

COVID FORECASTING

Per Januari 2022



PERKENALKAN

Saya Kresna Dwipayana S H. Sebelumnya, saya ingin mengucapkan terima kasih atas kesempatannya dalam mengikuti seleksi Jakarta Smart City Data Science Trainee ini.



PROBLEM DEFINITION

Pada proyek ini saya ingin membuat sebuah perkiraan atau Forecasting mengenai perkiraan jumlah kasus positif covid 19, khususnya di Indonesia. Hal ini menjadi motivasi mengingat pada juli 2021 lalu, indonesia mengalami lonjakan kasus covid 19 tipe delta yang memakan banyak korban, yang mana umumnya tingkat kematian yang tinggi salah satunya akibat tingkat kasus positif yang tinggi namun tidak diantisipasi dengan tingkat ketersediaan kasur rumah sakit yang memadai. Sekiranya proyek ini dapat menjadi antisipasi mengenai kasus positif yang akan terjadi kedepannya.

Tujuan

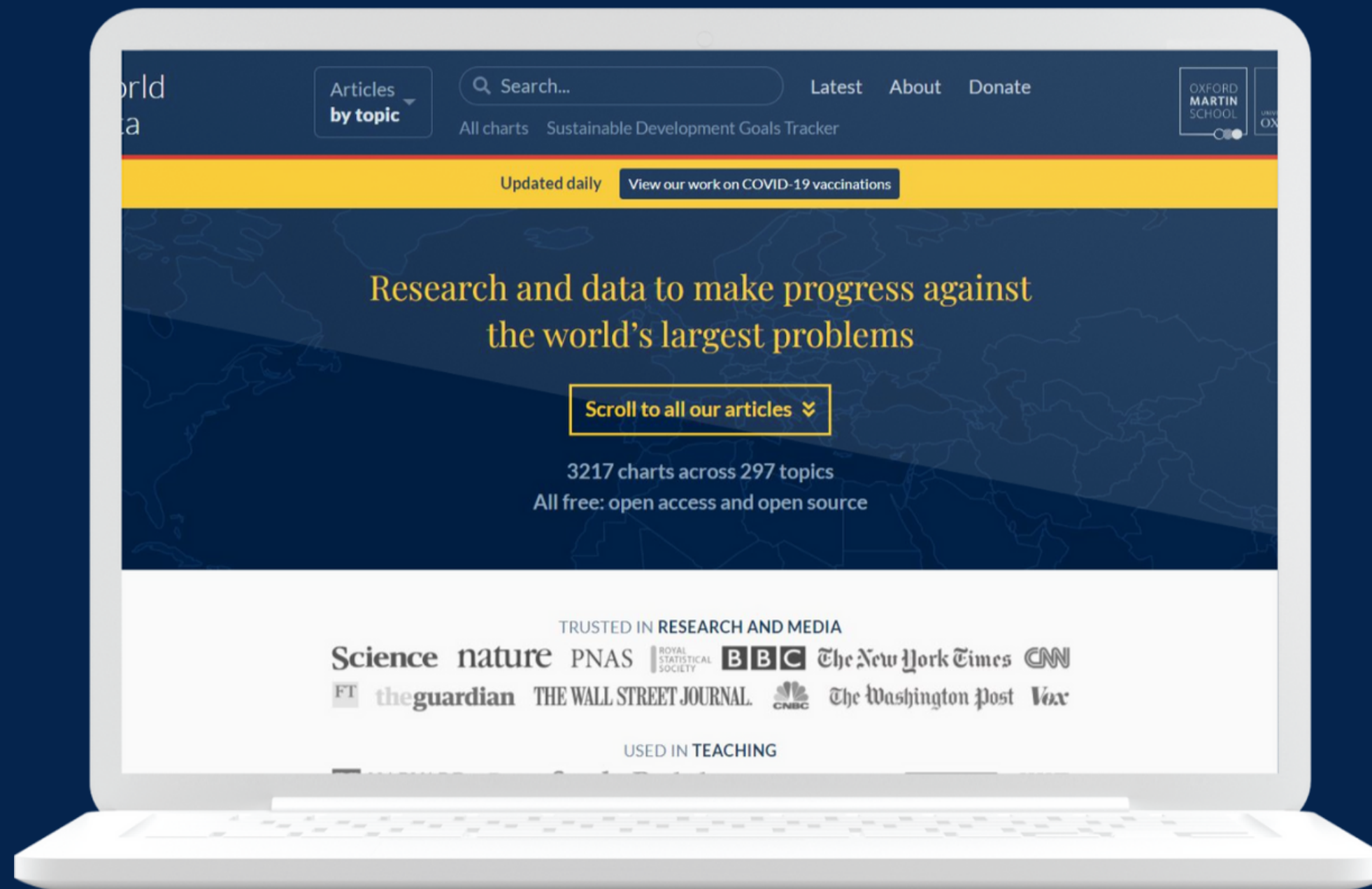
Memberikan informasi kepada para stackholder untuk mengantisipasi kasus covid19 yang akan terjadi di Indonesia pada 6 bulan ke depan untuk meminimalisir kasus kematian akibat covid19.



