

# JAWABAN PROBLEM SET

DATA SCIENCE ACADEMY

## Dream Team

ANGGOTA KELOMPOK

Filbert Felim

Hafidz Shidqi

Alisha Listya

2021

# Jawaban Non-Dataset

1. *Outlier* atau Pencilan adalah pengamatan yang sangat menyimpang dari pengamatan lain sehingga menimbulkan kecurigaan bahwa itu dihasilkan oleh mekanisme yang berbeda (Hawkins, 1980). Secara sederhana, *Outlier* dalam teori statistik merupakan data yang memiliki ciri khas unik sehingga tampak beda dari data lainnya. *Outlier* tampil dalam bentuk magnitudo yang bernilai ekstrim pada variabel kombinasi dan tunggal. Titik yang berada lebih dari 1,5 kali jangkauan interkuartil di atas kuartil ke-3 (tiga) atau di bawah kuartil pertama merupakan sebuah *outlier*

Implikasi adanya *outlier* dalam data ialah dapat mengubah konklusi dari penelitian yang diambil dari hasil analisis statistik. Maka dari itu, kita harus menangani *outlier* ini. Beberapa metode dalam melakukan manajemen data terhadap kasus *outlier* ialah :

1. Mengevaluasi kembali data yang telah diperoleh, seperti populasi target, subjek penelitian, pertanyaan, dan metodologi penelitian dengan tepat.
2. Mengevaluasi kembali apakah ada kesalahan saat memproses data seperti mengukur, mengamati, dan mengentri data
3. Mengecheck kembali apakah ada perbedaan substansi tentang suatu pengamatan seperti apakah itu orang atau barang
2. Korelasi dalam ilmu statistik dapat diartikan sebagai metode statistika yang digunakan untuk mengukur keeratan hubungan antara dua variabel (Astuti, 2017). Dikatakan berkorelasi jika magnitudo variabel yang satu diikuti oleh variabel lain, baik bersifat sebanding atau disebut sebagai korelasi positif maupun berbanding terbalik yang biasa disebut sebagai korelasi negatif. Hasil dari koefisien korelasi merupakan dalam bentuk angka, koefisien korelasi positif memiliki magnitudo antara 0 sampai 1 dan koefisien korelasi negatif memiliki magnitudo antara -1 sampai 0.

Koefisien korelasi dapat dihitung menggunakan metode koefisien korelasi Pearson yang diberi simbol  $r_{xy}$  atau  $r_{yx}$  yang dirumuskan (Pearson, 1896) :

$$r_{yx} \text{ atau } r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Dimana :

$r_{xy}$  atau  $r_{yx}$  = korelasi antara x dan y

$x_i$  = nilai (magnitudo) x ke- $i$

$y_i$  = nilai (magnitudo) y ke- $i$

$\underline{x}$  = mean (rata-rata) x

$\underline{y}$  = mean (rata-rata) y

$n$  = banyaknya nilai

Berikut merupakan hasil analisis dari rumus koefisien korelasi Pearson:

1. Jika koefisien korelasi bersifat positif yang mana magnitudonya bernilai diantara 0 sampai 1 ( $0 < r_{xy} \leq 1$ ) maka pastinya pembilang dari  $r_{xy}$  yaitu  $\sum_{i=1}^n (x_i - \underline{x})(y_i - \underline{y})$  bernilai positif. Ada beberapa kemungkinan yang membuat pembilang dari  $r_{xy}$  bernilai positif, yaitu :

- a.  $\sum_{i=1}^n (x_i - \bar{x}) > 0$  dan  $\sum_{i=1}^n (y_i - \bar{y}) > 0$
- b.  $\sum_{i=1}^n (x_i - \bar{x}) < 0$  dan  $\sum_{i=1}^n (y_i - \bar{y}) < 0$

Dengan analisis matematika, kita bisa menarik kesimpulan bahwasanya jika dua buah variabel memiliki koefisien korelasi positif, maka kemungkinan besar mean dari x dan y magnitudonya sebanding (jika mean x besar maka mean y juga besar, dan sebaliknya)

2. Jika koefisien korelasi bersifat negatif yang mana magnitudonya bernilai diantara -1 sampai 0 ( $-1 \leq r_{xy} < 0$ ) maka pastinya pembilang dari  $r_{xy}$  yaitu

$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  bernilai negatif. Ada beberapa kemungkinan yang membuat pembilang dari  $r_{xy}$  bernilai negatif, yaitu :

- a.  $\sum_{i=1}^n (x_i - \bar{x}) > 0$  dan  $\sum_{i=1}^n (y_i - \bar{y}) < 0$
- b.  $\sum_{i=1}^n (x_i - \bar{x}) < 0$  dan  $\sum_{i=1}^n (y_i - \bar{y}) > 0$

Dari rumus tersebut dapat ditarik kesimpulan bahwasanya jika dua buah variabel memiliki koefisien korelasi negatif, maka kemungkinan besar mean dari  $x$  dan  $y$  magnitudonya berbanding terbalik (jika mean  $x$  besar maka mean  $y$  kecil, dan sebaliknya)

3. *Machine Learning* (ML) atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi (Ahmad, 2017). Secara umum, *Machine Learning* (ML) dibagi menjadi tiga algoritma, yakni :

1. *Supervised Learning*

*Supervised Learning* merupakan salah satu kategori dari *Machine Learning* yang menggunakan data berlabel untuk menganalisa data lama dan membuat konklusi yang berguna untuk memetakan data-data baru. Salah satu implikasinya yaitu digunakan pada aplikasi yang dapat memprediksi atau memperkirakan kejadian yang akan datang berdasarkan data historis. Contoh penerapan dari *Supervised Learning* ialah *Convolutional Neural Networks (CNN)*.

2. *Unsupervised Learning*

Berbeda dari *Supervised Learning*, *Unsupervised Learning* tidak menggunakan data berlabel sehingga lebih cocok untuk permasalahan yang tidak terstruktur. *Unsupervised Learning* biasanya digunakan untuk mengolah data transaksional seperti mengidentifikasi segmen konsumen dengan atribut yang sama dan mengelompokkannya agar menjadi data yang terstruktur sehingga dapat diolah dengan mudah. Contoh penerapan *Unsupervised Learning* ialah *clustering*

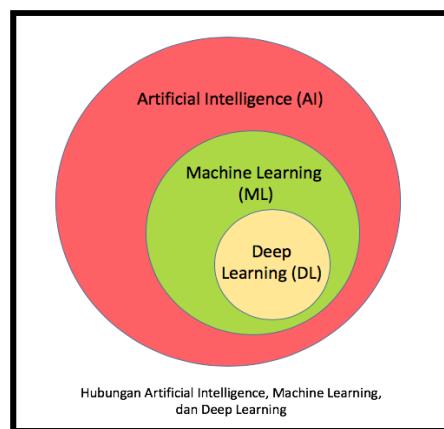
3. *Semi-Supervised Learning*

Mirip dengan *Supervised Learning* yaitu menggunakan data berlabel, akan tetapi *Semi-Supervised Learning* hanya digunakan data berlabel dalam jumlah kecil dan tidak untuk mengembangkan algoritma pemrograman. Contoh penggunaan *Supervised Learning*

ialah pada *webcam* atau *smartphone* yang berfungsi untuk mengidentifikasi wajah seseorang.

