

DATA SCIENCE ACADEMY
COMPFEST 13



Oleh Tim
“CascadePEOW”

JAWABAN SOAL *NON-DATASET*

1. Jelaskan secara teori statistik mengenai *outlier* (pencilan), implikasinya dalam analisis data, serta bagaimana melakukan manajemen data terhadap kasus *outlier*.

Pencilan (*outlier*) merupakan suatu nilai yang berbeda jauh dari observasi lain dalam sebuah populasi. Hal ini dapat mengacaukan distribusi pada data yang sedang dianalisis. Dalam mendeteksi dan menganalisis *outlier*, kita harus mengetahui kategori *outlier* yang kita temui. Adapun terdapat 3 kategori *outlier* yaitu:

- ❖ Global Outliers. Sebuah data dikatakan sebagai *outlier* global jika nilainya menyimpang sangat jauh terhadap data lainnya (biasanya berupa nilai ekstrim).
- ❖ Contextual Outliers. Sebuah data dikatakan sebagai *outlier* kontekstual jika nilainya menyimpang secara signifikan terhadap data lainnya dalam satu konteks yang sama.
- ❖ Collective Outliers. Suatu subset data dikatakan sebagai *outlier* kolektif jika kumpulan nilainya menyimpang secara signifikan terhadap keseluruhan *dataset*.

Dalam melakukan manajemen data terhadap *outlier*, kita harus mengkaji ulang data tersebut sehingga tidak menimbulkan bias pada proses analisisnya. Adapun terdapat 4 metode untuk melakukan manajemen data terhadap kasus *outlier* yaitu:

- ❖ Membuang data (*drop data*) yang dianggap sebagai *outlier* agar tidak mengacaukan proses analisis.
- ❖ Membatasi data. Contohnya menggunakan metode IQR dalam menentukan batas atas (*upper bound*) dan batas bawah (*lower bound*).
- ❖ Mengganti nilai *outlier* dengan nilai rata-rata atau nilai median data.
- ❖ Mentransformasikan data. Pencilan (*outlier*) dapat menyebabkan distribusi data menjadi tidak normal sehingga diperlukan transformasi untuk menormalkan distribusi data tersebut.

2. Jelaskan konsep dan prinsip korelasi, lalu sebisa mungkin kaitkan dengan dasar-dasar statistik serta implikasinya terhadap konsep/teori statistik lain.

Korelasi dalam lingkup statistika menyatakan relasi antara 2 buah variabel yang dianalisis dalam suatu data. Prinsip korelasi tersebut ditandai dengan adanya koefisien korelasi. Nilai dari koefisien korelasi (r) dapat dinyatakan sebagai berikut:

- ❖ $r = 1$, berarti hubungan antara X dan Y sempurna dan positif (semakin mendekati 1 berarti hubungan semakin kuat dan positif).
- ❖ $r = -1$, berarti hubungan antara X dan Y sempurna dan negatif (semakin mendekati -1 berarti hubungan semakin kuat dan negatif).
- ❖ $r = 0$, berarti tidak ada hubungan antara X dan Y.

Salah satu dasar statistik yang umumnya dipakai dalam menentukan korelasi adalah kovarians yang merupakan suatu ukuran yang berfungsi untuk mencari hubungan antara variabel X dan Y dengan memusatkan data. Kovarians sendiri dihitung dengan mengurangi setiap data dengan nilai rata-rata pada masing-masing variabel X dan Y. Nilai yang telah dipusatkan tersebut kemudian dijumlahkan lalu dikali sehingga mendapatkan nilai kovarians.

$$\text{Cov}(X, Y) = \frac{\sum (X_i - E(X)) (Y_i - E(Y))}{N} \dots (\text{population})$$

$$\text{Cov}(X, Y) = \frac{\sum (X_i - E(X)) (Y_i - E(Y))}{N - 1} \dots (\text{sample})$$

Pada statistika parametrik, biasanya koefisien korelasi antar dua variabel dihitung menggunakan metode pearson. Akan tetapi pada statistika non parametrik, biasanya digunakan metode peringkat spearman, kendall tau, gamma, atau somers.

- ❖ Koefisien korelasi pearson merupakan suatu nilai yang berfungsi untuk mengukur linearitas pada variabel X dan Y. Adapun koefisien korelasi dihitung dengan membagi kovarians terhadap perkalian antara standar deviasi variabel X dan Y. Adapun rumus untuk menghitungnya adalah:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- ❖ Koefisien korelasi peringkat spearman digunakan untuk mengukur hubungan monotonik antar 2 variabel pada statistik non parametrik (ordinal). Terdapat 3 output yaitu *strictly increase*, *strictly decrease*, atau independen. Pada metode ini, dicari rank dari masing-masing variabel dengan cara membagi kovarians dengan standar deviasi. Adapun rumus untuk menghitungnya adalah:

$$\rho_{\text{rank}_X, \rho_{\text{rank}_Y}} = \frac{\text{Cov}(\text{rank}_X, \text{rank}_Y)}{\sigma_{\text{rank}_X} \sigma_{\text{rank}_Y}} \dots (1)$$

$$\rho_s = 1 - \frac{6 \sum (\text{rank}_X - \text{rank}_Y)^2}{n(n^2 - 1)} \dots (2)$$

Note: Rumus 2 hanya berlaku jika nilai dari rank X tidak sama dengan rank Y.

- ❖ Koefisien korelasi peringkat kendall tau memiliki kemiripan dengan metode spearman. Pada metode kendall tau, yang diberi rank hanya salah satu variabel saja sedangkan variabel yang lain dijadikan sebagai acuan untuk menentukan kesearahan dengan variabel yang diberi rank. Metode ini cocok digunakan untuk data diskrit. Adapun rumus untuk menghitungnya adalah:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}$$

Note:

- C menyatakan jumlah pasangan data yang searah (konkordan).
- D menyatakan jumlah pasangan data yang tidak searah (diskordan).
- Pasangan (Xi, Yi) dan (Xj, Yj) disebut konkordan jika (Xi - Xj)(Yi - Yj) > 0. Artinya jika Xi > Xj maka Yi > Yj atau jika Xi < Xj maka Yi < Yj sehingga (X - X) dan (Y - Y) memiliki hasil kali yang selalu positif.
- Pasangan (Xi, Yi) dan (Xj, Yj) disebut diskordan jika (Xi - Xj)(Yi - Yj) < 0. Artinya jika Xi > Xj maka Yi < Yj atau jika Xi < Xj maka Yi > Yj sehingga (X - X) dan (Y - Y) memiliki hasil kali yang selalu negatif.
- ❖ Metode gamma memiliki dasar yang sama dengan metode kendall tau yaitu dengan menghitung banyaknya pasangan konkordan dan diskordan. Metode ini akan lebih efektif jika terdapat pasangan data yang banyak mengandung nilai yang sama. Adapun rumus untuk menghitungnya adalah:

$$G = \frac{C - D}{C + D} \dots (1)$$

$$C = \sum_{i,j} n_{ij} N_{ij}^+ \dots (2) \quad N_{ij}^+ = \sum_{p=i+1}^j \sum_{q=j+1}^k n_{pq} \dots (4)$$

$$D = \sum_{i,j} n_{ij} N_{ij}^- \dots (3) \quad N_{ij}^- = \sum_{p=i+1}^j \sum_{q=1}^{j-1} n_{pq} \dots (5)$$

Note:

- C menyatakan jumlah pasangan data yang searah (konkordan).
- D menyatakan jumlah pasangan data yang tidak searah (diskordan).
- ❖ Metode somers memiliki dasar yang sama dengan metode kendall tau dan gamma. Dengan metode ini, kita dapat mengukur kekuatan korelasi jika kedua variabel memiliki skala ordinal dan ditampilkan dalam bentuk tabel kontingensi. Adapun rumus untuk menghitungnya adalah:

$$d_{yx} = \frac{2(C - D)}{N^2 + \sum_{i=1}^k C_i^2}$$

Note:

- C menyatakan jumlah pasangan data yang searah (konkordan).
- D menyatakan jumlah pasangan data yang tidak searah (diskordan).
- C_i menyatakan frekuensi marginal dari pengamatan.
- d_{yx} menyatakan selisih proporsi pasangan konkordan dan diskordan diantara pasangan data pengamatan yang berangka sama untuk variabel X.

