

## Bagian 2 :Data Science Fundamentals (non-dataset)

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

- Outlier adalah suatu nilai data yang memiliki jarak abnormal dari nilai-nilai lain dalam sampel acak dari suatu populasi data tersebut.
- Kebanyakan statistik parametrik seperti mean, standar deviasi, korelasi dan statistik prosedur seperti linear regresi ANOVA sangat sensitif terhadap outliers, jika tidak ditangani maka outliers akan mengacaukan analisis kita.
- Manajemen terhadap outliers perlu dilakukan. Dalam melakukan manajemen tersebut kita tidak boleh serta merta menghapus outliers dari data kita, karena bisa saja outliers tersebut mengandung informasi yang berguna proses analisis data. Ada beberapa cara manajemen outliers:

### 1. Mengubah Nilai

Ketika outliers disebabkan karena adanya kesalahan pada data. Kita dapat menggantinya dengan suatu nilai. Pengubahan nilai biasanya menggunakan nilai mean, median atau menggunakan model regresi untuk memprediksi nilai yang hilang

### 2. Transformasi Data

Transformasi data berguna ketika kita berhadapan dengan kumpulan data yang sangat miring akibat outliers. Dengan mentransformasikan variabel, kita dapat menghilangkan outlier misalnya mengambil natural log dari suatu nilai mengurangi variasi yang disebabkan oleh nilai ekstrim.

### 3. Menghapus Nilai

Anda dapat menghapus outlier jika anda tahu bahwa outlier salah atau jika alasan outlier dibuat tidak akan pernah terjadi di masa mendatang. Misalnya, ada kumpulan data usia orang dan usia biasa terletak antara 0 hingga 90 tetapi ada entri data dari usia 150 yang hampir tidak mungkin. Jadi, kita bisa dengan aman menghapus nilainya yaitu 150.

### 4. Cap the outliers

Cara lain untuk memanajemen outliers dengan membatasi nilai dataset. Dengan analogi orang-orang di atas tingkat pendapatan tertentu berperilaku dengan cara yang sama seperti mereka yang berpenghasilan lebih rendah. Dengan begitu kita bisa menggunakan pembatasan persentil. Nilai yang lebih kecil dari nilai pada persentil ke-1 diganti dengan nilai pada persentil ke-1, dan nilai yang lebih besar dari nilai pada persentil ke-99 diganti dengan nilai pada persentil ke-99.

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

- Korelasi secara umum merupakan suatu hubungan atau relasi, namun dalam prinsip statistik korelasi merupakan analisis yang digunakan untuk mencari hubungan atau relasi antar dua variabel yang bersifat kuantitatif. Hubungan antar dua variabel tersebut dapat terjadi karena adanya hubungan sebab akibat ataupun dapat terjadi tanpa adanya sebab akibat. Dua variabel dikatakan berkorelasi apabila perubahan pada variabel yang satu akan diikuti perubahan pada variabel yang lain secara teratur dengan arah yang sama (korelasi positif) atau berlawanan (korelasi negatif).

Kedua variabel yang dibandingkan satu sama lain dalam korelasi dapat dibedakan menjadi variabel independen dan variabel dependen. Variabel independen merupakan sebuah variabel yang perubahannya cenderung di luar kendali manusia. Sementara itu variabel dependen adalah variabel yang dapat berubah sebagai akibat dari perubahan variabel independen. Hubungan antara korelasi dua variabel yang dapat diilustrasikan seperti pertumbuhan tanaman dengan variabel sinar matahari dan tinggi tanaman. Sinar matahari merupakan variabel independen karena intensitas cahaya yang dihasilkan oleh matahari tidak dapat diatur oleh manusia. Sedangkan tinggi tanaman merupakan variabel dependen karena perubahan tinggi tanaman dipengaruhi langsung oleh intensitas cahaya matahari sebagai variabel independen.

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

- Machine learning merupakan salah satu bagian dari pembelajaran kecerdasan buatan atau biasa disebut AI, dimana metode dari machine learning membuat sebuah sistem yang dapat belajar dan mengolah informasi sendiri tanpa harus dilakukan program berulang kali. Salah satu penerapan dari metode machine learning yang biasa digunakan di dalam kehidupan sehari-hari yaitu penggunaan pendeteksi bahasa dan kalimat, filter spam chat dan email, pencarian kata atau data mining, pencarian berdasarkan pengelompokan atau clustering.

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

- Artificial Intelligence merupakan suatu bidang ilmu komputer yang menciptakan suatu sistem atau mesin yang dapat belajar dan beradaptasi layaknya seorang manusia. Dimana Artificial Intelligence memungkinkan bagi

mesin untuk belajar dari pengalaman dimana menyesuaikan dengan input baru yang dilakukan manusia .sedangkan Machine Learning merupakan cabang pembelajaran dari Artificial Intelligence dimana Machine Learning membuat sistem dapat belajar dari data, mengidentifikasi pola, dan membuat keputusan dengan intervensi manusia yang manual. Dengan Machine Learning, komputer dapat menangani situasi baru melalui pelatihan mandiri, pengalaman, analisis, dan observasi.sedangkan deep learning sendiri merupakan sebuah teknik untuk mengolah machine learning menggunakan prinsip dari otak manusia yaitu prinsip jaringan saraf dari neuron atau biasa disebut Neural Network.

