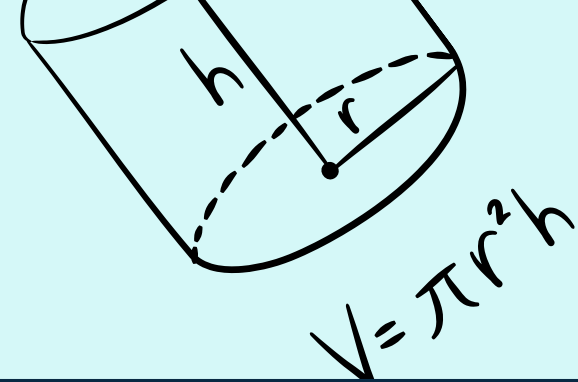



$$V = Lwh$$


$$V = \pi r^2 h$$

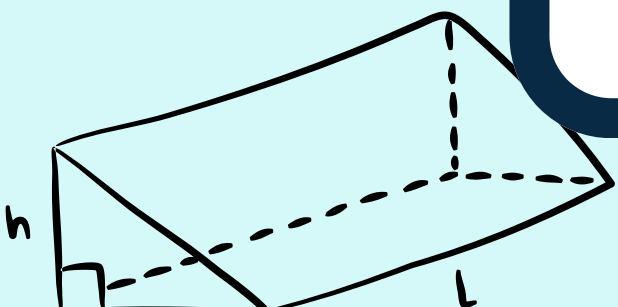
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Statistika & Probabilitas

# SIMPLE LINEAR REGRESION

$$y = mx + b$$

$$a = \frac{V_f - V_i}{t}$$


$$V = \frac{1}{2} bhl$$

$$\frac{x}{a} + \frac{y}{b} = 1$$

$$ax^2 + bx + c = 0$$



$$V = \frac{4}{3} \pi r^3$$

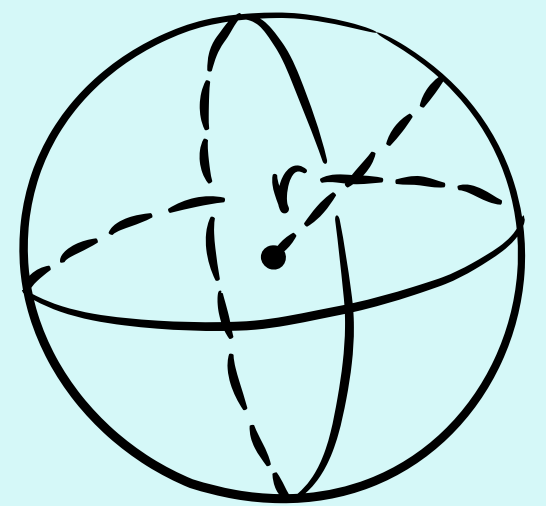
# NAMA KELOMPOK :

---

1. Achmad Dwiki Nomansyah (3122500054)
2. Aaron Febrian Prakoso (3122500060)

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$y = mx + b$$



$$V = \frac{4}{3} \pi r^3$$

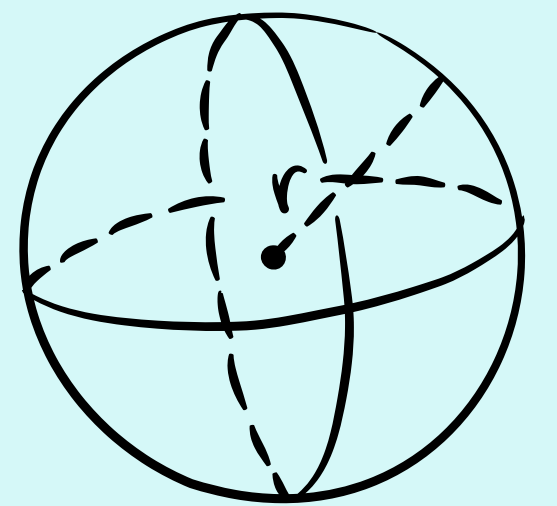
# STUDI KASUS:

---

Implementasi Simple Linear  
Regresion terhadap data gaji  
karyawan suatu perusahaan  
berdasarkan lama pengalaman  
masa kerjanya.

