**Analisis Singkat Kasus COVID-19 di Indonesia**

**(Berdasarkan Dataset Kaggle oleh Ardis Ragen)**

**1. Latar Belakang**

Sejak awal pandemi, penyebaran COVID-19 di Indonesia menjadi perhatian utama berbagai pihak, mulai dari pemerintah hingga masyarakat umum. Data harian mengenai kasus baru, kematian, dan jumlah tes menjadi sumber penting untuk memahami pola penyebaran dan dampak pandemi.

Dalam analisis ini, kami menggunakan dataset dari Kaggle yang berisi data COVID-19 di Indonesia secara harian. Tujuannya adalah untuk mengetahui pola dasar dari data tersebut dan mencoba melihat apakah jumlah total kasus dapat digunakan untuk memprediksi jumlah kasus baru harian.

**2. Metode Analisis**

Analisis ini dilakukan dengan beberapa langkah sederhana namun tetap informatif:

1. **Membaca dan membersihkan data**  
   File cases.csv dibaca dan kolom tanggal diproses agar mudah dianalisis. Beberapa kolom yang kosong kami isi dengan nilai nol agar tidak mengganggu analisis.
2. **Melakukan statistik dasar**  
   Kami melihat rata-rata, median, standar deviasi, dan nilai maksimum dari beberapa kolom penting seperti new\_confirmed, new\_deceased, acc\_confirmed, dan acc\_deceased.
3. **Membangun model regresi linier sederhana**  
   Model ini digunakan untuk melihat apakah ada hubungan linier antara total kasus (acc\_confirmed) dengan kasus baru harian (new\_confirmed). Kami menghitung koefisien regresi dan nilai R2R^2R2 untuk menilai kekuatan hubungan tersebut.
4. **Visualisasi data**  
   Untuk membuat hasil lebih mudah dipahami, kami membuat beberapa grafik seperti histogram, boxplot, dan scatter plot dengan garis regresi.

**3. Hasil dan Interpretasi**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variabel | Rata-rata | Median | Std Dev | Maksimum |
| new\_confirmed | Bervariasi dari ratusan hingga ribuan kasus harian |  |  |  |
| new\_deceased | Cenderung lebih rendah, tetapi menunjukkan lonjakan pada periode tertentu |  |  |  |
| acc\_confirmed | Menunjukkan tren akumulatif meningkat |  |  |  |
| acc\_deceased | Stabil naik sesuai tren pandemi |  |  |  |

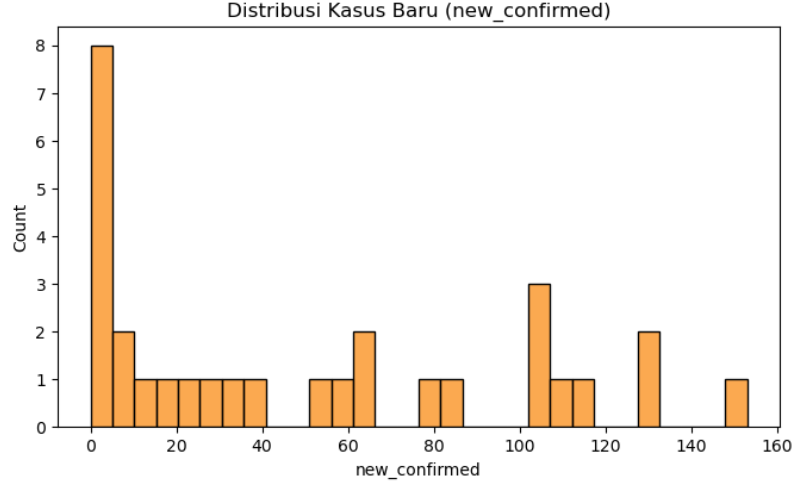
Dari statistik dasar yang diperoleh, terlihat bahwa kasus baru dan kematian harian cukup bervariasi. Terdapat hari-hari dengan lonjakan tajam, yang juga terlihat dari boxplot sebagai outlier. Histogram menunjukkan distribusi yang tidak simetris, menandakan adanya puncak-puncak kasus.