3. Sebutkan teori dasar *machine learning* yang kalian ketahui, lalu jelaskan dalam bahasa sederhana mengenai teori tersebut dan implikasinya.

- ❖ **Pengambilan data:** metode *Machine Learning* (ML) dapat bekerja untuk memprediksi dan menentukan suatu kejadian melalui data yang diperoleh melalui eksperimen. Data-data yang diambil mengandung karakteristik/parameter penting terhadap suatu hal yang ingin diprediksi seperti halnya jika kita ingin komputer dapat membedakan jenis-jenis kucing, maka data yang diperlukan adalah warna bulu, ukuran, dan lainnya. Melalui data tersebut, dibentuklah suatu model yang bertindak seperti suatu fungsi

yang ketika diberikan input berupa data baru, maka akan memberikan output berupa hasil prediksi. Data-data yang diperoleh harus terlebih dahulu dibersihkan dari nilai-nilai yang tidak diinginkan agar dapat membuat prediksi menjadi lebih baik. Himpunan data yang digunakan untuk membangun model disebut juga sebagai *training set*. Selain itu, terdapat pula *testing set*, yakni himpunan data yang digunakan untuk menguji model yang telah dibuat. *Testing set* hanya dapat dilihat dan digunakan setelah model selesai dibuat agar tidak ada asumsi awal atau bias.

❖ **Pembuatan model:** proses pembuatan model dari *training set* yang diperoleh merupakan suatu tahap untuk bisa mendapatkan hasil prediksi dari input data baru yang diberikan. Pemodelan ini dilakukan menggunakan formula matematis yang ada serta dibutuhkan pengetahuan mengenai teori statistik dan ilmu komputer. Pemilihan model tergantung dari jenis data yang diperoleh pada saat pengambilan data. Metode pembuatan model secara umum terbagi atas 3 jenis yaitu:

➤ Supervised Learning

Dengan menggunakan metode ini, *training set* yang diperoleh berupa data yang telah dilabeli dengan output sebuah variabel yang sudah ditargetkan. Contohnya adalah seperti jika ada label mamalia yang berisi data seperti kucing, anjing, sapi dan label unggas yang berisi data seperti bebek dan ayam. Melalui data-data tersebut, akan dibentuk model klasifikasi berdasarkan label yang sudah ada. Proses klasifikasi pada komputer dilakukan dengan menggunakan beberapa algoritma yang ada yaitu *Decision Tree*, *Naive Bayes*, *Support Vector Machine*, *Artificial Neural Network*, dan *Nearest Neighbour Classifier*.

➤ Unsupervised Learning

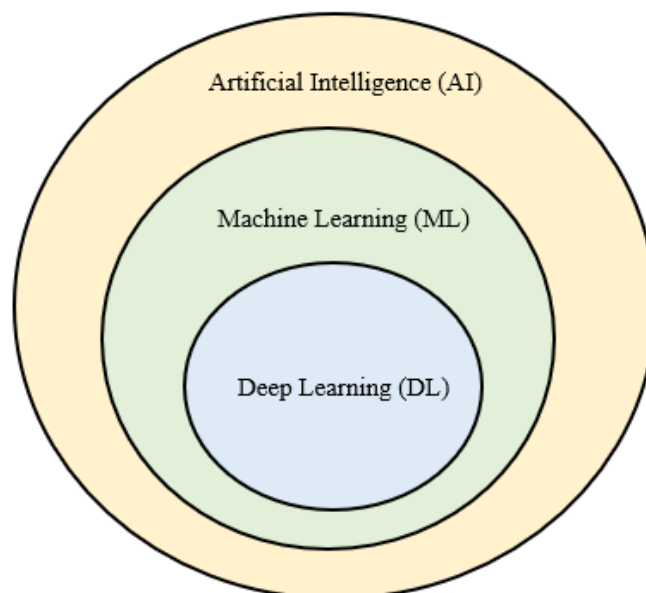
Dengan menggunakan metode ini, data yang diolah untuk dibentuk menjadi model tidak terdapat label sehingga melalui metode ini cenderung untuk membuat label itu sendiri. Tujuan metode ini adalah untuk mengelompokkan *training set* berdasarkan karakteristik atau pola yang ada. Proses tersebut dapat dilakukan dengan menggunakan algoritma yang ada yaitu *K-Means*, *Hierarchical Clustering*, *K-Nearest Neighbors*, *Apriori Algorithm*, dan *Neural Network*.

➤ Reinforcement Learning

Sistem yang telah dibuat atau disebut juga dengan agen akan mengumpulkan data melalui lingkungan sekitarnya layaknya seorang bayi yang baru lahir. Agen akan melakukan adaptasi secara terus menerus terhadap perubahan yang terjadi. Proses tersebut dapat dilakukan dengan menggunakan algoritma yang ada yaitu *Q-Learning* dan *SARSA*.

❖ **Pengetesan model dan evaluasi:** setelah model telah dibuat, maka selanjutnya model tersebut akan diuji tingkat keakuratan prediksinya menggunakan data yang ada. Jika hasil yang diperoleh belum akurat, maka perlu ditinjau kembali proses pengambilan data yang ada dengan meninjau karakteristik lain yang dapat meningkatkan keakuratan atau justru menghapus karakteristik tertentu. Selain itu, algoritma yang digunakan perlu ditinjau kembali dan dapat diganti dengan metode lain.

4. Menggunakan bahasa kalian sendiri, jelaskan kaitan antara *artificial intelligence*, *machine learning*, dan *deep learning*.



Berdasarkan gambar di atas, dapat dilihat bahwa sistem Deep Learning (DL) adalah subset dari sistem Machine Learning (ML) dan tentu saja subset dari Artificial Intelligence (AI). Artificial Intelligence mencakup segala sesuatu dengan tujuan agar suatu mesin dapat berpikir serta bertindak secara otomatis dan rasional (layaknya

manusia). Sistem yang bertambah pintar secara terus menerus tanpa campur tangan manusia disebut sebagai Machine Learning. Adapun ML mengacu pada sistem AI yang dapat belajar sendiri seiring dengan bertambah besarnya *training dataset* berdasarkan algoritma tertentu. Deep Learning sendiri merupakan penerapan sistem Machine Learning dengan ukuran dataset yang sangat besar. Struktur dan fungsi pada metode DL terinspirasi dari cara kerja sel saraf di otak manusia. Pengetahuan yang cukup dibutuhkan agar AI dapat bekerja secara maksimal. Adapun pengetahuan tersebut digali secara terus menerus melalui proses *training* menggunakan metode Machine Learning (ML).

5. Apakah yang kalian ketahui mengenai interpretasi data? Bagaimana signifikansi dan tantangannya? Bagaimana kaitan interpretasi data dengan *data story telling* dan *decision making*?

Interpretasi data adalah cara seseorang dalam memudahkan orang lain memahami data yang telah ia analisis. Dengan kata lain, interpretasi data dapat memberikan arti dan signifikansi terhadap analisis yang dilakukan kepada orang lain. Tantangan yang biasanya dihadapi dalam proses interpretasi data adalah pemaparan yang terlalu rumit dan membosankan. Hal ini dapat menyebabkan kekeliruan interpretasi data sehingga pemahaman orang lain akan keliru.

Data Storytelling merupakan salah satu metode interpretasi data dengan bercerita. Menurut riset yang dilakukan di Stanford Business School, orang mengingat cerita 22 kali lebih banyak dibandingkan fakta. Artinya adalah jika kita ingin membuat interpretasi data hanya dengan fakta-fakta atau statistik tertentu, kemungkinan orang untuk mengingatnya secara mendetail akan kecil. Dengan demikian dapat disimpulkan bahwa orang awam akan lebih mudah mencerna, memahami, dan mengingat data yang sebenarnya membingungkan bagi mereka karena disajikan dalam bentuk cerita.

