

# **ANALISIS HUBUNGAN ANTARA TINGKAT BANJIR DENGAN TINGKAT PENDIDIKAN DI DKI JAKARTA**

**OLEH**  
**SUSU BERUANG**

# DAFTAR ISI

Rumusan Masalah	3
Hipotesis	5
Metodologi dan Variabel	6
Analisis Awal	7
Analisis Mendalam	8
Kesimpulan dan Rekomendasi	9

# RUMUSAN MASALAH

Apakah **tingkat pendidikan berpengaruh pada tingkat banjir?**

# HIPOTESIS

$$\Sigma X_1^2 - (\Sigma X_1)^2/n$$

$$\Sigma X_2^2 - (\Sigma X_2)^2/n$$

$$\Sigma Y^2 - (\Sigma Y)^2/n$$

$$\Sigma X_1 Y - (\Sigma X_1)(\Sigma Y)/n$$

$$\Sigma X_2 Y - (\Sigma X_2)(\Sigma Y)/n$$

$$\Sigma X_1 X_2 - (\Sigma X_1)(\Sigma X_2)/n$$

## REGRESSION

HO (Hipotesis nol) = Tingkat pendidikan tidak berpengaruh terhadap tingkat banjir

H1 (Hipotesis alternatif) = Tingkat pendidikan berpengaruh terhadap tingkat banjir

Diprediksi dari analisis regresi yang dilakukan, didapatkan bahwa  
**tingkat pendidikan tidak berpengaruh pada tingkat banjir**

# METODOLOGI DAN VARIABEL

## A. Metodologi

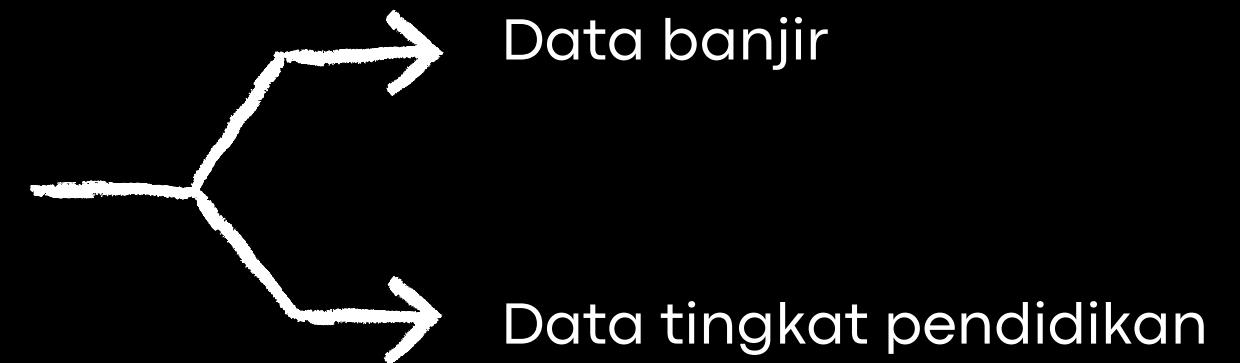
- Regression
- Alternative hypothesis
- Time series analysis

## B. Variabel

- Total hari mengalami banjir dalam satu tahun
- Tingkat pendidikan per tahun

# ANALISIS AWAL

Variabel yang digunakan



## A. Data Banjir

Data yang diperoleh dari Open Data Jakarta dikelompokkan berdasarkan tahun dan setiap bulannya berada dalam sheet yang terpisah sehingga untuk analisis lebih lanjut, data masih perlu digabungkan. Dalam kasus ini, data digabungkan per tahun karena kami ingin membandingkan data dari tahun ke tahun. Rentang tahun yang kami pilih adalah 2017–2020.

Data yang diambil dari Open Data Jakarta hanya berisi nama-nama daerah yang terkena banjir. Oleh karena itu, kami menambahkan nama-nama daerah yang tidak terdapat di dalam data dari internet (sumber: jakarta.go.id) sehingga total akhir nama-nama daerah di DKI Jakarta (populasi) ada 267 kelurahan.

