



## IMPLEMENTASI MODEL REGRESI LINEAR SEDERHANA UNTUK PREDIKSI GAJI BERDASARKAN PENGALAMAN LAMA BEKERJA

Yayan Adrianova Eka Tuah<sup>1</sup>, Anyan<sup>2</sup>

<sup>1,2</sup>Program Studi Pendidikan Komputer, STKIP Persada Khatulistiwa Sintang  
Email: [yayan.adrianova.eka.tuah89@gmail.com](mailto:yayan.adrianova.eka.tuah89@gmail.com)<sup>1</sup>, [anyanright@gmail.com](mailto:anyanright@gmail.com)<sup>2</sup>

Informasi Artikel	ABSTRACT
<p><b>Riwayat artikel :</b> Disubmit : 27 Nopember 2020 Direvisi : 11 Desember 2020 Diterima : 28 Desember 2020 Dipublikasi : 30 Desember 2020</p>	<p><i>The company cannot be separated from the workforce. Even though the company has advanced technology and large capital. One of the important factors to boost the performance of the workforce is the provision of appropriate salaries based on the length of time they work. This study aims to determine the prediction of employee salaries based on years of service. In this study, the factors that were tested included independent variables in the form of salary and years of service. Then the dependent variable is employee performance. This type of research is ex-post facto research to find out events that have occurred in the company as predictors of employee performance. Methods of data collection using a questionnaire (questionnaire). Instrument validation uses expert opinion and product-moment correlation. The data analysis technique uses linear regression analysis in python machine learning to determine the effect of the independent variable on the dependent variable. The results obtained from this study are a positive and significant influence between salary and years of service on employee performance. The conclusion that can be drawn is that the independent variable can be used as a predictor of the dependent variable. This means that the greater the salary and the longer the working period, the better the employee's performance will be.</i></p>
<p><b>Keywords:</b> Salary, Work experience, Performance, Linear Regression</p>	
<p><b>Kata Kunci:</b> Gaji, Pengalaman Bekerja, Kinerja, Regresi Linear</p>	<p style="text-align: center;"><b>ABSTRAK</b></p> <p>Perusahaan tidak dapat dipisahkan dengan tenaga kerja. Sekalipun perusahaan mempunyai teknologi canggih dan modal besar. Salah satu faktor penting untuk mendongkrak kinerja tenaga kerja adalah pemberian gaji yang sesuai berdasarkan lama waktu mereka bekerja. Penelitian ini bertujuan untuk mengetahui prediksi gaji karyawan berdasarkan tahun lama masa kerja. Dalam penelitian ini, faktor-faktor yang dilakukan pengujian di antaranya variabel independen berupa gaji dan masa kerja. Kemudian variabel dependen berupa kinerja karyawan. Jenis penelitian ini adalah penelitian ex-post facto untuk mengetahui peristiwa yang sudah terjadi dalam perusahaan sebagai prediktor kinerja karyawan. Metode pengumpulan data menggunakan metode angket (kuesioner). Validasi instrumen menggunakan pendapat ahli dan korelasi product moment. Teknik analisis data menggunakan analisis regresi linear dalam python machine learning untuk mengetahui pengaruh variable bebas terhadap variable terikat. Hasil yang didapatkan dari penelitian ini adalah adanya pengaruh positif dan signifikan antara gaji dan masa kerja terhadap kinerja karyawan. Simpulan yang dapat diambil adalah variable bebas dapat digunakan sebagai prediktor variable terikat. Artinya semakin besar gaji dan semakin lama masa kerja maka kinerja karyawan akan semakin baik.</p>