Objektif

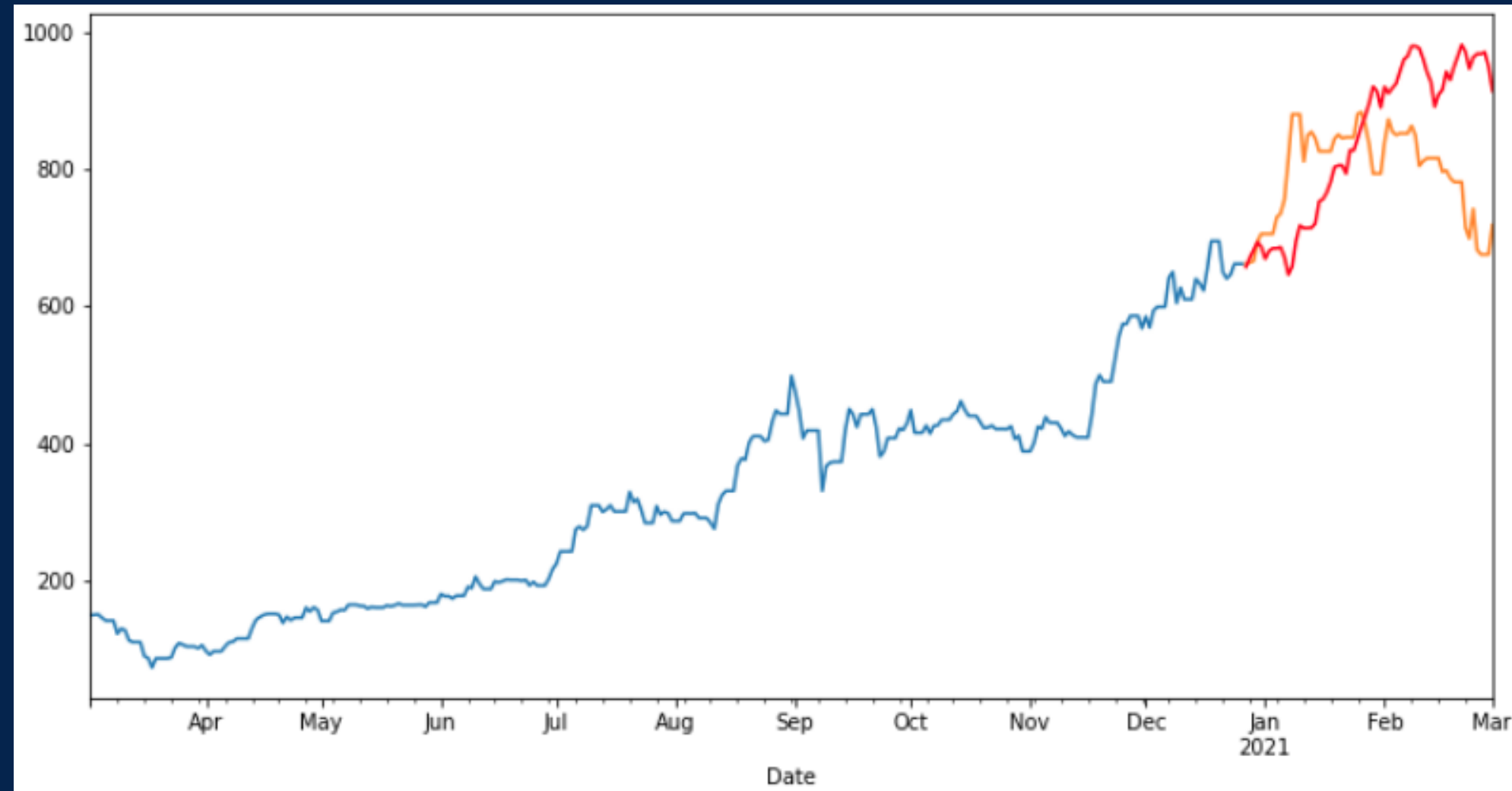
Membuat prediksi perkiraan kasus covid19 yang akan terjadi di masa mendatang dalam jangka pendek.