4. Sebelumnya menentukan kaitan antara *Artificial Intelligence*, *Machine Learning*, dan *Deep Learning*, berikut definisi dari ketiganya :

- Kecerdasan Buatan atau Artificial Intelligence (AI) adalah teknik yang digunakan untuk meniru kecerdasan yang dimiliki oleh makhluk hidup maupun benda mati untuk menyelesaikan sebuah persoalan (Abu Ahmad, 2017)
- *Machine Learning* (ML) termasuk bagian dari *Artificial Intelligence* (AI) yang di dalamnya menggunakan metode statistik yang memungkinkan mesin untuk meningkatkan *experiencenya*.
- *Deep Learning* (DL) termasuk bagian dari *Machine Learning* (ML) dan merupakan perkembangan dari *neural network multiple layer* yang mampu belajar secara mendalam. Singkatnya, *Deep Learning* (DL) selangkah lebih maju dari *Machine Learning* (ML) karena *Deep Learning* (DL) secara langsung melakukan representasi data tanpa memerlukan aturan *code* pemograman dari manusia



Gambar 4. 1 Hubungan Artificial Intelligence, Machine Learning, daan Deep Learning

Dari definisi di atas, kita bisa mengibaratkan sebagai sebuah kota yang berada dalam lingkup provinsi dan provinsi tersebut berada dalam lingkup sebuah pulau. *Deep Learning* (DL) dianalogikan sebagai sebuah kota, *Machine Learning* (ML) diibaratkan sebagai sebuah provinsi, dan *Artificial Intelligence* (AI) diibaratkan sebagai sebuah pulau.

5. Interpretasi adalah proses memberi arti dan signifikansi terhadap analisis yang dilakukan, menjelaskan pola-pola deskriptif, mencari hubungan dan keterkaitan antar deskripsi-deskripsi data yang ada (Barnsley & Ellis, 1992). Secara umum, interpretasi data merupakan upaya memaknai data dengan cara mengulas kembali data berdasarkan perspektif lain, membandingkan hasil penelitian yang pernah dilakukan, menggabungkan hasil analisis dengan pernyataan tertentu

Interpretasi data sangat penting untuk mengevaluasi proses dan hasil revisi sebagai pembelajaran, akan tetapi dalam menginterpretasi data juga terdapat beberapa tantangan, yaitu peneliti harus melibatkan hal-hal penting dalam suatu penelitian yakni berupa diskusi, konklusi, dan implikasi seperti refleksi data yang telah diperoleh, jawaban atas pertanyaan peneliti, perspektif peneliti yang kontradiktif dengan literatur-literatur yang ada, batasan penelitian, serta saran dan masukkan untuk penelitian selanjutnya.

Interpretasi data dan data storytelling memiliki hubungan timbal balik. Dengan adanya interpretasi data, data storytelling dapat dilakukan dengan mudah dan terstruktur karena sebelumnya sudah menginterpretasi data sehingga menjadi paham tentang seluk-beluk data yang akan divisualisasikan.

# **LAPORAN DATASET**

**DATA SCIENCE ACADEMY**

## **Dream Team**

**ANGGOTA KELOMPOK**

**Filbert Felim**

**Hafidz Shidqi**

**Alisha Listya**

**2021**

## **Daftar Pustaka**

I.	Latar Belakang .....	8
II.	Jawaban Soal.....	9
III.	Hasil Analisis Tambahan .....	11
III.1	Problem Statement .....	11
III.2	Hypothesis.....	11
III.3	Exploratory Data Analysis .....	12
III.4	Initial findings .....	17
IV.	Deep Dive Analysis .....	20
V.	Conclusion and Recommendation .....	21
VI.	Daftar Pustaka.....	22
VII.	Lampiran .....	23

## **I. Latar Belakang**

Demi mengendalikan tingginya kasus COVID-19, pemerintah memutuskan pemberlakuan pembatasan kegiatan masyarakat (PPKM) darurat. PPKM yang dilaksanakan tanggal 3-20 Juli ini dilaksanakan terbatas di beberapa daerah yang memenuhi parameter yang telah ditetapkan oleh pemerintah. Parameter tersebut mencakup kasus aktif, tingkat kematian, tingkat kesembuhan ataupun keterisian RS.

Salah satu daerah yang memenuhi parameter PPKM adalah provinsi DKI Jakarta. Menurut data provinsi dari kementerian kesehatan tertanggal 9 Juli 2021, jumlah terkonfirmasi positif COVID-19 di Jakarta sejumlah 636.383 orang atau sebesar 25.9% kasus di Indonesia (Satuan Tugas Penanganan COVID-19, 2021). Angka yang kian meningkat setiap harinya menyebabkan Jakarta menjadi daerah paling rawan COVID-19 di Indonesia. Hal ini menjadi salah satu hal yang memiliki tingkat urgensi yang tinggi bagi pemerintah provinsi DKI Jakarta.

Tujuan penelitian ini adalah untuk memetakan laju COVID-19 di Jakarta, serta dapat menjadi bahan pertimbangan terhadap kebijakan yang akan ditetapkan pemerintah kedepannya. Dataset yang akan digunakan dalam penelitian ini adalah dataset yang disediakan oleh pemerintah DKI Jakarta, mencakup informasi mengenai jumlah terkonfirmasi positif COVID-19, jumlah kematian, jumlah sembuh, jumlah dirawat setiap harinya sejak tanggal 1 Maret 2020 dan diperbarui secara berkala. Untuk mencapai tujuan tersebut, penelitian ini memanfaatkan penggunaan data science untuk menganalisis dataset yang telah diberikan. Dataset tersebut dianalisis dengan metode EDA (Exploratory Data Analysis). Exploratory Data Analysis adalah metode yang dikembangkan dari analisis deskriptif, dimana didapatkan suatu penemuan, tren, korelasi, ataupun hubungan antara data untuk menghasilkan ide ataupun hipotesis.

