

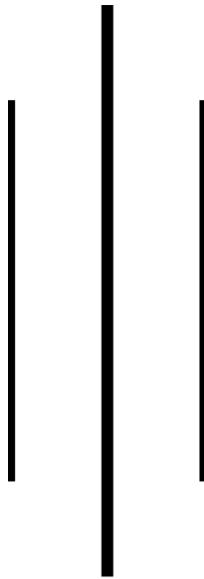
## Bagian 2: non-dataset

1. *Outlier* adalah suatu nilai yang dianggap tidak normal. *Outlier* atau yang biasa disebut juga dengan nilai pencilan bisa dikatakan nilai tersebut mempunyai perbedaan yang sangat jauh dengan pusat data. Sedangkan menurut (Rosa et al., 2015), *outlier* merupakan data yang dianggap tidak mirip atau tidak konsisten dengan data lainnya. *Outlier* atau nilai pencilan ini dapat menyebabkan distorsi terhadap nilai yang asli. Implikasi nilai *outlier* atau pencilan pada analisis data akan mengakibatkan ketidaksesuaian data analisa dan ketidakakuratan model *machine learning* yang sedang dikerjakan. Cara manajemen data untuk menangani kasus *outlier* ini bisa dengan mengganti nilai *outliers* menggunakan nilai *lower bound* dan *upper bound*, atau bisa juga dengan menggantinya dengan nilai mean, modus atau median (Chandra, 2019).
2. Korelasi atau dapat diartikan juga sebagai hubungan, dalam kamus besar bahasa indonesia artinya adalah hubungan timbal balik atau sebab akibat. Korelasi bisa juga untuk menentukan derajat asosiasi menurut (Asuero et al., 2006) Pada dunia statistika, korelasi adalah salah satu tehnik yang digunakan untuk mencari hubungan 2 variabel yang bersifat kuantitatif. Implikasi korelasi dalam teori statistika adalah untuk mengukur atau mengetahui seberapa kuat hubungan antar variabel. Manfaat dari mengetahui seberapa kuat hubungan antar variabel tersebut bisa untuk mendeskripsikan gambaran yang lebih bermanfaat dari data-data yang ada. Salah satu contohnya yaitu suatu perusahaan menggunakan korelasi dalam statistika untuk menentukan hubungan antara variabel pendapatan perusahaan dan variabel gaji karyawan. Dari analisis korelasi antara kedua variabel tersebut, maka perusahaan tersebut dapat membuat suatu keputusan.
3. *Machine learning* atau pembelajaran mesin merupakan salah satu cabang dari ilmu kecerdasan buatan yang dikembangkan untuk bisa belajar sendiri tanpa arahan dari penggunanya. Ilmu statistika, matematika, dan data mining diperlukan dalam *machine learning* untuk membuat sistem dapat belajar dan menganalisa sendiri tanpa perlu diprogram ulang atau diperintah. Meskipun dari disiplin ilmu statistika dan *machine learning* kelihatan sangat berbeda, namun dua disiplin ilmu ini ada kaitannya. *Machine learning* dan statistika adalah dua hal yang saling melengkapi (Cunningham, n.d.). Jadi untuk membuat *machine learning* diperlukan ilmu dari statistika untuk membuat peluang, pemodelan untuk membuat prediksi data baru. Sebaliknya, statistika juga dapat menggunakan *machine learning* untuk mempermudah pekerjaan yang dapat digunakan untuk mengambil kesimpulan dari data yang digunakan.

4. Pada dasarnya *artificial intelligence*, *machine learning*, dan *deep learning* adalah saling berkaitan (Trijaya, n.d.). *Artificial intelligence* atau kecerdasan buatan merupakan gabungan dari beberapa cabang, salah satunya yaitu *machine learning* atau pembelajaran mesin. Sedangkan *machine learning* sendiri mempunyai pecabangan diantaranya adalah *neural network*. *Neural network* atau jaringan syaraf tiruan tersebut merupakan teknik dalam *machine learning* yang menirukan syaraf pada manusia. Syaraf sendiri adalah bagian fundamental dari otak. Dalam *neural network* tersebut ada percabangan teknik yang dinamakan *deep learning*. Jadi *Deep learning* adalah teknik percabangan dari *Neural network*, sedangkan *neural network* merupakan percabangan dari *machine learning*, dan *machine learning* sendiri adalah teknik percabangan dari *artificial intelligence* atau kecerdasan buatan.
5. Intrepetasi data adalah menjelaskan kepada orang lain mengenai hasil dari pengolahan data yang kita dapatkan melalui sudut pandang penafsir. Interpretasi data bisa juga diartikan sebagai proses meninjau data melalui beberapa proses yang telah ditentukan sebelumnya yang akan membantu membuat kesimpulan. Namun menurut ahli Intrepetasi data merupakan deskripsi dan ekspresi dalam suatu penelitian kualitatif. Intepretasi data juga dapat diartikan sebagi usaha untuk menggali pengetahuan tentang data atau peristiwa melalui pemikiran yang lebih mendalam. Tantangan dari intreptasi data yaitu seseorang yang akan menafsirkan atau mengintrepetasikan data seharusnya mempunyai dasar pengetahuan yang kuat dan pengalaman yang matang agar dapat memahami dan mengambil makna dari hasil penafsiran atau intrepetasi tersebut. Intreptasi data dengan *storry telling* dan *decision making* sangat erat hubungannya. Kaitannya dengan *story telling* yaitu seseorang yang akan mengintrepetasi data harus bisa menjelaskan data yang dihasiljan kepada orang lain dengan baik agar orang lain dapat mudah memahami dan tidak salah paham dari apa yang dijelaskan. Setelah menjelaskan kepada orang lain atau *story telling*, hal yang perlu dilakukan yaitu mengambil keputusan atau *decision making*. Pengambilan keputusan diperlukan dalam intrepetasi data untuk mengambil sebuah kesimpulan. Setelah menjelaskan mengenai data yang dihasilkan, langkah selanjutnya yaitu mengambil kesimpulan. Dari kesimpulan tersebut bisa digunakan untuk mengambil keputusan atau *decision making*.

