

Nama Kelompok 5:

- Juan Wens Sanctung Rahawarin
- Athallah Rafi Y
- Rizky Maulana
- Dimas Adi Perdana
- Maiharsa Muara

Tugas Kelompok

1. Apa Kelebihan dan Kelemahan linear regression?

➤ Kelebihan:

- Metode ini mudah dipahami dan diimplementasikan, serta tidak memerlukan data dalam bentuk tertentu.
- Metode ini dapat menentukan seberapa kuat pengaruh variabel independen terhadap variabel dependen.
- Metode ini dapat memprediksi nilai variabel dependen di masa depan berdasarkan variabel independen.

➤ Kelemahan:

- Metode ini hanya dapat memodelkan hubungan linier antara variabel, sehingga tidak cocok untuk data yang memiliki hubungan non-linier.
- Metode ini sensitif terhadap outlier dan multicollinearity, yaitu kondisi di mana variabel independen saling berkorelasi tinggi.
- Metode ini tidak dapat menjamin bahwa hubungan yang ditemukan memiliki makna kausal, melainkan hanya korelasional.

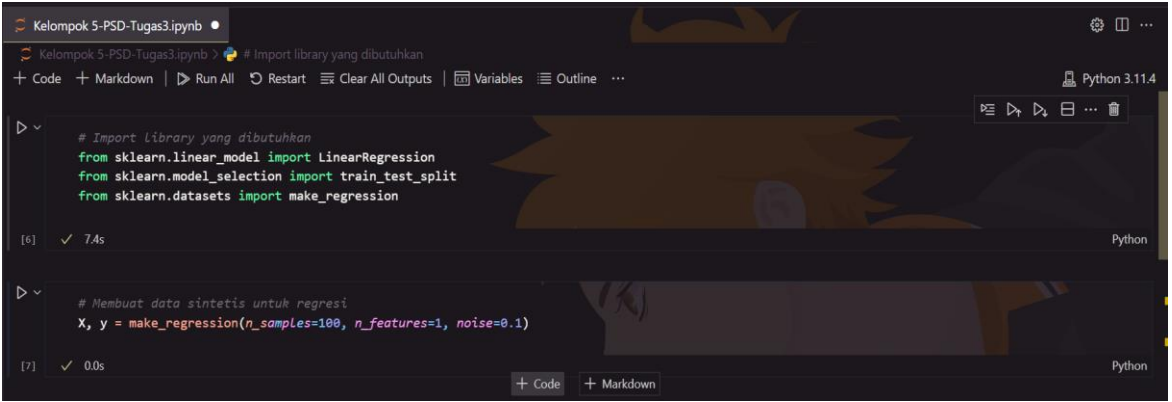
2. Apa faktor yang perlu diperhatikan jika ingin menggunakan linear regression

- Jumlah sampel yang digunakan harus sama untuk setiap variabel.
- Jumlah variabel bebas (X) harus satu, karena linear regression sederhana hanya dapat memodelkan hubungan antara satu variabel bebas dan satu variabel terikat.
- Nilai residual harus berdistribusi normal, yaitu tidak ada pola tertentu dalam perbedaan antara nilai observasi dan nilai prediksi.
- Terdapat hubungan yang linear antara variabel bebas (X) dan variabel terikat (Y), yaitu dapat digambarkan dengan garis lurus.
- Tidak terjadi gejala heteroskedastisitas, yaitu varians dari nilai residual tidak bergantung pada nilai variabel bebas.
- Tidak terjadi gejala autokorelasi, yaitu nilai residual tidak berkorelasi dengan nilai residual sebelumnya. Hal ini berlaku untuk data runtut waktu atau time series.
- Data yang digunakan harus valid dan reliabel, yaitu mampu mengukur apa yang seharusnya diukur dengan konsisten. Jika data berasal dari kuesioner, maka perlu dilakukan uji validitas dan reliabilitas terlebih dahulu.

- Data yang berskala ordinal harus diubah menjadi skala interval, yaitu memiliki jarak yang sama antara setiap nilai. Hal ini dapat dilakukan dengan metode successive interval atau metode lainnya.
3. Hal apa saja yang membuat linear regression gagal memprediksi data uji?
- Data uji tidak memiliki hubungan linier antara X dan Y, sehingga model yang dibuat tidak sesuai dengan pola data.
 - Data uji memiliki outlier atau nilai ekstrem yang menyimpang dari sebaran data normal, sehingga mempengaruhi estimasi parameter model.
 - Data uji memiliki multicollinearity atau korelasi tinggi antara variabel independen, sehingga menyebabkan model menjadi tidak stabil dan rentan terhadap perubahan data.
 - Data uji memiliki heteroskedasticity atau varians error yang tidak konstan, sehingga mengganggu asumsi kesalahan standar dan uji hipotesis.
 - Data uji memiliki autokorelasi atau ketergantungan antara error pada observasi berurutan, sehingga mengurangi efisiensi model dan menghasilkan estimasi yang bias.
 - Data uji memiliki missing value atau nilai yang hilang, sehingga mengurangi jumlah informasi yang tersedia untuk membuat model.