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$y = mx + b$$



$$V = \frac{4}{3} \pi r^3$$

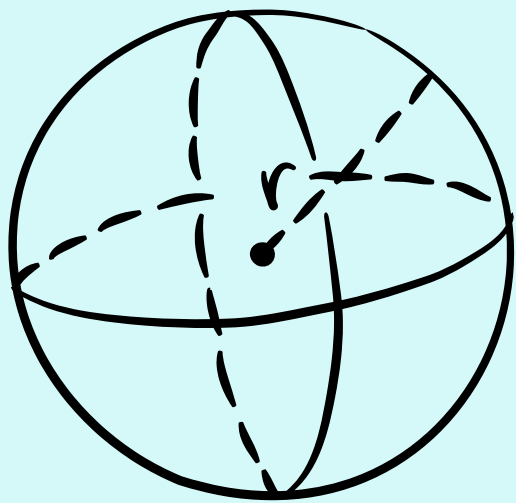
TABEL DATA GAJI KARYAWAN BERDASARKAN LAMA  
PENGALAMAN MASA KERJANYA:

YearsExperience	Salary
1.1	39343.00
1.3	46205.00
1.5	37731.00
2.0	43525.00
2.2	39891.00
2.9	56642.00
3.0	60150.00
3.2	54445.00
3.2	64445.00
3.7	57189.00
3.9	63218.00
4.0	55794.00
4.0	56957.00
4.1	57081.00
4.5	61111.00

YearsExperience	Salary
4.9	67938.00
5.1	66029.00
5.3	83088.00
5.9	81363.00
6.0	93940.00
6.8	91738.00
7.1	98273.00
7.9	101302.00
8.2	113812.00
8.7	109431.00
9.0	105582.00
9.5	116969.00
9.6	112635.00
10.3	122391.00
10.5	121872.00

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$y = mx + b$$

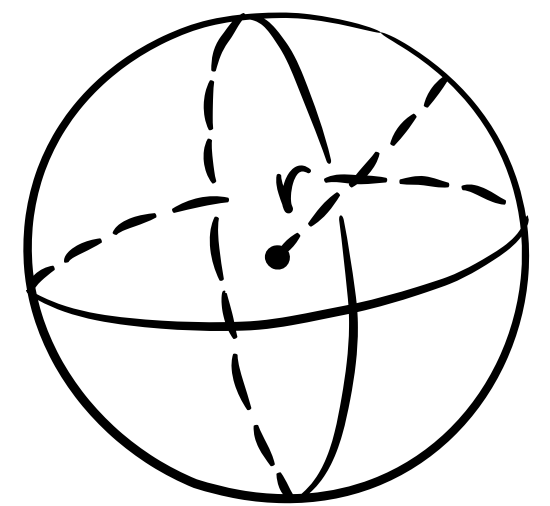


$$V = \frac{4}{3} \pi r^3$$

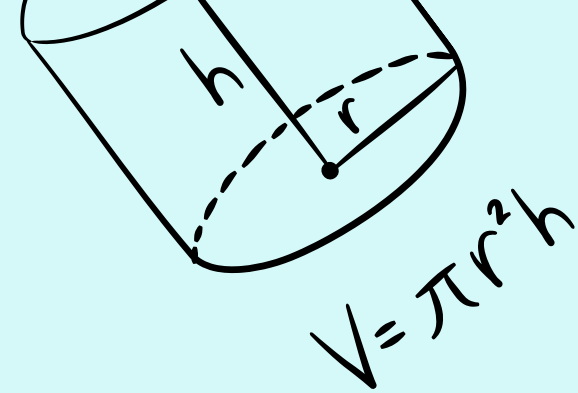
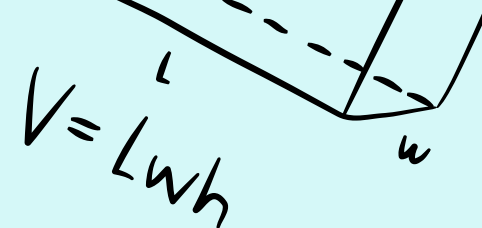
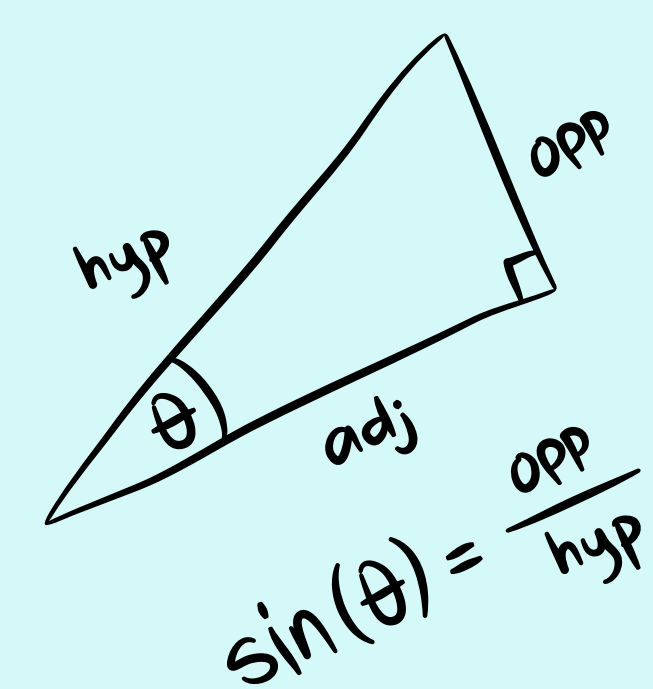
**Dari data tersebut, diketahui bahwa semakin lama pengalaman masa karyawan bekerja di perusahaan tersebut, maka semakin besar pula gajinya.**

