**Statistik Komputasi**

**Metode Prediksi Regresi Linier**

**Ditulis oleh**

**Aryajaya Alamsyah, S.Kom.**

**Kompetensi Dasar:**

1. Mahasiswa mampu menghitung nilai korelasi, nilai determinasi dan signifikansi
2. Mahasiswa mampu membuat model regresi
3. Mahasiswa mampu menghitung nilai akurasi (nilai residual, dan nilai rsme)

**Target Pembelajaran:**

1. Mahasiswa mampu membedakan hitungan statistik seperti regresi linier, regresi linier berganda, logistik regresi, arima, sarima dan hitungan neural network (deep learning) seperti multi layer perceptron (MLP), recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU) untuk membuat model prediksi.
2. Mahasiswa mengetahui hitungan statistik untuk prediksi umumnya bersifat:

* memiliki jumlah variabel dan data yang sedikit.
* Korelasi antar variabel bersifat linier
* Ketika nilai korelasi antar variabel kuat dan berpengaruh secara nyata maka kecenderungan hasil regresi akan baik.
* Prediksi dengan hitungan statistik hanya mampu memprediksi apabila tidak terdapat nilai aktual yang ekstrim.
* Dan lain-lain.

1. Mahasiswa mengetahui hitungan neural network (deep learning) untuk prediksi umumnya bersifat:

* Memiliki jumlah variabel dan data yang tak hingga sampe ke-N.
* Oleh karena itu dilakukan seleksi fitur dan reduksi data untuk mereduksi dimensi data tersebut dengan tetap mempertahankan karakteristik data tersebut.
* Korelasi antar variabel dapat bersifat linier dan non-linier.
* Ketika nilai korelasi cenderung rendah, maka neural network mampu membuat prediksi dengan baik karena terdapat fungsi feed forward, back propagation, koreksi inputan bias, dll.
* Prediksi neural network mampu memprediksi apabila terdapat nilai aktual yang ekstrim. Karena neural network mampu mengingat nilai ekstrim tersebut terjadi di periode waktu tertentu dan disebabkan oleh variabel tertentu.

**Perbedaan Korelasi dengan Regresi**

Analisis regresi adalah analisis yang bertujuan untuk mengetahui hubungan secara linier antara satu atau lebih variabel bebas dengan variabel terikat. Analisis regresi terbagi menjadi dua yaitu Regresi Linier Sederhana (RLS) dan Regresi Linier Berganda (RLB). RLS adalah regresi dengan menggunakan satu variabel bebas dan satu variabel terikat. RLB adalah regresi dengan menggunakan lebih dari satu variabel bebas dan satu variabel terikat.

Selain menggunakan RLS dan RLB dapat dgunakan analisis korelasi untuk mengetahui hubungan secara linier anatar variabel bebas dan terikat. Namun pada analisis korelasi tidak memerlukan hubungan sebab-akibat sehingga tidak dapat membedakan antara variabel bebas dan variabel terikat.

**Penjelasan Regresi Linier**

RLS adalah salah satu metode statistik untuk mengetahui sejauh mana hubungan sebab-akibat antara variabel bebas dengan variabel (bersifat univariate) terikat melalui persamaan regresi.

Rumus RLS adalah:

Dimana rumus untuk dan adalah

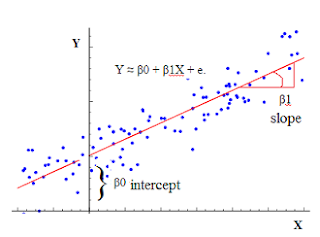
adalah variabel terikat

adalah parameter intercept

adalah parameter slope

adalah variabel bebas dengan i = 1, 2, 3, …, n (n adalah jumlah data)

adalah kesalahan acak (pengaruh variabel bebas lain selain variabel X)



