

Pengolahan Big Data Tinggi Muka Air Sungai

Di Jakarta Bulan Mei Tahun 2020



Disusun Oleh :

NIM : 145410006

NAMA : RAHMAT HADI SUKO WIJOYO

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER AKAKOM
YOGYAKARTA**

2020

1. Untuk dataset yang akan saya gunakan, saya mengunduhnya pada halaman :

<https://data.jakarta.go.id/apps/download/data/data-tinggi-muka-air-mei-2020.csv>

Dataset tersebut berisikan tentang data tinggi air pada tiap pintu air yang beroperasi di Jakarta, dimana setiap data direkam pada jeda waktu tertentu setiap harinya.

2. Proses *pre-processing* akan saya jelaskan dibawah.

```
In [8]: 1 import pandas as pd
        2 import numpy as np
        3 import seaborn as sns
        4 sns.set()
        5 import matplotlib.pyplot as plt
        6 %matplotlib inline
        7
        8 df=pd.read_csv('data-tinggi-muka-air.csv')
        9 df.head()
```

```
Out[8]:
```

| | nama_pintu_air | lokasi | latitude | longitude | tanggal | tinggi_air | status_siaga |
|---|---------------------------|----------|-----------|------------|---------------------|------------|-----------------|
| 0 | Bendung. Cibalok - Gadag | Ciliwung | -6.653417 | 106.869551 | 2020-05-04 14:50:00 | -80 | Status : Normal |
| 1 | Bendung. Katulampa (Hulu) | Ciliwung | -6.633167 | 106.836806 | 2020-05-04 14:50:00 | 240 | Status : Normal |
| 2 | PS. Depok | Ciliwung | -6.400514 | 106.832143 | 2020-05-04 14:50:00 | 1110 | Status : Normal |
| 3 | PA. Manggarai | Ciliwung | -6.207825 | 106.848458 | 2020-05-04 14:50:00 | 6150 | Status : Normal |
| 4 | PS. Krukut Hulu | Krukut | -6.343900 | 106.798874 | 2020-05-04 14:06:22 | 550 | Status : Normal |

Langkah awal yaitu meng-*import tools* yang akan saya gunakan, yaitu pandas, numpy, seaborn dan matplotlib. Selanjutnya adalah membaca dataset yang telah saya unduh tadi.

Dataset tersebut berisikan nama_pintu_air, lokasi, latitude, longitude, tanggal, tinggi_air dan status_siaga.

```
In [32]: 1 df.drop(['latitude', 'longitude'], axis=1, inplace=True)
```

Karena latitude dan longitude tidak akan saya gunakan dalam pengolahan data, maka 2 atribut tersebut akan saya hapus menggunakan *df.drop*.

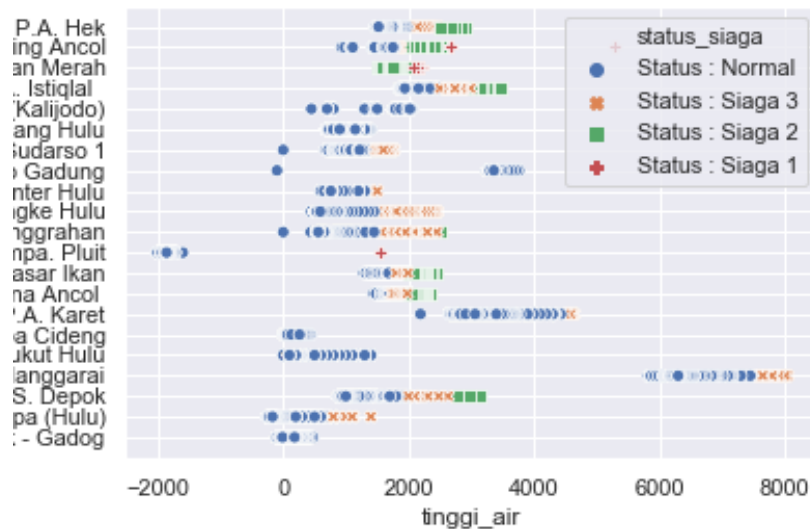
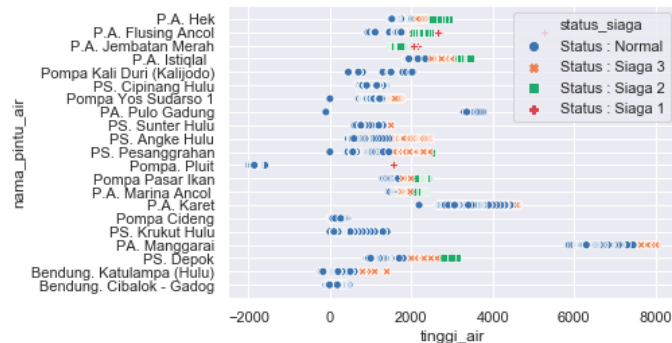
```
In [33]: 1 df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20912 entries, 0 to 20911
Data columns (total 5 columns):
 nama_pintu_air    20912 non-null object
 lokasi           20912 non-null object
 tanggal          20912 non-null object
 tinggi_air        20912 non-null int64
 status_siaga      20912 non-null object
dtypes: int64(1), object(4)
memory usage: 817.0+ KB
```

Lakukan pengecekan dataset menggunakan *df.info()*. Info diatas menunjukkan bahwa atribut latitude dan longitude telah terhapus, dan tidak ada data yang kosong / *null*. Untuk jumlah data keseluruhan terdapat 20912 record.

- Visualisasi dataset saya menggunakan scatterplot dari seaborn, dimana penggunaannya lebih mudah dibandingkan dengan scatterplot dari matplotlib.

```
In [44]: 1 ax=sns.scatterplot(x='tinggi_air', y='nama_pintu_air', hue='status_siaga', style='status_siaga', data=df)
2 plt.savefig("plottinggiair.png")
```



Pada scatterplot diatas, sumbu x berisikan data dari tinggi_air, sumbu y dengan data nama_pintu_air dan legenda berisikan tentang status_siaga. Dari hasil visualisasi data tersebut dapat kita lihat bahwa tinggi_air tertinggi terdapat pada P.A Manggarai pada status Siaga 3. Sedangkan pada status Siaga 1, tinggi_air tertinggi ditemukan pada P.A Flusing Ancol.

Dengan data yang telah kita visualisasikan tersebut, dapat diambil kesimpulan bahwa tinggi_air pada tiap nama_pintu_air memiliki batas status_siaga masing-masing, dan memiliki korelasi yang kuat terhadap penetapan status banjir di Jakarta.