## PENDAHULUAN

Trend perkembangan ilmu pengetahuan dan teknologi saat ini adalah revolusi industri 4.0. Integrasi teknologi digital berbasis *cyber physical system* pada hampir semua bidang pekerjaan menjadi ciri hadirnya revolusi industri tersebut (Schwab, 2016). Teknologi ini memungkinkan terhubungnya manusia dengan manusia, manusia dengan mesin, dan mesin dengan mesin (Devi et al., 2020; Xu et al., 2018). Perubahan karakteristik pekerjaan menjadi dampak tersendiri dari hadirnya revolusi industri 4.0 (Shahroom & Hussin, 2018). Karakteristik pekerjaan yang berubah akan mendisrupsi pekerjaan yang telah ada dan menggantikannya dengan pekerjaan dengan karakteristik baru (Kergroach, 2017). Karakteristik baru pada pekerjaan juga membutuhkan kompetensi baru pada para pekerja (Manda & Dhaou, 2019; Sharma, 2019). Dengan demikian, sangat penting bagi suatu negara untuk mengembangkan kompetensi berbasis industri 4.0 pada sumber daya manusianya. Sekarang ini, dalam dunia bisnis, setiap perusahaan harus bersiap untuk saling bersaing. Manajemen yang efektif dan keunggulan perlu dimiliki oleh setiap perusahaan untuk menyambut persaingan itu. Dalam hal ini, salah satu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja dari karyawannya. Faktor yang berpengaruh secara internal dapat terlihat dari penentuan gaji yang tepat oleh sebuah perusahaan kepada karyawan. Di samping itu juga perusahaan juga harus mau mengeluarkan bonus bagi para karyawannya yang berhasil bekerja maksimal sesuai yang dibutuhkan oleh perusahaan. Sementara itu, faktor yang berpengaruh secara eksternal dapat terlihat dari bagaimana sebuah perusahaan mampu memberikan pengembangan dan pelatihan pada para karyawannya. Dalam hal ini, perusahaan tidak dapat dipisahkan dengan tenaga kerja. Sekalipun perusahaan mempunyai teknologi canggih dan modal besar. Dengan tidak terdapatnya tenaga manusia sebagai penggerak, perusahaan tidak akan mampu meraih tujuan yang direncanakan. Oleh karenanya, diharapkan sumber daya manusia dapat mengalami peningkatan tahun demi tahun, seiring dengan berkembangnya teknologi dan majunya zaman. Salah satu masalah yang dihadapi adalah bagaimana mengetahui gaji yang akan dikeluarkan berdasarkan lama bekerja. Dalam penelitian ini, usulan penggunaan metodenya adalah dengan menggunakan perhitungan regresi sederhana. Pemilihan metode ini bertujuan agar nilai variabel terikat dapat diketahui dengan cara melakukan perhitungan pengaruh dari beberapa variable bebas. Pengembangan sistem ini, harapannya memberikan bantuan dalam memberikan prediksi kisaran gaji yang akan ditentukan berdasarkan berapa lama karyawan tersebut bekerja.



## METODE PENELITIAN

Penelitian ini merupakan jenis penelitian deskriptif kualitatif yang menggunakan regresi linear sederhana dengan menggunakan Bahasa *Python*. Penelitian deskriptif adalah penelitian yang memberikan atau menggambarkan suatu keadaan atau fenomena yang ada dengan menggunakan proses ilmiah untuk menjawab permasalahan dunia nyata. Analisis data dilakukan secara kualitatif, termasuk mencari permasalahan yang lebih mendalam tentang suatu fenomena, fakta, atau peristiwa. Fakta, masalah, gejala dan peristiwa hanya dapat dipahami oleh peneliti bila diamati dan tidak dibatasi oleh ide-ide yang dangkal. Oleh karena itu, peneliti deskriptif kualitatif sering digunakan untuk menganalisis peristiwa, fenomena, atau situasi sosial.

### Pemrograman Python

Python merupakan bahasa pemrograman interpretatif multiguna yang menggunakan filosofi rancangannya berfokus pada tingkat keterbacaan kode. Sebagai bahasa pemrograman, Python menggabungkan kemampuan, kapabilitas dan sintaksis kode serta fungsi pustaka yang bermutu tinggi (Asadi, 2015). Secara umum, bahasa pemrograman hasil rancangan dari Guido van Rossum ini sesungguhnya sangat banyak dipakai oleh masyarakat umum dalam pembuatan program. Misalnya, aplikasi smartphone, program CLI, program GUI (desktop), game, web, IoT, program untuk hacking dan sebagainya (*A Python Book: Beginning Python, Advanced Python, and Python Exercises*. web.archive.org). Aplikasi kasir adalah program kasir, fungsinya adalah sebagai alat bantu dalam proses transaksi penjualan bisnis. Fungsi pokok aplikasi kasir, yang merupakan bentuk penggantian posisi mesin kasir konvensional, adalah membantu proses transaksi seperti pembayaran secara integratif. Aplikasi ini menggunakan metode *Modular programming* dimana *Modular programming* adalah teknik pemrograman dengan memisahkan kode menjadi beberapa bagian yang disebut dengan *modules*. Tujuannya untuk mengurangi ketergantungan antar modul. Teknik ini biasanya digunakan untuk software berskala besar. Software ini memiliki banyak modul di *built* sendiri-sendiri yang kemudian digabungkan untuk menjalankan fungsionalitas software tersebut. Dalam bahasa python, modul diimplementasikan dalam file berekstensi *.py* yang berisi sekumpulan kode untuk menjalankan *task* tertentu. Modul dapat berisi class, function, statement, atau kode-kode yang lain. Salah satu keuntungan penggunaan modul adalah *reusable*. Artinya modul ini dapat dipanggil oleh modul lain yang berbeda-beda.