**Bagian 1: Analisis Dataset**

**Laporan Analisis Covid-19 di Jakarta**



**Disusun Oleh:**

Kelompok Moxa

**Anggota:**

Reo Sahobby

Naufal Afkar A

Rozian Firmansyah

**CompFest Academy and Competition**

**2021**

## Latar Belakang

Dunia saat ini sedang dalam masa darurat pandemi virus corona atau COVID-19. Berdasarkan data dari (Septiani, 2021), pada bulan Januari, hanya ada 11 negara yang masih belum ada kasus Covid-19. Covid-19 sendiri adalah penyakit baru yang berasal dari virus SARS-Cov-2. Kasus pertama virus ini terjadi di Wuhan, China pada bulan Desember tahun 2019 (Rezki & Yunus, 2020). Sementara di Indonesia Kasus pertama virus corona dilaporkan terjadi pada tanggal 2 Maret 2020 tepatnya di kota Depok. Dilaporkan terdapat dua orang yang terpapar virus corona setelah melakukan kontak dengan Warga WNA asal Jepang yang positif mengidap covid-19 (Fadli, 2021). Penularan dari virus corona terjadi melalui proses droplet percikan kecil-kecil dari dinding saluran pernapasan seseorang yang sakit pada saat batuk ataupun flu (Rezki & Yunus, 2020). Sedangkan menurut WHO, virus Corona bisa menular secara *Airborne* atau lewat partikel kecil yang ada di udara dan yang terakhir yaitu lewat permukaan barang-barang yang terkontaminasi droplet dan terdapat kontak fisik dengan orang yang normal (WHO, 2020). Seperti kebanyakan virus lainnya, corona juga menimbulkan beberapa gejala bagi orang yang positif terjangkit, berikut ini gejala yang sering muncul yang disebabkan virus corona yang bersumber dari (Kompas, 2021), pertama adalah hilangnya penciuman penderita, kemudian penderita lebih mudah merasa lelah, sakit tenggorokan, demam disertai sakit kepala hingga diare, mual dan muntah.

Seiring berjalannya waktu ternyata tidak semua pasien positif corona ditandai dengan gejala atau disebut dengan orang tanpa gejala (OTG), meskipun tanpa gejala, penderita juga masih dapat menularkan virus tersebut, maka dari itu pencegahan penularan akan virus corona sulit dilakukan. Sebenarnya pemerintah sudah melakukan beberapa kebijakan untuk menghentikan perkembangan virus corona seperti adanya himbauan untuk *work from home* (WFH), *Lockdown* hingga PPKM. Tetapi semua upaya tersebut belum bisa berhasil 100% mengatasi permasalahan pertumbuhan kasus Covid-19. Berdasarkan data dari situs Covid19.go.id sampai dengan tanggal 14 bulan Juli 2021 terdapat 2.670.046 kasus yang terkonfirmasi dengan sebanyak 2.157.363 yang sembuh dan korban meninggal sebanyak 69.210 atau 38% persen dari kasus positif. Berdasarkan Laporan dari (Okezone, 2021), provinsi yang paling banyak terjadi kasus Covid-19 adalah provinsi DKI Jakarta. DKI Jakarta menjadi provinsi dengan tingkat kasus Covid-19 tertinggi dikarenakan beberapa faktor, salah satu faktornya adalah kota Jakarta yang menjadi pintu utama keluar masuknya turis manca negara sehingga interaksi dengan warga negara asing menjadi lebih sering terjadi, selain itu masyarakat kota Jakarta juga memiliki mobilitas yang tinggi yang menyebabkan mudahnya terjadi penyebaran Covid-19.

Bersumber dari dataset yang diberikan, per tanggal 14 Juli 2021 total pasien yang positif Covid-19 mencapai 700.910 kasus dengan rincian 592.556 sembuh, 9.603 meninggal, 78.571 melakukan isolasi mandiri dan sebanyak 21.180 orang memerlukan perawatan. Hal ini menunjukkan selama sebulan kebelakang kasus Covid-19 mengalami kenaikan lebih dari 60 persen dengan sebelumnya terdapat total kasus terkonfirmasi sebanyak 450.793 kasus. Selain informasi mengenai kasus terkonfirmasi positif, dalam dataset juga terdapat data mendetail mengenai kasus positif harian, sembuh harian, penderita tanpa gejala dan bergejala serta data yang telah terinput kedalam database secara harian. Dari semua paparan diatas, kami akan menganalisis kasus Covid-19 di provinsi DKI Jakarta dengan sumber data dari dataset yang diberikan untuk mengetahui lebih dalam persebaran kasus Covid-19 di DKI Jakarta secara lebih mendalam.

### Jawaban Soal

1. Dari dataset yang disediakan, temukan nilai mean, median, dan modus dari positif COVID-19 harian Jakarta.

```
In [6]: df.agg({"positif_harian":["max", np.median, "min", "mean"]})
Out[6]:
```

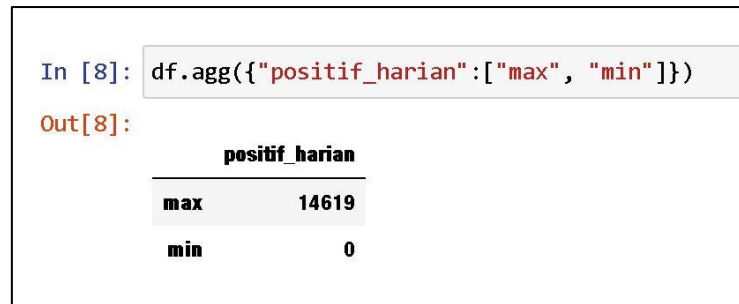
	positif_harian
max	14619.000000
median	884.000000
min	0.000000
mean	1356.835671

```
In [7]: df["positif_harian"].mode()[0]
Out[7]: 0
```

**Gambar 1. 1** Mencari Agg pada kolom positif\_harian

Dari *syntax* yang ditampilkan pada Gambar 1.1 di atas. Dapat diketahui bahwa nilai *mean* atau rata-rata kasus positif harian di Jakarta adalah 1356 kasus, nilai *median* untuk kolom positif\_harian adalah 884, dan *modus* (data yang paling sering muncul) pada kolom positif\_harian adalah 0.