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 merupakan sebuah bentuk dari kegiatan untuk melakukan penggabungan terhadap hasil analisis dengan berbagai macam pertanyaan, kriteria, maupun pada sebuah standar tertentu untuk dapat menciptakan sebuah makna dari adanya sebuah data yang dimana telah dikumpulkan oleh seseorang guna untuk mencari sebuah jawaban terhadap permasalahan yang dimana terdapat didalam sebuah penelitian yang dimana sedang diperbaiki. Tantangan dari interpretasi data yang biasa sering terjadi yaitu data scientist harus memahami data yang akan diolah dan mengambil keputusan untuk melakukan metode pengolahan agar sistem dapat melakukan suatu prediksi atau sistematika machine learning. Dimana setiap pengambilan keputusan yang dilakukan akan berpengaruh pada pengolahan sistem yang dimana hasil yang didapatkan harus bisa dipahami dan dijelaskan secara jelas terhadap audiens yang dituju.

## **Bagian 1 :Data Science Fundamentals (dataset)**

### **b. Latar Belakang**

Laporan ini merupakan pemenuhan tugas seleksi tahap pertama COMPFEST untuk program bootcamp Data Science Academy yang diselenggarakan oleh mahasiswa Universitas Indonesia. Untuk memenuhi tugas ini, kami menggunakan bahasa pemrograman Python, dimana untuk menyelesaikan pengolahan data yang akan dibuat, data yang digunakan merupakan data covid yang telah diberikan oleh panitia sebanyak 10 data sheet tentang kasus penyebaran corona di Jakarta dan Indonesia dengan nama file dataset “Daily Update Data Agregat Covid-19 Jakarta.csv” dengan beberapa variabel prediktor corona meliputi jumlah

positif, kematian, tanggal, meninggal, sembuh, vaksinasi total maupun vaksinasi harian dan lain lain. Variabel yang nantinya diberikan nantinya akan kami olah untuk memprediksi tanggal dimana herd immunity Indonesia mencapai 60% dan 80% serta mengetahui apakah vaksin dapat mempengaruhi tingkat penyebaran virus corona di Indonesia.

Dengan adanya vaksin COVID-19 yang mulai diluncurkan, banyak yang bertanya-tanya "Kapan semuanya akan kembali normal"? Meskipun tidak ada satu definisi khusus tentang apa yang sebenarnya normal, banyak yang setuju bahwa begitu kekebalan kelompok tercapai, itu akan cukup aman untuk kembali ke kehidupan semi normal. Herd immunity adalah bentuk perlindungan tidak langsung dari penyakit menular yang dapat terjadi ketika presentasi yang cukup dari suatu populasi telah menjadi kebal. Distribusi vaksin meningkat dengan cepat. Peningkatan ini dapat dimodelkan menggunakan model regresi, yang memungkinkan kita untuk memprediksi kapan kekebalan kelompok akan tercapai di Indonesia.

### c. Jawaban Soal

#### Bagian 1 : Analisis Dataset

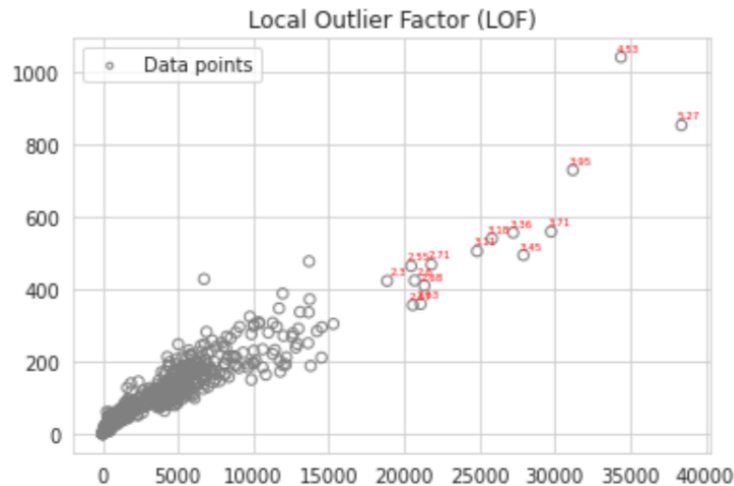
1. Dari dataset diambil sheet data Jakarta dan didapati nilai mean, median, dan modus dari positif COVID-19 harian Jakarta sebagai berikut:

```
mean positif harian    : 1259.1454545454546
median positif harian  : 867.0
modus positif harian   : 0      0
```

2. Dari dataset yang digunakan nilai minimal dan maksimal dari positif COVID-19 harian Jakarta didapati hasil sebagai berikut:

```
min positif harian    : 0
max positif harian    : 12974
```

3. Dari dataset yang digunakan dengan mengambil variabel positif harian dan meninggal harian dan didapati nilai-nilai outlier yang ada yaitu di kisaran value 15000 keatas



positif outliers : [20574, 18872, 21095, 21342, 20694, 20467, 21807, 24836, 25830, 27913, 27233, 29745, 31189, 34379, 38391]  
 meninggal outliers : [355, 422, 358, 409, 423, 463, 467, 504, 539, 493, 555, 558, 728, 1040, 852]

Untuk mendeteksi outliers kami menggunakan metode Local Outliers Factor(LOF), dimana LOF menghitung deviasi kerapatan lokal dari titik data tertentu terhadap tetangganya. Sebuah data dianggap sebagai outlier jika memiliki kerapatan yang jauh lebih rendah daripada tetangganya.

4. Dari dataset yang disediakan, kami melakukan pengolahan korelasi antara variabel meninggal harian (Jakarta) dengan variabel positif harian (Jakarta) dimana jika jumlah pertambahan kasus positif harian di jakarta bertambah maka kemungkinan kasih meninggal akan terus bertambah sehingga menimbulkan suatu korelasi dan dimana setelah melakukan pengolahan didapat hasil nilai korelasi sebagai berikut:

**Spearman's correlation coefficient: 0.778**

Kami menggunakan Spearman's Rank Correlation Coefficient dimana didapat angka 0.778 artinya terdapat korelasi yang tinggi antara variabel 'meninggal harian (jakarta)' dan 'positif harian (jakarta)'.