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$y = mx + b$$



$$V = \frac{4}{3} \pi r^3$$

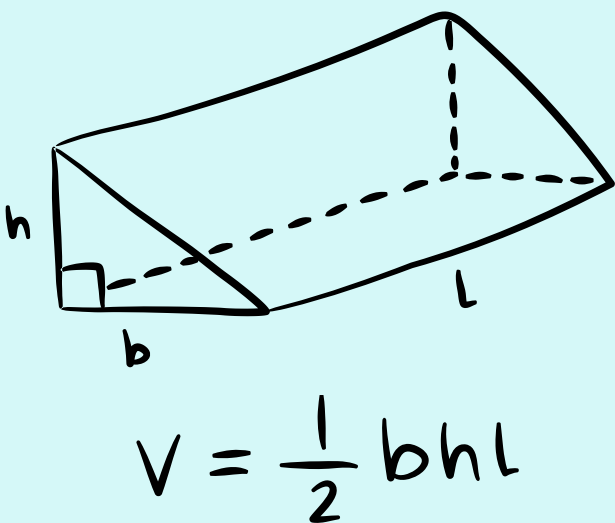


$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

# IMPLEMENTASI MENGGUNAKAN SIMPLE REGRESI LINEAR

a =

y + b



$$\frac{x}{a} + \frac{y}{b} = 1$$

$$ax^2 + bx + c = 0$$



# MELAKUKAN IMPORT FILE DATA (CSV) UNTUK MENAMPILKAN DATA YANG ADA PADA FILE TERSEBUT

✓ 2s # untuk melakukan impor file csv dan menampilkan datanya

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import seaborn as sns
```

```
from sklearn.linear_model import LinearRegression # penggunaan regresi linear untuk memodelkan hubungan antara variabel dalam data 'Salary_dat
```

```
from sklearn.model_selection import train_test_split # pemisahan data ke dalam data latih dan data uji
```

```
df = pd.read_csv('Salary_data.csv', delimiter=';') # delimiter untuk memisahkan dua kolom
```

```
df.head()
```



	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

# MELAKUKAN PERHITUNGAN JUMLAH BARIS DAN KOLOM TERHADAP DATAFRAME, SERTA MENAMPILKAN RINCIAN DATAFRAME TERSEBUT

```
[2] # mengetahui jumlah baris dan kolom dalam dataframe
df.shape
```

```
(30, 2)
```

```
# memberikan informasi secara rinci tentang dataframe
# termasuk tipe data setiap kolom, nilai non null, dan penggunaan memori
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   YearsExperience  30 non-null    float64
1   Salary          30 non-null    float64
dtypes: float64(2)
memory usage: 608.0 bytes
```



MELAKUKAN PERHITUNGAN JUMLAH DATA, MEAN, STANDAR DEVIASI, NILAI MINIMUM, KUARTIL BAWAH, MEDIAN, KUARTIL ATAS, DAN NILAI MAXIMUM