Untuk mengatasi hal-hal di atas, ada beberapa langkah yang dapat dilakukan, seperti:

- Melakukan transformasi data untuk membuat hubungan antara X dan Y menjadi lebih linier, misalnya dengan menggunakan logaritma, pangkat, atau akar.
 - Melakukan deteksi dan penanganan outlier, misalnya dengan menggunakan metode boxplot, z-score, atau IQR.
 - Melakukan deteksi dan penanganan multicollinearity, misalnya dengan menggunakan metode VIF, tolerance, atau ridge regression.
 - Melakukan deteksi dan penanganan heteroskedasticity, misalnya dengan menggunakan metode White test, Breusch-Pagan test, atau weighted least squares.
 - Melakukan deteksi dan penanganan autokorelasi, misalnya dengan menggunakan metode Durbin-Watson test, Cochrane-Orcutt method, atau generalized least squares.
 - Melakukan deteksi dan penanganan missing value, misalnya dengan menggunakan metode listwise deletion, pairwise deletion, atau imputation.
4. Buatlah satu contoh program sederhana (python) untuk kasus dengan linear regression. Lengkapi dengan komentar penjelasan dalam bahasa Indonesia pada Jupyter Notebook nya.



```
Kelompok 5-PSD-Tugas3.ipynb
Kelompok 5-PSD-Tugas3.ipynb > # Import library yang dibutuhkan
+ Code + Markdown | Run All | Restart | Clear All Outputs | Variables | Outline ... Python 3.11.4

# Import library yang dibutuhkan
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression

[6] ✓ 7.4s Python

# Membuat data sintetis untuk regresi
X, y = make_regression(n_samples=100, n_features=1, noise=0.1)

[7] ✓ 0.0s Python
```

```
Kelompok 5-PSD-Tugas3.ipynb
Kelompok 5-PSD-Tugas3.ipynb > # Import library yang dibutuhkan
+ Code + Markdown | ▶ Run All ⌂ Restart ≡ Clear All Outputs | 📄 Variables ≡ Outline ... Python 3.11.4

# Membagi data menjadi data Latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

[8] ✓ 0.0s Python

# Membuat objek model regresi linear
model = LinearRegression()

[9] ✓ 0.0s Python

# Melatih model dengan data Latih
model.fit(X_train, y_train)

[10] ✓ 0.0s Python
...
LinearRegression
LinearRegression()
```

```
Kelompok 5-PSD-Tugas3.ipynb
Kelompok 5-PSD-Tugas3.ipynb > # Import library yang dibutuhkan
+ Code + Markdown | ▶ Run All ⌂ Restart ≡ Clear All Outputs | 📄 Variables ≡ Outline ... Python 3.11.4

# Memprediksi data uji
predictions = model.predict(X_test)

[11] ✓ 0.0s Python

# Menampilkan prediksi
print(predictions)

[12] ✓ 0.0s Python
...
[ 44.00588749 -157.60444996  22.88092539  32.97900507 -122.82224604
 -49.13896273  -4.65602312  -97.29845635  73.17464753  -93.03391366
 -66.35307126   5.59024576  -85.46691628  38.58873122  -87.76107204
 -210.85592698 116.50213243  90.69977187  -3.40770007  24.53887644]
```

```
Kelompok 5-PSD-Tugas3.ipynb
Kelompok 5-PSD-Tugas3.ipynb > # Import library yang dibutuhkan
+ Code + Markdown | ▶ Run All ⌂ Restart ≡ Clear All Outputs | 📄 Variables ≡ Outline ... Python 3.11.4

Komentar:

Pertama, kita mengimpor library yang dibutuhkan dan membuat data sintesis untuk regresi. Kemudian, kita membagi data menjadi data latih dan data uji. Selanjutnya, kita membuat objek model regresi linear dan melatihnya dengan data latih. Akhirnya, kita menggunakan model yang telah dilatih untuk memprediksi data uji dan menampilkan prediksi tersebut.
```