#### d. Hasil Analisis Tambahan

Hasil Analisis lanjutan yang kami lakukan terhadap data covid di jakarta yaitu kami melakukan prediksi dimana kapan indonesia akan mencapai tingkat Herd immunity. Pertama kami melakukan pengolahan data dan memilih atribut yang digunakan dan menjadikan satu file dalam file csv yang nantinya akan kami olah menggunakan model prediksi yang kami gunakan.

Model prediksi yang kami gunakan yaitu regresi menggunakan model eksponensial atau polinomial agar sesuai dengan model matematika untuk mengamati data tentang infeksi dan vaksinasi.

- Persentase populasi yang divaksinasi lengkap seiring berjalannya waktu untuk memprediksi kapan Herd immunity akan tercapai
- Kasus harian COVID-19 seiring berjalannya waktu (untuk memeriksa dinamika penyebaran penyakit)
- Kasus COVID-19 harian karena persentase populasi yang divaksinasi lengkap meningkat (untuk menentukan bagaimana kekebalan oleh vaksin mempengaruhi penyebaran penyakit)

Model-model ini mengasumsikan bahwa tingkat kasus dan vaksinasi COVID-19 saat ini/yang diamati sebelumnya akan berlanjut di masa depan.

#### a. Problem Statement

Kondisi penyebaran Covid-19 belakangan ini kian tak terkendalikan. Hal ini ditengarai oleh varian delta yang diklaim lebih menular. Hal ini memperpanjang derita panjang yang dialami seluruh sektor nasional. Salah satu solusi yang paling efektif dalam menyelesaikan kondisi pandemi Covid-19 ini adalah mencapai Herd Immunity. Kami akan melakukan prediksi kapan Indonesia akan mencapai herd immunity berdasarkan model regresi polinomial. Kami juga akan menganalisis berapa minimal persentase populasi yang harus divaksin agar kasus harian covid bisa menurun dan mencapai angka nol.

#### b. Hypothesis

- Kasus harian COVID-19 seiring berjalannya waktu terus meningkat sehingga menimbulkan jumlah korban yang terus bertambah setiap hari.
- Kasus Covid-19 di Indonesia akan turun secara signifikan seiring meningkatnya vaksinasi terhadap masyarakat Indonesia.

#### c. Exploratory Data Analysis

Penyebaran corona di Jakarta serta Indonesia dengan nama file dataset "Daily Update Data covid di Indonesia dengan beberapa variabel prediktor corona meliputi jumlah positif yang dialami pasien, kematian, tanggal, meninggal, sembuh, vaksinasi total maupun vaksinasi harian dan lain-lain. Variabel yang nantinya diberikan nantinya akan kami olah untuk memprediksi tanggal dimana herd immunity Indonesia mencapai 60% dan 80% serta mengetahui apakah vaksin dapat mempengaruhi tingkat pertumbuhan virus corona di Indonesia.

1. Memasukkan data yang ingin diolah, disini kami menamai data kami dengan nama “compfest56.csv” yang berisikan atribut-atribut terpilih untuk mengolah data nantinya.
2. Mengecek nama kolom, jumlah baris, dan tipe data