## II. Jawaban Soal

1. Dari dataset yang disediakan, temukan nilai mean, median, dan modus dari positif COVID19 harian Jakarta! Dari dataset yang telah disediakan, per tanggal 9 Juli 2021, nilai mean, median, dan modus positif COVID-19 harian Jakarta adalah berturut-turut 1283.042339, 867, dan 0.

	Positif Harian (Jakarta)
count	496.000000
mean	1283.042339
std	1791.071185
min	0.000000
25%	211.250000
50%	867.000000
75%	1430.250000
max	13112.000000

2. Dari dataset yang disediakan, temukan nilai minimal dan maksimal dari positif COVID19 harian Jakarta! Per tanggal 9 Juli 2021, nilai minimal dan nilai maksimal dari positif COVID-19 harian Jakarta adalah berturut-turut 0 dan 13112.

Tabel 1. 1 Deskripsi Positif Harian (Jakarta)

3. Dari dataset yang disediakan, temukan nilai-nilai outlier yang ada (menggunakan variabel yang kalian tentukan)! Dari dataset yang disediakan, variabel yang dianalisis adalah ‘Positif Harian (Jakarta)’. Pada variabel ini, terdapat 40 nilai outlier, dengan rincian sebagai berikut:

	Tanggal	Outlier Positif Harian (Jakarta)
0	[2021-01-13T00:00:00.000000000]	3476
1	[2021-01-16T00:00:00.000000000]	3536
2	[2021-01-17T00:00:00.000000000]	3395
3	[2021-01-20T00:00:00.000000000]	3786
4	[2021-01-22T00:00:00.000000000]	3792
5	[2021-01-23T00:00:00.000000000]	3285
6	[2021-01-24T00:00:00.000000000]	3512
7	[2021-01-29T00:00:00.000000000]	3448
8	[2021-01-30T00:00:00.000000000]	3491
9	[2021-01-31T00:00:00.000000000]	3474
10	[2021-02-01T00:00:00.000000000]	3614
11	[2021-02-02T00:00:00.000000000]	3362
12	[2021-02-03T00:00:00.000000000]	3567
13	[2021-02-04T00:00:00.000000000]	3632
14	[2021-02-05T00:00:00.000000000]	3340
15	[2021-02-07T00:00:00.000000000]	4213
16	[2021-02-09T00:00:00.000000000]	3437
17	[2021-02-10T00:00:00.000000000]	3309
18	[2021-02-12T00:00:00.000000000]	3810
19	[2021-06-17T00:00:00.000000000]	4144
21	[2021-06-19T00:00:00.000000000]	4895
22	[2021-06-20T00:00:00.000000000]	5582
23	[2021-06-21T00:00:00.000000000]	5014
24	[2021-06-23T00:00:00.000000000]	4693
25	[2021-06-24T00:00:00.000000000]	7505
26	[2021-06-25T00:00:00.000000000]	6934
27	[2021-06-26T00:00:00.000000000]	9271
28	[2021-06-27T00:00:00.000000000]	9394
29	[2021-06-28T00:00:00.000000000]	8348
30	[2021-06-29T00:00:00.000000000]	7379
31	[2021-06-30T00:00:00.000000000]	7680
32	[2021-07-01T00:00:00.000000000]	7541
33	[2021-07-02T00:00:00.000000000]	9399
34	[2021-07-03T00:00:00.000000000]	9702
35	[2021-07-04T00:00:00.000000000]	10485
36	[2021-07-05T00:00:00.000000000]	10903
37	[2021-07-06T00:00:00.000000000]	9439
38	[2021-07-07T00:00:00.000000000]	9366
39	[2021-07-08T00:00:00.000000000]	12974

Tabel 3. 1 Outlier Positif Harian (Jakarta)

4. Dari dataset yang disediakan, usulkan dua buah variabel dan berikan analisis korelasi antara kedua variabel tersebut. Jelaskan apa kesimpulan yang dapat diambil berdasarkan analisis kalian.

Variabel yang diusulkan adalah ‘Positif Harian (Jakarta)’ dan ‘Meninggal Harian (Jakarta)’. Nilai korelasinya adalah [0.79]. Hal tersebut berarti kaitan antara kedua variabel cukup tinggi. Dari hasil tersebut, dapat disimpulkan bahwa kasus positif harian COVID-19 Jakarta memiliki korelasi positif dengan kasus meninggal harian Jakarta di mana jika kasus positif harian Jakarta naik, maka kemungkinan besar kasus meninggal harian Jakarta juga akan naik, begitu juga sebaliknya.

```
In [76]: 1 df['Positif Harian (Jakarta)'].corr(df['Meninggal Harian (Jakarta)'])  
Out[76]: 0.7895845711212867
```

Gambar 4. 1 Korelasi Positif Harian (Jakarta) dengan Meninggal Harian (Jakarta)

### **III. Hasil Analisis Tambahan**

#### **III.1 Problem Statement**

Berdasarkan latar belakang masalah yang ada, maka permasalahan dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana mengekstraksi data COVID-19 di Jakarta?
2. Bagaimana menganalisis data COVID-19 di Jakarta?
3. Bagaimana analisis korelasi antar variabel dalam data COVID-19 di Jakarta?
4. Bagaimana visualisasi data COVID-19 di Jakarta?

#### **III.2 Hypothesis**

Berdasarkan perumusan masalah, maka hipotesis dalam penelitian ini dirumuskan sebagai berikut:

1. Data dapat dianalisis secara deskriptif dan eksploratif, dengan menemukan tren, korelasi, dan hubungan untuk menghasilkan ide dan kesimpulan (Leek dan Peng, 2015). Data yang akan digunakan merupakan data diskrit (berupa numerik, misalnya jumlah positif harian Jakarta) dan data kontinu (misalnya jumlah positif kumulatif Jakarta) dalam interval waktu tertentu.
2. Analisis korelasi adalah metode statistika yang digunakan untuk menentukan suatu besaran yang menyatakan bagaimana kuat hubungan suatu variabel dengan variabel lain dengan tidak mempersoalkan apakah suatu variabel tertentu tergantung pada variabel lain (Sekaran, 2010). Dengan memanfaatkan pertumbuhan data di dalam dataset, analisis korelasi pada penelitian ini dapat dilakukan dengan metode Pearson.