```
# count (jumlah data), mean (rata-rata), std (standar deviasi), min (nilai minimum), 25% (kuartil bawah),  
# 50% (median atau nilai tengah), 75% (kuartil atas), dan max (nilai maksimum).  
df.describe()
```

	YearsExperience	Salary
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

# MEMBUAT FIGURE DENGAN DUA SUBPLOT, YAITU COUNTPLOT DAN BOXPLOT

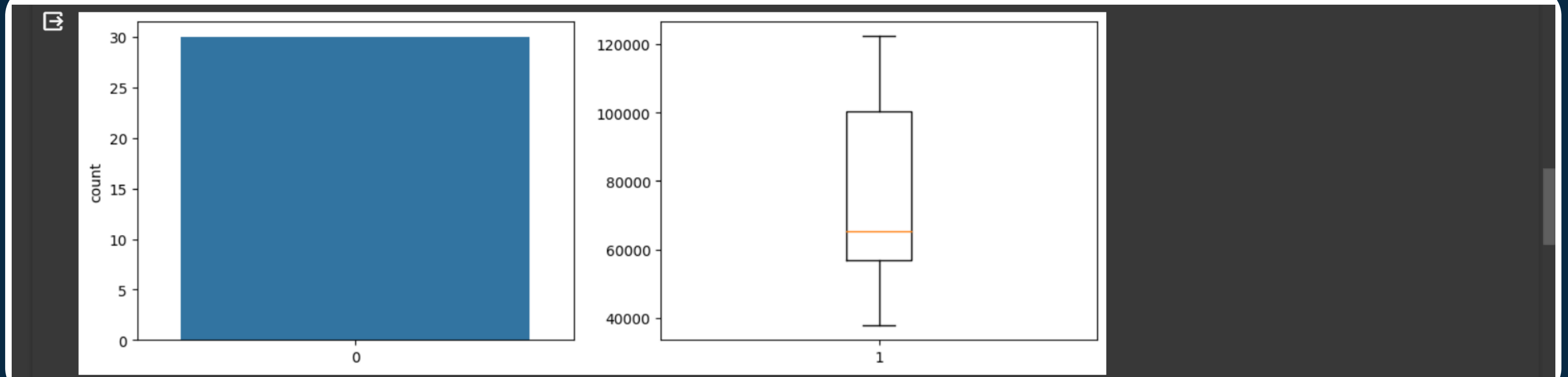
```
✓ 1s # Membuat figure dengan dua subplot
f = plt.figure(figsize=(12, 4))

# Subplot pertama: Countplot untuk 'YearsExperience'
f.add_subplot(1, 2, 1)
sns.countplot(df['YearsExperience'])

# Subplot kedua: Boxplot untuk 'Salary'
f.add_subplot(1, 2, 2)
plt.boxplot(df['Salary'])

# Menampilkan plot
plt.show()
```

# MEMBUAT FIGURE DENGAN DUA SUBPLOT, YAITU COUNTPLOT DAN BOXPLOT



# MELAKUKAN REGRESI LINEAR SEDERHANA PADA DATA

```
# melakukan regresi linier sederhana pada data
X = df["YearsExperience"] # kolom YearsExperience sebagai variabel independen
Y = df["Salary"] # kolom Salary sebagai variabel dependen
X_b = np.c_[np.ones((len(X), 1)), X] # menyiapkan matriks X dengan kolom konstan
beta_values = np.linalg.inv(X_b.T.dot(X_b)).dot(X_b.T).dot(Y)
print(beta_values)
```

```
[25792.20019867  9449.96232146]
```

# MENGGUNAKAN MODEL UNTUK MEMPREDIKSI SALARY BERDASARKAN YEARS EXPERIENCE

```
# Menggunakan model untuk memprediksi Salary berdasarkan YearsExperience
X_new = np.array([[1.1]]) # Prediksi gaji untuk YearsExperience 1.1
X_new_b = np.c_[np.ones((1, 1)), X_new] # Menyiapkan matriks untuk prediksi
y_predict = X_new_b.dot(beta_values) # Melakukan prediksi

# Menampilkan hasil prediksi
print(f"Years Experience = {X_new[0][0]}; Predicted Salary is {y_predict[0]:.2f}, actual Salary is 39343.00")