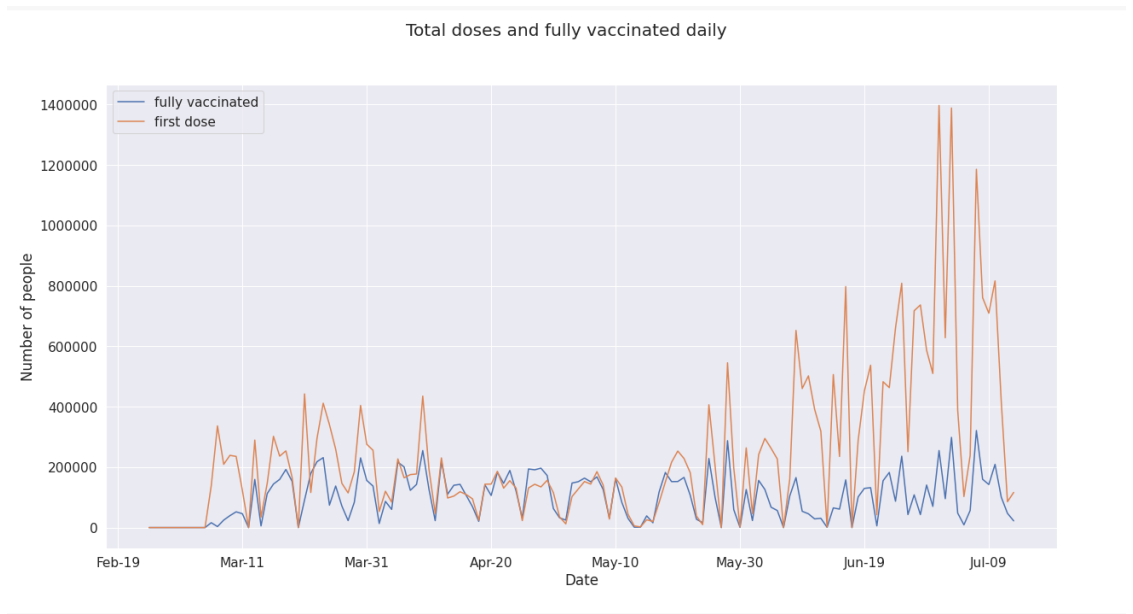
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 27 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Tanggal                                   500 non-null    object
1   Meninggal (Indonesia)                   500 non-null    int64
2   Sembuh (Indonesia)                     500 non-null    int64
3   Dirawat (Indonesia)                    500 non-null    int64
4   Positif (Indonesia)                     500 non-null    int64
5   Meninggal (Jakarta)                    500 non-null    int64
6   Sembuh (Jakarta)                       500 non-null    int64
7   Self-Isolation (Jakarta)                500 non-null    int64
8   Dirawat (Jakarta)                      500 non-null    int64
9   Positif (Jakarta)                      500 non-null    int64
10  Positif Aktif (Jakarta)                  500 non-null    int64
11  Meninggal Harian (Indonesia)             500 non-null    int64
12  Sembuh Harian (Indonesia)                500 non-null    int64
13  Dirawat Harian (Indonesia)               500 non-null    int64
14  Positif Harian (Indonesia)                500 non-null    int64
15  vaksinasi 1 total                       500 non-null    int64
16  vaksinasi 2 total                       500 non-null    int64
17  vaksinasi 1 harian                      500 non-null    int64
18  vaksinasi 2 harian                      500 non-null    int64
19  persentase full vaksin                  500 non-null    float64
20  Meninggal Harian (Jakarta)               500 non-null    int64
21  Sembuh Harian (Jakarta)                  500 non-null    int64
22  Self-Isolation Harian (Jakarta)          500 non-null    int64
```

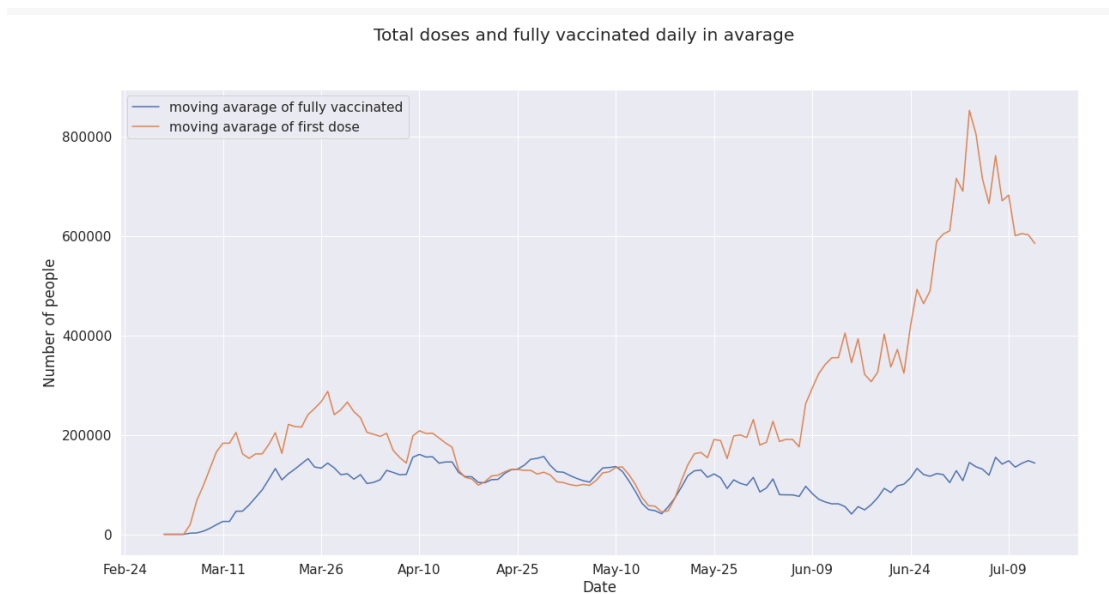
3. Melihat rangkuman statistik dari dataset

	Meninggal (Indonesia)	Sembuh (Indonesia)	Dirawat (Indonesia)	Positif (Indonesia)	Meninggal (Jakarta)	Sembuh (Jakarta)	Self- Isolation (Jakarta)	Dirawat (Jakarta)	Positif (Jakarta)
count	500.000000	5.000000e+02	500.000000	5.000000e+02	500.000000	500.000000	500.000000	500.000000	500.000000
mean	20617.126000	6.174448e+05	79307.998000	7.173699e+05	3143.072000	162773.85400	7968.400000	3856.596000	177741.922000
std	19051.514963	6.475310e+05	67226.340474	7.216135e+05	2728.124864	167405.55207	11109.473759	4448.687168	178593.530143
min	0.000000	0.000000e+00	0.000000	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3075.750000	2.805625e+04	30648.250000	6.178025e+04	649.500000	7310.000000	2998.250000	1865.500000	11985.250000
50%	14395.000000	3.589235e+05	63477.500000	4.276850e+05	2339.500000	99318.000000	5010.500000	2833.000000	109747.000000
75%	37961.250000	1.218476e+06	111038.000000	1.399864e+06	5928.000000	341625.000000	8900.750000	4383.500000	354553.500000