### III.3 Exploratory Data Analysis

#### 1. Sampel Data

Berikut merupakan sampel dari data yang telah diberikan:

Tabel III. 1 Tabel Sampel Data

	Tanggal	Meninggal (Indonesia)	Sembuh (Indonesia)	Dirawat (Indonesia)	Positif (Indonesia)	Meninggal (Jakarta)	Sembuh (Jakarta)	Self- Isolation (Jakarta)	Dirawat (Jakarta)	Positif (Jakarta)	Positif Aktif (Jakarta)	Meninggal Harian (Indonesia)
0	2020-03-01	0	0	0	0	0	0	0	0	0	0	0
1	2020-03-02	0	0	2	2	0	0	0	0	0	0	0
2	2020-03-03	0	0	2	2	1	0	0	2	3	2	0
3	2020-03-04	0	0	2	2	1	0	0	2	3	2	0
4	2020-03-05	0	0	2	2	3	0	0	4	7	4	0
5	2020-03-06	0	0	4	4	3	0	0	4	7	4	0
6	2020-03-07	0	0	4	4	3	0	0	4	7	4	0
7	2020-03-08	0	0	6	6	3	0	0	4	7	4	0

#### 2. Deskripsi Data

Berikut merupakan deskripsi dari data yang telah diberikan:

Tabel III. 2 Tabel Deskripsi Data

	Meninggal (Indonesia)	Sembuh (Indonesia)	Dirawat (Indonesia)	Positif (Indonesia)	Meninggal (Jakarta)	Sembuh (Jakarta)	Self- Isolation (Jakarta)	Dirawat (Jakarta)	Positif (Jakarta)	Positif Aktif (Jakarta)
count	496.000000	4.960000e+02	496.000000	4.960000e+02	496.000000	496.000000	496.000000	496.000000	496.000000	496.000000
mean	20244.088710	6.054969e+05	76846.850806	7.025879e+05	3092.300403	159484.324597	7471.185484	3728.209677	173776.020161	11199.395161
std	18666.948686	6.362355e+05	61609.312166	7.053707e+05	2679.533189	163991.059480	9664.903042	4220.597572	173728.366933	13625.996373
min	0.000000	0.000000e+00	0.000000	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3023.750000	2.734275e+04	30003.250000	6.036975e+04	647.500000	7049.500000	2991.500000	1859.500000	11787.250000	4482.750000
50%	14202.500000	3.513895e+05	63372.500000	4.200530e+05	2307.500000	97367.500000	4993.000000	2818.000000	108233.000000	7757.500000
75%	37336.250000	1.196837e+06	110469.750000	1.381386e+06	5802.000000	337802.250000	8824.750000	4314.000000	350870.750000	12752.750000
max	64631.000000	2.023548e+06	367733.000000	2.455912e+06	9306.000000	526941.000000	73239.000000	30418.000000	636389.000000	102082.000000

### 3. Karakteristik Dataset

Berikut merupakan karakteristik dari data yang telah diberikan:

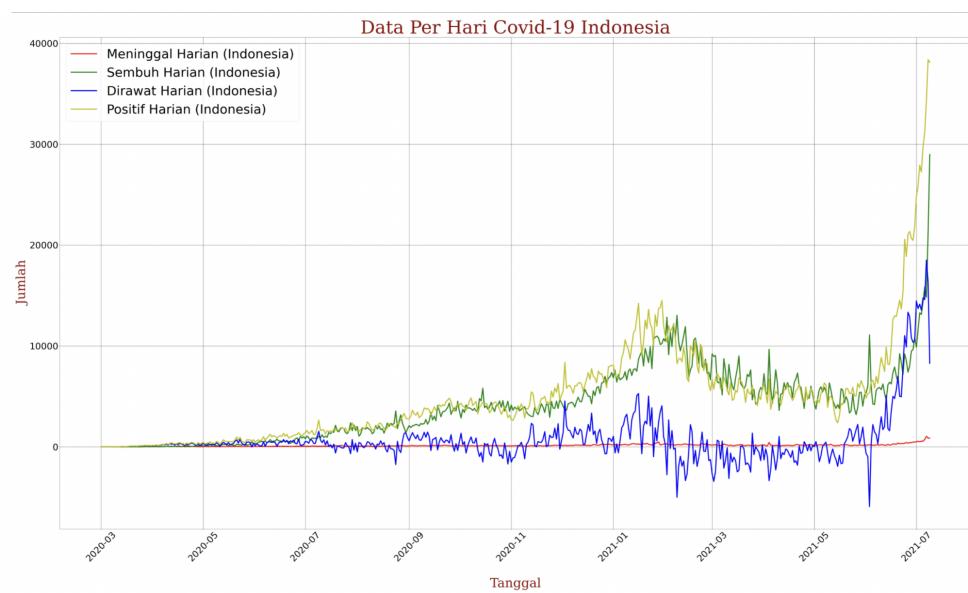
Nama Dataset	Rincian Data	
	Jumlah variabel Observasi	Jumlah Baris Observasi
Data Indonesia dan Jakarta	21	496
Data Jakarta	17	497
Data Suspek dll (Jakarta)	19	359
RDT Ab	6	471
RDT Ag	22	192
Hasil Lab	9	497
Vaksinasi	45	130
Pemakaman	5	496
Vaksinasi Wilayah	86	46
Daily Vaksinasi Wilayah	4	28

#### 4. Visualisasi Data

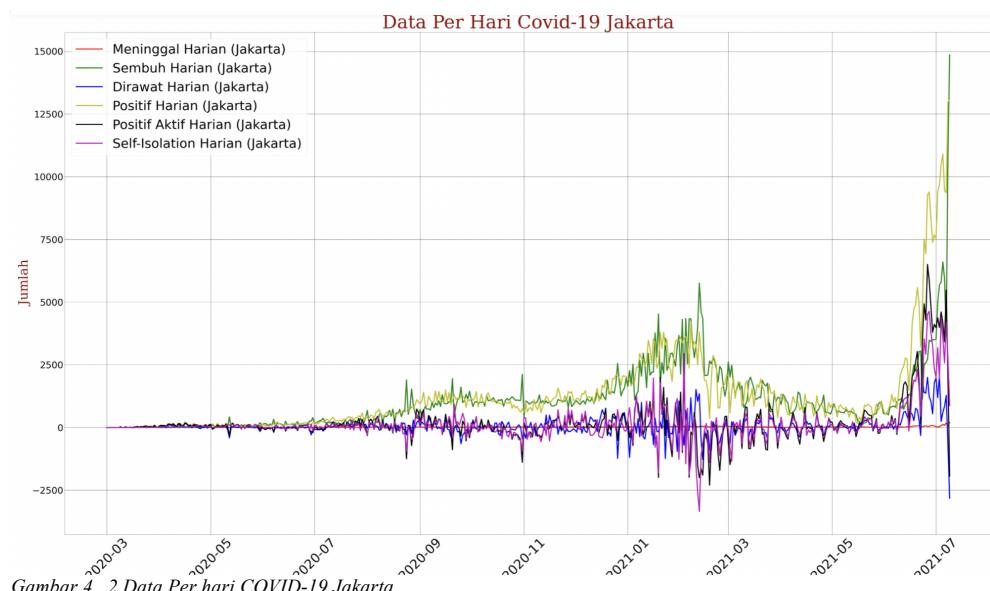
Pada bagian ini, data COVID-19 di Jakarta dan Indonesia divisualisasikan menjadi grafik garis, ataupun *heat-map* menggunakan library *matplotlib* dan *seaborn*.