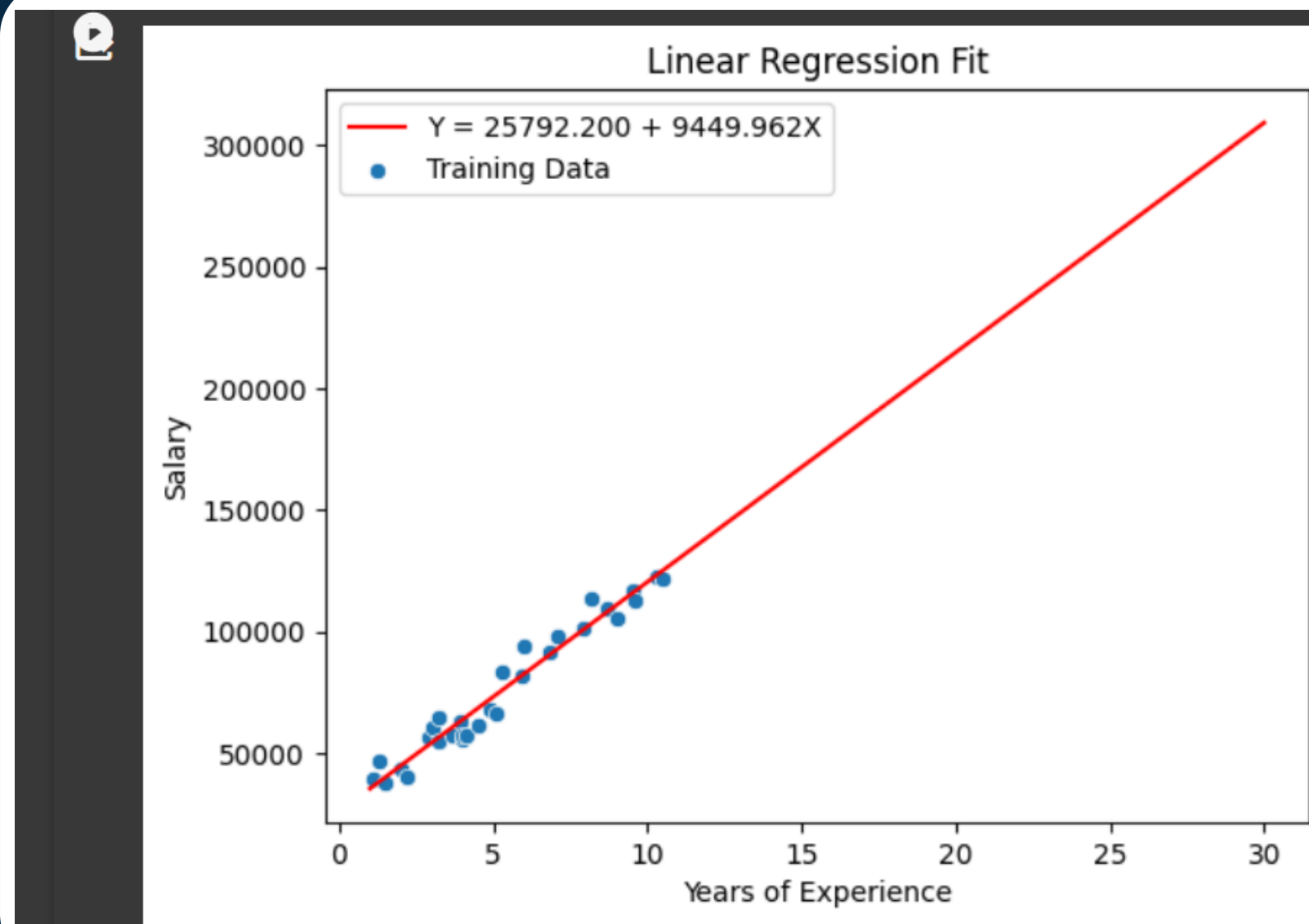
Years Experience = 1.1; Predicted Salary is 36187.16, actual Salary is 39343.00
```

$$\text{PREDICTED SALARY} = \beta_0 + \beta_1 \times \text{YEARS EXPERIENCE}$$

## MELAKUKAN PLOTTING YANG AKAN MEMPERLIHATKAN HUBUNGAN ANTARA 'YEAREXPERIENCE' DAN 'SALARY'

