**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 *outlier*s 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 tekhnik 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 inteligence* 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 pengatahuan 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. Setalah 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**

Shape

Description automatically generated with low confidence

**Disusun Oleh:**

Kelompok Moxa

**Anggota:**

Reo Sahobby

Naufal Afkar A

Rozian F

**CompFest Academy and Compettition**

**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. Coivd-19 sendiri adalah penyakit baru yang berasala 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 dikota 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 lainya, 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 berjalanya 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 dalah 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 menunjukan 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 perseberan kasus Coivd-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.

![Graphical user interface, text, application, email

Description automatically generated]()

**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.

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

![Graphical user interface, text, application

Description automatically generated]()

**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*).

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

Graphical user interface, application

Description automatically generated

**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.

Chart, histogram

Description automatically generated

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

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

(1.1)

(1.2)

(1.3)

Keterangan:

kuartil 1 (25% data)

kuartil 3 (75% data)

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

![Text

Description automatically generated]()

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

Hasil dari perhitungan yang ditampilkan pada Gambar 1.5 di atas adalalah 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 dibersikan *outlier* mempunyai rentang data dari 0 sampai 17524. Grafik distibusi kolom self\_isolation setelah dihapus *outlier* dapat dilihat pada Gambar 1.6 di bawah ini.

Chart, histogram

Description automatically generated

**Gambar 1. 6** DistPlot self\_isolation setelah dihapus outlier

1. 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.

Chart, treemap chart

Description automatically generated

**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.

Chart, scatter chart

Description automatically generated

**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 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**

1. Problem Statement

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

* Kasus positif aktif yang meningkat lagi dari pada tahun 2021 terutama pada bulan Juni dan Juli.

1. Hypothesis

Ini belum...

1. Exploratory Data Analisis

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.

Chart, line chart

Description automatically generated

**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. Sedangkat tingkat kesembuhan dapat dilihat pada Gambar 1.10 di bawah ini.

Chart, line chart

Description automatically generated

**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 kesembuh 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.

![Graphical user interface, application, table

Description automatically generated]()

**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.

Chart, scatter chart

Description automatically generated

**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.

Chart, scatter chart

Description automatically generated

**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.

1. Initial Findings

Ini belum...

1. Deep Dive Analysis
2. Conclusion and Reccomendation

**Kesimpulan**

**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 Gelaja 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-inteligence-machine-learning-deep-learning/