##### 4.1 Data COVID-19 di Jakarta dan Indonesia

Berikut merupakan data perkembangan kasus COVID-19 di Indonesia dan Jakarta. Data tersebut divisualisasikan menjadi grafik garis. Angka positif menunjukkan bahwa adanya pertambahan jumlah kasus, sedangkan angka negatif menunjukkan adanya penurunan kasus.



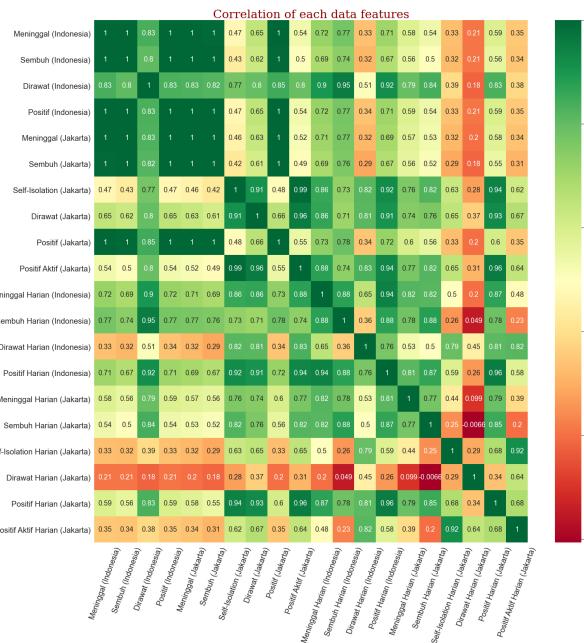
Gambar 4. 1 Data Per hari COVID-19 Indonesia



Gambar 4. 2 Data Per hari COVID-19 Jakarta

## 4.2 Korelasi Data

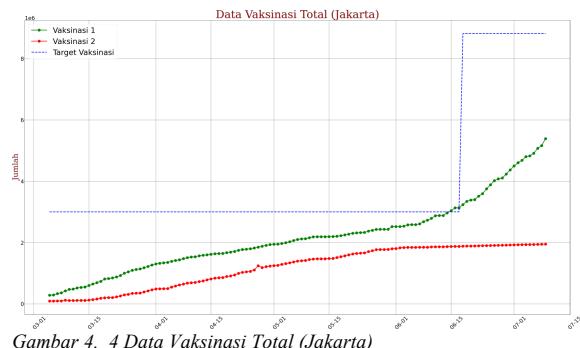
Berikut merupakan visualisasi data korelasi kasus data COVID-19 di Indonesia dan Jakarta menggunakan *heat-map*. Warna hijau menunjukkan korelasi tinggi (korelasi = 1) dan warna merah menunjukkan bahwa data tidak berkorelasi (korelasi = 0).



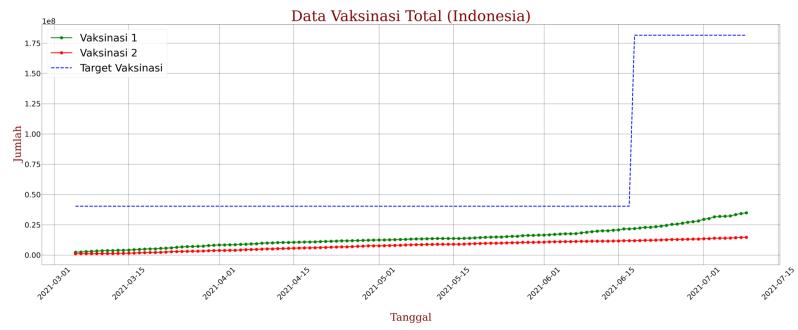
Gambar 4. 3 Korelasi Data

## 4.3 Data vaksinasi

Berikut merupakan data vaksinasi total di Jakarta dan Indonesia. Garis berwarna biru menunjukkan target vaksinasi harian, sedangkan warna merah dan hijau menunjukkan jumlah vaksinasi harian.



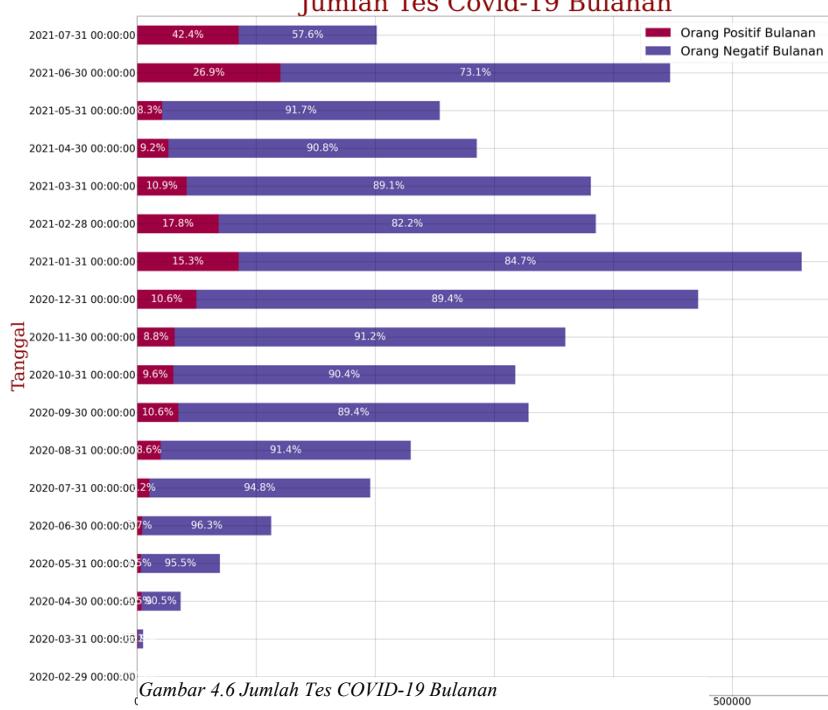
Gambar 4. 4 Data Vaksinasi Total (Jakarta)