### Regresi Linear

Regresi linear (*linear regression*) adalah metode yang dipakai untuk memperoleh model hubungan antara variabel dependen dan variabel independent (Walpole et al., 2012). Terdapat 2 jenis regresi linear yaitu, regresi linear sederhana (*simple linear regression*) dan regresi linear ganda (*multiple linear*



*regression*). Perbedaan terdapat pada banyaknya variabel independen. Regresi linear sederhana adalah metode yang dipakai untuk mendapatkan model hubungan antara satu variabel dependen dan satu variabel independen, sedangkan regresi linear ganda metode yang digunakan untuk memperoleh model hubungan antara satu variabel dependen dengan lebih dari satu variabel independen (Walpole et al., 2012). Penganalisisan regresi sederhana adalah hubungan antara variabel berjumlah dua, khususnya *independent variable* (variabel bebas), dan *dependent variable* (variabel tak bebas), sementara regresi dobel atau berganda adalah hubungan antara setidaknya tiga variabel, khususnya sekitar dua variabel bebas dengan satu variabel tak bebas.

a) Regresi Linier Sederhana

Penganalisisan regresi sederhana dipakai dalam menumbuhkan suatu persamaan yang menampilkan bagaimana hubungan antara variabel bebas (independen) dan variabel terikat (dependen), dan untuk mengantisipasi nilai variabel terikat jika nilai variabel bebasnya bertambah atau berkurang. Persamaan yang memaparkan bagaimana hubungan antara faktor X dan variabel Y dan model kesalahan dikenal sebagai model regresi. Model regresi yang dipakai dalam regresi linear sederhana adalah:

:  $Y =$

dengan:

$b_0$  dan  $b_1$  = parameter model

X = variabel independen

a = parameter intercept

b = parameter koefisien regresi variabel bebas

b) Scikit-learn

Scikit-learn merupakan modul Python yang menyediakan pelbagai jenis algoritma machine learning. Bentuk library python merupakan bentuk yang tersedia dalam Scikit-learn. Scikit-learn memanfaatkan task-oriented interface yang konsisten sehingga memudahkan dalam membandingkan antarmetode (Pedregosa, et al., 2011). Dengan prediksi gaji terhadap lamanya bekerja dapat dilakukan dengan model regresi linear sederhana, dimana makna dari regresi sendiri adalah alat estimasi yang dipakai untuk melakukan pengukuran ada atau tidaknya keterkaitan antarvariabel. Dalam penganalisisan regresi, kondisi regresi atau kondisi penilai dibentuk untuk memperjelas contoh keterkaitan variabel apakah ada benaran keterkaitannya satu sama lain. Keterkaitan yang diperoleh secara umum mengungkapkan hubungan praktis antara variabel-variabel yang dimaksud. 'Regresi' sendiri kali pertama dipopulerkan oleh seorang pakar yang namanya adalah Fancis Galton pada tahun 1886. Menurut Galton (F.



Galton, 1886), analisis regresi berkaitan dengan studi ketergantungan dari suatu variabel yang disebut variabel tak bebas (*dependent variable*), pada satu atau variabel yang menjelaskan dengan tujuan untuk memprediksi ataupun meramal nilai akhir dari variabel tak bebas apabila sudah diketahui nilai variabel yang menjelaskannya. Variabel yang menerangkan sering disebut variabel bebas (*independent variable*). Untuk mempelajari hubungan-hubungan antara beberapa variabel, analisis regresi dapat dilihat dari dua bentuk yaitu: (1). Analisis Regresi Sederhana (*Simple Regression*) (2). Analisis Regresi Berganda (*Multiple Regression*).

## HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam hal ini adalah pengaruh gaji seseorang karyawan terhadap waktu kerja karyawan di perusahaan. Kumpulan data diambil, diurutkan dan dikonversikan ke angka dan Bahasa Indonesia untuk memudahkan dalam melakukan analisis terhadap pengaruh gaji karyawan. (Jogiyanto, 1999:129), mendefinisikan analisis sistem merupakan proses mengurai dari suatu sistem informasi yang utuh ke dalam bagian-bagian elemennya bertujuan untuk melakukan pengidentifikasian dan evaluasi masalah-masalah yang terjadi, kesempatan-kesempatan, hambatan-hambatan dan keperluan-keperluan yang diinginkan sehingga dapat dijadikan sebagai usulan untuk memperbaiki. Mulyadi (2001:377) mengemukakan, “gaji merupakan pembayaran atas penyerahan jasa yang dilakukan oleh karyawan baik yang mempunyai jabatan maupun karyawan pelaksana”. Seorang karyawan menghasilkan gaji berupa:

### 1. Gaji Pokok

Ukuran pemberian gaji kepada pekerja adalah sesuai dengan posisi dan administrasi yang diberikan kepada organisasi dan kompensasi dasar telah diselesaikan ketika pekerja melakukan pekerjaan pertama kali.

### 2. Insentif

#### a. Uang makan dan transportasi

Merupakan ekspansi biaya seorang pekerja di samping gaji pokok dan ditentukan tergantung pada level dan posisi sesuai keahliannya dengan menghitungnya setiap hari tetapi diberikan kepada masing-masing saat gaji dibayarkan.