# ANALISIS AWAL

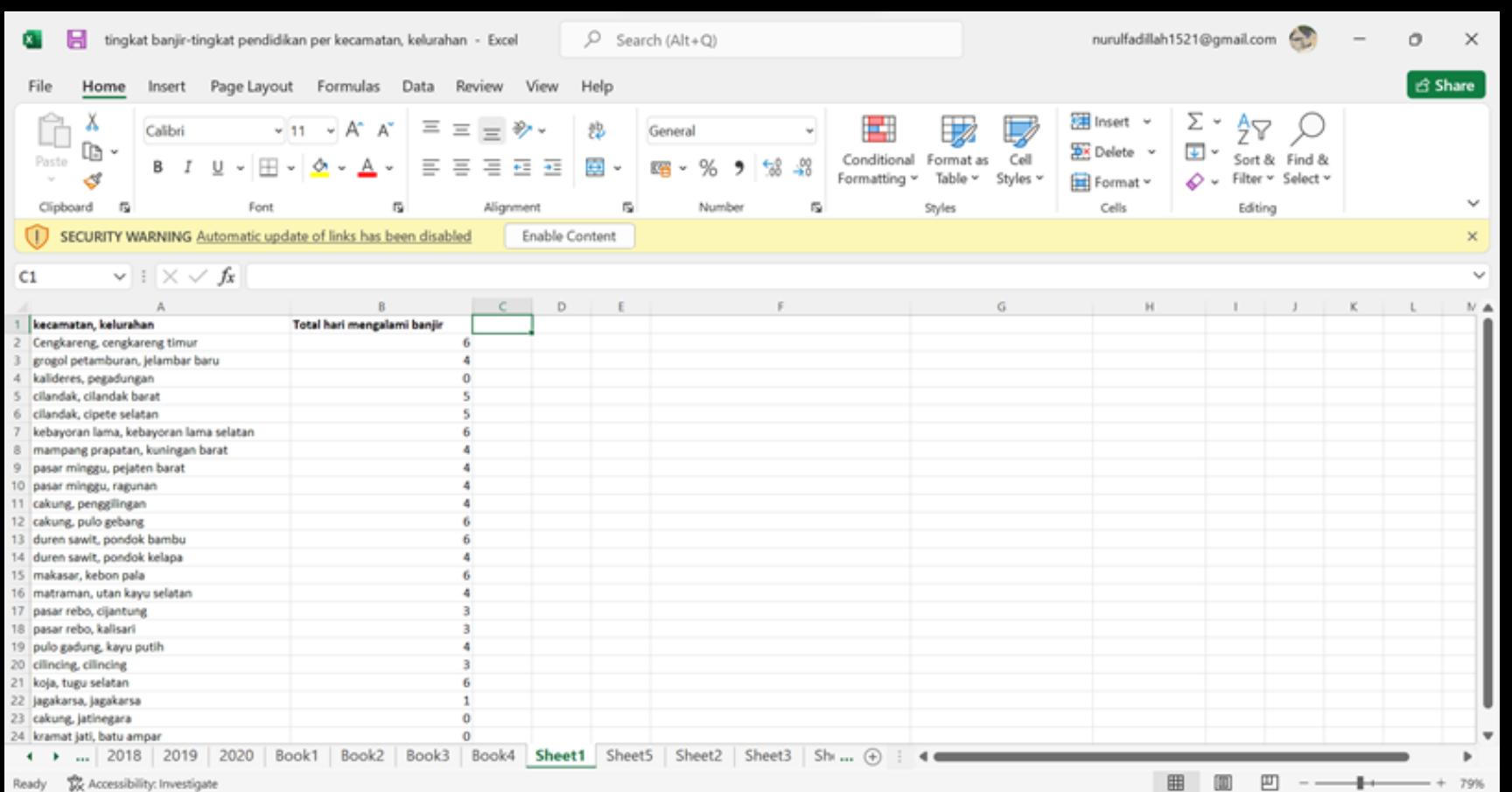
Kami memilih 23 kelurahan sebagai sampel dari penelitian ini. Daerah yang digunakan sebagai sampel adalah daerah yang memiliki total hari mengalami banjir > 0 selama minimal dua kali dalam setahun. Akhir dari proses ini adalah data yang terdiri atas lima kolom yang masing-masing Bernama: "Kota Administrasi", "Kecamatan, Kelurahan", "Bulan Mengalami Banjir (dalam Bulan Tersebut)", dan "Total Hari Mengalami Banjir", seperti yang terlihat pada Gambar 1.

	Bulan mengalami banjir	Berapa hari (dlm bulan tsb)	Total hari mengalami banjir	
1	kecamatan, kelurahan			
2	Cengkareng, cengkareng timur			
3	grogol petamburan, jelambar baru			
4	kalideres, pegadungan			
5	cilandak, cilandak barat			
6	cilandak, cipete selatan			
7	kebayoran lama, kebayoran lama selatan			
8	mampang prapatan, kuningan barat			
9	pasar minggu, pejaten barat			
10	pasar minggu, ragunan			
11	cakung, penggilingan			
12	cakung, pulo gebang			
13	duren sawit, pondok bambu			
14	duren sawit, pondok kelapa			
15	makasar, kebo pala			
16	matraman, utan kayu selatan			
17	pasar rebo, cijantung			
18	pasar rebo, kalisari			
19	pulo gadung, kayu putih			
20	cilincing, cilincing			

Gambar 1 Template data

# ANALISIS AWAL

Data-data dikelompokkan berdasarkan tahun. Selanjutnya kami melengkapi data untuk mengisi semua kolom. Setelah data-data tersebut rampung, kami merangkum data menjadi data yang hanya terdiri atas dua kolom, yaitu kolom "Kecamatan, Kelurahan" dan kolom "Total Hari Mengalami Banjir" yang dikelompokkan berdasarkan tahun seperti yang terdapat dalam Gambar 2.



A	B	C
1 kecamatan, kelurahan	Total hari mengalami banjir	
2 CengKareng, cengKareng timur	6	
3 grogol petamburan, jelambar baru	4	
4 kalideres, pegadungan	0	
5 cilandak, cilandak barat	5	
6 cilandak, cipete selatan	5	
7 kebayoran lama, kebayoran lama selatan	6	
8 rambah prapatan, kunilagan barat	4	
9 pasar minggu, pejaten barat	4	
10 pasar minggu, ragunan	4	
11 cakung, penggilingan	4	
12 cakung, pulo gebang	6	
13 duren sawit, pondok bambu	6	
14 duren sawit, pondok kelapa	4	
15 makasar, kebon pala	6	
16 matraman, utan kayu selatan	4	
17 pasar rebo, cijantung	3	
18 pasar rebo, kalsari	3	
19 pulo gadung, kayu putih	4	
20 cilincing, cilincing	3	
21 koja, tugu selatan	6	
22 jagakarsa, jagakarsa	1	
23 cakung, jatinegara	0	
24 kramat jati, batu ampar	0	