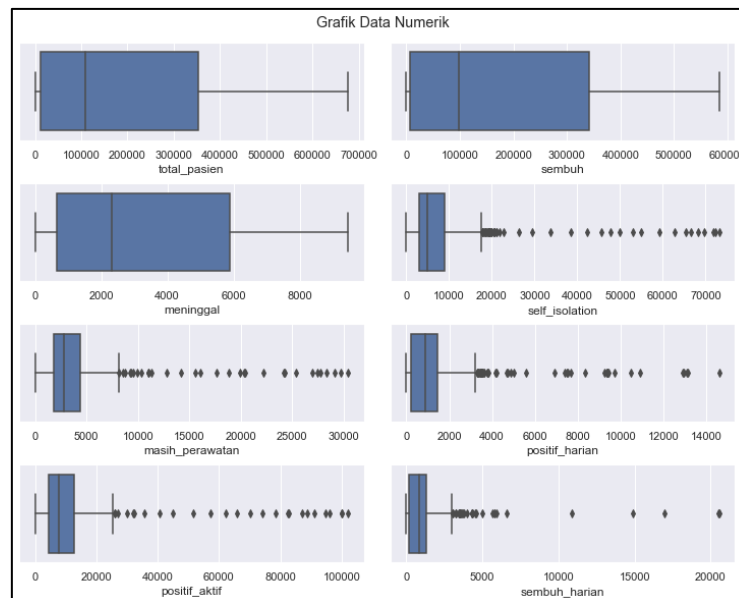
2. Dari dataset yang disediakan, temukan nilai minimal dan maksimal dari positif COVID-19 harian Jakarta.



**Gambar 1.2** Nilai minimal dan maksimal kasus positif harian

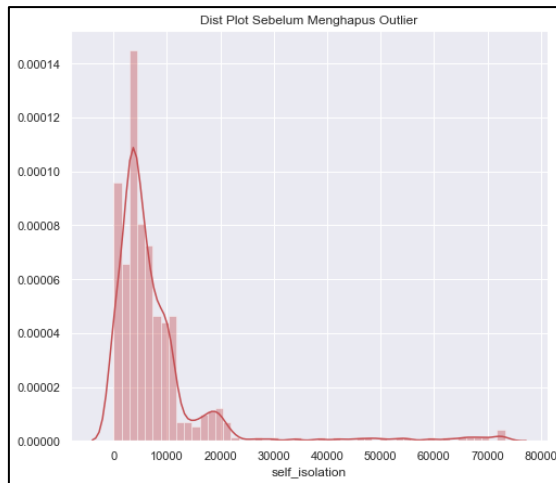
Dari syntax yang ditampilkan pada Gambar 1.2 di atas. Nilai maksimal untuk kasus positif harian di Jakarta adalah 14619 kasus, dan nilai minimal atau kasus terendah positif harian adalah 0 (*tidak ada kasus positif harian*).

3. Dari dataset yang disediakan, temukan nilai-nilai *outlier* yang ada (menggunakan variabel yang kalian tentukan).



**Gambar 1.3** BoxPlot data numerik

Berdasarkan Gambar 1.3 di atas, dari beberapa kolom yang ditampilkan dalam bentuk *boxplot*, ternyata terdapat beberapa kolom yang memiliki *outlier* diantaranya adalah kolom *self\_isolation*, *masih\_perawatan*, *positif\_harian*, *positif\_aktif*, dan kolom *sembuh\_harian*. Oleh karena itu, kami memilih menganalisis *oulier* yang ada di kolom *self\_isolation*. Grafik distribusi dari kolom *self\_isolation* ditampilkan pada Gambar 1.4 di bawah ini.



**Gambar 1. 4** DistPlot kolom *self\_isolation*

Untuk melihat batas yang dirasa merupakan outlier, maka kami menggunakan IQR untuk menghitungnya batas atas dan batas bawahnya. Rumus mencari nilai IQR dan batas-batasnya adalah sebagai berikut.

$$IQR = Q3 - Q1 \dots\dots\dots (1.1)$$

$$batas_{bawah} = Q1 + (1.5 * IQR) \dots\dots\dots (1.2)$$

$$batas_{atas} = Q3 + (1.5 * IQR) \dots\dots\dots (1.3)$$

Keterangan:

$Q1$  = kuartil 1 (25% data)

$Q3$  = kuartil 3 (75% data)

Sesuai dengan rumus di atas, kami melakukan perhitungan terhadap kolom *self\_isolation*.

```
In [13]: q1 = df["self_isolation"].quantile(0.25)
q3 = df["self_isolation"].quantile(0.75)
IQR = q3 - q1
batas_bawah = q1 - (1.5 * IQR)
batas_atas = q3 + (1.5 * IQR)
print(f"nilai Q1 = {q1}\nnilai Q3 = {q3}\nnilai IQR = {IQR}\nnilai batas bawah = {batas_bawah}\nnilai batas atas = {batas_atas}")

nilai Q1 = 2997.5
nilai Q3 = 8868.0
nilai IQR = 5870.5
Batas Bawah = -5808.25
Batas Atas = 17673.75

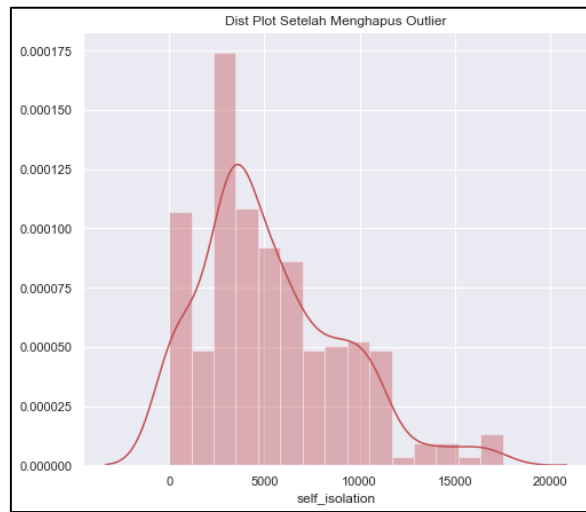
In [14]: df[(df["self_isolation"]>=batas_bawah) & (df["self_isolation"]<=batas_atas)]

Out[14]: max    17524
min         0
Name: self_isolation, dtype: int64
```

**Gambar 1. 5** Perhitungan IQR dan Batas

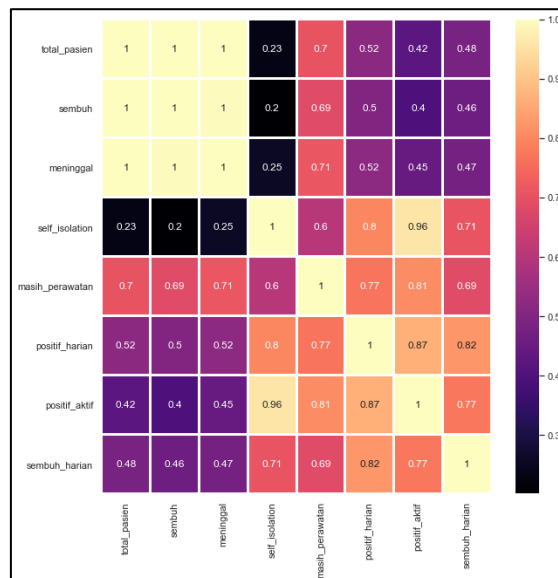
Hasil dari perhitungan yang ditampilkan pada Gambar 1.5 di atas adalah nilai IQR adalah 6870, batas atas adalah 17673, dan batas bawah adalah -5808 atau 0 karena

tidak ada data yang negatif. Sedangkan data pada kolom `self_isolation` yang sudah dibersihkan *outlier* mempunyai rentang data dari 0 sampai 17524. Grafik distribusi kolom `self_isolation` setelah dihapus *outlier* dapat dilihat pada Gambar 1.6 di bawah ini.