Decision making adalah tahapan yang paling vital dalam berbagai bidang pekerjaan, tak terkecuali bisnis. Seorang *data scientist* harus bisa menyampaikan *insight* bisnis yang mereka temukan dengan baik ke seluruh anggota tim dengan bahasa yang mudah dipahami. Dengan demikian, *data storytelling* merupakan kemampuan yang wajib dimiliki olehnya karena hasil analisis data yang mereka interpretasikan berpengaruh besar terhadap pengambilan keputusan bagi keseluruhan tim.

DATA SCIENCE ACADEMY
COMPFEST 13



Oleh Tim
“CascadePEOW”

LAPORAN ANALISIS PROBLEM SET
SOAL SELEKSI

LATAR BELAKANG

Perkembangan teknologi informasi membuat produksi data meningkat setiap harinya dan diperkirakan akan terdapat 163 Triliun GB data di seluruh dunia pada 2025 mendatang. Jika diolah dengan baik, data akan memberikan banyak pengetahuan atau inovasi, khususnya bagi perkembangan industri, bisnis, maupun sektor lainnya. Namun, pengolahan data memerlukan kemampuan analisis dan tingkat ketelitian yang tinggi sehingga *output* yang dihasilkan menjadi lebih akurat. Data Science Academy (DSA) hadir sebagai solusi bagi seseorang yang ingin memiliki *career path* dalam bidang *data science*.

Data Science Academy (DSA) adalah sebuah rangkaian acara tahunan yang diselenggarakan oleh COMPFEST 13 dengan tujuan untuk memberikan *training academy* bagi seseorang yang ingin berkarir sebagai *data scientist*. Selama rangkaian acara, peserta akan mendapatkan materi dari pembicara yang berpengalaman di sesi camp dan mentoring pada bagian kedua. Peserta yang telah mengikuti rangkaian acara ini diharapkan akan memiliki *skill set* yang baik sebagai seorang *data scientist* sehingga dapat memenuhi kebutuhan yang ada dalam industri.

Dengan diadakannya acara ini, kami merasa tertantang sehingga kami membentuk sebuah tim bernama “CascadePEOW” yang beranggotakan 3 orang, yaitu Alvin Filipi, Louis Mayco Dillon Wijaya, dan Wilson Alfrido Situmorang. Kami semua berasal dari Fakultas Teknik Universitas Indonesia dengan jurusan Teknik Elektro.

Data yang digunakan dalam proses penyisihan kali ini adalah data COVID-19 di Jakarta (dataset). Dataset ini didapatkan pada website corona.jakarta.go.id. Adapun dataset tersebut berformat xlsx dan memuat 10 buah *sheet* yaitu Data COVID-19 di Indonesia dan Jakarta, Data COVID-19 Jakarta, Data Suspek Jakarta, Rapid Diagnostic Test Antibodi, Rapid Diagnostic Test Antigen, Hasil Lab, Vaksinasi, Pemakaman, Vaksinasi Wilayah, dan Vaksinasi Wilayah Harian Jakarta. Melalui data-data tersebut, kami berharap dapat memvisualisasikan gambaran persebaran COVID-19 dan memperoleh langkah-langkah baru yang inovatif dalam menangani permasalahan COVID-19 di Jakarta.

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

Mean : 1.202,23 kasus/hari \cong 1.202 kasus/hari

Median : 859 kasus/hari

Modus : 661,67 kasus/hari \cong 662 kasus/hari

```
Range = Data_Jak['positive_daily'].describe()['max'] - Data_Jak['positive_daily'].describe()['min']
Class = np.ceil(1 + 3.3*np.log10(Data_Jak['positive_daily'].describe()['count']))
Interval = np.ceil(Range/Class)
Data_Jak_Class = []
Tb = Data_Jak['positive_daily'].describe()['min'] - 0.5
Ta = 0
add = 0 # for checking
class_range = []
Data_Jak_Class_Daily = []
print("Kelas\t\t\t\t\tNilai")
for i in range(int(Class)):
    Ta = Tb + Interval + 1
    Data_Jak_Class.append(Data_Jak[Data_Jak['positive_daily'].between(Tb,Ta)])
# print("converting Class " + str(i) + " - > " + str(Tb) + " to " + str(Ta))
    class_range.append(str(Tb + 0.5) + " - " + str(Ta - 0.5))
    Data_Jak_Class_Daily.append(Data_Jak_Class[i].shape[0])
    print(str(class_range[i]) + " \t\t" + str(Data_Jak_Class_Daily[i]))
    Tb = Ta
```

Kelas	Nilai
0.0 - 1091.0	315
1092.0 - 2183.0	111
2184.0 - 3275.0	29
3276.0 - 4367.0	20
4368.0 - 5459.0	4
5460.0 - 6551.0	1
6552.0 - 7643.0	4
7644.0 - 8735.0	2
8736.0 - 9827.0	4
9828.0 - 10919.0	2

```
Mo = -.5 + (315/(315+315-111))*Interval
```

```
print("Mean\t: " + str(Data_Jak['positive_daily'].describe()['mean']))
print("Median\t: " + str(Data_Jak['positive_daily'].describe()['50%']))
print("Modus\t: " + str(Mo))
```

```
Mean    : 1202.2317073170732
Median  : 859.0
Modus   : 661.6676300578034
```

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

Minimal : 0 kasus/hari

Maksimal : 10.903 kasus/hari

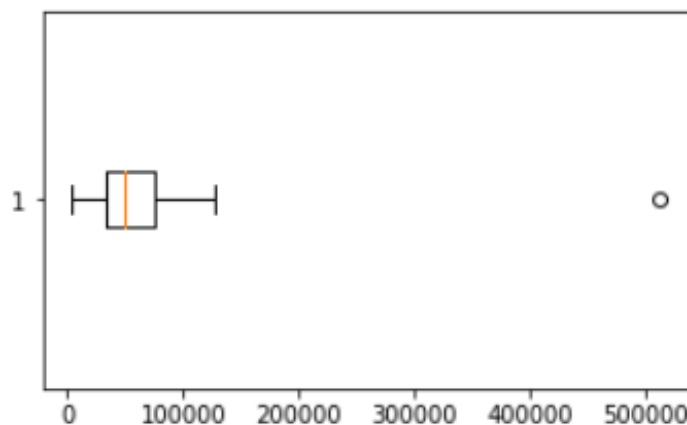
```
print("Minimal\t: " + str(Data_Jak['positive_daily'].describe()['min']))
print("Maximal\t: " + str(Data_Jak['positive_daily'].describe()['max']))

Minimal : 0.0
Maximal : 10903.0
```

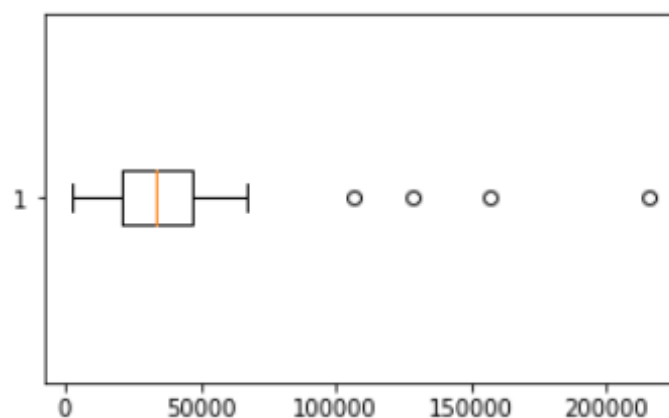
3. Dari *dataset* yang disediakan, temukan nilai-nilai *outlier* yang ada (menggunakan variabel yang kalian tentukan).