Gambar 2 Data banjir tahun 2017 yang telah dirangkum

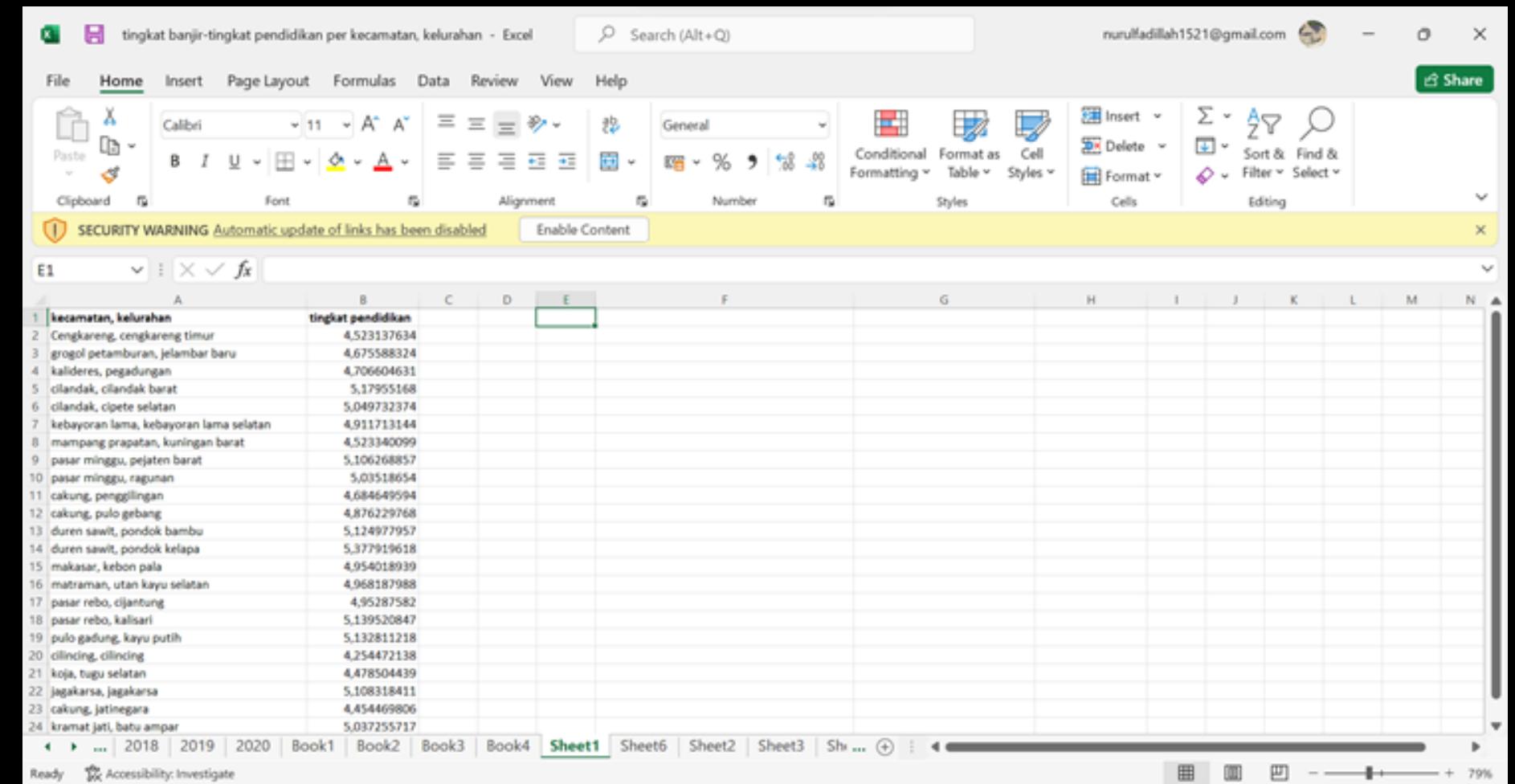
# **ANALISIS AWAL**

Setelah semua data rampung, kami menyimpan terlebih dahulu data tersebut untuk nantinya dibuat grafik bersamaan (combo chart) dengan data tingkat pendidikan.

## **B. Data Tingkat Pendidikan**

•Kami mengumpulkan data pendidikan dari 23 kelurahan (sampel) yang ada di DKI Jakarta dengan kategori: tidak/belum bersekolah, belum tamat SD, tamat SD, SLTP/sederajat, SLTA/sederajat, diploma I/II, diploma III, diploma IV, strata II, strata III. Adapun yang menjadi tolak ukur tingkat pendidikan adalah indeks pendidikan, seperti yang dapat dilihat pada Gambar 3.