#### b. Uang lembur

Menurut KEPMEN Pasal 1 Ayat 1 (2004:1), “waktu kerja lembur adalah waktu kerja yang melebihi 7 (tujuh) jam sehari dan 40 (empat puluh) jam 1 (satu) minggu untuk 6 (enam) hari kerja dalam 1 (satu) minggu atau 8 (delapan) jam sehari dan 40 (empat puluh) jam 1



(satu) minggu untuk 5 (lima) hari kerja dalam 1 (satu) minggu atau waktu kerja pada hari istirahat mingguan dan atau pada hari libur resmi yang ditetapkan.”

Tabel 1. Pengalaman Kerja Karyawan

Tahun_bekerja	Gaji
1.1	39343
1.3	46205
1.5	37731
2	43525
2.2	39891
2.9	56642
3	60150
3.2	54445
3.2	64445
3.7	57189
3.9	63218
4	55794
4	56957
4.1	57081
4.5	61111
4.9	67938
5.1	66029
5.3	83088
5.9	81363
6	93940
6.8	91738
7.1	98273
7.9	101302
8.2	113812
8.7	109431
9	105582
9.5	116969
9.6	112635
10.3	122391
10.5	121872

Pengalaman kerja karyawan yang dikumpulkan merupakan tahun bekerja dan gaji. Tahun bekerja yang diperoleh merupakan wawancara hasil perkiraan tahun bekerja dimulai dari tahun pertama bekerja.

### Data Preprocessing

Hasil dari pengumpulan data yang dilakukan akan didaftar ke dalam tabel. File berformat CSV merupakan data yang dipergunakan sebagai data training dalam melakukan penghitungan terhadap prediksi. File berformat csv tersebut di-save ke dalam folder dengan nama static dan dilabeli dengan Daftar\_gaji.csv. Perubahan nilai data disesuaikan dengan mengacu tabel 1.



```
#Melihat Data Set Teratas
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math

dataset = pd.read_csv("D:\\Daftar gaji.csv")
dataset.head()
```

Gambar 1. Skript *import* dataset gaji karyawan pada *Python*

	Tahun_bekerja	Gaji
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

Gambar 2. Hasil keluaran program

`dataset.describe()` berfungsi untuk melakukan analisis deskriptif secara otomatis terhadap dataset yang dipilih.

## Pengujian

Pembahasan pengujian dalam penelitian yang dilakukan ini adalah melakukan perbandingan pengaruh variabel bebas terhadap variabel terikat dengan mengacu korelasi antara tiap-tiap variabel bebas (x) dan variabel terikat (y). Pengujian dengan melakukan pengukuran tingkat ketepatan pada model multiple linear regression. Pengujian Model Linear regression. Untuk melakukan pengujian data secara menyeluruh, metode perhitungan akurasi sendiri akan diberlakukan. Penerapan metode melakukan perhitungan nilai total jumlah prediksi yang benar pada data memanfaatkan multiple linear regression. Untuk menetapkan koefisien-koefisien regresi ( ) dapat diketahui dengan memakai kode berikut:



```
# Impor library yang
dibutuhkan import numpy as
np import matplotlib.pyplot
as plt import pandas as pd

# Impor dataset
dataset = pd.read_csv('D:\\Daftar gaji.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values

# Membagi data menjadi Training Set dan Test Set from
sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=1 / 3, random_state=0)

# Fitting Simple Linear Regression terhadap Training
set from sklearn.linear_model import LinearRegression

regressor = LinearRegression()
regressor.fit(X_train, y_train)

# Memprediksi hasil Test Set
y_pred = regressor.predict(X_test)

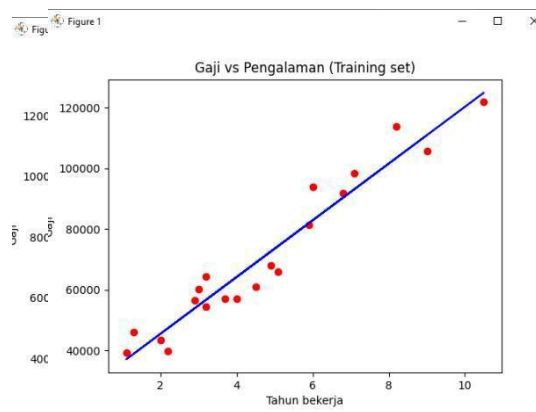
# Visualisasi hasil Training Set plt.scatter(X_train,
y_train, color='red') plt.plot(X_train,
regressor.predict(X_train), color='blue') plt.title('Gaji vs
Pengalaman (Training set)') plt.xlabel('Tahun bekerja')

plt.ylabel('Gaji')
plt.show()

# Visualisasi hasil Test Set plt.scatter(X_test, y_test,
color='red') plt.plot(X_train, regressor.predict(X_train),
color='blue') plt.title('Gaji vs Pengalaman (Test set)')
plt.xlabel('Tahun bekerja')

plt.ylabel('Gaji')
plt.show()
```

Gambar 3. Skrip pengujian model regresi linear



Gambar 4. Grafik prediksi gaji berdasarkan lamanya bekerja

Keterangan:

- Line 2 hingga 4 adalah mengimport library yang dibutuhkan.