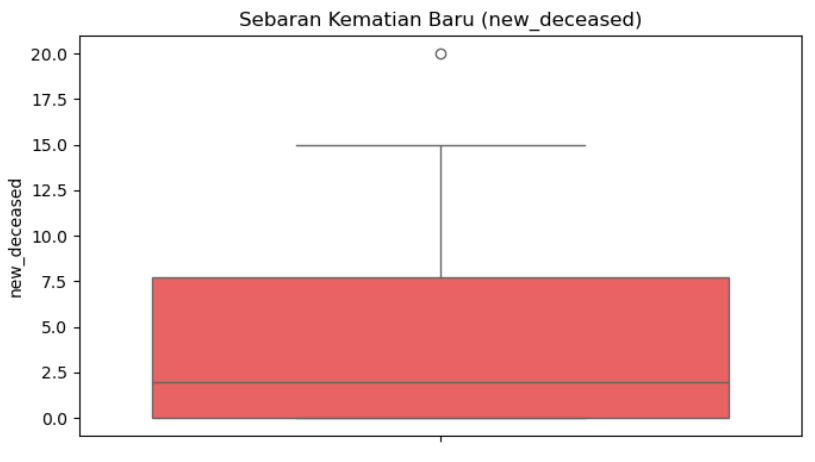
Untuk model regresinya, kami dapatkan bahwa:

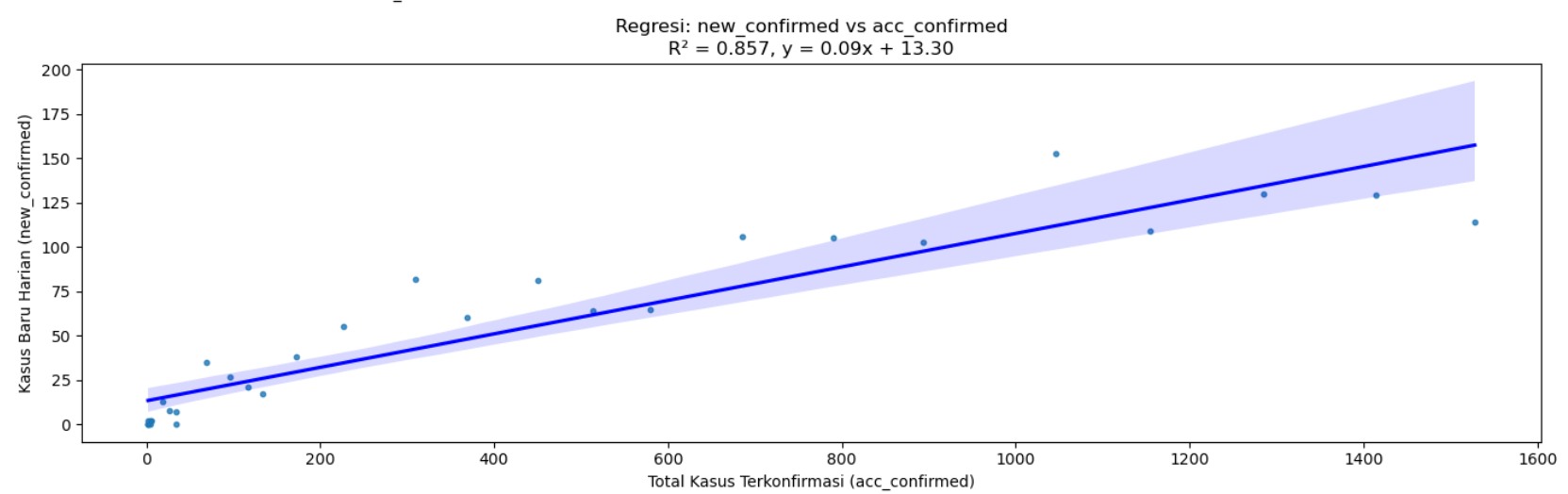
**new\_confirmed = coef: .2f × acc\_confirmed + intercept: .2f**

Dengan nilai R2R^2R2 sebesar **{r2:.4f}**, ini berarti total kasus memang punya pengaruh terhadap jumlah kasus baru, tapi tidak bisa sepenuhnya menjelaskan variasinya. Bisa jadi ada faktor lain seperti kebijakan pemerintah, tingkat vaksinasi, atau mobilitas penduduk yang juga berperan.

Visualisasi yang dibuat cukup menggambarkan kondisi datanya. Scatter plot antara total kasus dan kasus baru memperlihatkan tren naik, meskipun sebarannya masih cukup luas.







**4. Kesimpulan dan Rekomendasi**

Dari analisis ini bisa disimpulkan bahwa:

* Ada pola dasar yang bisa dilihat dari data COVID-19 harian di Indonesia, meskipun cukup bervariasi.
* Total kasus memiliki hubungan dengan kasus baru harian, tapi tidak cukup kuat untuk dijadikan satu-satunya prediktor.
* Visualisasi membantu memperjelas gambaran umum data dan bisa sangat bermanfaat untuk edukasi publik atau pelaporan internal.

**Rekomendasi ke depan:**

* Analisis lebih lanjut bisa melibatkan faktor tambahan seperti jumlah testing, vaksinasi, atau mobilitas masyarakat.
* Model prediksi dapat ditingkatkan dengan menggunakan metode seperti regresi multivariat, model time series, atau machine learning.
* Penyajian data sebaiknya dibuat lebih interaktif (misalnya dashboard) agar lebih mudah digunakan oleh pihak non-teknis.