**Gambar 1. 6** DistPlot `self_isolation` setelah dihapus outlier

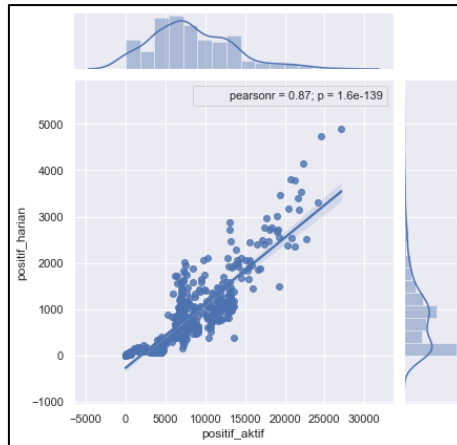
4. Dari dataset yang disediakan, usulkan dua buah variabel dan berikan analisis korelasi antara kedua variabel tersebut. Jelaskan apa kesimpulan yang dapat diambil berdasarkan analisis kalian.



**Gambar 1. 7** Heatmap data covid-19 Jakarta

Berdasarkan grafik heatmap yang dapat dilihat pada Gambar 1.7 di atas, diketahui bahwa kolom yang memiliki korelasi paling erat (*berdasarkan data*) adalah kolom `positif_harian` dan `positif_aktif`. Oleh karena itu, kami memilih kolom tersebut untuk kami analisis. Grafik untuk kedua kolom dapat dilihat pada gambar di bawah ini.





**Gambar 1. 8** *JointPlot kolom positif\_harian dan positif\_aktif*

Dari Gambar 1.8 di atas dapat dilihat bahwa pertambahan jumlah positif aktif selaras dengan bertambahnya jumlah positif harian yang ada. Dari grafik dapat diketahui bahwa korelasi dari kedua variabel tersebut mencapai 0.87 yang berarti memiliki korelasi yang kuat. Kesimpulan yang kami berikan dari analisis grafik di atas adalah kasus positif harian dan positif aktif yang meningkat secara bersamaan. Selama jumlah kasus positif harian yang ada naik, maka kasus positif aktif juga akan naik. Dan juga menandakan pasien yang positif aktif belum banyak yang sembuh.

## Hasil Analisis Tambahan

### a. Problem Statement

Berdasarkan datasets yang ada dan analisis dari data tersebut, permasalahan yang ada adalah:

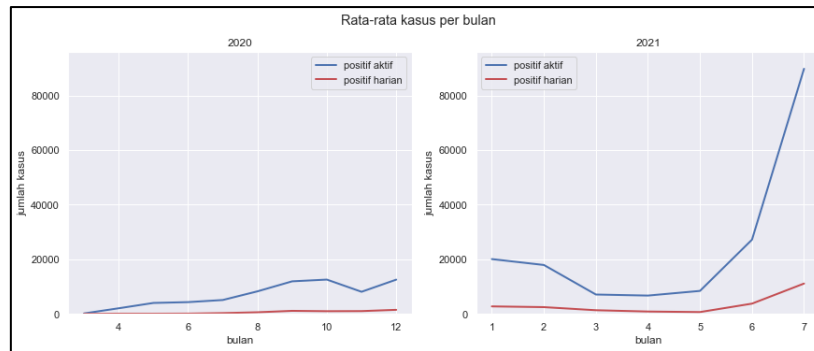
- Kasus positif aktif yang meningkat lagi dari pada tahun 2021 terutama pada bulan Mei hingga Juli.
- Pasien sembuh yang meningkat, namun sayangnya diikuti juga dengan peningkatan jumlah pasien positif dan pasien yang meninggal

### b. Hypothesis

Kasus Covid-19 di DKI Jakarta sudah mencapai puncak dan akan segera menurun dilihat dari kesesuaian antar pertambahan kasus positif dengan pasien sembuh harian.

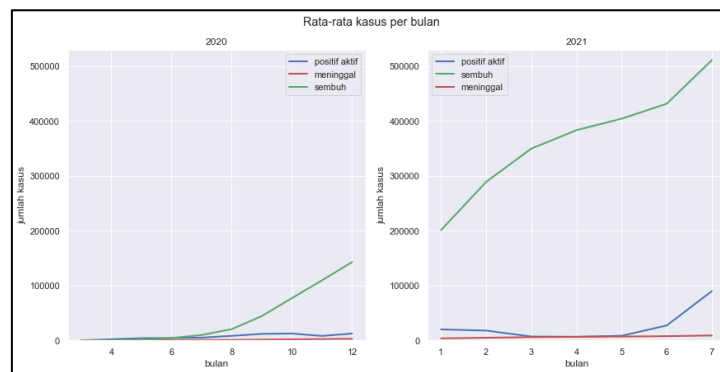
### c. Exploratory Data Analysis

Untuk melakukan analisis kami melakukan beberapa *plotting* terhadap data. Gambar di bawah ini adalah gambar grafik perkembangan covid-19 di Jakarta menurut bulan, dan dibagi menurut tahun.



**Gambar 1. 9** Perkembangan Covid-19 di Jakarta

Berdasarkan grafik pada gambar di atas dapat dikatakan bahwa kasus covid-19 di Jakarta mengalami kenaikan jumlah kasus yang sangat tinggi. Sedangkan tingkat kesembuhan dapat dilihat pada Gambar 1.10 di bawah ini.