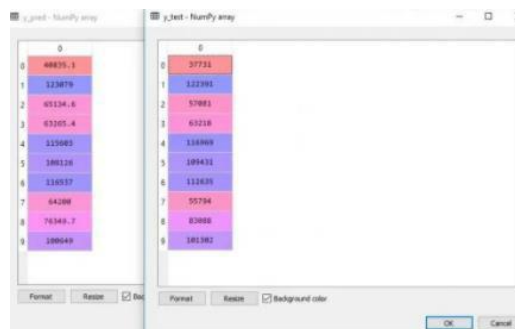


- Line 7 mengimport data ke python.
- Line 8 menetapkan variabel independent (sumbu x) yakni kolom pertama/1 (Tahun\_bekerja). Perlu diberikan perhatian, saat melakukan slicing memakai method `.values`, hal ini hanya akan memotong data tanpa headernya. Jika tanpa `.values` maka header dari slicing akan diturutsertakan.
- Line 9 menetapkan variabel dependen (sumbu y) yakni kolom ke2 (Gaji).
- Line 12 mengimport library untuk memisahkan menjadi train set dan test.
- Line 13 memisahkan menjadi train set dan test, dengan train setnya adalah 2/3 dari dataset yang ada. Anda bisa mengklik bagian Variable explorer di spyder untuk mengetahui hasil bagi antara train dan test setnya.
- Line 16 mengimport class `LinearRegression` dari library `sklearn.linear_model` yang dibutuhkan dalam pembuatan model regresi.
- Line 17 menciptakan objek regressor sebagai fungsi dari `LinearRegression`. Hanya dengan menuliskan `LinearRegression()`, model regresi sudah dipersiapkan. Jika Anda kursor diarahkan pada `LinearRegression` kemudian mengklik CTRL + i maka akan tampil object inspector untuk `LinearRegression`. Di tahap ini, Anda bisa mengetahui parameter apa saja yang dibutuhkan.
- Line 18 menciptakan model regresi untuk train set dengan menulis `regressor.fit(X_train, y_train)`. Untuk mengetahui parameter apa saja yang dibutuhkan dalam metode `fit()`, hanya dengan mengarahkan kursor pada `.fit()` lalu mengklik CTRL+i.
- Line ke 18 adalah proses pembuatan model machine learning regresi. Berarti, dalam hal ini model tengah belajar untuk menemukan hubungan antara `X_train` dan `y_train`.
- Keluaran pada python akan terlihat sebagai berikut.  
*Out (12): Linear Regression(copy\_X=True, fit\_intercept=True, n\_jobs=1, normalize=False).*  
 Setelah machine learning mempelajari hubungan antara `X_train` dan `y_train`, maka tahapan berikutnya adalah melakukan percobaan dengan membuat prediksi ke depan. Prediksi ini memakai hubungan yang telah dipelajari oleh model. Perintah untuk melaksanakan prediksi ini dilakukan pada line 21.
- Line 21 memprediksi dengan memakai metode `.predict`. Dalam rangka mengetahui parameter apa saja, dapat dilakukan melalui object inspector. Parameter yang dibutuhkan adalah variabel independent, dalam hal ini adalah `X_test` dan bukan `X_train`. Kenapa seperti itu? Sebab ingin membuat prediksi data baru. Jika memakai `X_train` perlu membuat prediksi berdasarkan



pemahaman  $X_{train}$ , sedangkan pemahaman itu sendiri dibuat mengacu  $X_{train}$ . Memakai  $X_{test}$ . Nanti, prediksi ini dibandingkan dengan  $y_{test}$ . Apabila hasil yang didapatkan mendekati (jarak tidak terlalu jauh), maka model dianggap baik.

Setelah itu, jika telah selesai dilakukan, berikut hasil tersebut bisa dibandingkan antara hasil  $y_{pred}$  dan  $y_{test}$ , dengan  $y_{pred}$  merupakan prediksi model dan  $y_{test}$  merupakan data sebenarnya.



Gambar 5. Hasil  $y_{pred}$  dan  $y_{test}$

Agar dapat melakukan perbandingan hasil antara  $y_{pred}$  dan  $y_{test}$ , hasilnya sangat baik apabila dilakukan pembuatan berupa ilustrasi visual. Hal yang harus diperhatikan, supaya penampilan visualnya baik di python, maka block semua line dari line 24 hingga line 29 lalu mengklik CTRL + ENTER.