# ANALISIS AWAL

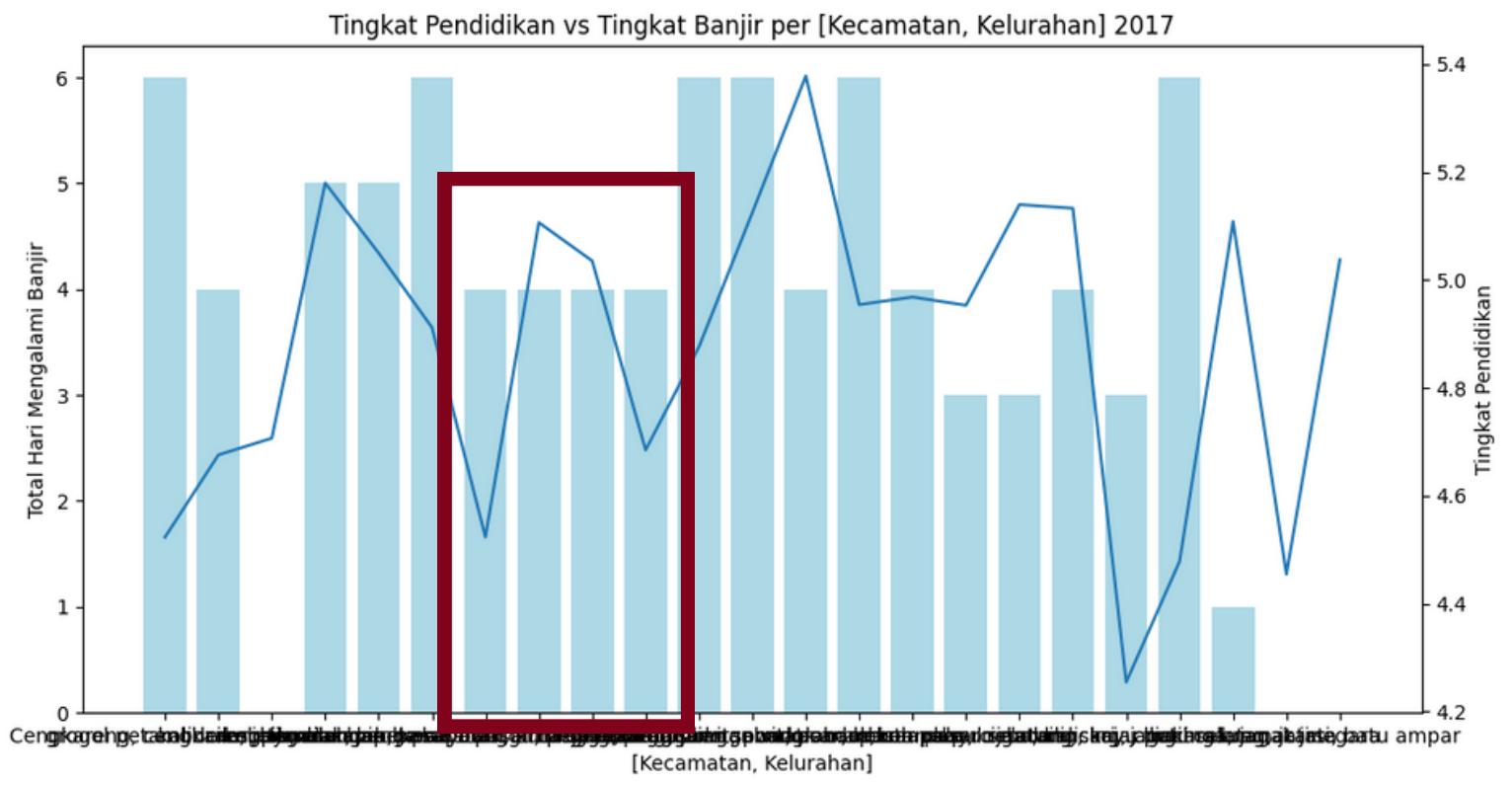


The screenshot shows a Microsoft Excel spreadsheet titled "tingkat banjir-tingkat pendidikan per kecamatan, kelurahan - Excel". The data is organized into two columns: "kecamatan, kelurahan" (districts) and "tingkat pendidikan" (education level). The districts listed are: Cengkareng, cengkareng timur, grogol petamburan, jelambar baru, kalideres, pegadungan, cilandak, cilandak barat, cilandak, cipete selatan, kebayoran lama, kebayoran lama selatan, mampang prapatan, kuningan barat, pasar minggu, pejaten barat, pasar minggu, ragunan, cakung, penggilingan, cakung, pulo gebang, duren sawit, pondok bambu, duren sawit, pondok kelapa, makasar, kebon pala, matraman, utan kayu selatan, pasar rebo, cijantung, pasar rebo, kalsari, pulo gadung, kayu putih, cilincing, cilincing, koja, rugu selatan, jagakarsa, jagakarsa, cakung, jatinegara, kramat jati, batu ampar. The "tingkat pendidikan" column contains numerical values ranging from 4,523137634 to 5,132811218.