DATA SOURCE



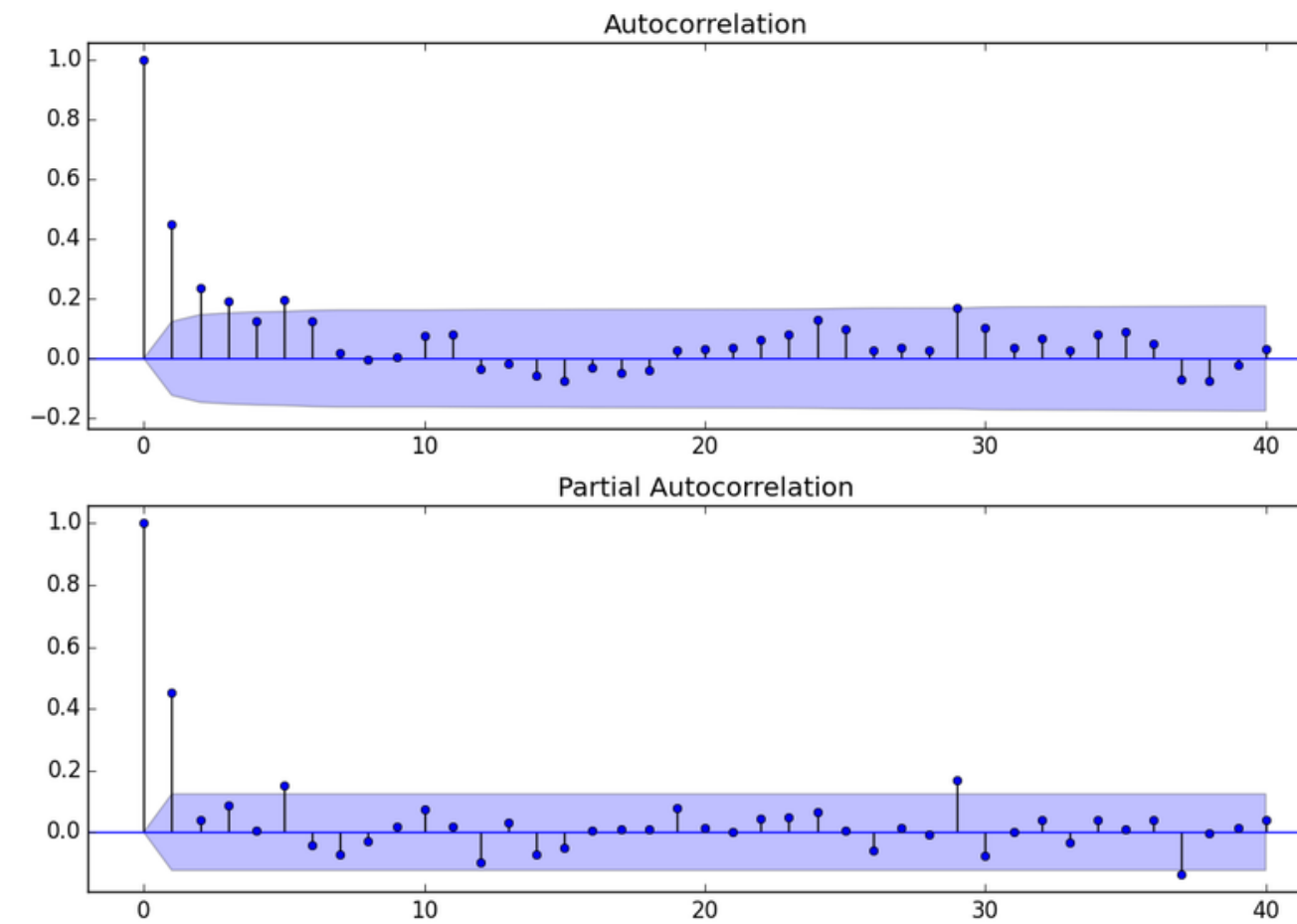
Saya mendapatkan dataset ini berasal dari website **Ourworldindata.org**, yakni dataset berisi seluruh data yang berkaitan tentang covid19. Saya mendownload dataset per tanggal 29 Januari 2022.