Nilai-nilai *outlier* didapatkan pada *datasheet* Vaksinasi Wilayah. Berdasarkan rumus perhitungan *outlier* menggunakan IQR, didapatkan nilai-nilai *outlier* sebagai berikut:

- Pada tanggal 13 Januari - 11 Juni 2021, untuk cakupan suntik 1 di daerah Tanah Abang didapatkan 1 nilai *outlier* yaitu 511.564. Hal ini terbukti melalui visualisasi *box plot* berikut ini:



- Pada tanggal 13 Januari - 11 Juni 2021, untuk cakupan suntik 1 didapatkan 4 nilai *outlier* yaitu 156.991 di daerah Palmerah, 128.537 di daerah Tanah Abang, 215.545 di daerah Kebayoran Baru, dan 106.825 di daerah Gambir. Hal ini terbukti melalui visualisasi *box plot* berikut ini:

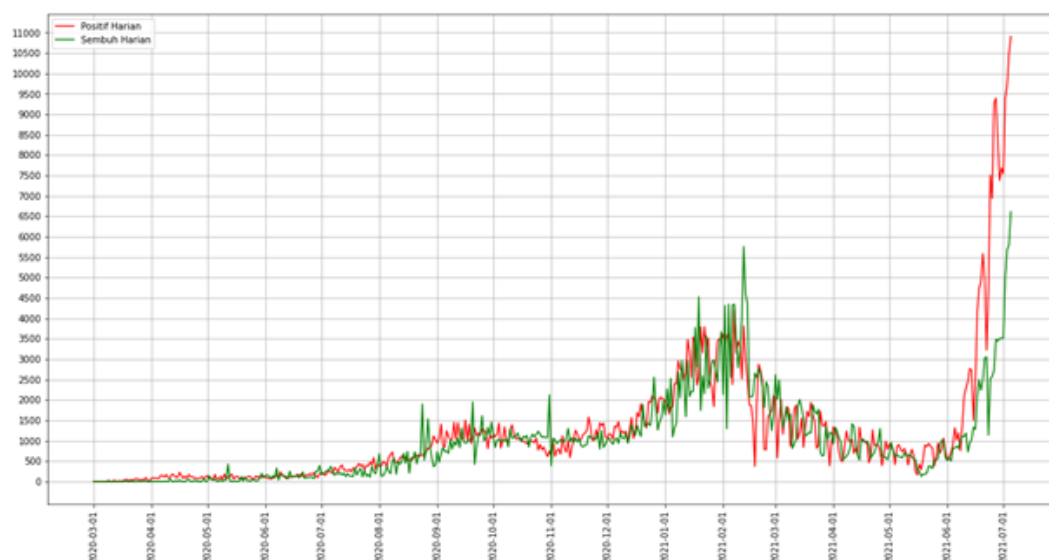


4. Dari *dataset* yang disediakan, usulkan dua buah variabel dan berikan analisis korelasi antara kedua variabel tersebut. Jelaskan apa kesimpulan yang dapat diambil berdasarkan analisis kalian.

Kami mengusulkan variabel Positif Harian dan Sembuh Harian pada *datasheet* “Data Jakarta”. Hubungan antara kedua variabel tersebut dicari menggunakan koefisien korelasi spearman dan didapatkan angka sebesar 0,901728.

	Positif Harian (Jakarta)	Sembuh Harian (Jakarta)
Positif Harian (Jakarta)	1.000000	0.901728
Sembuh Harian (Jakarta)	0.901728	1.000000

Nilai koefisien korelasi tersebut cukup besar yang mengindikasikan adanya hubungan positif yang sangat kuat antara variabel Positif Harian dan Sembuh Harian. Semakin banyak orang yang terkena COVID-19, semakin banyak pula orang yang mendapatkan perawatan sehingga peluang untuk sembuh akan semakin besar. Akan tetapi, tak memungkiri juga ada pasien yang meninggal. Hal yang harus diperhatikan adalah seberapa besar tingkat kesembuhan pasien COVID-19 tiap harinya untuk dapat menekan peningkatan kasus aktif COVID-19 dan mengurangi tingkat kematian akibat COVID-19.



Pada rentang Maret 2020 - Juni 2021, terlihat bahwa variabel positif harian dan sembuh harian memiliki tren yang sama. Meskipun kasus positif COVID-19 harian meningkat, hal ini diimbangi juga dengan peningkatan jumlah kasus sembuh harian.

Perbedaan nilai antara sembuh harian dan positif harian pada interval waktu tersebut juga cenderung lebih kecil jika dibandingkan dengan Juni - Juli 2021 yang terus membesar. Hal tersebut dapat terjadi karena terjadinya peningkatan kasus positif harian yang sangat besar dalam interval 1 bulan tersebut, tetapi tingkat kesembuhan tidak dapat mengimbangi besar kasus positif harian. Perubahan angka kasus positif COVID-19 tersebut cenderung sangat tinggi dan pada interval waktu yang lebih singkat jika dibandingkan pada bulan sebelumnya yang cenderung perlahan. Peningkatan kasus dalam waktu yang cukup singkat itu mempengaruhi tingkat kesembuhan pasien COVID-19. Semakin banyak yang terkena COVID-19, maka fasilitas kesehatan seperti rumah sakit menjadi penuh. Hal ini dapat menghambat proses penyembuhan pasien serta justru dapat meningkatkan kasus kematian karena lambatnya penanganan. Fasilitas kesehatan adalah satu faktor dari tingkat kesembuhan yang ada selain daripada kemampuan tubuh pasien itu sendiri. Hingga saat ini, tercatat bahwa tingkat kelangkaan kamar tidur rumah sakit di Jakarta melebihi 90%.

Kesimpulan yang dapat diperoleh melalui hubungan kedua variabel tersebut adalah pada tingkat peningkatan kasus positif COVID-19 yang perlahan dalam waktu yang lama, tingkat kesembuhan masih dapat mengimbangi. Sedangkan jika peningkatan kasus positif menjadi sangat tinggi dan dalam interval waktu yang lebih singkat, tingkat kesembuhan tidak dapat mengimbangnya. Hal ini menyebabkan terjadinya akumulasi kasus aktif COVID-19.

Berdasarkan hal tersebut, diperlukan kesadaran masyarakat untuk mematuhi protokol kesehatan. Hal-hal berikut ini dapat terjadi jika tercipta kesadaran masyarakat yang tinggi:

1. Grafik kasus positif COVID-19 harian di Jakarta menjadi lebih landai.
2. Penanganan medis oleh tenaga kesehatan akan lebih maksimal. Menurut Badan Pusat Statistik, terdapat 63.098 tenaga kesehatan di Jakarta pada tahun 2020. Angka kasus COVID-19 harian yang semakin tinggi tentu akan membuat pekerjaan mereka menjadi tidak maksimal.
3. Dengan adanya penurunan jumlah kasus COVID-19, maka ketersediaan rumah sakit akan meningkat

HASIL ANALISA TAMBAHAN

5. Dari *dataset* yang disediakan, buatlah analisis dengan runtunan berikut:

a. Problem Statement

Jumlah kasus positif COVID-19 harian di Jakarta mengalami kenaikan yang signifikan dalam 1 bulan terakhir (Juni - Juli 2021). Faktor penyebab kenaikan jumlah kasus positif COVID-19 harian dapat dipengaruhi oleh beberapa hal. Berdasarkan permasalahan tersebut, kami dapat merumuskan beberapa masalah yaitu:

- Adakah hubungan antara jumlah spesimen yang di tes terhadap kenaikan jumlah kasus positif COVID-19 harian di Jakarta?
- Adakah hubungan timbal balik antara jumlah pasien yang menjalani *self isolation* dengan peningkatan kasus positif COVID-19 harian di Jakarta?
- Adakah pengaruh vaksinasi terhadap kenaikan angka positif COVID-19 harian di Jakarta?
- Adakah faktor lain yang dapat memberikan dampak signifikan terhadap kenaikan kasus positif COVID-19 harian di Jakarta?

b. Hypothesis

H0: Tidak ada korelasi yang kuat antara kenaikan kasus positif COVID-19 harian terhadap variabel jumlah tes spesimen harian, *self isolation* harian, dan program vaksinasi di Jakarta.