**Gambar 1. 10** Tingkat kesembuhan

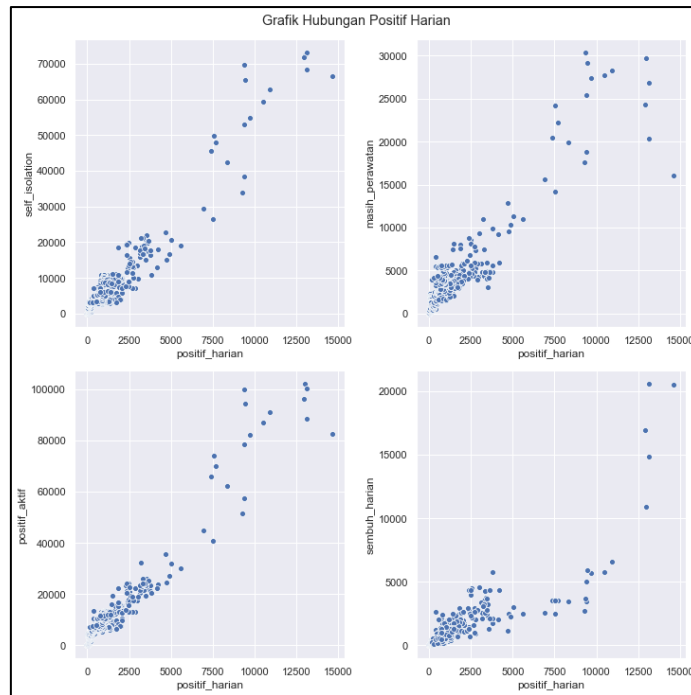
Dari gambar 1.10 di atas bisa dilihat bahwa ternyata jumlah kesembuhan yang ada ternyata sangat banyak, bahkan melebihi jumlah kasus aktif yang ada, karena jumlah kesembuhan tersebut gabungan dari pasien positif aktif maupun kesembuhan dari pasien tanpa gejala maupun dengan gejala ringan lainnya. Data positif harian, positif aktif dan kesembuhan dalam bentuk tabel dapat dilihat pada Gambar 1.11 di bawah ini.

[32]:

		positif_harian	positif_aktif	sembuh
year	month			
2021	7	11132.750000	89767.166667	510709.666667
	6	3780.300000	27204.466667	431337.766667
	1	2773.645161	20088.290323	201060.680645
	2	2600.607143	17931.821429	288962.428571
2020	12	1612.064516	12664.516129	142770.096774
2021	3	1365.161290	7115.096774	349673.387097
2020	9	1135.300000	11944.100000	44344.766667
	11	1042.133333	8118.000000	109508.166667
	10	1007.387097	12682.064516	77101.516129
2021	4	885.500000	6732.966667	383145.233333
	5	691.580645	8441.483871	404089.129032
2020	8	616.387097	8281.451613	20555.741935
	7	320.161290	5108.161290	9914.677419
	6	133.466667	4332.333333	4229.966667
	4	113.233333	2084.466667	194.600000
	5	101.096774	4033.225806	1236.645161
	3	23.903226	188.774194	12.806452

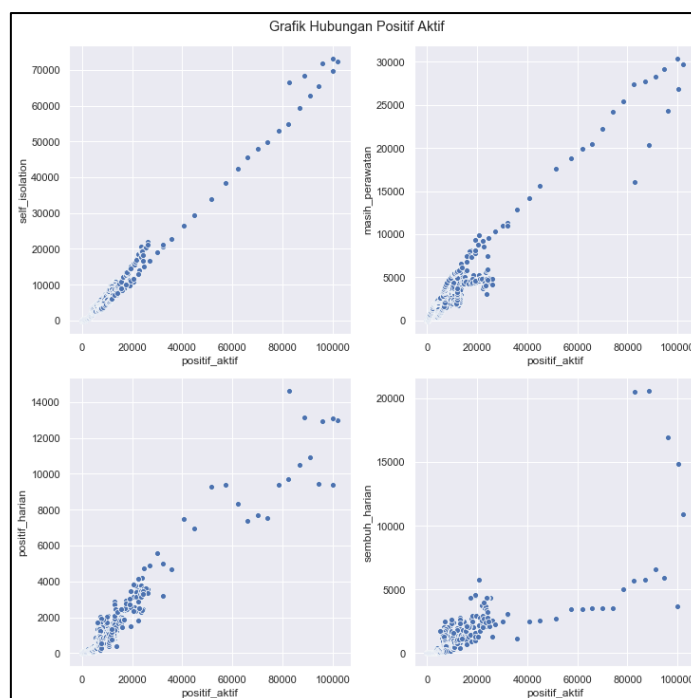
Gambar 1. 11 Data Dibagi per tahun dan bulan

Sedangkan kemungkinan korelasi data yang ada dari data positif harian ditampilkan pada Gambar 1.12 di bawah ini.



**Gambar 1. 12** Grafik data positif harian

Dari grafik di atas terlihat hampir semua data yang ada memiliki korelasi yang baik, yaitu semakin banyak nilai kasus positif harian (*dalam sumbu x*) maka semakin banyak pula nilai yang ditampilkan pada sumbu y. Mulai dari jumlah data *self\_isolation* sampai data kesembuhan yang bertambah walaupun sempat ada kenaikan data positif harian namun kesembuhan tidak mengalami pertambahan. Untuk data yang berkorelasi dengan positif\_aktif ditampilkan pada Gambar 1.13 di bawah ini.



**Gambar 1. 13** Grafik kasus positif harian

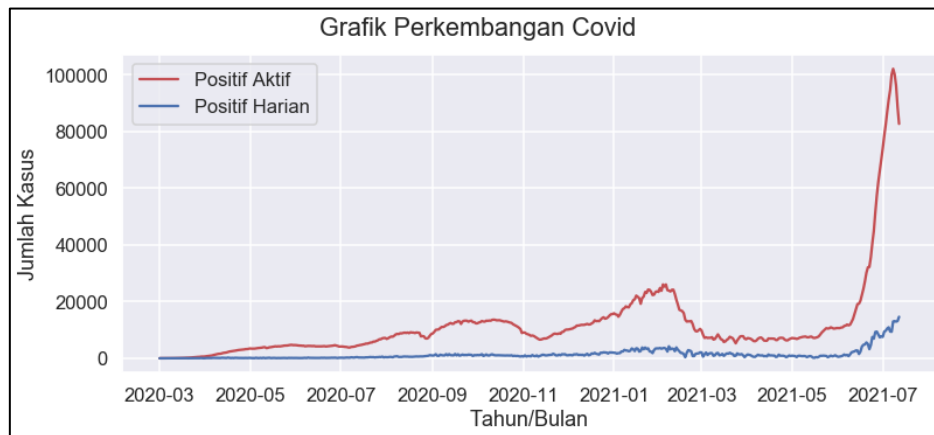
Dari Gambar 1.13 di atas terlihat bahwa data positif aktif semakin naik, maka data lain (*ditampilkan dalam sumbu y*) juga ikut naik. Walaupun dalam data kesembuhan terdapat data positif aktif yang mengalami kenaikan namun data kesembuhan tidak naik.