- Line 24 membuat scatter plot  $X_{train}$  dan  $y_{train}$  dengan warna merah untuk data poinnya. Tentu saja warnanya bisa diubah sesuka hati.
- Line 25 membuat line plotnya (garis regresi) berwarna biru. Metode plot() memerlukan parameter pertama, yakni data poin untuk sumbu x, dan parameter kedua merupakan data poin untuk sumbu y. Data poin sumbu x adalah  $X_{train}$  sebab hendak melihat model regresi dari training set, sedangkan data poin sumbu y merupakan prediksi dari  $X_{train}$  dengan perintah regressor.predict( $X_{train}$ ). Harus diperhatikan, data poin sumbu y bukan  $y_{predict}$  atau regressor.predict( $X_{test}$ ), sebab, kali ini bertujuan membuat plot regresi untuk  $X_{train}$ , bukan  $X_{test}$ . Misalnya, Anda melakukan kesalahan, memakai regressor.predict( $X_{test}$ ), maka garis regresinya tidak akan muncul. Kenapa? Sebab, jumlah data poin  $X_{train}$  (20 baris) dan  $X_{test}$  (10 baris) sudah memiliki perbedaan, maka tidak mungkin bisa dibuat garis regresinya.
- Line 26 membuat judul yang akan tampil pada bagian teratas grafik.
- Line 27 membuat label untuk sumbu x.
- Line 28 membuat label untuk sumbu y.
- Line 29 melakukan eksekusi dan menampilkan hasil dari semua perintah dari Line 24 hingga Line 28.



```

# Mengimport dataset
dataset = read.csv('Daftar gaji.csv')

# Membagi dataset menjadi Training set and Test set
# install.packages('caTools') <-- jika belum mempunyai library nva
hapus tanda #
library(caTools
) set.seed(123)
split = sample.split(dataset$Gaji, SplitRatio = 2 /
3) training_set = subset(dataset, split == TRUE)
test_set = subset(dataset, split == FALSE)

# Feature Scaling
# training_set = scale(training_set)
# test_set = scale(test_set)

# Fitting Simple Linear Regression ke the Training set
regressor = lm(Gaji
~ Tahun bekerja, data = training_set)

# Memprediksi hasil test set
y_pred = predict(regressor, newdata=test_set)

# Visualising the Training set
results library(ggplot2)
ggplot() +
geom_point(aes(x=training_set$Tahun bekerja, y = training_set$Gaji),
colour = 'red') +
geom_line(aes(x=training_set$Tahun bekerja, y = predict(regressor,
newdata=training_set)), colour = 'blue') +
qgtitle('Gaji vs Tahun bekerja (Training set)') +
xlab('Tahun bekerja') +
ylab('Gaji')

# Visualising the Test set
results library(ggplot2)
ggplot() + geom_point(aes(x=test_set$Tahun bekerja, y =
test_set$Gaji), colour = 'red') +
geom_line(aes(x=training_set$Tahun bekerja, y = predict(regressor,
newdata=training_set)), colour = 'blue') +
qgtitle('Gaji vs Tahun bekerja (Test set)') + xlab('Tahun bekerja')
+ ylab('Gaji')

# Analisis statistik model regresi
summary(regressor)

```

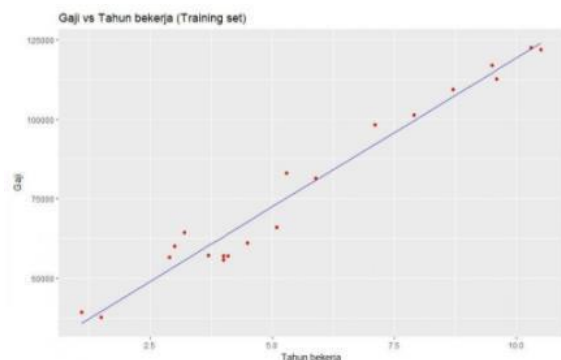
Gambar 6. Skrip analisis static model regresi

Keterangan:

- Line 2 mengimport dataset.
- Line 6 mengimport librarycaTools untuk melakukan pembagian data ke test set dan train.
- Line 7 menetapkan bilangan random (contohnya kali ini bilangan 123. Tentu saja angkanya bebas).
- Line 8 memberikan definisi variabel split untuk membagi test set dan train. Agar dapat melihat parameter apa saja yang dibutuhkan, cursor hanya harus diarahkan pada sample.split, lalu mengklik F1. Setelah itu memunculkan dokumentasi. Parameter pertama adalah variabel dependen (*Salary*), lalu rasio split secara default sebesar 2/3.
- Line 9 menetapkan bagian training set, dari hasil Boolean TRUE dari variabel split.