H1: Ada korelasi yang kuat antara kenaikan kasus positif COVID-19 harian terhadap variabel jumlah tes spesimen harian, *self isolation* harian, dan program vaksinasi di Jakarta.

Kriteria: Koefisien korelasi dari variabel jumlah tes spesimen harian, *self isolation* harian, dan program vaksinasi terhadap positif COVID-19 harian di Jakarta memiliki nilai $> 0,50$.

c. Exploratory Data Analysis

Pada tahapan eksplorasi data ini, akan dilakukan analisis korelasi kuantitatif dengan menggunakan analisis bivariat, analisis multivariat, dan koefisien korelasi spearman. Menurut Jonathan Sarwono (2009), nilai korelasi dikelompokkan menjadi 6 yaitu:

Nilai	Keterangan
0	Tidak ada korelasi
0 - 0,25	Sangat lemah
0,25 - 0,50	Cukup
0,50 - 0,75	Kuat
0,75 - 0,99	Sangat kuat
1	Sempurna

Berdasarkan data yang diberikan dan ditambah dengan sedikit modifikasi, kami akan mencari korelasi dari setiap variabel yang telah kami tentukan yaitu:

Nama Variabel	Keterangan
Positive_Daily	Jumlah positif COVID-19 harian di Jakarta
Positive	Jumlah kasus positif berdasarkan hasil tes lab harian
Negative	Jumlah kasus negatif berdasarkan hasil tes lab harian
Total_Test_Daily	Jumlah tes harian di Jakarta (penjumlahan antara variabel Positive dan Negative)
Self_Isolation_Daily	Jumlah orang yang melakukan <i>self isolation</i> harian
Recover_Daily	Jumlah orang yang sembuh harian
Total Vaksin Harian	Jumlah pemberian vaksin 1 dan vaksin 2 harian
Positif H+7 Harian	Jumlah kasus positif selama 7 hari

	kedepan setelah hari H
Positif H+14 Harian	Jumlah kasus positif selama 14 hari kedepan setelah hari H

- **Data Cleansing**

Sebelum melakukan analisis korelasi kuantitatif pada variabel-variabel *dataset*, kami melakukan eksplorasi data dan menemukan kesalahan perhitungan. Adapun kesalahan tersebut kami temukan pada variabel “Total Spesimen di Test” di *datasheet* Hasil Lab yang merepresentasikan jumlah antara variabel positif dan negatif pada *datasheet* Hasil Lab. Kesalahan perhitungan terdapat pada data bertanggal 18 Mei 2021.

```
Hasil_Lab['Total_Test_Daily'] = Hasil_Lab['Positif'] + Hasil_Lab['Negatif']

Hasil_Lab["Fit"] = (Hasil_Lab['Total_Test_Daily'] == Hasil_Lab['Total Spesimen di Test'])
Hasil_Lab.Fit.value_counts()

True      492
False      1
Name: Fit, dtype: int64

DifferenceValue_Count = Hasil_Lab['Total Spesimen di Test'] == Hasil_Lab['Total_Test_Daily']

FalseCount_HasilLab = DifferenceValue_Count[DifferenceValue_Count == False]
FalseCount_HasilLab

444      False
dtype: bool

Hasil_Lab['Tanggal'][444]

Timestamp('2021-05-18 00:00:00')
```

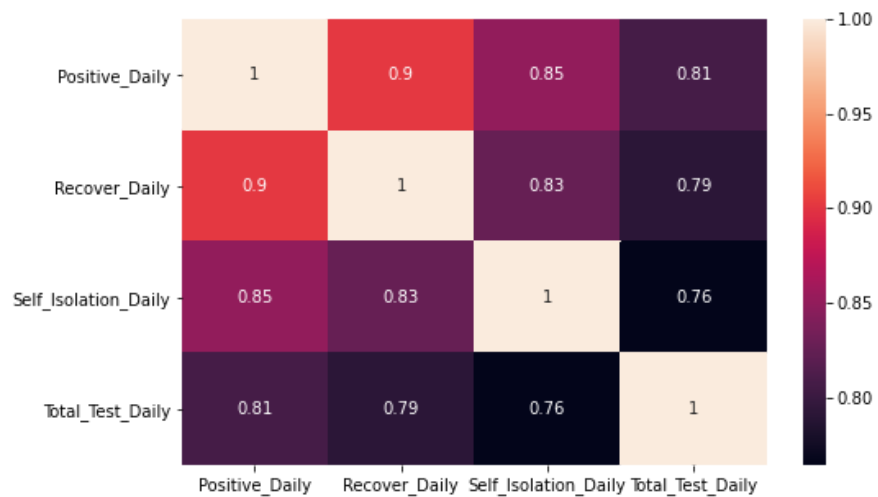
Hal ini mungkin terjadi karena adanya kesalahan dalam penginputan hasil hitungan data pada *dataset* dan dapat menyebabkan adanya perbedaan nilai rata-rata dan standar deviasi pada variabel “Total Spesimen di Test” sehingga persebaran distribusi data tidak sesuai dengan yang seharusnya dihasilkan. Adapun perbandingan nilai mean dan standar deviasi sebelum dan sesudah dilakukan *data cleansing* yaitu:

Mean		Standar Deviasi	
Sebelum	Sesudah	Sebelum	Sesudah
11.853,99	11.862,68	7.251,11	7.247,15

Setelah itu, data dideklarasikan ke dalam *dataset* COVID19JKT dengan nama variabel yaitu Total_Test_Daily.

- **Analisis Multivariat**

Pada analisis multivariat, dilakukan uji coba kuantitatif antara 4 buah variabel yaitu Self_Isolation_Daily, Positive_Daily, dan Total_Test_Daily. Hubungan antara keempat variabel tersebut ditentukan melalui koefisien korelasi dengan metode spearman. Adapun koefisien korelasi yang didapat dapat dilihat pada visualisasi heatmap berikut ini:



Berdasarkan visualisasi heatmap di atas, dapat disimpulkan bahwa:

1. Variabel Self_Isolation_Daily memiliki pengaruh yang sangat kuat terhadap Positive_Daily dengan koefisien korelasi sebesar 0,85.
2. Variabel Total_Test_Daily memiliki pengaruh yang sangat kuat terhadap Positive_Daily dengan koefisien korelasi sebesar 0,81.

- **Analisis Bivariat**

Analisis bivariat diterapkan untuk mencari korelasi antara data warga yang tervaksinasi dengan angka positif COVID-19 harian di Jakarta. Hubungan antara kedua variabel tersebut dicari menggunakan koefisien korelasi spearman. Adapun koefisien korelasi yang didapat yaitu:

1. Pada interval 5 Maret - 16 Juli 2021 (periode vaksinasi keseluruhan), didapatkan koefisien korelasi sebesar 0,39. Artinya adalah terdapat hubungan cukup (moderat) antara Total Vaksin Harian dengan Positif Harian H+7.
2. Pada interval 1 Juni - 16 Juli 2021 (periode vaksinasi massal), didapatkan koefisien korelasi sebesar 0,58. Artinya adalah terdapat hubungan kuat antara Total Vaksin Harian dengan Positif Harian H+7.

d. Initial Findings

Berdasarkan usaha eksplorasi data yang telah kami lakukan, kami menemukan beberapa poin penting yaitu:

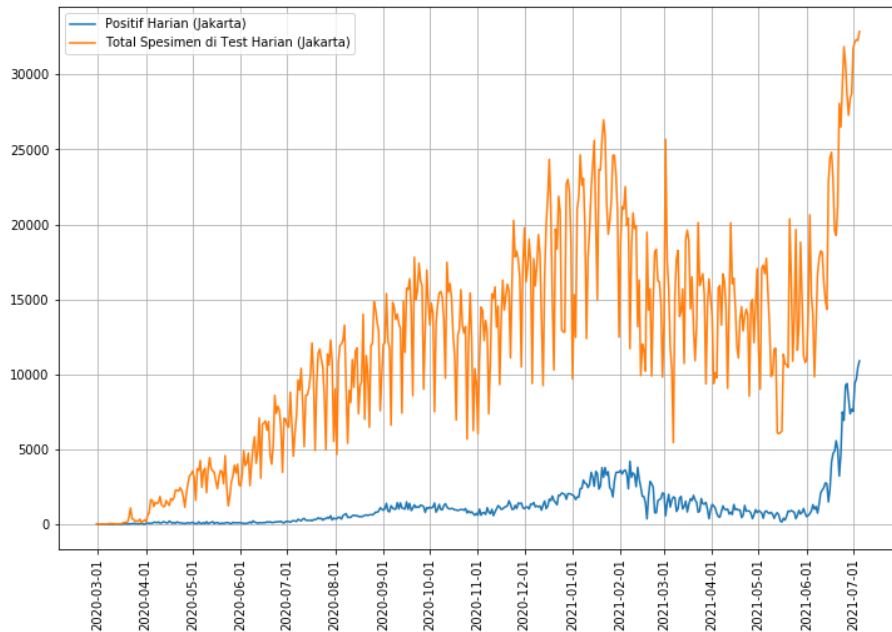
- Ada korelasi yang sangat kuat antara kenaikan kasus positif COVID-19 harian terhadap variabel *self isolation* harian di Jakarta.
- Ada korelasi yang sangat kuat antara kenaikan kasus positif COVID-19 harian terhadap variabel jumlah tes spesimen harian di Jakarta.
- Ada korelasi yang kuat antara variabel vaksinasi harian dengan positif H+7 harian di Jakarta pada selang waktu 1 Juni hingga 16 Juli 2021.

Akan tetapi, faktor-faktor penyebab kenaikan kasus positif COVID-19 harian lainnya masih belum ditemukan.

e. Deep Dive Analysis

Berdasarkan temuan sementara pada bagian Initial Findings, kami kemudian melakukan analisis yang lebih dalam mengenai hubungan antar variabel tersebut.

- **Bagaimana hubungan antara jumlah spesimen yang di tes terhadap kenaikan jumlah kasus positif COVID-19 harian di Jakarta?**



Tahun	Bulan	Total Tes	Total Positif
2020	Maret	5.047	741
	April	51.198	3.397
	Mei	99.311	3.134
	Juni	163.350	4.004
	Juli	270.032	9.925
	Agustus	321.372	19.108
	September	410.016	34.059
	Oktober	394.446	31.129
	November	422.913	31.264
	Desember	545.533	46.874
2021	Januari	641.944	85.983
	Februari	455.628	70.017

	Maret	442.728	42.320
	April	390.011	26.565
	Mei	355.049	21.439
	Juni	622.118	113.409
	Juli	191.983	48.030

Total tes memiliki peran yang sangat besar dalam pendeteksian COVID-19. Berdasarkan visualisasi grafik dan tabel di atas, dapat dilihat bahwa terdapat pola pergerakan yang sama antara variabel *Positive_Daily* dan *Total_Test_Daily*. Pola tersebut akan terlihat lebih jelas dalam skala bulanan pada tabel yang tertera. Kami mendapatkan bahwa total kasus positif COVID-19 secara bulanan sebenarnya berpola *uptrend* namun tidak diimbangi dengan jumlah tes bulanan di Jakarta.

Dengan cakupan tes dan *tracing* yang diperluas, kasus-kasus baru yang sebelumnya tidak tercatat menjadi tercatat. Hal ini terjadi pada rentang Mei - Juni 2021. Dapat terlihat pada tabel bahwa terdapat peningkatan jumlah tes diimbangi dengan kenaikan jumlah positif bulanan pula pada rentang waktu tersebut. Artinya, hipotesis sementara yang kami dapat pada bagian Initial Findings terbukti.

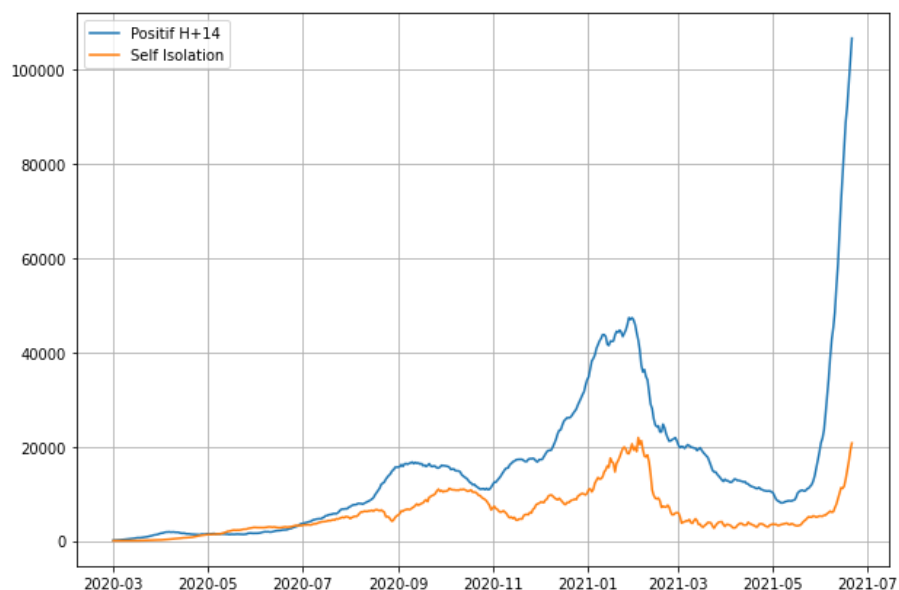
Pada rentang waktu Januari - Mei 2021, terlihat pola pergerakan yang *sideways* di antara kedua variabel tersebut. Kami kemudian menelusuri apa yang terjadi dan mengasumsikan bahwa terdapat faktor-faktor lain yang mempengaruhi pergerakan pola kedua variabel pada rentang waktu tersebut. Adapun faktor-faktor lain yang kami maksud yaitu:

1. Hari libur. Terdapat kenaikan yang signifikan pasca libur panjang seperti Nataru (Desember 2021) atau Idul Fitri (Mei 2021). Kami melihat bahwa terdapat akumulasi kasus positif COVID-19 pada masa liburan panjang. Pada masa libur tersebut, mobilitas masyarakat meningkat sehingga peluang penularan akan meningkat. Adapun operasional laboratorium

juga dipengaruhi oleh hari libur sehingga jumlah tes mengalami penurunan pada masa libur tersebut.

2. Mutasi virus. Mobilitas masyarakat yang meningkat seperti dijelaskan pada poin pertama memungkinkan terjadinya mutasi virus pada individu baru. Setiap individu memiliki sistem kekebalan tubuh yang berbeda sehingga virus bermutasi untuk beradaptasi terhadap lingkungan barunya. Dilaporkan bahwa varian baru yaitu varian Delta, dapat menular 10 kali lebih cepat.