max	68219.000000	2.139601e+06	407709.000000	2.615529e+06	9541.000000	589486.000000	73239.000000	30418.000000	689243.000000

4. Melakukan visualisasi untuk melihat perbandingan antara jumlah harian orang yang divaksin dosis pertama dengan orang yang sudah divaksin secara lengkap “Dosis total dan vaksin lengkap setiap hari”,



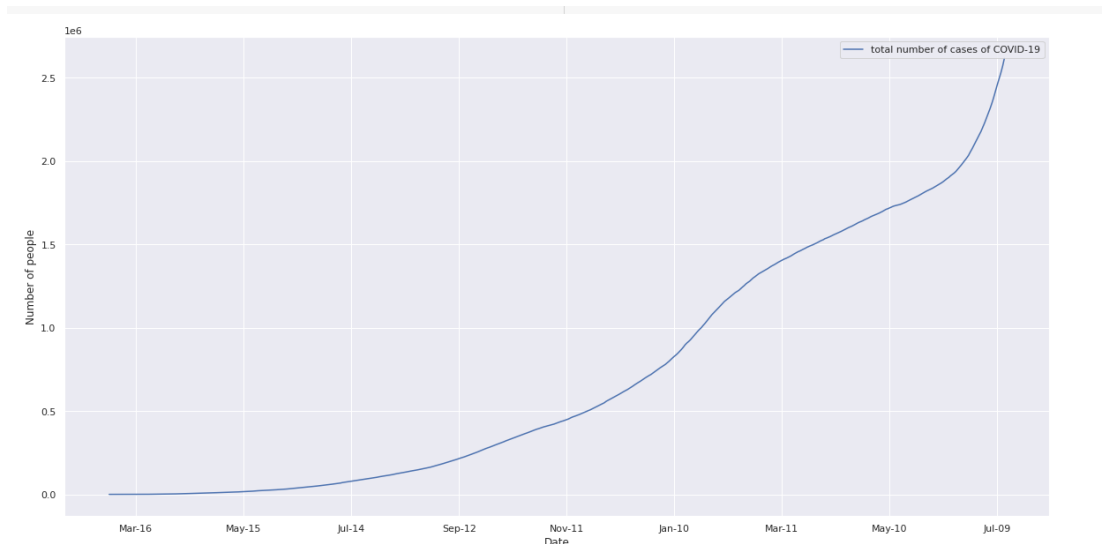
Namun dikarenakan adanya perubahan nilai yang sangat ekstrem, kami akan menggunakan metode moving avarage dengan ‘np.convolve’ pada Python. Berikut hasilnya



Dari grafik diatas terlihat pergerakan vaksinasi dosis pertama mengalami peningkatan yang cukup signifikan mulai pertengahan bulan Mei hingga Juli, sedangkan untuk vaksinasi dosis 2/fully vaccinated tidak mengalami peningkatan yang cukup signifikan dan cenderung stagnan di bawah angka 200.000 dosis.

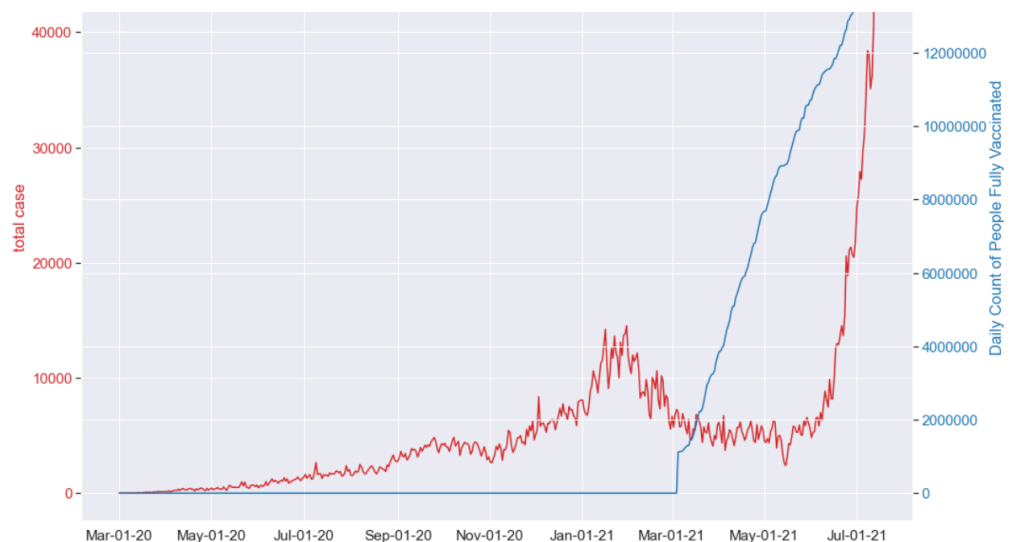
“Pertumbuhan kasus corona di Indonesia”.





Visualisasi diatas merupakan hasil visualisasi pertumbuhan kasus di Indonesia setiap harinya dimulai saat covid pertama kali terdeteksi di indonesia pada Maret 2020 hingga Juli 2021.

5. Berikut kami memvisualisasikan korelasi antara variabel kasus harian dan total orang yang sudah divakasi full.



Terlihat dari grafik tersebut vaksinasi tidak menyebabkan angka kasus harian corona menurun.

#### d. Initial Findings

Temuan yang kami dapatkan adalah pergerakan vaksinasi dosis pertama mengalami peningkatan yang cukup signifikan mulai pertengahan bulan Mei hingga Juli,

sedangkan untuk vaksinasi dosis 2/fully vaccinated tidak mengalami peningkatan yang cukup signifikan dan cenderung stagnan di bawah angka 200.000 dosis. Lalu kami menganalisis bahwa vaksinasi belum bisa menurunkan angka harian covid-19.

#### e. Deep Dive Analysis

**Sentence Learning:** Kelompok kami mengevaluasi data yang diberikan dan memproses data hingga menjadi sedemikian rupa. Setelah mendalami data yang telah diberikan mengenai kasus covid-19 yang ada di Indonesia, kami dapat menemukan beberapa variabel data dapat diolah menjadi sesuatu yang informasi. Kami mengerjakan suatu model untuk menghasilkan prediksi herd immunity di Indonesia serta dampak vaksin terhadap penurunan angka covid-19 di Indonesia.

**Action:** Membuat laporan yang dapat digunakan untuk melakukan analisis dan evaluasi untuk memberikan penjelasan terhadap riset kasus, memberikan akses ke semua data yang terlibat pada pengolahan kasus ini, menunjukkan prediksi dari suatu target yang akan dicapai pada setiap langkah pengujian dengan kemampuan untuk melakukan analisis mendalam. Dimana didapati Hasilnya, kami menemukan suatu prediksi tentang kasus covid-19 dengan menggunakan pemodelan regresi. Berikut langkah pemodelan yang telah kami buat:

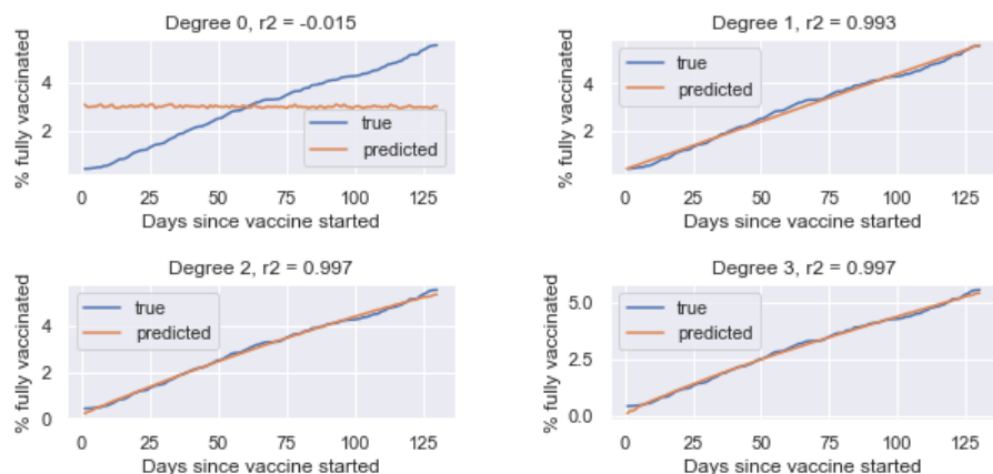