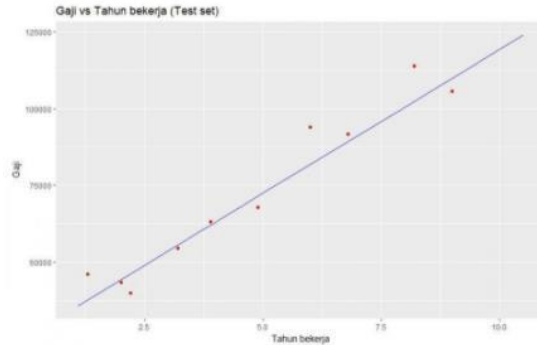
- Line 10 menetapkan bagian testiset, dari hasil booleani FALSE dari variabel split.
- Line 12 hingga Line 14 merupakan perintah feature scaling jika dibutuhkan. Namun, kali ini tidak membutuhkannya.
- Line 17 memberikan definisi objek regressor sebagai suatu model regresi. Di R, perintah membuat model regresi adalah dengan memakai fungsi lm(). Parameter pertama adalah variabel dependent (Gaji) diikuti dengan simbol ~ (disebut tilde). Lalu, setelah simbol ~ diikuti dengan variabel independen yaitu Tahun\_bekerja, dan definisikan data diambil dari training\_set.
- Line 20 merupakan prediksi pembelajaran machine learning untuk testiset. Line 20 ini sama dengan Line 21 y\_pred dalam perintah bahasa python yang telah disebutkan.
- Line 23 mengimport library ggplot2 untuk membuat visualisasi model regresi. ggplot2 adalah library yang terkenal di R, sebab dapat menampilkan visual yang beragam dan powerful. Harus diberikan perhatian, sebenarnya ggplot ditulis dalam satu baris, tetapi Anda melanjutkan melanjutkan penulisan di bawahnya jika layar komputer yang digunakan tidak dapat menampilkan semua perintah dalam satu baris. Apabila mengklik ENTER, maka otomatis akan dialihkan ke baris di bawahnya, dan R otomatis juga akan memberikan identitas (sedikit menjorok kanan).
- Line 24 perintah awal ggplot, dituruti dengan parameter berikutnya dengan simbol + (plus). Perintah ggplot() adalah menyiapkan objek visualisasi, yang apabila dituruti simbol +, maka seluruh komponen visualisasi setelah simbol + akan dieksekusi bersama.
- Line 25 membuat scatter plot dengan perintah geom\_point dituruti dengan aes (peningkatan dari aesthetic) untuk menetapkan mana sumbu x dan mana sumbu y, dan warna apa yang diinginkan.
- Line 26 membuat plot line dengan perintah geom\_line. Parameter sama dengan geom\_point.
- Line 27 memberi judul untuk grafik
- Line 28 memberi label untuk sumbu x
- Line 29 memberi label untuk sumbu y



Gambar 7. Hasil Regresi Training Set



- Line 32 hingga Line 38 menunjukkan grafik test set.



Gambar 8. Hasil Regresi *Test Set*

Agar dapat menjamin keakuratannya, hal yang diperlukan adalah pengujian statistik. Maksudnya adalah, R memang dibuat sebagai software statistik, sehingga pengujiannya hanya dengan menuliskan perintah `summary` pada line 41.

- Line 41 perintah mengetahui seberapa baik (fit) model regresi yang dibuat. Hanya dengan menuliskan `summary` (kemudian mengikut model regresi)

Eksekusi dari model regresi, yakni :

```
Call:
lm(formula = Gaji ~ Tahun_bekerja, data = training_set)

Residuals:
    Min       1Q   Median       3Q      Max
-7325.1 -3814.4  427.7  3559.7  8884.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  25592      2646   9.672 1.49e-08 ***
Tahun_bekerja  9365       421  22.245 1.52e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5391 on 18 degrees of freedom
Multiple R-squared:  0.9649,    Adjusted R-squared:  0.963
F-statistic: 494.8 on 1 and 18 DF,  p-value: 1.524e-14
```

Gambar 9. Eksekusi Model Regresi

Residual merupakan banyaknya data poin yang tidak dapat diberikan penjelasan dengan model. Gambarannya secara umum, makin kecil residual, makin kuatlah model regresi memberikan penjelasan hubungan antara variabel dependent dan independent. Min sebagai nilai minimal, 1Q merupakan residual kuartal pertama (persentil 25), median merupakan nilai tengah, 3Q adalah persentil 75, dan Max merupakan nilai paling besar. Bagian Coefficients dijelaskan sebagai berikut:

- Intercept adalah koefisien dari sebuah model regresi. Dalam hal ini, berarti tiap nilai X nol (karyawan belum berpengalaman kerja), ia menerima gaji sebanyak 25592 tiap tahun. Angka ini dapat dilihat pada bagian Estimate dari Intercept.



