 50 juta dolar, 100 juta dolar, dan 150 juta dolar, kita telah memprediksi biaya advertising yang dibutuhkan:
    - ☐ Untuk Sales 50 juta dolar: Prediksi biaya advertising adalah 70.64773341965265 million \$
    - ☐ Untuk Sales 100 juta dolar: Prediksi biaya advertising adalah 153.75066092686419 million \$
    - ☐ Untuk Sales 150 juta dolar: Prediksi biaya advertising adalah 236.85358843407573 million \$
  - b. Evaluasi Model: RMSE dan R2 Score  
Setelah melatih model dan melakukan prediksi, kita juga mengevaluasi kinerja model menggunakan dua metrik utama:
    - **Root Mean Squared Error (RMSE):** 12.8197563902471
    - **R-squared (R2 Score):** 0.5875233784338036

c. Interpretasi Evaluasi

- RMSE (Root Mean Squared Error): Jika RMSE adalah 12.82 juta dolar, ini menunjukkan bahwa rata-rata kesalahan prediksi model terhadap data pengujian adalah sekitar 12.82 juta dolar; dengan kata lain, biaya iklan yang diprediksi oleh model rata-rata dapat memiliki kesalahan sekitar 12.82 juta dolar dari nilai sebenarnya. Semakin rendah nilai RMSE, semakin akurat model dalam memprediksi biaya iklan.
- R2 Score: Nilai R2 sebesar 0.59 menunjukkan bahwa sekitar 59% variabilitas biaya iklan dapat dijelaskan oleh variabilitas penjualan berdasarkan model ini, sementara nilai R2 yang lebih tinggi menunjukkan bahwa model dapat menjelaskan lebih banyak variabilitas penjualan daripada variabilitas biaya iklan. Meskipun nilai R2 sebesar 0.59 menunjukkan bahwa ada hubungan antara penjualan dan biaya iklan, ada juga 41% variabilitas tambahan yang tidak dapat dijelaskan.

Kesimpulan :

Anda dapat memberi tahu klien berapa biaya iklan yang seharusnya mereka alokasikan untuk mencapai target penjualan tertentu dengan menggunakan hasil prediksi dan evaluasi dari model regresi linear sederhana ini. Evaluasi model juga membantu Anda memahami seberapa baik model dapat memprediksi berdasarkan data yang ada. Pemangku kepentingan membutuhkan interpretasi ini untuk membuat keputusan periklanan yang lebih baik.

Link Github : <https://github.com/MahdaMuhammad/HM-ML>

Link Google Colab :

<https://colab.research.google.com/drive/1Mg1oIe6ReI03mZBZXzn8efbSqSES346d?usp=sharing>