METODE



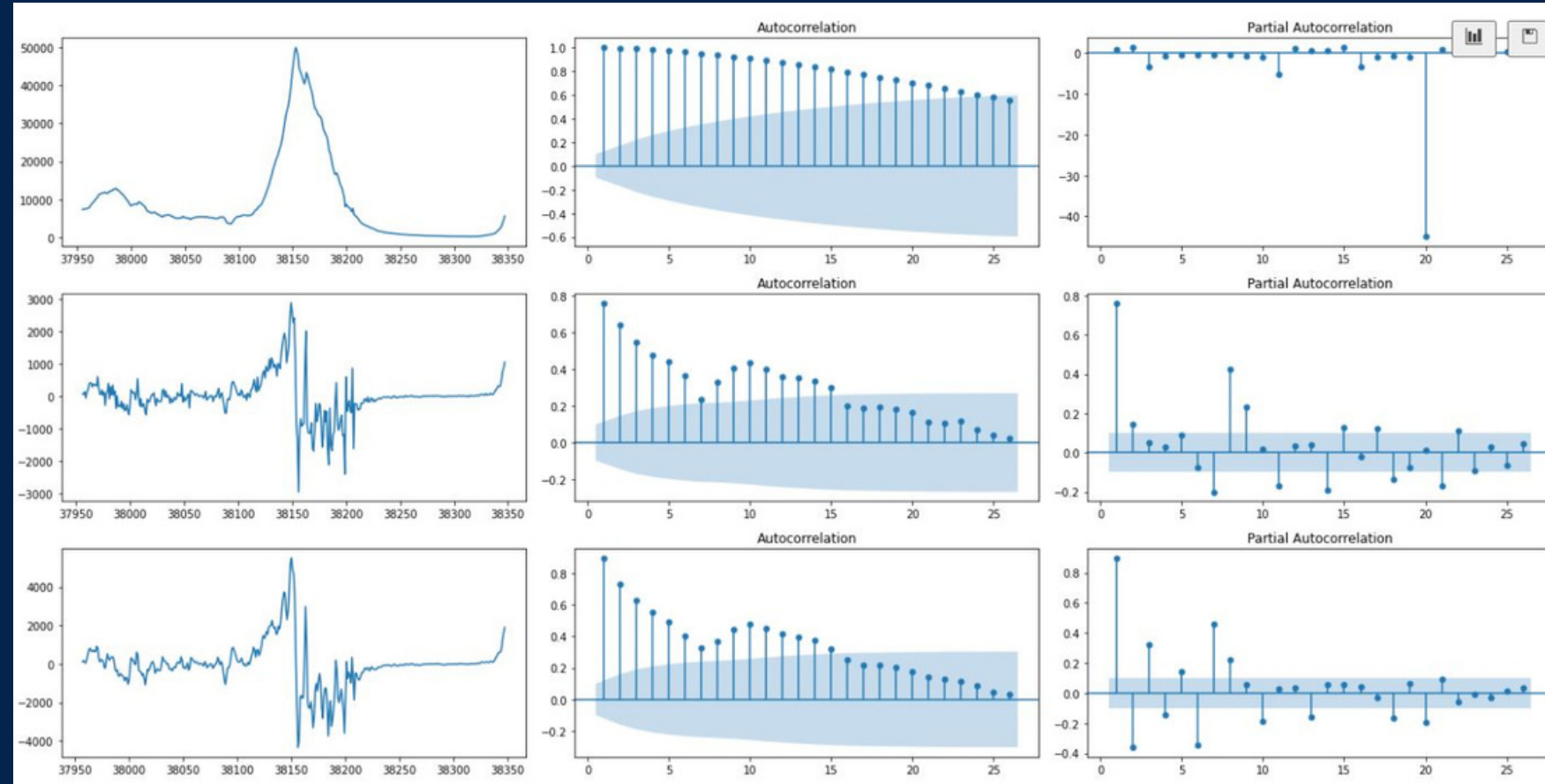
Pada proyek ini, dalam membuat forecasting saya menggunakan metode ARIMA. Yakni merupakan metode prediksi dengan memanfaatkan data periode sebelumnya untuk memprediksi masa yang akan datang. Selengkapnya dapat di lihat pada notebook modeling.

ACF – PACF



Plot ACF – PACF digunakan sebagai pengamatan untuk dapat menentukan hyperparameter yang tepat untuk model. Plot ACF berfungsi untuk mengukur korelasi antara time series dengan lag-nya, sedangkan plot PACF, yakni untuk mengukur korelasi antara time series dengan lag-nya secara parsial.

HYPERPARAMETER



Setelah membuat plot ACF – PACF, kita dapat menentukan beberapa hyperparameter yang harus ditentukan untuk mendapatkan model ARIMA yang sesuai dengan dataset, yakni :

- p : menyatakan berapa banyak data dari observasi sebelumnya yang akan digunakan ke dalam forecasting dan mewakili autoregressive
- d : menyatakan ordo differencing yang kita perlukan dan merepresentasikan bagian integrated
- q : nilai yang digunakan untuk mengontrol autokorelasi residual dalam pemodelan serta mewakili moving average

HASIL FORECASTING



Setelah melihat plot ACF - PACF dan menentukan hyperparameter, terdapat 2 kandidat model yakni ARIMA (1,2,2) dan ARIMA (1,2,1). Berdasarkan nilai AIC, performa terbaik adalah model ARIMA (1,2,1). Selanjutnya kita fit ke dalam model dan menghasilkan prediksi yang kita tentukan untuk 6 bulan ke depan, yakni sampai bulan juli 2022. Hasilnya terprediksi angka kasus positif covid untuk 6 bulan ke depan berada pada rentang 5000 sampai 7000 kasus setiap bulannya. Saya juga membuat batas bawah dan batas atas sebagai antisipasi jika prediksi angka positif covid meleset ataupun melonjak sewaktu - waktu.

TERIMA KASIH