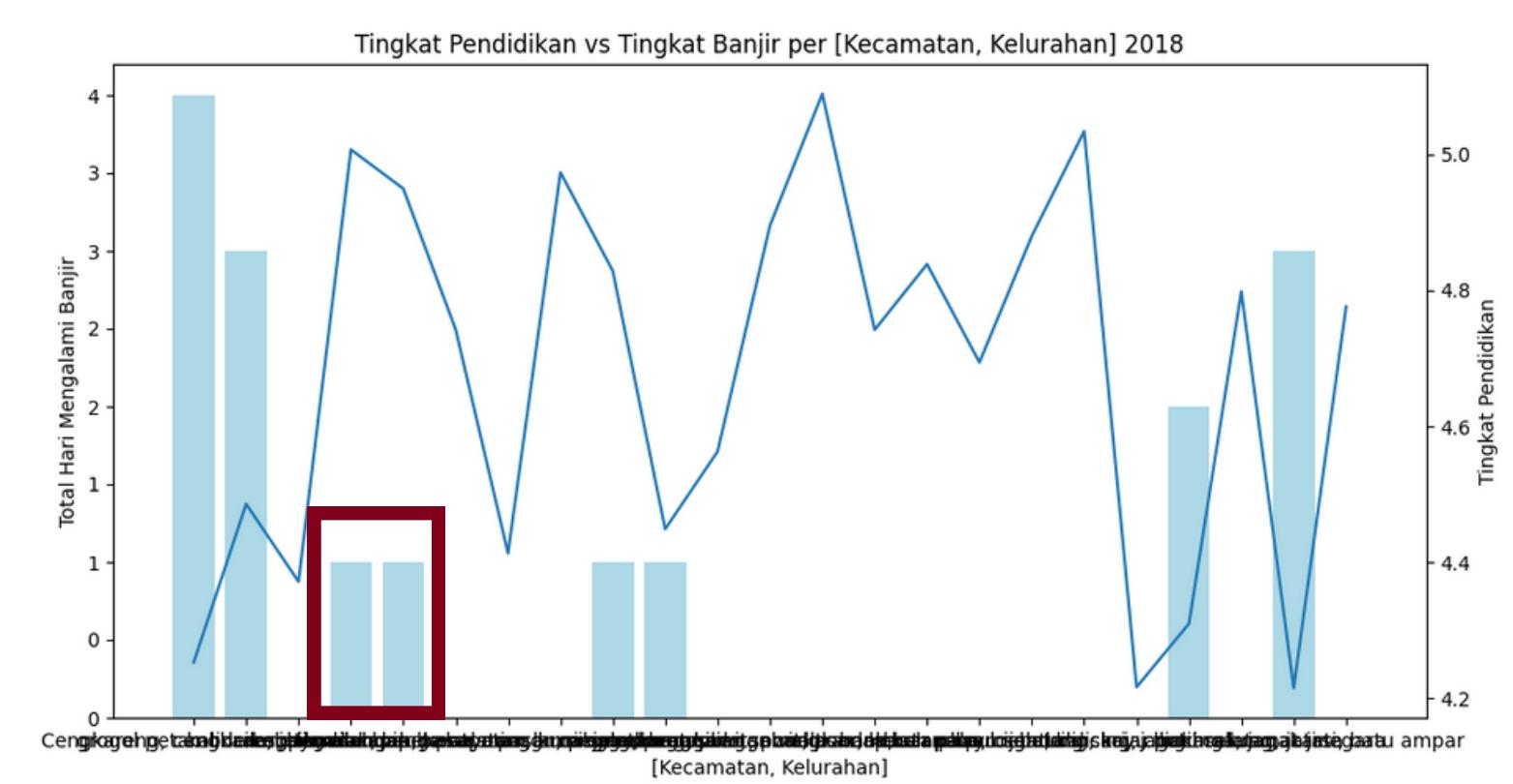
kecamatan, kelurahan	tingkat pendidikan
Cengkareng, cengkareng timur	4,523137634
grogol petamburan, jelambar baru	4,675588324
kalideres, pegadungan	4,706604631
cilandak, cilandak barat	5,17955168
cilandak, cipete selatan	5,049732374
kebayoran lama, kebayoran lama selatan	4,911713144
mampang prapatan, kuningan barat	4,523340099
pasar minggu, pejaten barat	5,106268857
pasar minggu, ragunan	5,03518654
cakung, penggilingan	4,684649594
cakung, pulo gebang	4,876229768
duren sawit, pondok bambu	5,124977957
duren sawit, pondok kelapa	5,377919618
makasar, kebon pala	4,954018939
matraman, utan kayu selatan	4,968187988
pasar rebo, cijantung	4,95287582
pasar rebo, kalsari	5,139520847
pulo gadung, kayu putih	5,132811218
cilincing, cilincing	4,254472138
koja, rugu selatan	4,478504439
jagakarsa, jagakarsa	5,108318411
cakung, jatinegara	4,454469806
kramat jati, batu ampar	5,037255717