- Lalu, di bagian Tahun\_bekerja nilai 9365 mengungkapkan bahwa tiap penambahan satu tahun kerja, gaji yang diterima akan mengalami penambahan sebanyak 9365 tiap tahun.
- Persamaan regresi dapat dituliskan dalam bentuk berikut:

$$Gaji = 25592 + 9365 * Tahun Bekerja$$

- Std.Error (standar error) merupakan penyimpangan data prediksi dari data sebenarnya. Makin kecil std. error, tentu modelnya pun dianggap makin baik.
- t value merupakan pengujian statistik dari distribusi t. Nilai t di atas 2 atau kurang dari -2 umumnya diasumsikan signifikan.
- Pr adalah p value. Variabel independen (Tahun\_bekerja) mempunyai nilai p di bawah 0.05, hal ini menandakan ia signifikan melakukan prediksi variabel dependen (Gaji).
- Apabila ingin mengetahui seberapa baik, lihatlah bagian Multiple R-squared yang merupakan nilai dengan besarnya kisaran 0 hingga 1. Nilai nol (0), berarti error (jarak antara nilai prediksi dan nilai sesungguhnya) terlalu besar, sehingga model yang dipakai dianggap sangat buruk. Sebaliknya, apabila nilainya 1, modelnya dianggap sempurna, berarti prediksi yang diberikan dianggap akurat tanpa error sedikit pun.

Hal yang pasti, model regresi tidak mungkin mempunyai nilai 1. Sebab, apabila seperti yang disebutkan, modelnya diartikan terlalu sempurna. Model mempunyai nilai 0.96 tentu, modelnya sangat baik. Apabila melakukan pendekatan regresi, lalu nilainya sangatlah buruk, misalnya di bawah 0.5 atau bahkan dekat dengan angka nol, hal ini dapat dijamin keterkaitan antara variabel dependen dan independen tidaklah linear. Apabila benar seperti yang telah disebutkan, hal yang harus dilakukan adalah dengan *fitting* (melihat hubungan variabel dependen dengan independen) dengan model regresi polinomial, atau nonlinear.

## SIMPULAN

Berdasarkan pengujian menggunakan fungsi linear regression untuk memprediksi besarnya gaji (Y) dari pengalaman bekerja (X) diketahui bahwa semakin bertambahnya pengalaman bekerja seorang karyawan maka gaji karyawan tersebut juga akan bertambah setiap tahunnya. Hasil tersebut menunjukkan bahwa pengalaman bekerja memberikan pengaruh positif terhadap besarnya gaji karyawan. Hasil tersebut disimpulkan bahwa semakin lama pengalaman bekerja seorang karyawan maka semakin berpengaruh positif terhadap gaji yang akan diterima. Sehingga fungsi linear regression layak digunakan untuk memprediksi pengaruh pengalaman kerja terhadap gaji karyawan.



## DAFTAR RUJUKAN

- Asadi, A. (2015). The Python Book. In *Imagine Publishing Ltd*.
- Devi, M., Annamalai, M. A. R., & Veeramuthu, S. P. (2020). Literature education and industrial revolution 4.0. *Universal Journal of Educational Research*, 8(3), 1027–1036. <https://doi.org/10.13189/ujer.2020.080337>
- F. Galton. (1886). Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, Vol. 15, 246–263. <https://doi.org/10.2307/2841583>
- F. Pedregosa, G. Varouquaux, A. Gramfort, V. Michael, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cornapeau, M. Brucher dan M. P, A. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830. <https://doi.org/10.1289/EHP4713>
- Jogiyanto, H. (1999). *Analisis & Desain Sistem Informasi* (2nd ed.). CV ANDI OFFSET.
- Kergroach, S. (2017). Industry 4.0: New challenges and opportunities for the labour market. *Foresight and STI Governance*, 11(4). <https://doi.org/10.17323/2500-2597.2017.4.6.8>
- Manda, M. I., & Dhaou, S. Ben. (2019). Responding to the challenges and opportunities in the 4th industrial revolution in developing countries. *ACM International Conference Proceeding Series, Part F1481*. <https://doi.org/10.1145/3326365.3326398>
- Mulyadi. (2001). Sistem Akuntansi. In *Edisi 3*. Salemba Empat.
- Schwab, K. (2016). *The Fourth Industrial Revolution*. World Economic Forum.
- Shahroom, A. A., & Hussin, N. (2018). Industrial Revolution 4.0 and Education. *International Journal of Academic Research in Business and Social Sciences*. <https://doi.org/10.6007/ijarbss/v8-i9/4593>
- Sharma, P. (2019). Digital Revolution of Education 4.0. *International Journal of Engineering and Advanced Technology*, 9(2), 3558–3564. <https://doi.org/10.35940/ijeat.a1293.129219>
- Walpole, E. R., Myers, H. R., Myers, L. S., & Ye, K. (2012). *Probability and Statistics For Engineers and Scientists*. Mass: Pearson Education.
- Xu, M., David, J. M., & Kim, S. H. (2018). The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research*, 9(2). <https://doi.org/10.5430/ijfr.v9n2p90>