- **Bagaimana hubungan timbal balik antara jumlah pasien yang menjalani *self isolation* dengan peningkatan kasus positif COVID-19 harian di Jakarta?**

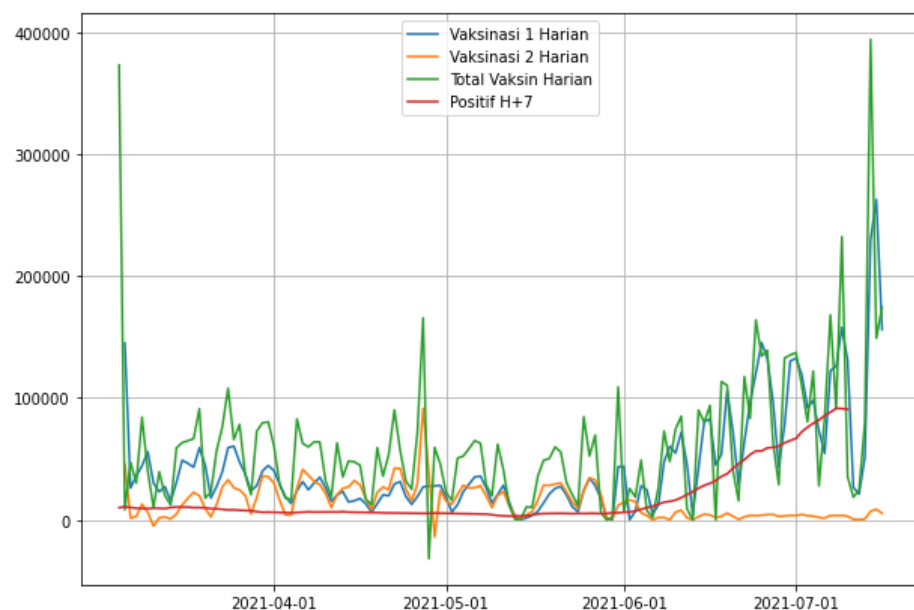


Berdasarkan analisa yang dilakukan, kami menggunakan dua buah variabel yaitu *Self_Isolation_Daily* dan *Positif H+14*. Hal tersebut dilakukan dengan pertimbangan standar prosedur durasi *self isolation* di Jakarta. *Self Isolation* sendiri merupakan salah satu metode penanganan COVID-19 di Jakarta. Ketika seseorang mendapatkan hasil tes dan dinyatakan positif COVID-19 namun dengan gejala yang tidak akut ataupun dengan pertimbangan medis lain, ia diizinkan untuk melakukan perawatan secara pribadi di rumah dengan prosedur *self*

isolation yang berlaku. Adapun beberapa poin penting yang kami dapatkan yaitu:

1. Ketika jumlah *self isolation* harian menurun, maka kasus Positif Harian H+14 akan cenderung naik dalam selang waktu tertentu.
2. Ketika *self isolation* mengalami kenaikan, akan sulit disimpulkan bahwa kasus positif harian H+14 mengalami penurunan, Hal tersebut disebabkan karena terdapat faktor lain yang mempengaruhi angka kenaikan kasus positif harian selain dari angka *self isolation* harian.
3. Berdasarkan grafik, metode *self isolation* yang diberlakukan di Jakarta terbukti mampu menekan laju kenaikan COVID-19 harian.

- **Bagaimana pengaruh vaksinasi massal terhadap kenaikan angka positif COVID-19 harian di Jakarta?**



Durasi yang diperlukan oleh tubuh manusia dalam meningkatkan imunitas setelah vaksinasi adalah 7-14 hari. Berdasarkan informasi tersebut, Kementerian Kesehatan Republik Indonesia menganjurkan untuk melakukan isolasi mandiri (*self isolation*) selama 7-14 hari setelah mendapat vaksinasi. Hal ini justru dapat memperbesar peluang terinfeksi COVID-19 jika tidak melakukan *self isolation* dalam kurun

waktu tersebut. Maka dari itu, kami ingin mencari tahu korelasi antara angka positif harian kasus COVID-19 di Jakarta selama 7 hari kedepan setelah mendapatkan vaksin (baik vaksin dosis pertama maupun vaksin dosis kedua) dengan penerimaan vaksin harian (Total Vaksin Harian).

Berdasarkan grafik tersebut, terdapat lonjakan nilai vaksin dosis 1 pada bulan Juni 2021 di Jakarta. Hal tersebut diperkirakan terjadi ketika mulai diberlakukannya program vaksinasi massal pada interval 1 Juni - 16 Juli 2021, Vaksinasi massal merupakan program vaksinasi terbuka untuk seluruh warga Jakarta yang berusia 18 tahun ke atas dan tidak terbatas oleh prioritas vaksinasi lagi. Hal tersebut menyebabkan lonjakan tinggi pada angka vaksin dosis 1 dan penurunan angka vaksin dosis 2. Selain itu, terdapat peningkatan angka pada variabel Positif H+7. Adapun beberapa poin penting yang kami dapatkan yaitu:

1. Terdapat pola pergerakan yang sama pada variabel Vaksinasi 1 Harian dengan Total Vaksin Harian. Hal tersebut membuktikan bahwa dalam periode vaksinasi massal tersebut (1 Juni - 16 Juli 2021), terdapat kesesuaian pada target yang dituju, yaitu warga di atas 18 tahun yang belum mendapatkan vaksinasi sebelumnya.
2. Hubungan yang kuat juga terjadi antara Total Vaksin Harian serta Vaksinasi 1 Harian dengan Positif H+7. Jumlah kasus positif H+7 akan meningkat apabila prosedur pasca vaksinasi tidak dilakukan dengan standar yang sudah ditetapkan.
3. Terdapat juga faktor eksternal lain seperti kondisi pembatasan sosial yang sedang berlaku ketika pengambilan data, tingkat efektivitas vaksin yang disediakan oleh pemerintah daerah Jakarta, serta jumlah warga yang sudah melalui masa pemulihan pasca vaksinasi.

f. Conclusion and Recommendation

- Berdasarkan analisis tambahan yang telah kami buat, terdapat beberapa kesimpulan yang dapat diambil yaitu:
 1. Tingkat mobilitas masyarakat berperan besar dalam peningkatan jumlah kasus positif COVID-19 harian di Jakarta.
 2. Mutasi virus dapat terjadi jika rasio penularan meningkat.
 3. Hari libur nasional seperti Nataru dan Idul Fitri dapat menyebabkan peningkatan akumulasi kasus positif COVID-19 di Jakarta.
 4. Pemberlakuan vaksinasi massal dapat mengakibatkan penurunan imunitas secara massal dalam kurun waktu tertentu.
 5. Self Isolation terbukti mampu menekan penularan virus SARS-CoV-19.
 6. Kesadaran masyarakat akan pentingnya mematuhi protokol kesehatan masih kurang.
- Berdasarkan analisis tambahan yang telah kami buat, terdapat beberapa saran yang dapat kami usulkan yaitu:
 1. Mengubah protokol 6M menjadi 7M yaitu Mencuci tangan dengan sabun, Menggunakan masker dengan benar, Menjaga jarak, Menghindari keramaian, Menghindari makan bersama, Mengurangi mobilitas, dan Menjaga imunitas tubuh.
 2. Menerapkan aturan PPKM di Jakarta hingga grafik COVID-19 harian menjadi landai.
 3. Jumlah tes harian dan *tracing* di Jakarta ditingkatkan sehingga pelacakan positif COVID-19 menjadi lebih maksimal.

6. Terkait pertanyaan sebelumnya, jelaskan dengan lebih detail usaha exploratory data analysis (EDA) yang kalian lakukan dan mengapa kalian melakukan teknik EDA tersebut.

Usaha EDA (Exploratory Data Analysis) yang kami lakukan bertujuan untuk mencari korelasi kuantitatif dari beberapa variabel ujicoba yang telah kami tentukan. Adapun beberapa teknik yang kami gunakan adalah *data cleansing*, koefisien korelasi spearman, analisis bivariat, analisis multivariat.

- Data Cleansing bertujuan untuk memvalidasi data-data pada variabel yang ingin kami ujicoba. Pada tahapan ini, kami mencari data yang memiliki kesalahan perhitungan numerik dan menggantinya dengan nilai yang seharusnya dihasilkan.
- Terdapat beberapa cara penentuan koefisien korelasi seperti pearson, spearman, kendall tau, somers, dan gamma. Pada laporan ini, penentuan korelasi dilakukan dengan metode spearman karena data-data COVID-19 yang didapatkan bersifat non parametrik.
- Analisis bivariat dilakukan untuk mencari hubungan antara pengaruh angka vaksinasi COVID-19 terhadap kenaikan angka positif harian, ditambah dengan pergeseran interval yang bertujuan untuk melihat korelasi pada interval-interval tertentu.
- Analisis multivariat dilakukan untuk mencari hubungan antara 3 buah variabel yaitu positif harian, *self isolation* harian, dan jumlah tes harian.