Gambar 1. Ilustrasi Persamaan Regresi Linier

**Studi Kasus Regresi Linier**

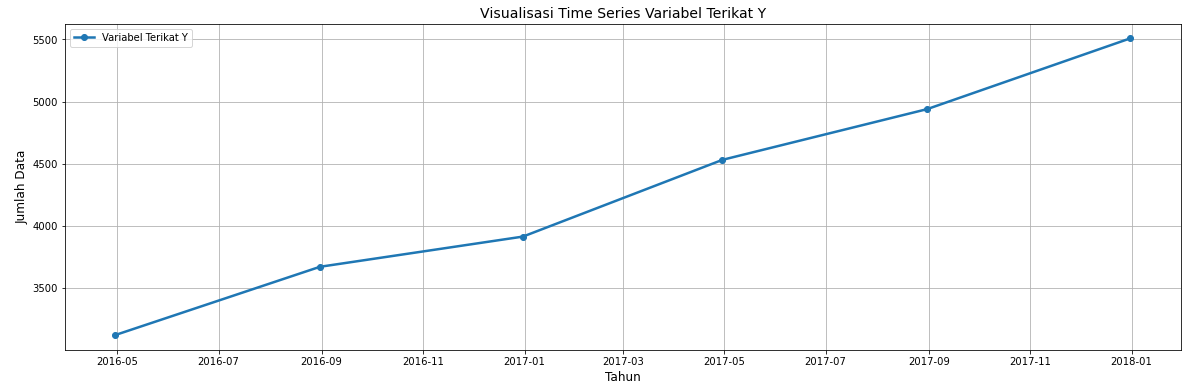
Tabel 1. Dataset Dummy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Ke | tahun | x1 | x2 | y |
| 1 | 01/12/2017 | 2.75 | 5.3 | 1464 |
| 2 | 01/11/2017 | 2.50 | 5.3 | 1394 |
| 3 | 01/10/2017 | 2.50 | 5.3 | 1357 |
| 4 | 01/09/2017 | 2.50 | 5.3 | 1293 |
| 5 | 01/08/2017 | 2.50 | 5.4 | 1256 |
| 6 | 01/07/2017 | 2.50 | 5.6 | 1254 |
| 7 | 01/06/2017 | 2.50 | 5.5 | 1234 |
| 8 | 01/05/2017 | 2.25 | 5.5 | 1195 |
| 9 | 01/04/2017 | 2.25 | 5.5 | 1159 |
| 10 | 01/03/2017 | 2.25 | 5.6 | 1167 |
| 11 | 01/02/2017 | 2.00 | 5.7 | 1130 |
| 12 | 01/01/2017 | 2.00 | 5.9 | 1075 |
| 13 | 01/12/2016 | 2.00 | 6.0 | 1047 |
| 14 | 01/11/2016 | 1.75 | 5.9 | 965 |
| 15 | 01/10/2016 | 1.75 | 5.8 | 943 |
| 16 | 01/09/2016 | 1.75 | 6.1 | 958 |
| 17 | 01/08/2016 | 1.75 | 6.2 | 971 |
| 18 | 01/07/2016 | 1.75 | 6.1 | 949 |
| 19 | 01/06/2016 | 1.75 | 6.1 | 884 |
| 20 | 01/05/2016 | 1.75 | 6.1 | 866 |
| 21 | 01/04/2016 | 1.75 | 5.9 | 876 |
| 22 | 01/03/2016 | 1.75 | 6.2 | 822 |
| 23 | 01/02/2016 | 1.75 | 6.2 | 704 |
| 24 | 01/01/2016 | 1.75 | 6.1 | 719 |

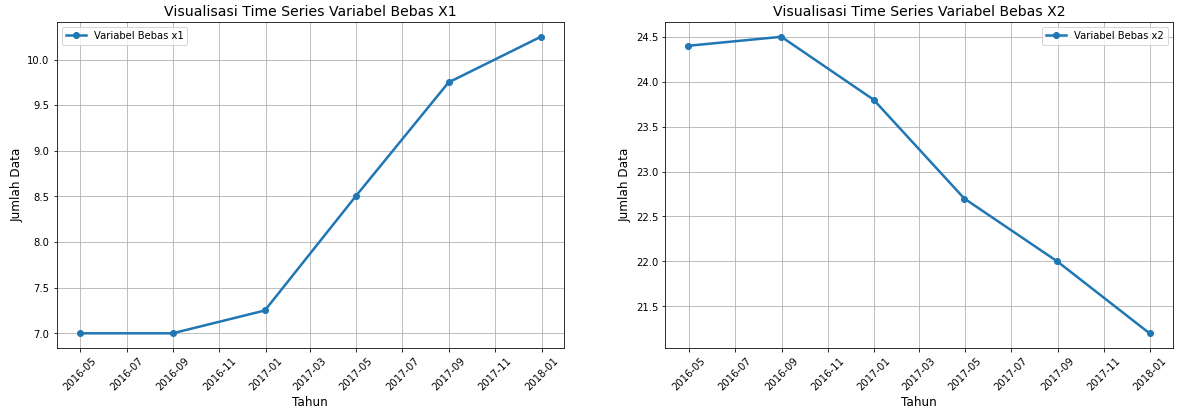
Langkah 1. Tentukan variabel bebas dan terikat.