1. Melakukan proses pemodelan dengan menggunakan regresi eksponensial atau polinomial agar sesuai dengan model matematika untuk mengamati data tentang infeksi dan vaksinasi. Dimana kami menggunakan hubungan antara variabel sehingga mendapatkan prediksi yang kami inginkan, disini kami mencoba memodelkan 3 hubungan:
  - Persentase populasi yang divaksinasi lengkap seiring berjalannya waktu (untuk memprediksi kapan kekebalan kelompok akan tercapai).
  - Kasus harian COVID-19 seiring berjalannya waktu terus meningkat sehingga menimbulkan jumlah korban yang terus bertambah setiap hari.
  - Kasus COVID-19 harian karena persentase populasi yang divaksinasi lengkap meningkat (untuk menentukan bagaimana kekebalan oleh vaksin mempengaruhi penyebaran penyakit).
2. Kami memproyeksikan data ke model polinomial dengan derajat tertentu, lalu menginisialisasi k-fold dan regresi linier serta menginisialisasi inisialisasi array kosong untuk `y_pred`. Lalu melakukan pemisahan data dengan membagi menjadi train dan testing yang kemudian menyesuaikan model linier dengan data train serta memproses `predict y` nya. Selanjutnya menghitung `r2` yang divalidasi silang, proses tadi akan menghasilkan return

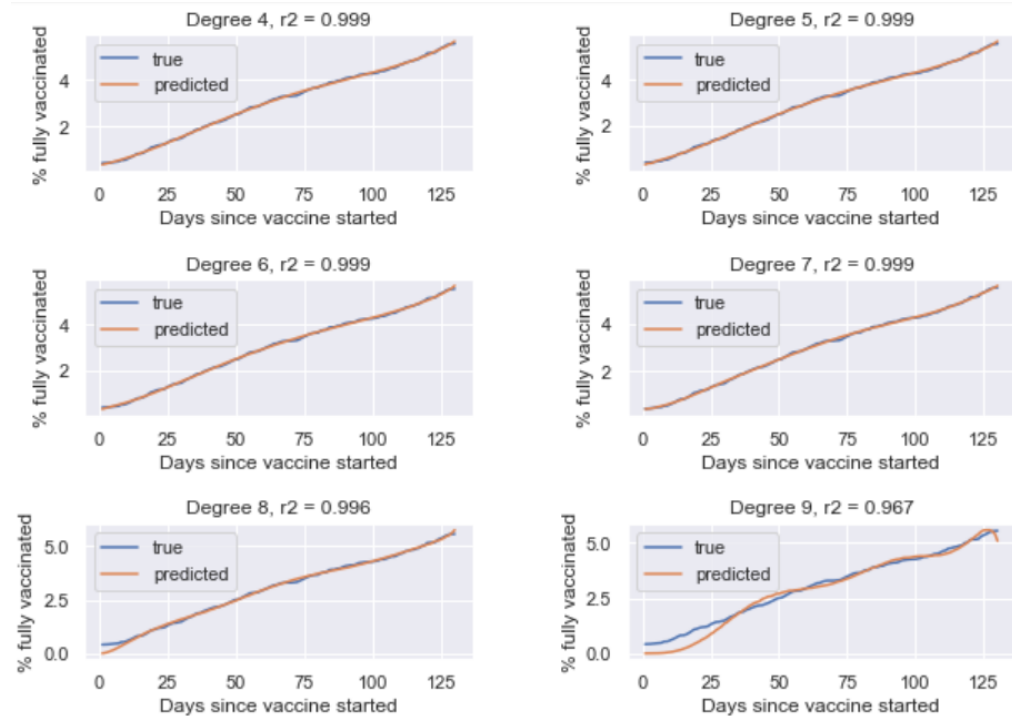
`y_pred`, `crossval_r2` yang dimasukkan ke dalam function bernama **`crossval_poly_reg`**.

3. Membuat function **`test_poly_degrees`** supaya dapat menjalankan model polinomial yang divalidasi silang untuk derajat yang berbeda dan mengembalikan hasil yang diplot. Dimana pertama diawal dengan mengatur jumlah plot secara vertikal dilanjutkan menginisialisasi subplot serta melacak nilai `r2`. Kemudian melanjutkan dengan uji model polinomial hingga derajat maksimal dimana ada proses looping harus dilalui yaitu langkah pertama yang harus dilanjutkan adalah mendapatkan `y_pred` dan `r2` yang divalidasi silang untuk `degree`, langkah kedua melacak `r2` per derajat lalu mendapatkan indeks subplot untuk memplot nilai sebenarnya dan prediksi untuk model ini lalu langkah terakhir memformat plotnya. Setelah uji model polinomial hingga derajat maksimal, dilanjutkan proses menambahkan label pada sumbu, akses dan plot `r2` per informasi derajat, dan terakhir memformat seluruh gambar. Dimana hasil dari function ini berupa `r2_dict_list`
6. Melakukan *slicing* data dimana kami mengambil data mulai dari baris 370 yaitu awal program vaksinasi dan disimpan pada sebuah variabel yang kami beri nama `post_vak_df`.
7. Melakukan plotting terhadap data menggunakan function model regresi polinomial yang sebelumnya sudah dibuat.

---

Polynomials for % vaccinated over time

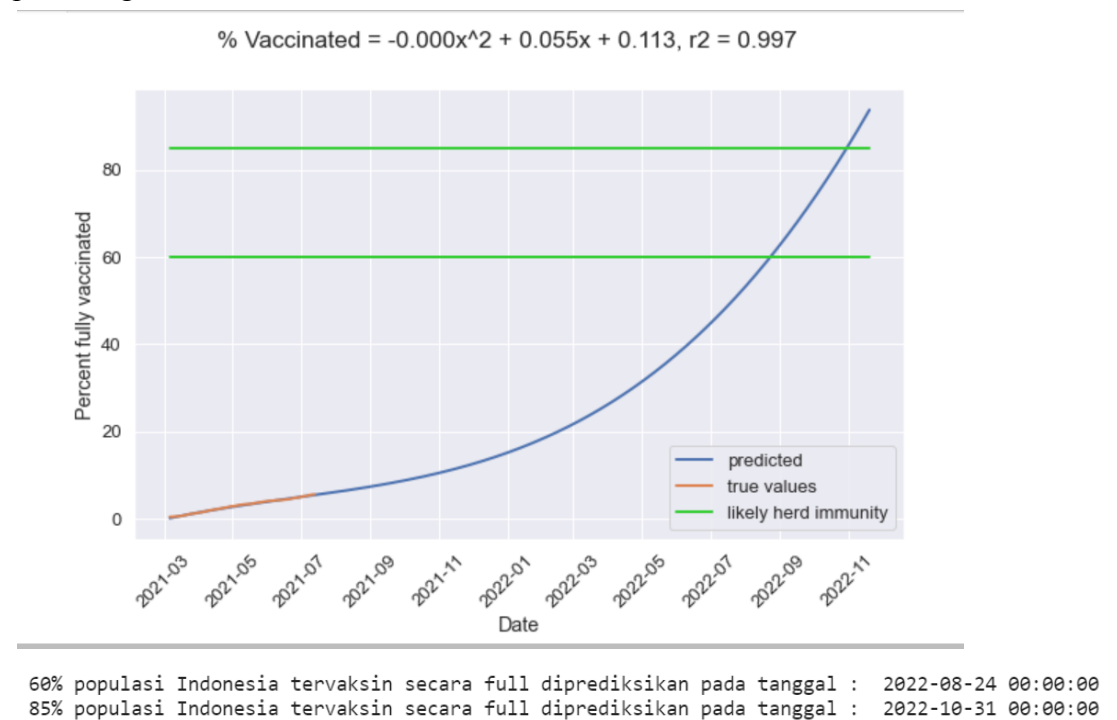