d. Initial Findings

Dari hasil EDA yang kami lakukan dapat disimpulkan bahwa kasus kenaikan Covid di DKI Jakarta sangat tinggi dilihat dari pertambahan kasus positif harian maupun positif aktif serta terdapat hubungan ketika jumlah kesembuhan sangat banyak maka kasus positif aktif juga meningkat disertai data penunjang lain seperti *self isolation* dll yang juga meningkat maka dapat disimpulkan bahwa kenaikan kasus positif selaras dengan kasus kesembuhan, sehingga diprediksi kedepannya kasus Covid-19 akan segera menurun,

e. Deep Dive Analysis

Sebelum memulai analisis yang lebih lanjut, lebih baik kita melihat perkembangan kasus covid yang ada di Jakarta. Grafik perkembangan dapat dilihat pada Gambar 1.14 di bawah ini.

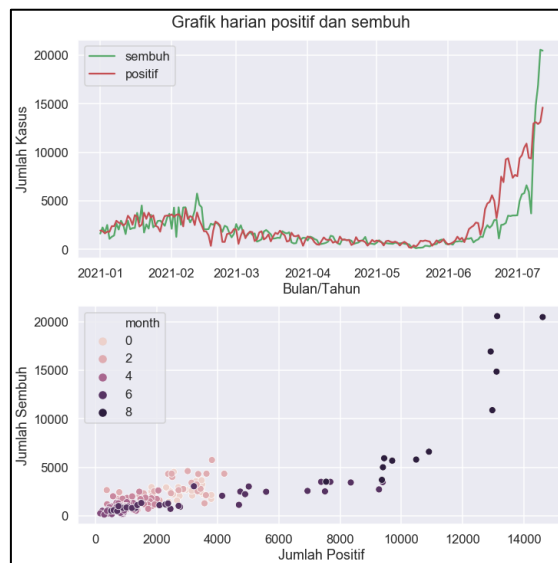


**Gambar 1. 14** Grafik Perkembangan Kasus

Dapat dilihat dari Gambar 1.14 di atas, bahwa jumlah kasus positif aktif yang ada di Jakarta semakin hari mengalami kenaikan yang sangat pesat. Terlihat kenaikan yang sangat signifikan terjadi pada bulan Mei Tahun 2021. Untuk kasus covid di Jakarta pada tahun 2020 juga mengalami kenaikan dan penurunan, namun di tahun 2021 mengalami kenaikan yang sangat tinggi bahkan melebihi jumlah kenaikan tertinggi di tahun 2020. Untuk menganalisis waktu yang lebih *update*, maka selanjutnya kita hanya akan menggunakan data pada tahun 2021.

Selanjutnya setelah mengetahui perkembangan kasus yang terjadi selama tahun 2021, mari kita melakukan analisis terhadap jumlah kasus yang semakin naik dengan

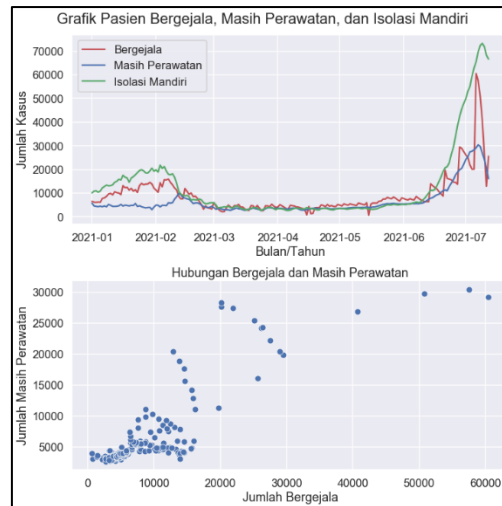
jumlah pasien. Bagaimana grafiknya dan adakah hubungan diantara dua parameter tersebut. untuk melihat grafik data positif harian dan data pasien sembuh harian dapat dilihat pada Gambar 3.15 di bawah ini.



**Gambar 1. 15** Grafik Data Positif Harian dan Sembuh Harian

Dari Gambar 1.15 di atas dapat dilihat bahwa kenaikan jumlah pasien positif harian yang sangat signifikan tentunya masih diikuti dengan kenaikan pasien yang sembuh. Walaupun masih sama seperti grafik sebelumnya bahwa kenaikan kasus baik positif ataupun pasien sembuh terjadi paling banyak di bulan Mei hingga Juli. Pada Gambar 1.15 di atas juga terlihat bahwa hubungan antara jumlah positif dan jumlah pasien sembuh memiliki hubungan yang lumayan kuat, dapat dikatakan bahwa semakin banyak kasus positif harian masih diimbangi dengan semakin banyaknya jumlah pasien yang sembuh.

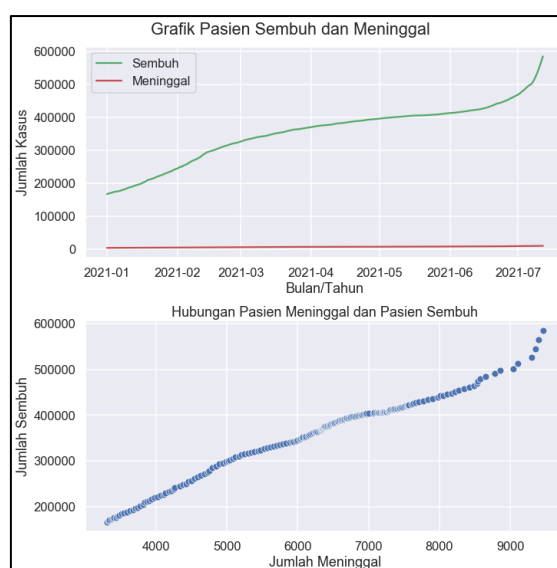
Kemudian kita berpindah ke pasien yang dinyatakan terkena covid. Dalam data yang ada parameter yang menyatakan pasien terkena covid diantaranya adalah jumlah pasien bergejala, jumlah pasien yang masih perawatan, dan jumlah pasien yang isolasi mandiri. Untuk lebih lanjut mari kita lihat grafik pada Gambar 1.16 di bawah ini.