```
# Melakukan plotting yang akan memperlihatkan hubungan antara 'YearsExperience' dan 'Salary's
X_plot = np.array([[1, 30]]) # Dari 1 sampai 30 untuk plot garis regresi
X_plot_b = np.c_[np.ones((2, 1)), X_plot.T]
Y_plot = X_plot_b.dot(beta_values)
Equation_line = "Y = {:.3f} + {:.3f}X".format(beta_values[0], beta_values[1])
plt.plot(X_plot.T, Y_plot, "r-", label=Equation_line) # Plot garis regresi
sns.scatterplot(data=df, x='YearsExperience', y='Salary', label="Training Data") # Plot data sebenarnya
plt.legend()
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.title('Linear Regression Fit')
plt.show()
```

# MELAKUKAN PLOTTING YANG AKAN MEMPERLIHATKAN HUBUNGAN ANTARA 'YEAREXPERIENCE' DAN 'SALARY'



$a =$

$$V = \frac{1}{2} b h l$$

$$\frac{x}{a} + \frac{y}{b} = 1$$

$$ax^2 + bx + c = 0$$

$$V = \frac{4}{3} \pi r^3$$

$+ b$



# MELAKUKAN REGRESI LINEAR DENGAN MENGGUNAKAN PUSTAKA SCIKIT-LEARN UNTUK MENGETAHUI SEBERAPA BAIK MODEL REGRESI LINEAR

```
# Menggunakan pustaka scikit-learn untuk melakukan regresi linier.
from sklearn.linear_model import LinearRegression
reg = LinearRegression() # Menggunakan Linear Regression dari scikit-learn
#df = pd.read_csv("Salary_data.csv")

# Mengambil kolom YearsExperience dan Salary sebagai variabel X dan Y
X = df["YearsExperience"]
Y = df["Salary"]
# Mengubah bentuk X dan Y menjadi 2 dimensi (array 2D) karena itulah yang diharapkan oleh LinearRegression di scikit-learn
X = X.values.reshape(-1,1)
Y = Y.values.reshape(-1,1)
# Melatih model regresi linier
reg.fit(X, Y)
print("The value obtained for beta_0 is: ", reg.intercept_)
print("The value obtained for beta_1 is: ", reg.coef_)
r_squared = reg.score(X, Y)
print("R-squared: ", r_squared)
```

```
The value obtained for beta_0 is: [25792.20019867]
The value obtained for beta_1 is: [[9449.96232146]]
R-squared: 0.9569566641435086
```

$$V = \frac{1}{2} b h l$$

$$\frac{1}{a} + \frac{1}{b} = 1$$

$$ax^2 + bx + c = 0$$

$$V = \frac{1}{3} \pi r^3$$



# PREDIKSI GAJI BERDASARKAN YEARS EXPERIENCE 30 DAN 1

```
[11] # Prediksi gaji berdasarkan YearsExperience 30 dan 1
      X_new = np.array([[30],[1]])
      print(reg.predict(X_new))

[[309291.06984232]
 [ 35242.16252012]]
```

# TERIMA KASIH

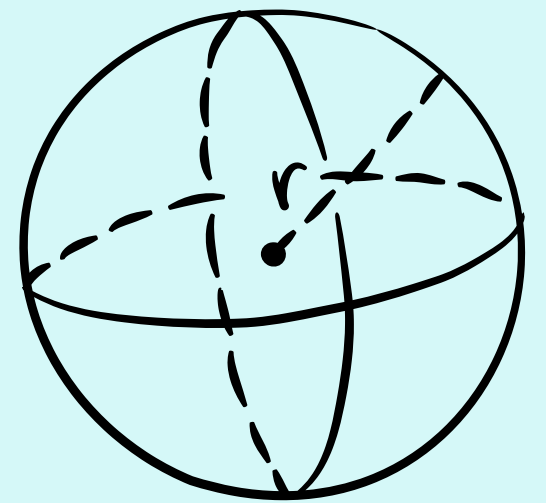
---

Link colab google :

[https://colab.research.google.com/drive/1TBgCOvzrubdH\\_4K0Jc0ldT30WienAEnp?usp=sharing](https://colab.research.google.com/drive/1TBgCOvzrubdH_4K0Jc0ldT30WienAEnp?usp=sharing)

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$y = mx + b$$



$$V = \frac{4}{3} \pi r^3$$