Variabel bebas x1 dan x2. Variabel terikat y

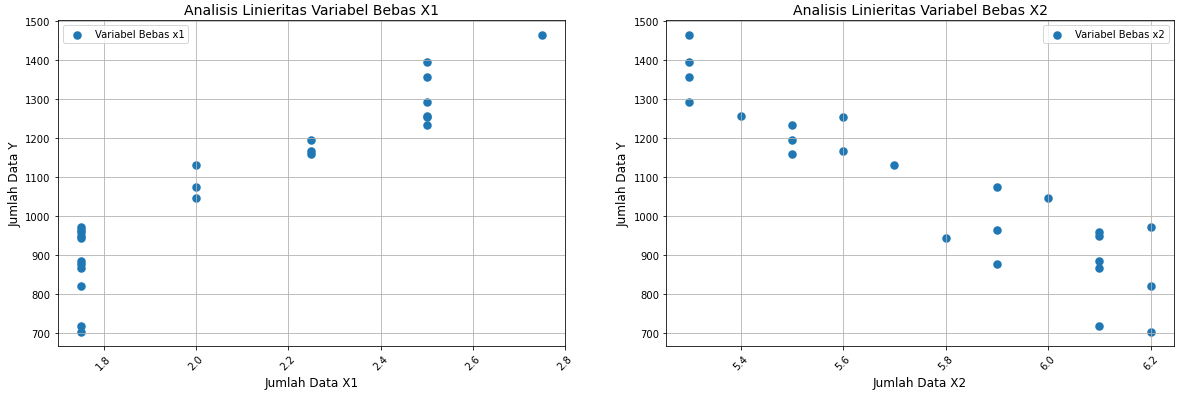
Langkah 2. Cek pergerakan nilai x1, x2 dan y (cara visualisasi data)



Gambar 2. Pergerakan nilai Y



Gambar 3. Pergerakan nilai x1(kiri) dan x2(kanan)



Gambar 4. Visual Linieritas x1 dengan y (kiri) dan x2 dengan y (kanan)

Diketahui bahwa

1. Pergerakan nilai y membentuk pola tren naik
2. Pergerakan nilai x1 pola tren naik dan x2 pola tren turun
3. Variabel bebas dan variabel terikat bersifat linier

Langkah 3. Hitung nilai regresi (menggunakan nilai x1 dan y) secara manual

Tabel 2. Hitungan awal regresi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Ke | x1 | y | x1^2 | y^2 | x1 y |
| 1 | 2,75 | 1464 | 7,56 | 2143296 | 4026,00 |
| 2 | 2,50 | 1394 | 6,25 | 1943236 | 3485,00 |
| 3 | 2,50 | 1357 | 6,25 | 1841449 | 3392,50 |
| 4 | 2,50 | 1293 | 6,25 | 1671849 | 3232,50 |
| 5 | 2,50 | 1256 | 6,25 | 1577536 | 3140,00 |
| 6 | 2,50 | 1254 | 6,25 | 1572516 | 3135,00 |
| 7 | 2,50 | 1234 | 6,25 | 1522756 | 3085,00 |
| 8 | 2,25 | 1195 | 5,06 | 1428025 | 2688,75 |
| 9 | 2,25 | 1159 | 5,06 | 1343281 | 2607,75 |
| 10 | 2,25 | 1167 | 5,06 | 1361889 | 2625,75 |
| 11 | 2,00 | 1130 | 4,00 | 1276900 | 2260,00 |
| 12 | 2,00 | 1075 | 4,00 | 1155625 | 2150,00 |
| 13 | 2,00 | 1047 | 4,00 | 1096209 | 2094,00 |
| 14 | 1,75 | 965 | 3,06 | 931225 | 1688,75 |
| 15 | 1,75 | 943 | 3,06 | 889249 | 1650,25 |
| 16 | 1,75 | 958 | 3,06 | 917764 | 1676,50 |
| 17 | 1,75 | 971 | 3,06 | 942841 | 1699,25 |
| 18 | 1,75 | 949 | 3,06 | 900601 | 1660,75 |
| 19 | 1,75 | 884 | 3,06 | 781456 | 1547,00 |
| 20 | 1,75 | 866 | 3,06 | 749956 | 1515,50 |
| 21 | 1,75 | 876 | 3,06 | 767376 | 1533,00 |
| 22 | 1,75 | 822 | 3,06 | 675684 | 1438,50 |
| 23 | 1,75 | 704 | 3,06 | 495616 | 1232,00 |
| 24 | 1,75 | 719 | 3,06 | 516961 | 1258,25 |
| Jumlah | 49,75 | 25.682,00 | 105,94 | 28.503.296,00 | 54.822,00 |
| Rataan | 2,07 | 1070,08 |  |  |  |