**Gambar 1. 16** Grafik Pasien Bergejala, Masih Perawatan, dan Isolasi Mandiri

Dari Gambar 1.16 di atas dapat dilihat bahwa pada awal tahun 2021 yaitu pada bulan January hingga February mengalami kenaikan pasien, dari data pada bulan tersebut terlihat bahwa pasien yang positif lebih banyak yang menjalani isolasi mandiri karena mungkin pasien tidak mengalami gejala yang terlalu parah. Untuk bulan Mei hingga Juli pasien mengalami kenaikan yang sangat signifikan baik pada pasien yang menjalani isolasi mandiri ataupun pasien yang menjalani rawat inap/perawatan. Pada Gambar 1.16 di atas juga terlihat bahwa dari kedua data tersebut tidak terlalu memiliki hubungan, walaupun kenaikan pasien juga diikuti dengan kenaikan pasien yang dirawat.

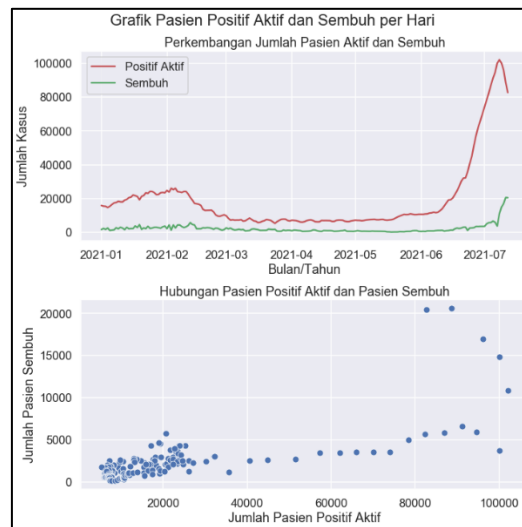
Untuk data jumlah pasien sembuh dan meninggal ditampilkan pada Gambar 1.17 di bawah ini.



**Gambar 1. 17** Grafik Pasien Sembuh dan Meninggal

Hasil dari Gambar 1.17 di atas dapat dikatakan jumlah pasien yang sembuh lebih banyak daripada jumlah pasien yang meninggal. Dari Gambar 1.17 tersebut juga dapat dilihat walaupun terdapat kenaikan jumlah pasien yang meninggal walaupun tidak terlalu signifikan, namun tetap diikuti dengan bertambahnya jumlah pasien yang sembuh.

Selanjutnya adalah data jumlah pasien positif aktif yaitu jumlah pasien yang masih didiagnosis positif, dan jumlah pasien sembuh per hari. Grafik tersebut dapat dilihat pada Gambar 1.18 di bawah ini.

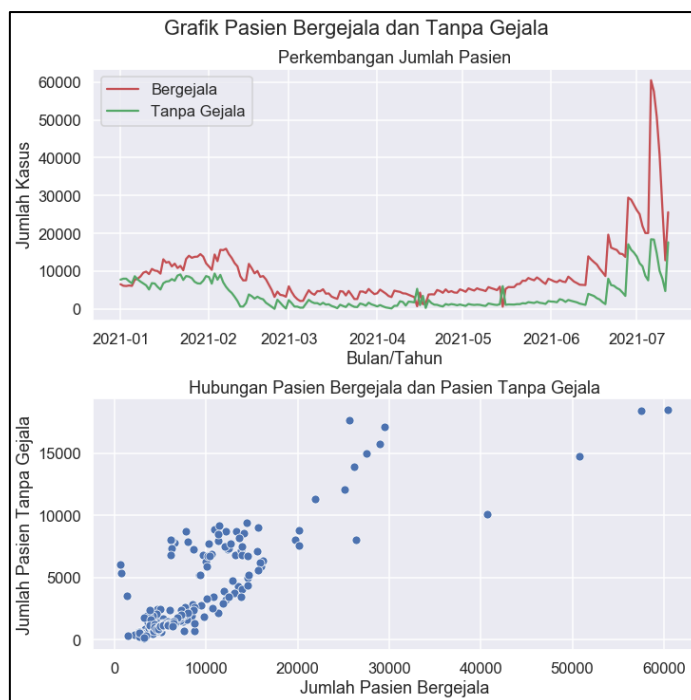


**Gambar 1. 18** Grafik Positif Aktif dan Sembuh Harian

Dari Gambar 1.18 di atas dapat dikatakan bahwa jumlah pasien aktif yang dihitung satu hari diikuti dengan kenaikan jumlah pasien yang sembuh. Dalam gambar tersebut juga dapat dilihat bahwa diantara kedua parameter tersebut tidak terlalu memiliki hubungan yang baik karena grafik *scatter plot* yang dihasilkan tidak terlalu memberikan pola yang jelas.

Kemudian, data terakhir adalah data pasien yang bergejala dan tidak bergejala. Dapat dilihat Pada Gambar 1.19 di bawah ini.





**Gambar 1. 19** Grafik Pasien Bergejala dan Tanpa Gejala

Dari Gambar 1.19 di atas terlihat bahwa jumlah pasien yang bergejala lebih banyak daripada jumlah pasien yang tidak bergejala. Dan dari gambar juga terlihat diantara kedua parameter tersebut tidak terlalu memiliki hubungan yang kuat.

#### f. Conclusion and Recommendation

Dari hasil EDA dan Deep Dive Analysis di atas dapat dilihat bahwa jumlah pasien atau penderita covid dikatakan meningkat sangat signifikan terutama di dua bulan terakhir ini. Walaupun jumlah kesembuhan harian yang ada juga menunjukkan grafik yang baik, dan menunjukkan hubungan yang baik juga. Yaitu kenaikan jumlah pasien yang terkena covid juga diimbangi dengan kenaikan jumlah pasien yang sembuh.

Sedangkan untuk rekomendasi dari kami terkait dengan meningkatnya jumlah kasus positif covid adalah, harus adanya penanganan yang baik dari semua lapisan masyarakat yang ada. Untuk menekan angka pertumbuhan yang semakin tinggi, perlu adanya kerja sama dari semua pihak, baik dari pemberi peraturan yang harus memberikan aturan yang tegas, dan juga pihak masyarakat yang harus menaati aturan demi kebaikan semua.

Sedangkan dalam segi teknologi, kami memberikan rekomendasi adanya sistem prediksi *forecasting* yang dapat digunakan untuk memprediksi jumlah covid di masa depan berdasarkan data yang ada. Sehingga data yang ada dapat digunakan lebih baik dan dengan adanya sistem prediksi tersebut dapat digunakan untuk memberi gambaran

atau membuat kebijakan yang seharusnya dilakukan jika prediksi jumlah kenaikan covid menurut sistem tersebut meningkat. Sehingga kita dapat mengantisipasi lonjakan jumlah kenaikan kasus ini.