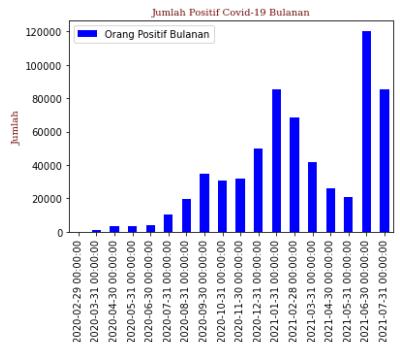
Gambar 4. 5 Data Vaksinasi Total (Jakarta)

## 7.4 Data Hasil Tes

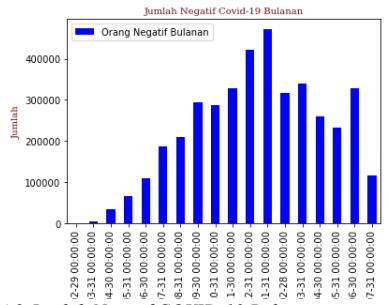
Berikut merupakan data hasil tes COVID-19 bulanan yang divisualisasikan dalam diagram batang.



Gambar 4.6 Jumlah Tes COVID-19 Bulanan



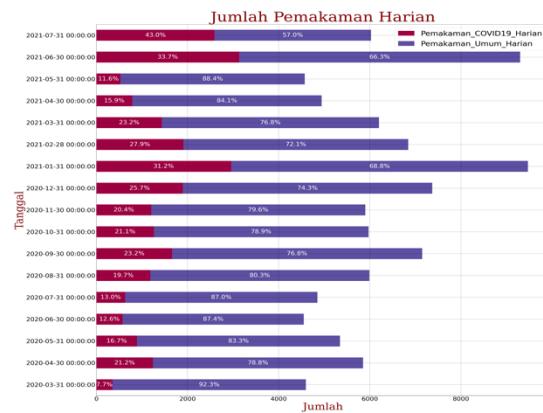
Gambar 4.7 Jumlah Positif COVID-19 Bulanan



Gambar 4.8 Jumlah Negatif COVID-19 Bulanan

## 7.5 Data pemakaman

Berikut merupakan data pemakaman yang divisualisasikan dalam diagram batang.



Gambar 4.9 Jumlah Pemakaman Harian

### III.4 Initial findings

Analisis dilakukan berdasarkan *Exploratory Data Analysis* (EDA) dataset COVID-19 di Jakarta dan Indonesia dengan mengambil data hariannya. Studi kasus yang dipilih dari dataset adalah kasus COVID-19 harian di Jakarta dalam 1 tahun terakhir.

#### 4.1 Dataset

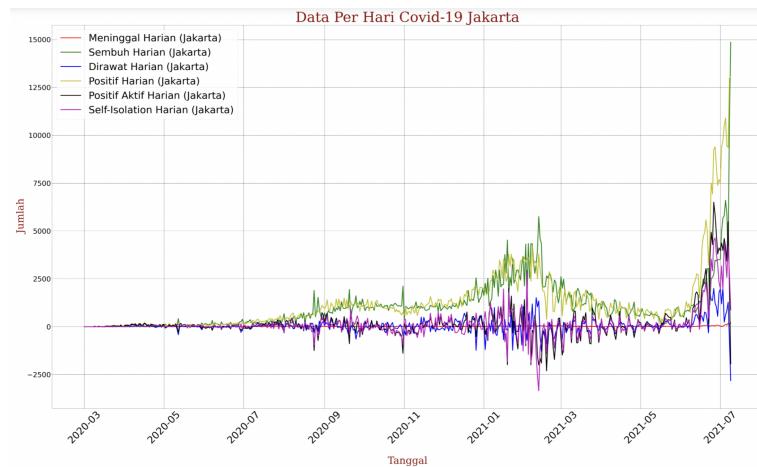
Dataset yang digunakan pada analisis pada studi kasus COVID-19 harian di Jakarta adalah dataset dari Pemerintah Provinsi DKI Jakarta pada kurun periode 01 Maret 2020 - 09 Juli 2021 dengan mengambil sampel data sebanyak 21 variabel dan 496 baris. Berikut adalah contoh dari dataset yang diambil sebagai bahan analisis.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Tanggal	Meninggal (Indonesia)	Sembuh (Indonesia)	Dirawat (Indonesia)	Positif (Indonesia)	Meninggal (Jakarta)	Sembuh (Jakarta)	Self-isolation (Jakarta)	Dirawat (Jakarta)	Positif (Jakarta)	Positif Aktif (Jakarta)	Meninggal Harian (Indonesia)	Sembuh Harian (Indonesia)
2	03/01/2020	0	0	0	0	0	0	0	0	0	0	0	0
3	03/02/2020	0	0	2	2	0	0	0	0	0	0	0	0
4	03/03/2020	0	0	2	2	1	0	0	2	3	2	0	0
5	03/04/2020	0	0	2	2	1	0	0	2	3	2	0	0
6	03/05/2020	0	0	2	2	3	0	0	4	7	4	0	0
7	03/06/2020	0	0	4	4	3	0	0	4	7	4	0	0
8	03/07/2020	0	0	4	4	3	0	0	4	7	4	0	0
9	03/08/2020	0	0	6	6	3	0	0	4	7	4	0	0
10	03/09/2020	0	0	19	19	3	0	0	31	34	31	0	0
11	03/10/2020	0	2	25	27	3	0	0	31	34	31	0	2
12	03/11/2020	1	2	31	34	3	0	0	33	36	33	1	0
13	03/12/2020	1	2	31	34	5	0	0	57	62	57	0	0
14	03/13/2020	4	2	63	69	7	0	0	65	72	65	3	0
15	03/14/2020	5	8	83	96	9	0	0	70	79	70	1	6
16	03/15/2020	5	8	104	117	11	0	0	84	95	84	0	0
17	03/16/2020	5	8	121	134	12	0	7	78	97	85	0	0
18	03/17/2020	5	9	158	172	12	0	27	83	122	110	0	1
19	03/18/2020	19	11	197	227	15	12	42	91	160	133	14	2
20	03/19/2020	25	15	269	309	19	13	57	121	210	178	6	4
21	03/20/2020	32	17	320	369	20	13	66	125	224	191	7	2
22	03/21/2020	38	20	392	450	23	17	71	157	268	228	6	3
23	03/22/2020	48	29	437	514	29	21	77	177	304	254	10	9
24	03/23/2020	49	30	500	579	29	22	79	225	355	304	1	1
25	03/24/2020	55	30	601	686	34	23	109	260	426	369	6	0
26	03/25/2020	58	31	701	790	43	27	112	290	472	402	3	1
27	03/26/2020	78	35	780	893	49	29	113	324	515	437	20	4
28	03/27/2020	87	46	913	1046	57	31	132	346	566	478	9	11
29	03/28/2020	102	59	994	1155	62	43	134	364	603	498	15	13
30	03/29/2020	114	64	1107	1285	67	48	151	435	701	586	12	5
31	03/30/2020	122	75	1217	1414	78	49	151	449	727	600	8	11
32	03/31/2020	136	81	1311	1528	84	49	157	451	741	608	14	6
33	04/01/2020	157	103	1417	1677	90	51	176	499	816	675	21	22
34	04/02/2020	170	112	1608	1700	66	64	186	566	800	760	42	6