Model yang paling tepat untuk % vaksinasi penuh vs waktu adalah derajat ke-3. Model ini memiliki  $r^2$  per derajat yang tinggi untuk derajat yang rendah.

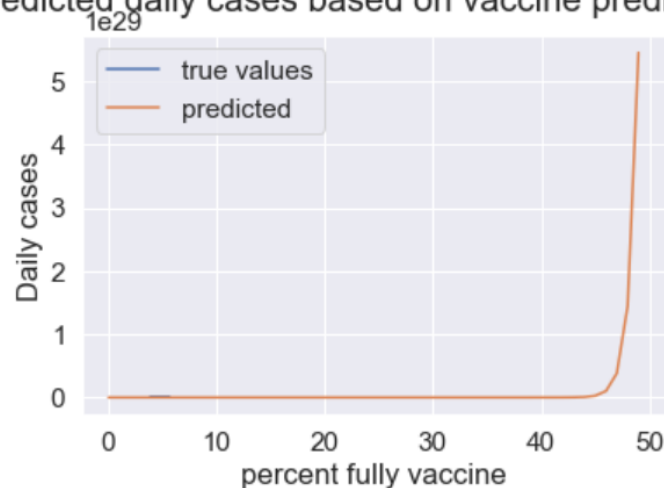
8. Melakukan prediksi polinomial berdasarkan hari dimana vaksinasi dilakukan dengan jumlah masyarakat yang telah melakukan vaksinasi penuh dengan dua dosis. Lalu melakukan visualisasi dari hasil yang didapat untuk bisa mendapatkan informasi yang mudah dipahami..
9. Sebelum proses menentukan herd immunity langkah yang harus dilalui adalah mengambil data lalu membuat variabel dengan x yang menandakan hari sejak dimulai/Days since start, lalu membuat variabel y yang menandakan presentasi vaksin dan x\_future yang menandakan mewakili hari 0-180 setelah variabel baru usai dibuat dilanjutkan dengan menyesuaikan data dengan polinomial derajat 2 kemudian melatih model regresi linier pada data yang ada dan menggunakannya untuk memprediksi tingkat vaksin di masa depan.
10. Menentukan persentase kekebalan kelompok atau Herd immunity yang diduga dapat melindungi diri dari covid -19, dimana kami menentukan persentase menjadi high and low predict Herd immunity. kami melakukan prediksi sesuai analisis data dimana vaksinasi dimulai yaitu pada tanggal 6 maret 2021. Setelah itu kami membuat rentang waktu prediksi dimana tercatat dalam data yang diolah kami memprediksi Herd immunity dari 200 hari setelah vaksin pertama dilakukan . setelah itu kami mencari nilai regresinya berdasarkan nilai prediksi dan value data training. Setelah itu kami menambahkan sebuah

visualisasi grafik garis yang menandai kisaran yang diharapkan dari kekebalan kelompok (60% - 85%) dari hasil regresi yang telah didapatkan pada perhitungan model.



11. Selanjutnya kami akan menganalisis bagaimana perubahan kasus harian covid-19 seiring dengan program vaksinasi.

#### Predicted daily cases based on vaccine predictions



Dari grafik diatas dapat disimpulkan kasus covid belum akan turun seiring pertambahan jumlah vaksinasi.

#### f. Conclusion and Recommendation.

##### Conclusion

Jadi setelah kami selesai melakukan pemodelan di dapatkan hasil dari prediksi kami 60% populasi Indonesia tervaksin secara full sehingga dapat dikatakan

mendapat Herd Immunity diprediksikan pada tanggal : 2022-08-24 dan 85% populasi Indonesia diprediksikan pada tanggal : 2022-10-31.

Setelah di lakukan prediksi hasil analisis, kami menyimpulkan bahwa dengan adanya vaksinasi dengan dosis full yang sudah dilakukan sampai bisa mencapai Herd Immunity belum tentu mengurangi pertumbuhan dari kasus positif kasus corona setiap hari, tampak dari grafik walaupun masyarakat Indonesia telah divaksin grafik pertumbuhan covid-19 tetap melonjak. Disini kami beranggapan dengan kondisi menyebarnya varian delta yang dibarengi dengan program vaksinasi yang mulai digencarkan oleh pemerintah membuat prediksi terhadap data sulit dilakukan, prediksi akan lebih baik jika data positif harian covid sudah menurun sehingga kita dapat melihat dampak vaksinasi terhadap angka positif harian Covid-19.