Gambar 3 Data tingkat pendidikan tahun 2017 yang telah dirangkum

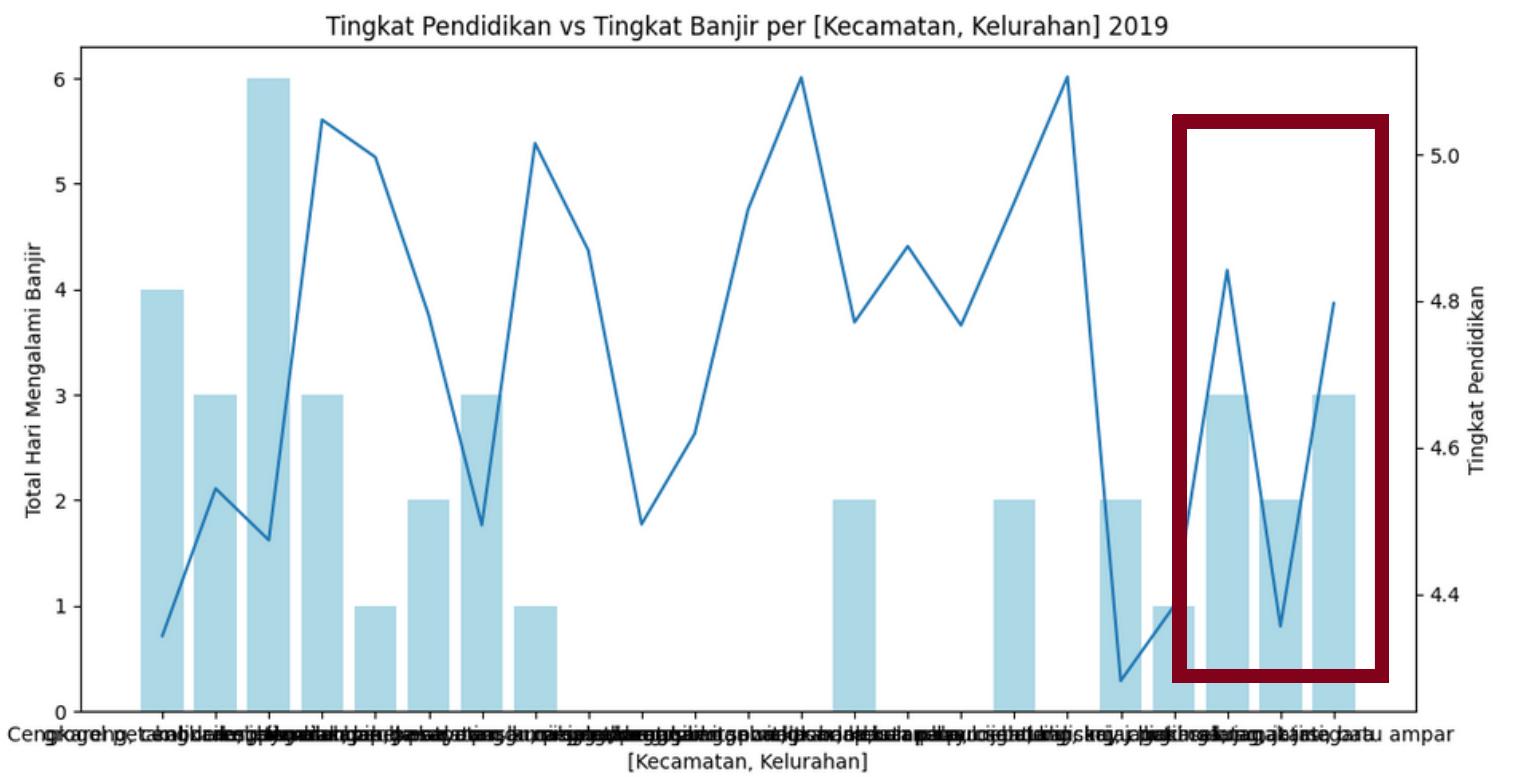
Indeks pendidikan ini digunakan sebagai acuan untuk menentukan tingkat pendidikan di kelurahan tersebut, maka data tingkat pendidikan siap digabungkan dengan data tingkat banjir untuk dibuat combo chart.