Rumus Dasar

Mencari nilai slop

Mencari nilai intercept

Persamaan Regresi Hitungan Manual

Y = -96,75 + 563,689X

Langkah 3. Hitung nilai regresi (menggunakan nilai x1 dan y) secara komputasi

# Deklarasi Pustaka Scikit Learn

**from** sklearn**.**linear\_model **import** LinearRegression

# transformasi data x1 menjadi array 2D

x1\_matrix **=** x1**.**values**.**reshape**(-**1**,**1**)**

x1\_matrix

# Output

array**([[**2.75**],**

**[**2.5 **],**

**[**2.5 **],**

**......,**

**[**1.75**],**

**[**1.75**],**

**[**1.75**]])**

# membuat model regresi

reg1 **=** LinearRegression**()**

reg1**.**fit**(**x1\_matrix**,**y**)**

# Output

LinearRegression**()**

# Menghitung nilai intercept

reg1**.**intercept\_

# Output

**-**99.46431881371632

# Menghitung nilai Slop

reg1**.**coef\_

# Output

array**([**564.20389249**])**

Persamaan Regresi Hitungan Komputasi

Y = -99.4643 + 564.2038X

Persamaan Regresi Hitungan Manual

Y = -96,75 + 563,689X

Perbedaan model regresi karena pembuatan kode program menggunakan pustaka scikit learn. Penentuan model regresi dengan menambahkan konsep Ordinary Least Square (penentuan model terbaik berdasarkan nilai eror terkecil).

Dokumentasi resmi scikit learn Regresi Linier dengan OLS

<https://scikit-learn.org/stable/modules/linear_model.html>

Dokumentasi resmi scikit learn penerapan studi kasus

<https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html>

Langkah 4. Membuat prediksi dengan persamaan regresi manual

Diketahui persamaannya

Y = -99.4643 + 564.2038x

Ketika x = 1.75 maka -99.45643 + 564,2038(1,75) = 887.8924

Ketika x = 2.00 maka -99.45643 + 564,2038(2.00) = 1.028,9434

Ketika x = 2.25 maka -99.45643 + 564,2038(2.25) = 1.169,9944

Ketika x = 2.50 maka -99.45643 + 564,2038(2.50) = 1.311,0454

Ketika x = 2.75 maka -99.45643 + 564,2038(2.75) = 1.452,0963

Langkah 4. Membuat prediksi dengan persamaan regresi komputasi

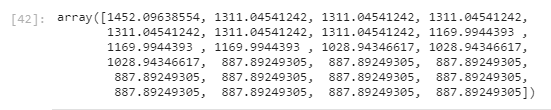
# Membuat model prediksi dengan inputan x

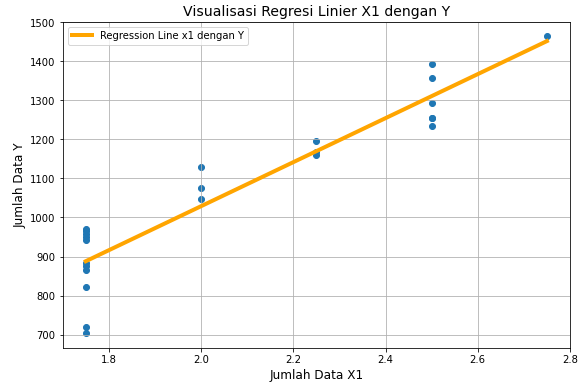
y1\_hat **=** reg1**.**predict**(**x1\_matrix**)**

# Menampilkan hasil prediksi

**print(**y1\_hat**)**

# Hasil prediksi





Gambar 5. Hasil Prediksi Regresi Linier

**Membuat Kesimpulan Akhir**

Hitung determinasi dan signifikansi x1 dan y menggunakan pustaka statmodels

# Deklarasi pustaka

**import** statsmodels**.**api **as** sm

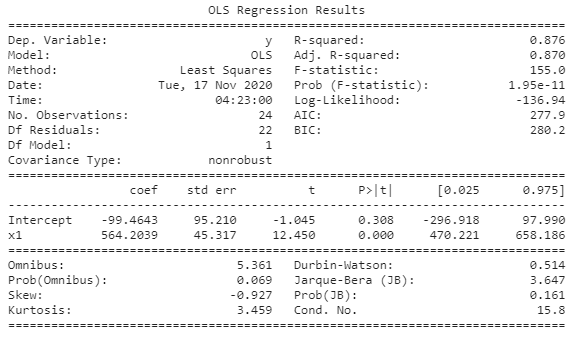
# Membuat model regresi linier sederhana

model\_lin **=** sm**.**OLS**.**from\_formula**(**"y ~ x1"**,** data**=**df**)**

result\_lin **=** model\_lin**.**fit**()**

# Menampilkan hasil regresi

**print(**result\_lin**.**summary**())**



Diketahui bahwa determinasi 0.876. Artinya 87,6% perubahan nilai y dipengaruhi oleh nilai x1. sisanya sekitar 12,4% dipengaruhi oleh variabel lain.

Dekatahui bahwa perubahan variabel x berpengaruh secara nyata dengan perubahan nilai y. Karena signifikansi (p-value) kurang dari 0.005.

**TUGAS.**

Bagaimana hitungan manual regresi linier apabila variabel x2 dan y?

Bagimana cara akurasi prediksi regresi linier dengan residual dan nilai rsme.?

Bagaimana kesimpulan yang didapat?