## Usaha EDA

Usaha kelompok kami untuk melakukan EDA adalah melakukan *plotting* untuk setiap parameter yang terlihat memiliki hubungan. Karena data yang kami gunakan hanya memiliki dua jenis tipe data yaitu data numerik dan data *datetime*, maka jenis analisis yang kami gunakan adalah membuat scatter plot untuk mencari pola dari dua parameter data, dan untuk data *datetime* kami melakukan konversi untuk dapat mengambil data *datetime* itu, misalnya kami memisahkan data berdasarkan tahun atau bulan.

Untuk menemukan parameter apa saja yang memiliki hubungan kami menggunakan *plotting* heatmap yang akan menghasilkan grafik tingkat hubungan dengan satuan nilai *pearson r*, dan kami melakukan analisis pada data yang sekiranya memiliki hubungan yang kuat.

Contoh *syntax* yang kami gunakan untuk melakukan *plotting* adalah sebagai berikut:

```
fig, ax = plt.subplots(figsize=(10, 10), nrows=2, constrained_layout=True)
fig.suptitle("Grafik Pasien Bergejala, Masih Perawatan, dan Isolasi Mandiri")
_ = ax[0].plot(df[df["year"] == 2021].groupby("tanggal")["bergejala"].mean(),
color="r", label="Bergejala")
_ = ax[0].plot(df[df["year"] == 2021].groupby("tanggal")["masih_perawatan"].mean(), color="b", label="Masih Perawatan")
_ = ax[0].plot(df[df["year"] == 2021].groupby("tanggal")["self_isolation"].mean(),
color="g", label="Isolasi Mandiri")
_ = ax[0].set_xlabel("Bulan/Tahun")
_ = ax[0].set_ylabel("Jumlah Kasus")
ax[0].legend()

_ = ax[1].set_title("Hubungan Bergejala dan Masih Perawatan")
_ = sns.scatterplot(data=df[df["year"] == 2021], x="bergejala",
y="masih_perawatan", ax=ax[1])
_ = ax[1].set_xlabel("Jumlah Bergejala")
_ = ax[1].set_ylabel("Jumlah Masih Perawatan")
plt.show()
```

hasil dari analisis tersebut nantinya kami gunakan untuk menemukan *insight* dan kesimpulan dari data analisis data yang didapatkan.

## Kesimpulan

Berdasarkan analisis data covid-19 di jakarta yang sudah dilakukan maka dapat disimpulkan bahwa kasus positif covid-19 di jakarta masih tinggi. Tetapi kenaikan yang sangat tinggi ini juga dibarengi oleh kasus kesembuhan yang tinggi pula, oleh karena itu diprediksi

kasus covid-19 di DKI Jakarta sudah mencapai puncaknya dan segera menurun dari hari kehari.. Analisis data ini juga dapat dimanfaatkan untuk memprediksi penambahan kasus positif covid-19 di Jakarta. Prediksi penambahan jumlah kasus positif covid-19 ini dapat digunakan untuk pihak-pihak yang berwenang untuk mempersiapkan apa saja yang akan dilakukan di waktu mendatang sebagai bentuk antisipasi penambahan jumlah kasus positif covid-19 di Jakarta. Selain untuk memprediksi jumlah kenaikan kasus positif covid-19 di Jakarta, dari data tersebut bisa digunakan untuk memvisualisasikan grafik pasien sembuh, pasien meninggal, pasien positif aktif, pasien bergejala, pasien tidak bergejala, pasien isolasi mandiri, dan lainnya. Dari visualisasi grafik tersebut bisa menjadi gambaran kasus covid-19 di Jakarta dan diharapkan bisa menjadi bahan pertimbangan untuk mengambil kebijakan agar kasus positif covid-19 di Jakarta dapat terkendali dengan cepat.

## Daftar Pustaka

- Chandra, A. (2019). *Memahami Data Dengan Exploratory Data Analysis*. Medium.  
<https://medium.com/data-folks-indonesia/memahami-data-dengan-exploratory-data-analysis-a53b230cce84>
- Fadli, dr. R. (2021). *Begini Kronologi Lengkap Virus Corona Masuk Indonesia*. Halodoc.  
<https://www.halodoc.com/artikel/kronologi-lengkap-virus-corona-masuk-indonesia>
- Kompas. (2021). *5 Gejala Ringan Tefinfeksi Covid-19 yang Perlu Anda Ketahui*. Kompas Media. <https://www.kompas.com/tren/read/2021/07/10/060600065/5-gejala-ringan-terinfeksi-covid-19-yang-perlu-anda-ketahui?page=all>
- Okezone, T. (2021). *5 Provinsi Penyumbang Kasus Positif Covid-19 Terbanyak Hari Ini*. Okezon TV. <https://nasional.okezone.com/read/2021/06/30/337/2433363/5-provinsi-penyumbang-kasus-positif-covid-19-terbanyak-hari-ini>
- Rezki, A., & Yunus, N. R. (2020). Kebijakan Pemberlakuan Lockdown Sebagai Antisipasi Penyebaran Corona Virus Covid-19. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(3).  
<https://doi.org/10.15408/sjsbs.v7i3.15048>
- Septiani, A. (2021). *11 Negara Ini Masih “Bebas” COVID-19, Mana Saja?* Detikhealth.  
<https://health.detik.com/berita-detikhealth/d-5338356/11-negara-ini-masih-bebas-covid-19-mana-saja>
- Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry*, 36(1), 41–59.  
<https://doi.org/10.1080/10408340500526766>
- Cunningham, S. J. (n.d.). *Machine Learning and Statistics: A matter of perspective*.
- Rosa, P. H. P., Kwa, F. E., & Rosa, P. H. P. (2015). Deteksi Outlier Menggunakan Algoritma Block- based Nested Loop ( *Studi Kasus : Data Akademik Mahasiswa ... Deteksi Outlier Menggunakan Algoritma Block-based Nested Loop ( Studi Kasus : Data Akademik Mahasiswa Prodi PS Universitas XYZ )* ). March, 3–6.
- Trijaya, T. A. M. (n.d.). *Mengenal Artificial Intelligence, Machine Learning, & Deep Learning*.  
<https://amt-it.com/mengenal-perbedaan-artificial-intelligence-machine-learning-deep-learning/>