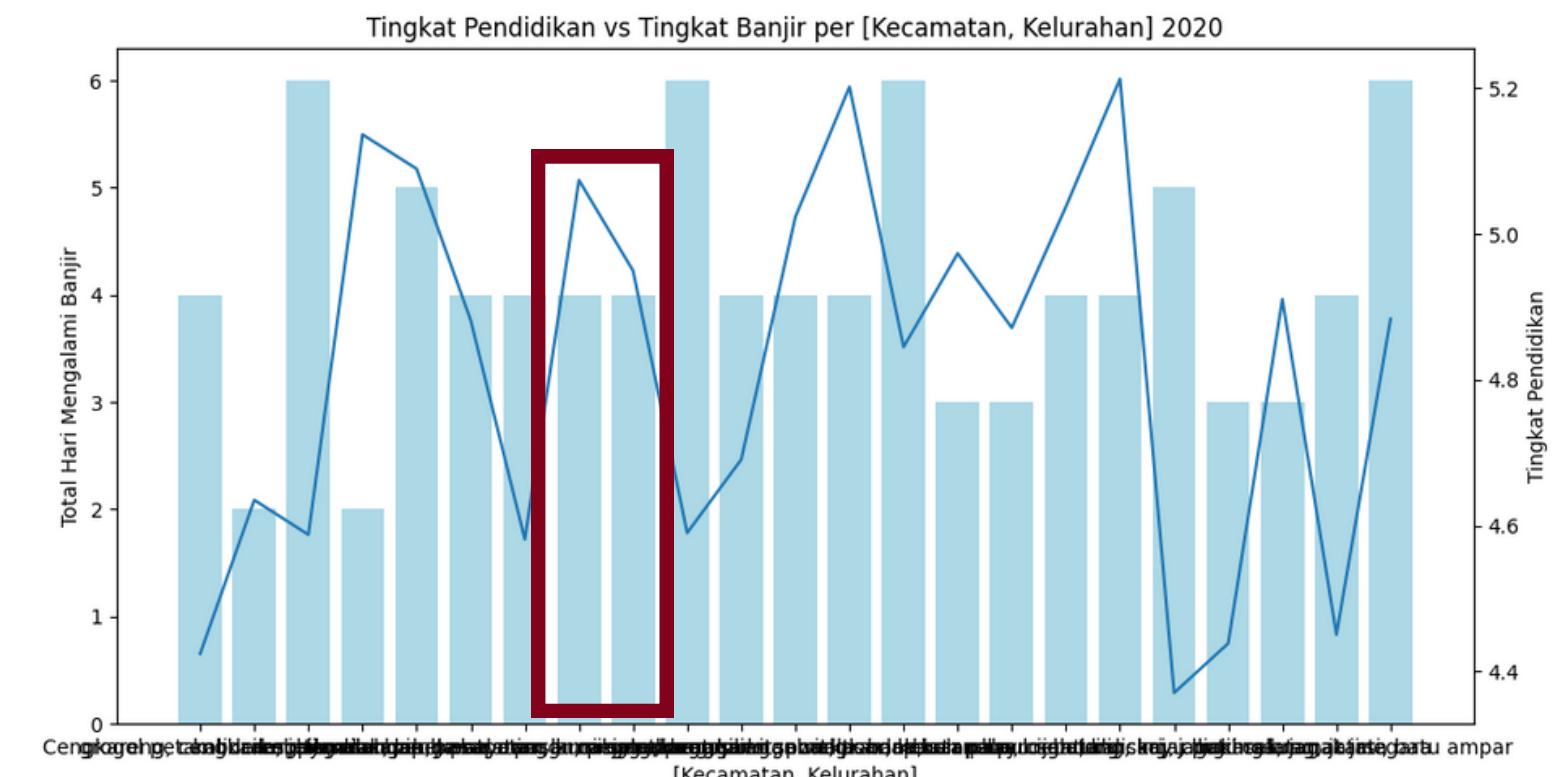
Gambar 4 Grafik 2017



Gambar 5 Grafik 2018

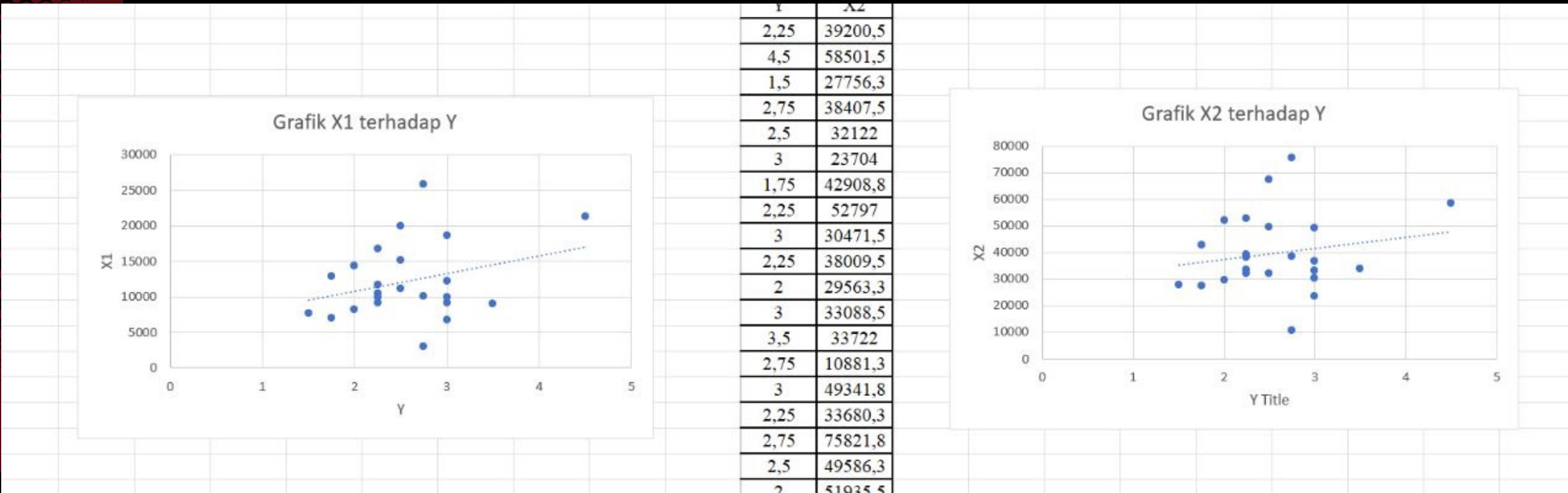


## Gambar 6 Grafik 2019



Gambar 7 Grafik 2020

# ANALISIS AWAL



Gambar 8 Regresi linear

Dari grafik dan regresi linear di atas, terlihat bahwa tidak ada hubungan antara tingkat banjir dengan tingkat pendidikan.

Dari Gambar 4, Gambar 5, Gambar 6, dan Gambar 7, didapatkan bahwa dengan frekuensi banjir yang sama, tingkat pendidikan berfluktuasi cukup jauh.

Dari regresi linear turut didapatkan hubungan yang lemah antara kedua variabel.

X1 = Tidak bersekolah-Tamat SD

X2 = SLTP sederajat-Strata III

(Perhitungan regresi linear dapat dilihat di slide selanjutnya)

# ANALISIS AWAL

Analisis Regresi keseluruhan 2017-2020

Y	$2,707103 + 0,000254X_1 - 0,000081X_2$
KORELASI GANDA	0,54
KONTRIBUSI KORELASI GANDA	29,30%
F HITUNG	2,07146293
F TABEL	$F\{(0,95)(2), (20)\} = 3,49$

X1 = Tidak bersekolah-Tamat SD

X2 = SLTP-Strata III

Karena  $F_{hitung} < F_{tabel}$ , maka  $H_0$  diterima dan  $H_a$  ditolak sehingga disimpulkan bahwa **tingkat pendidikan tidak berpengaruh pada tingkat banjir**

Pada persamaan regresi yang didapatkan, ditunjukkan bahwa nilai  $X_2$  adalah negatif (-) yang mengartikan bahwa  $X_2$  (SLTP-Strata III) **tidak menunjukkan pengaruh yang cukup signifikan terhadap tingkat banjir**

Seluruh perhitungan dapat dilihat pada link di bawah ini :

<https://ipb.link/susuberuang-datawarehouse>

# ANALISIS MENDALAM

Salah satu metode yang kami gunakan adalah metode time series analysis, artinya kami menganalisis data dari tahun ke tahun. Namun, sebagai contoh, kode yang ditampilkan hanya untuk data tahun 2017.