Gambar 4. 1 Dataset COVID-19 di Jakarta

#### 4.2. Kasus COVID-19 Harian di Jakarta

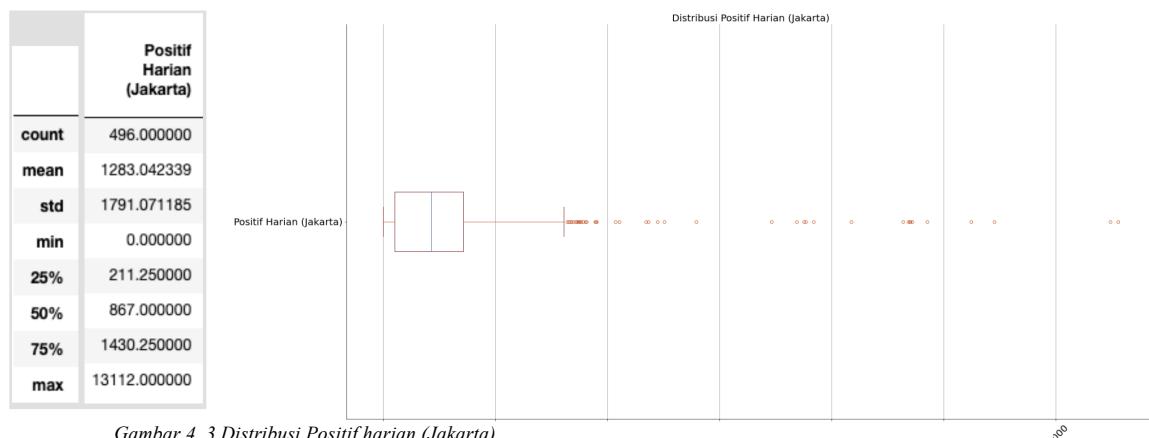
Pada bagian ini, akan dibahas data per-hari COVID-19 di Jakarta. Berikut visualisasi data per-hari COVID-19 Jakarta.



Gambar 4. 2 Data Per COVID-19 Jakarta

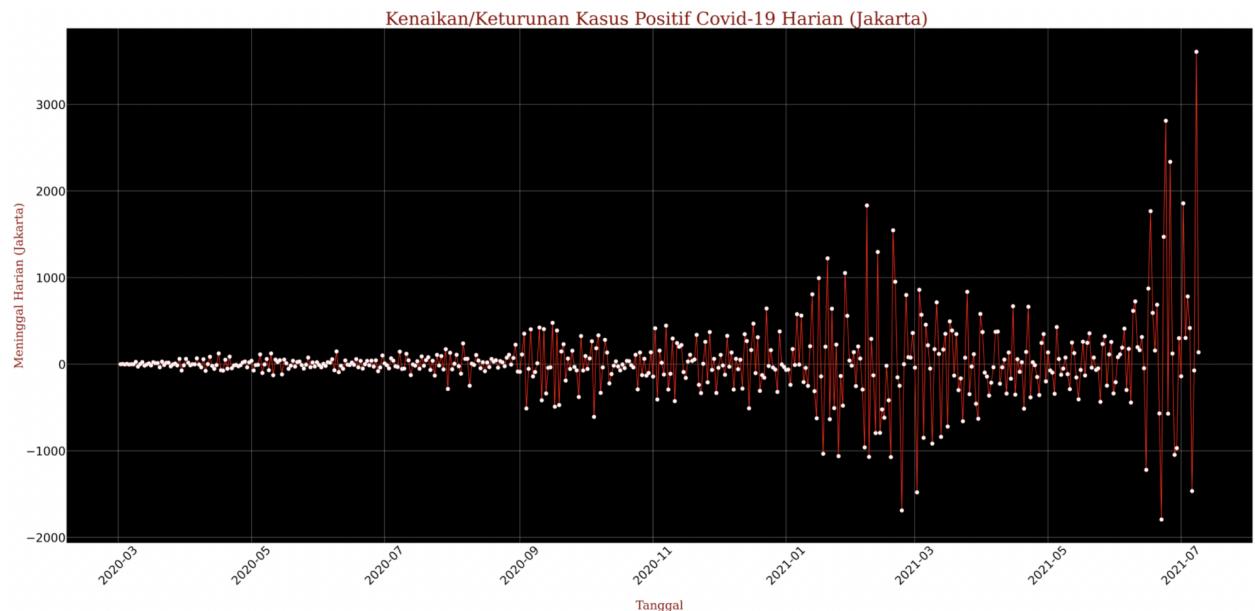
Dari data tersebut, terlihat bahwa jumlah kasus positif harian di Jakarta meningkat secara signifikan. Kenaikan terjadi paling tinggi pada hari ke 8 Juli 2021 dengan jumlah 3608 kasus. Pada data tersebut juga diketahui bahwa pada bulan Januari sampai Maret 2021 merupakan periode dimana terdapat sembuh harian terbanyak dan angka self-isolation berkurang (negatif).

Kasus positif harian di Jakarta sejak tanggal 01 Maret 2020 sampai 09 Juli 2021 memiliki nilai minimum 0 dan nilai maksimum 13112. Rata-rata kasus positif harian di Jakarta adalah sebanyak 1283 kasus. Berikut diagram *boxplot* distribusi kasus positif COVID-19 harian di Jakarta.