### Rekomendasi

Model yang lebih efektif untuk memperkirakan data ini adalah model SIR, yang lebih umum digunakan dalam pemodelan epidemiologi. Ini memperhitungkan individu dari tiga kelompok: rentan, kebal, dan pulih. Kelompok-kelompok ini kemudian digunakan untuk membuat prediksi tentang tren masa depan dalam penyebaran penyakit.

## **e. Kesimpulan**

Setelah kami melakukan Analisis dan proses preprocessing terhadap data set dimana kami mencari nilai min max suatu variabel, mencari nilai outlier, dan mencari korelasi antar variabel yang telah kami tentukan untuk melakukan analisa lebih lanjut. Kami melakukan Analisis tambahan menggunakan algoritma pemodelan regresi menggunakan bahasa pemrograman python dan mendapati hasil prediksi yang mampu menjawab hipotesis yang kami buat ,dimana kami mampu memprediksi kapan Indonesia dapat mencapai Herd immunity setelah melakukan vaksinasi dengan dosis full dan menjawab hipotesis dimana jumlah orang yang meninggal karena covid-19 akan terus meningkat seiring meningkatnya jumlah kasus yang ada di Indonesia sehingga dapat disimpulkan jumlah kematian berbanding lurus dengan jumlah kasus positif yang ada di Indonesia.

## **f. Daftar Pustaka**

[1] Sludden, J., Beaver, J., Park, E., Kumar, A., & Gao, J. (2014, February 4). *Introduction initial findings - ERIC*. <https://files.eric.ed.gov/fulltext/ED553142.Pdf>.

<https://files.eric.ed.gov/fulltext/ED553142.pdf>

[2] Ata, S. (2019, June 19). *How to perform deep dives into your business real life 5 examples*. Medium.Com.

<https://sinanata.medium.com/how-to-perform-deep-dives-into-your-business-5-real-life-examples-40fb51bcb11f>

[3] Kumar, P. (n.d.). *How to handle outliers*. Sigma Magic. Retrieved July 12, 2021, from

<https://www.sigmamagic.com/blogs/how-to-handle-outliers/>

[4] Mathematical Association of America (MAA). (2004, December). *The SIR Model for Spread of Disease - The Differential Equation Model*. Maa.Org.

<https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model>

**g. Lampiran (opsional).**

**Link kodingan soal dataset :**

[https://colab.research.google.com/drive/1PmPC6KSS3--vACYu8WeELW5oc9oT\\_58t?usp=sharing](https://colab.research.google.com/drive/1PmPC6KSS3--vACYu8WeELW5oc9oT_58t?usp=sharing)

**Link kodingan analisis tambahan:**

<https://colab.research.google.com/drive/1vC3zV679BArKaw3AtrcQjy80osvkIVHx?usp=sharing>