## 0. Import Important Libraries

pandas, scikit-learn, numpy, matplotlib, math, openpyxl

## 1. Loading the Data

```
df = pd.read_excel(r'D:\PROJECTS\DSA COMPFEST\TUGAS 1\referensi\tingkat_banjir-tingkat_pendidikan_per_kecamatan,_kelurahan.xlsx',  
                   sheet_name='Sheet1')
```

## 2. Understanding the Dataset

df.info()



RangeIndex: 23 entries, 0 to 22			
Data columns (total 3 columns):			
#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	kecamatan, kelurahan	23 non-null	object
1	Total hari mengalami banjir	23 non-null	int64
2	tingkat pendidikan	23 non-null	float64

dtypes: float64(1), int64(1), object(1)

Dari sini dapat dilihat tipe--tipe data dan terlihat juga bahwa tidak ada data yang hilang (lengkap kelurahan)

# ANALISIS MENDALAM

## 3. Data Preprocessing

Ubah tipe data dari "object" menjadi "category"

```
df['kecamatan, kelurahan'] = df['kecamatan, kelurahan'].astype('category')
```



0	kecamatan, kelurahan	23 non-null	category
---	----------------------	-------------	----------

## 4. Data Visualization

```
fig, ax1 = plt.subplots(figsize=(12,6))
ax1.bar(df['kecamatan, kelurahan'], df['Total hari mengalami banjir'], color='lightblue')
ax1.set_yticklabels(['{:,.0f}'.format(int(x)) for x in ax1.get_yticks().tolist()])
ax2 = ax1.twinx()
ax2.plot(df['kecamatan, kelurahan'], df['tingkat pendidikan'])
ax1.set_title('Tingkat Pendidikan vs Tingkat Banjir per [Kecamatan, Kelurahan] 2017')
ax1.set_xlabel('[Kecamatan, Kelurahan]')
ax1.set_ylabel('Total Hari Mengalami Banjir')
ax2.set_ylabel('Tingkat Pendidikan')
plt.show()
```

Lihat Gambar 4 untuk melihat output grafik.

# ANALISIS MENDALAM

Lakukan one-hot encoding kepada categorical data untuk merepresentasikan variabel kategorik menjadi nilai numerik

```
df = pd.get_dummies(df)
```

## 5. Building a Regression Model

- Tentukan variabel target dan variabel feature

Sebelum membuat model regresi, tentukan terlebih dahulu variabel apa yang akan menjadi variabel target (variabel dependen/variabel output (y)) dan variabel apa yang akan menjadi variabel feature (variabel independen/variabel input (x)).

Dalam kasus ini, jadikan variabel total hari mengalami banjir menjadi variabel output/variabel dependen sehingga variabel yang tersisa, yaitu variabel tingkat pendidikan, menjadi variabel input/variabel independen.

```
y = df["Total hari mengalami banjir"]
x = df.drop("Total hari mengalami banjir", axis=1)
```

- Bagi dataset menjadi data training dan data testing

Pembagiannya adalah 80% untuk data training dan 20% untuk data testing

# ANALISIS MENDALAM

```
from sklearn.model_selection import train_test_split  
x_train, x_test, y_train, y_test = train_test_split(  
    X, Y,  
    train_size = 0.8, #####--> 80% training dan 20% testing  
    random_state = 1)
```

- Buat model menggunakan data training

```
from sklearn.linear_model import LinearRegression  
lr = LinearRegression()  
lr.fit(x_train, y_train)
```

## 6. Model Evaluation

Evaluasi model dilakukan dengan menggunakan mean squared error (MSE)

```
y_pred = lr.predict(x_test)  
from sklearn.metrics import mean_squared_error  
  
import math  
c = math.sqrt(mean_squared_error(y_test, y_pred))  
print(c)
```



Nilai yang dihasilkan adalah 1.936289620456332. Dalam MSE, semakin kecil nilai yang dihasilkan, semakin baik modelnya. Maka, model yang telah dibuat cukup bagus untuk memprediksi

# ANALISIS MENDALAM

## 3. Model Prediction

Sebagai contoh, akan diprediksi baris pertama.

```
data_new = x_train[:1]
d = lr.predict(data_new)
print(d)
```



[3.](output = 3)

Bandingkan dengan nilai asli

```
e = y_train[:1]
print(e)
```



1633(output = 3)

Terlihat bahwa nilai yang diprediksi sama dengan nilai asli.

# **KESIMPULAN DAN REKOMENDASI**

## **KESIMPULAN**

Tidak ada hubungan antara tingkat banjir dengan tingkat pendidikan

## **REKOMENDASI**

Meskipun kesimpulan dari analisis yang telah dilakukan adalah tidak ada hubungan antara tingkat banjir dengan tingkat pendidikan, pendidikan tetap harus digencarkan karena tidak menutup kemungkinan bahwa terdapat situasi di mana ada hubungan antara tingkat pendidikan dengan tingkat banjir. Tingkat pendidikan yang tinggi menunjukkan indeks pendidikan yang tinggi dan indeks pendidikan yang tinggi bermakna kualitas pendidikan yang tinggi. Kualitas pendidikan yang tinggi membantu membangun kesadaran seseorang akan kepedulian lingkungan.