Gambar 4. 3 Distribusi Positif harian (Jakarta)

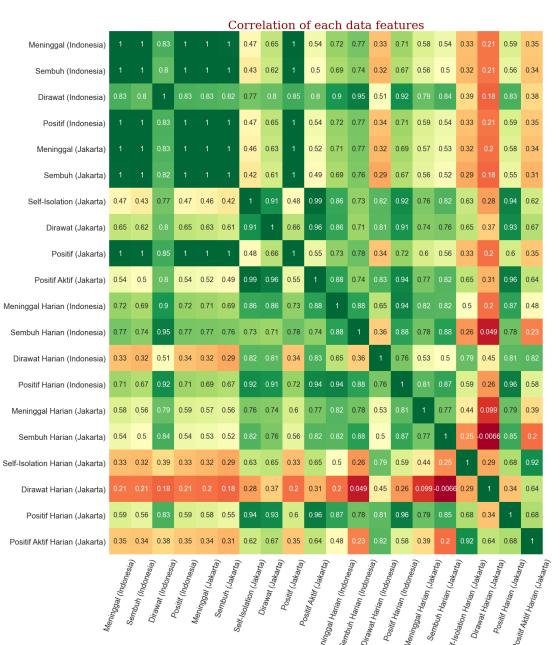
Berikut merupakan grafik fluktuasi kasus COVID-19 Harian di Jakarta.



Gambar 4. 4 Grafik fluktuasi COVID-19 Harian di Jakarta

#### 4.3 Analisis Korelasi Kasus COVID-19

Korelasi dalam ilmu statistik dapat diartikan sebagai metode statistika yang digunakan untuk mengukur keeratan hubungan antara dua variabel (Astuti, 2017). Dengan mengacu pada Dataset ‘Data Indonesia dan Jakarta’, didapat *heat-map* korelasi antar variabel seperti berikut:



Gambar 4. 5 Heat-map Korelasi Data

Warna hijau menunjukkan korelasi positif dengan nilai korelasi = 1, sedangkan warna merah menunjukkan nilai korelasi = 0. Berdasarkan diagram tersebut, diketahui bahwa variabel yang memiliki korelasi erat yaitu variabel Self-isolation (Jakarta) dengan variabel Positif Aktif (Jakarta) yang memiliki nilai korelasi 0.99. Variabel yang memiliki korelasi minimum yaitu variabel Sembuh Harian (Jakarta) dengan variabel Dirawat Harian (Jakarta) yang memiliki nilai korelasi -0.0066.

## **IV. Deep Dive Analysis**

Setelah dilakukan Exploratory Data Analysis, data dapat dijadikan bahan untuk melakukan analisis prediksi. Analisis prediksi merupakan analisis yang mengekstrak informasi dari dataset untuk menentukan pola dan memprediksi hasil maupun tren di masa depan. Analisis ini ditargetkan untuk menguji hipotesis tertentu atau eksplorasi untuk merumuskan hipotesis (Hastir et al., 2008; NRC, 2013). Analisis prediksi kasus COVID-19 di Jakarta dapat menggunakan beberapa metode machine learning, antara lain (Makovsaja 2018):

A. Supervised learning (pembelajaran terawasi)

Metode ini merupakan metode pembelajaran mesin dimana diperlukan tahapan pelatihan berdasarkan data berlabel. Algoritma ini akan menerima serangkaian input yang telah berlabel, selanjutnya algoritma mesin akan belajar. Dengan membandingkan output aktual dengan output yang benar untuk menemukan kesalahan. Pembelajaran terawasi menggunakan pola memprediksi kelas/label pada data tambahan yang tidak berlabel.

B. Unsupervised learning (pembelajaran tidak terawasi)

Metode ini digunakan dimana tidak diperlukan label historis. Sistem tidak diberi tahu jawaban yang benar. Tujuannya adalah mengeksplorasi data dan menemukan beberapa struktur di dalamnya. Salah satu contoh unsupervised learning adalah *clustering*.

C. Semi-supervised learning

Metode ini merupakan metode pembelajaran dimana hanya diperlukan sedikit data yang berlabel. Metode ini menutupi kelemahan pembelajaran terawasi yang membutuhkan banyak data berlabel agar performansinya tinggi.

D. Reinforcement learning

Metode machine learning merupakan metode yang dikembangkan untuk menangani analisis prediksi di luar sampel. Machine learning mengembangkan teknik agar komputer dapat belajar tanpa program secara eksplisit.

## **V. Conclusion and Recommendation**

Berdasarkan analisis diatas, berikut ini merupakan kesimpulan yang dapat diperoleh.

1. Dari data positif COVID-19 Harian di Jakarta dari rentang waktu 01 Maret 2020 sampai 09 Juli 2021, didapatkan nilai minimum dan maksimum berturut-turut 0 dan 13112. Rata-rata positif COVID-19 Harian di Jakarta adalah sebanyak 1283 kasus. Selain itu, diperoleh sebanyak 40 *outlier*.
2. Dua variabel yang memiliki korelasi tertinggi adalah variabel Self-isolation (Jakarta) dengan variabel Positif Aktif (Jakarta). Sedangkan variabel dengan korelasi terendah adalah variabel Sembuh Harian (Jakarta) dengan variabel Dirawat Harian (Jakarta).

Berdasarkan analisis dan kesimpulan tersebut, dapat diterapkan algoritma pembelajaran mesin untuk mendapatkan prediksi-prediksi yang selanjutnya dapat digunakan sebagai bahan pertimbangan untuk kebijakan pemerintah. Prediksi yang dibuat dapat merupakan prediksi resiko terpapar COVID-19 di Jakarta ataupun prediksi kapan kurva pasien positif COVID-19 melandai.

## **VI. Daftar Pustaka**

- Satuan Tugas Penanganan COVID-19. (2021, Juli 9). *Peta Sebaran*. Retrieved Juli 9, 2021, from Data Covid19 di Indonesia: <https://covid19.go.id/peta-sebaran>
- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. 3.
- Astuti, C. C. (2017). Analisis Korelasi untuk Mengetahui Keeratan Hubungan antara Keaktifan Mahasiswa dengan Hasil Belajar Akhir. *Journal of Information and Computer Technology Education*, 5.
- Hawkins, D. M. (1980). *Identifications of Outliers*.
- National Academies of Sciences, E. a. (2017). *Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions*. Washington, DC.
- Jakarta, P. D. (2021). *Daily Update Data Agregat Covid-19 Jakarta*. Jakarta.
- COVID-19, S. T. (2021, Juli 09). *Peta Sebaran*. Retrieved from Covid.go.id: <https://covid19.go.id/peta-sebaran>

## VII. Lampiran