Alasan utama kami menggunakan 2 tipe analisis (bivariat dan multivariat) adalah adanya perbedaan *time frame* pada variabel Total Vaksin Harian.

KESIMPULAN

Berdasarkan analisis yang kami lakukan ditambah dengan pencarian dari berbagai sumber terkait permasalahan COVID-19 di Jakarta, kami menarik beberapa kesimpulan yaitu:

- Kenaikan jumlah kasus positif COVID-19 harian di Jakarta dalam sebulan belakangan ini disebabkan oleh jumlah tes yang meningkat, *tracing* yang lebih agresif, program vaksinasi massal, dan kepatuhan pasien COVID-19 dalam menjalani *self isolation*.
- Berdasarkan *dataset* yang kami gunakan, rata-rata kasus positif COVID-19 di Jakarta adalah 1.202 per hari.
- Berdasarkan *dataset* yang kami gunakan, jumlah kasus positif COVID-19 harian tertinggi di Jakarta adalah 10.903 kasus.
- Akumulasi kasus positif COVID-19 di Jakarta akan mencapai puncaknya pada hari libur seperti Nataru dan Idul Fitri.
- Jumlah tes harian perlu ditingkatkan agar memperluas cakupan deteksi kasus COVID-19 di Jakarta.
- Pelaksanaan prosedur pasca vaksinasi massal perlu diperketat karena berpotensi dapat meningkatkan jumlah kasus positif harian selama selang waktu 7 hari kedepan.
- Protokol 7M dapat dilakukan sebagai upaya preventif dalam menghambat penularan virus SARS-CoV-19.
- Kesadaran masyarakat akan pentingnya mematuhi protokol kesehatan perlu ditingkatkan agar setidaknya peningkatan kasus positif harian COVID-19 tidak melonjak secara drastis dalam waktu yang singkat.
- Kenaikan jumlah kasus positif COVID-19 tidak diimbangi dengan kenaikan jumlah tenaga kesehatan di Jakarta.
- Kenaikan jumlah kasus positif COVID-19 secara signifikan dalam waktu yang singkat menyebabkan fasilitas kesehatan tidak mampu melayani pengobatan pasien COVID-19 secara maksimal.
- Kelangkaan kamar tidur rumah sakit di Jakarta akan terus meningkat seiring dengan peningkatan kasus positif harian.
- Mutasi virus akan terus terjadi jika mobilitas terus meningkat disertai dengan penurunan sistem kekebalan tubuh masyarakat.
- Varian Delta adalah varian virus baru yang diklaim dapat menularkan *strain* 10 kali lebih cepat.

DAFTAR PUSTAKA

- Arbi, Ivany A. (2021). *Covid-19 Menggila, RS Rujukan Covid-19 di Jabodetabek Lumpuh Halaman all - Kompas.com*. KOMPAS.com. Retrieved 13 July 2021, from <https://megapolitan.kompas.com/read/2021/06/29/05000081/covid-19-menggila-rs-rujukan-covid-19-di-jabodetabek-lumpuh?page=all>
- Arora, S. (2021). *Supervised vs Unsupervised vs Reinforcement - AITUDE*. AITUDE. Retrieved 13 July 2021, from <https://www.aitude.com/supervised-vs-unsupervised-vs-reinforcement/>
- Australian Government Department of Health. (2021). *How long does it take to have immunity after vaccination?*. Health.gov.au. Retrieved 15 July 2021, from <https://www.health.gov.au/initiatives-and-programs/COVID-19-vaccines/is-it-true/is-it-true-how-long-does-it-take-to-have-immunity-after-vaccination>
- Bohang, Fatimah K. (2021). *2025, Ada 163 Triliun GB Data di Seluruh Dunia*. KOMPAS.com. Retrieved 11 July 2021, from <https://tekno.kompas.com/read/2017/04/07/07421057/2025.ada.163.triliun.gb.data.di.seluruh.dunia>
- Bruce, P. (2020). *PRACTICAL STATISTICS FOR DATA SCIENTISTS*. O'REILLY MEDIA, INCORPORA.
- Cucis.ece.northwestern.edu. (2021). Retrieved 13 July 2021, from <http://cucis.ece.northwestern.edu/projects/DMS/publications/AnomalyDetection.pdf>
- Frost, J. (2021). *5 Ways to Find Outliers in Your Data - Statistics By Jim*. Statistics By Jim. Retrieved 18 July 2021, from <https://statisticsbyjim.com/basics/outliers/>
- Fundarika, Bimo A. & Efendi, Dini A. (2021). *Protokol Kesehatan Tambah Jadi 6M, Apa Saja?*. Suara.com. Retrieved 14 July 2021, from <https://www.suara.com/health/2021/06/23/195000/protokol-kesehatan-tambah-jadi-6m-apa-saja>
- Jonathan Sarwono, *Statistik Itu Mudah: Panduan Lengkap untuk Belajar Komputasi Statistik Menggunakan SPSS 16* (Yogyakarta: Penerbit Universitas Atma Jaya Yogyakarta, 2009).
- Kleinbaum, D. (2014). *Applied regression analysis and other multivariable methods*. Cengage Learning.

- Mashabi, Sania. (2021). *Dinkes DKI: Positivity Rate COVID-19 di DKI Lampau Batas WHO*. KOMPAS.com. Retrieved 13 July 2021, from <https://nasional.kompas.com/read/2020/08/05/13484941/dinkes-dki-positivity-rate-COVID-19-di-dki-lampau-batas-who>
- Ridwan, Muhammad. (2021). *Update Covid Jakarta: Ketersediaan Tempat Tidur Sangat Mepet | Jakarta*. Bisnis.com. Retrieved 16 July 2021, from <https://jakarta.bisnis.com/read/20210703/77/1413299/update-covid-jakarta-ketersediaan-tempat-tidur-sangat-mepet>
- Roy, R. (2021). *Understanding the difference between AI, ML and DL!!*. Medium. Retrieved 19 July 2021, from <https://towardsdatascience.com/understanding-the-difference-between-ai-ml-and-dl-ccb63252a6c>
- Song, X., Wu, M., Jermaine, C., & Ranka, S. (2007). Conditional Anomaly Detection. *IEEE Transactions On Knowledge And Data Engineering*, 19(5), 631-645. <https://doi.org/10.1109/tkde.2007.1009>
- US National Library of Medicine. (2012). *A guide to appropriate use of Correlation coefficient in medical research*. Ncbi.nlm.nih.gov. Retrieved 16 July 2021, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/>
- Wiryo, Singgih. (2021). *Jakarta Kekurangan 2.156 Tenaga Kesehatan untuk Tangani Covid-19*. KOMPAS.com. Retrieved 19 July 2021, from <https://megapolitan.kompas.com/read/2021/06/29/21360751/jakarta-kekurangan-2156-tenaga-kesehatan-untuk-tangani-covid-19>
- Ye, K., Myers, S., Myers, R., & Walpole, R. (2012). *Probability & statistics for engineers & scientists*. Pearson Education.
- Yunianto, Tri K. (2021). *Pemprov DKI Jakarta Klaim Laksanakan Tes PCR Melampaui Target WHO - Nasional*. Katadata.co.id. Retrieved 13 July 2021, from <https://katadata.co.id/febrinaiskana/berita/5ee98e29e5723/pemprov-dki-jakarta-klaim-laksanakan-tes-pcr-melampaui-target-who>

LAMPIRAN

- Dataset sebelum dimodifikasi (covid19.xlsx):
<https://drive.google.com/file/d/1xpVpNxGbtKz1iGPxebZho8gnpnbIr1SA/view?usp=sharing>
- Dataset setelah dimodifikasi (COVID19JKT.xlsx):
https://drive.google.com/file/d/1b5AQxa5tFV_wMzBg_kLAtKk2aDNBARbp/view?usp=sharing