**Nama** : Andreas Novian Dwi Triastanto

**NIM** : 23518002

**Kuliah** : IF5175 - Analisis Data & Bisnis

**Tanggal Pengumpulan** : 7 April 2020

**Laporan Projek Analisis Data Covid-19**

**Latar Belakang**

Covid-19 adalah penyakit menular yang disebabkan oleh sindrom pernafasan akut coronavirus 2 (SARS-CoV-2). Penyakit ini pertama kali diidentifikasi pada 2019 di Wuhan, ibukota provinsi Hubei, China. Di Indonesia, penyakit ini pertama kali memakan korban meninggal pada tanggal 11 Maret 2020 dan jumlahnya terus bertumbuh sejak itu. Menurut para ahli, penanganan jenazah korban meninggal akibat penyakit ini memerlukan penanganan khusus supaya virus tidak menular kepada orang di sekitarnya. Pertumbuhan jumlah korban meninggal perlu diantisipasi supaya beberapa persiapan yang diperlukan dapat dilakukan dengan baik.

**Permasalahan**

Berdasarkan latar belakang di atas, beberapa pertanyaan yang ingin dijawab adalah sebagai berikut:

1. Berapa prediksi jumlah kematian pada beberapa hari mendatang akibat dari Covid-19 di Indonesia?
2. Berapa prediksi jumlah kematian maksimal akibat dari Covid-19 di Indonesia?

**Dataset**

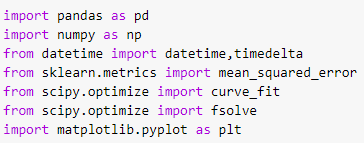
Untuk menjawab beberapa pertanyaan di atas, dibutuhkan data berupa jumlah kasus korban meninggal akibat Covid-19 di Indonesia. Dataset yang berisi situasi mengenai kasus Covid-19 di Indonesia tersedia di Kaggle[[1]](#footnote-1). Dataset tersebut berisi beberapa file tentang situasi penyebaran virus Covid-19 di Indonesia, seperti jumlah kasus positif harian, jumlah kasus per provinsi, dan data pasien positif. Dalam analisis ini, diambil file ‘cases.csv’ yang terdiri dari 15 kolom dengan jumlah baris yang bertambah setiap hari, dimulai dari tanggal 2 Maret 2020. Kolom yang dipakai untuk analisis ini adalah tanggal (date) dan jumlah kumulatif korban meninggal (acc\_deceased).

**Metode dan Teknik Analisis**

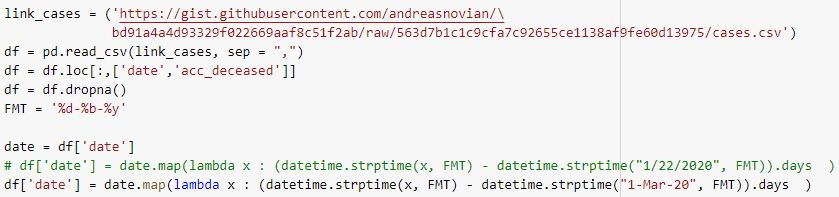
Untuk membuat model prediksi jumlah kematian akibat dari Covid-19, digunakan dua buah metode. Metode pertama adalah logistic model yang diadaptasi dari model buatan rizkyalifr[[2]](#footnote-2) yang dibuat untuk memprediksi jumlah kasus positif harian di Indonesia di Kaggle. Selain dapat memprediksi jumlah kasus positif harian, model tersebut juga mampu memprediksi jumlah maksimal kasus positif Covid-19 di Indonesia. Model kedua adalah model yang dibuat dengan menggunakan Prediction Tool: FBProphet yang diadaptasi dari model buatan ardisragen[[3]](#footnote-3) yang dibuat di Kaggle untuk memprediksi jumlah kasus positif harian di Indonesia.

Langkah-langkah pembuatan logistic model untuk memprediksi jumlah kematian akibat dari Covid-19 adalah sebagai berikut:

1. Melakukan import library yang dibutuhkan



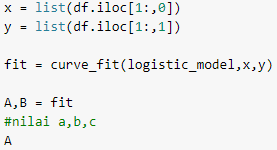
1. Membaca data dan mengubah format tanggal menjadi hitungan jumlah hari sejak 1 Maret 2020



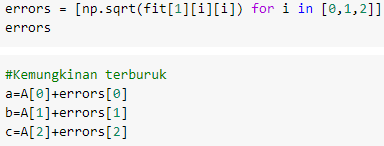
1. Membuat implementasi logistic model dengan variabel x adalah waktu, parameter a adalah kecepatan infeksi, parameter b adalah hari di mana jumlah maksimum terinfeksi terjadi, dan parameter c adalah jumlah orang yang terinfeksi di akhir masa infeksi



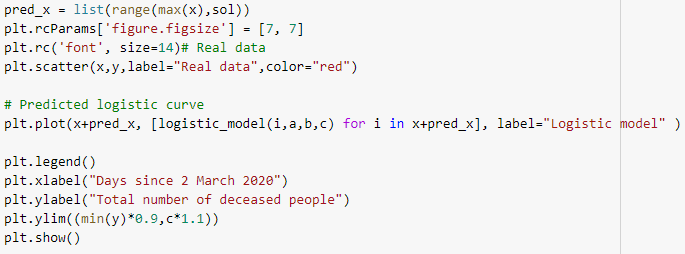
1. Melakukan proses fit fungsi ke data dengan curve\_fit



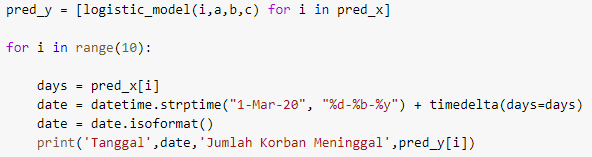
1. Menghitung error untuk ditambahkan menjadi kemungkinan terburuk



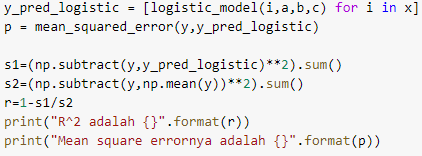
1. Melakukan plotting fungsi regresi dan data asli



1. Membuat prediksi untuk 10 hari ke depan



1. Menghitung nilai error R2 dan mean square error

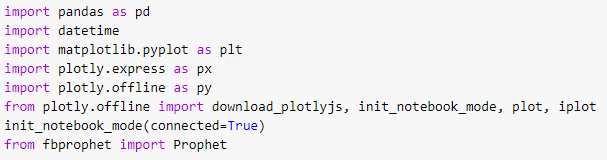


1. Melakukan prediksi jumlah maksimal korban meninggal



Langkah-langkah pembuatan model dengan Prediction Tool: FBProphet untuk memprediksi jumlah kematian akibat dari Covid-19 adalah sebagai berikut:

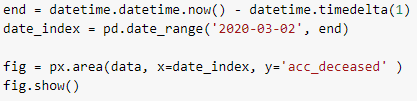
1. Melakukan import library yang dibutuhkan



1. Membaca data



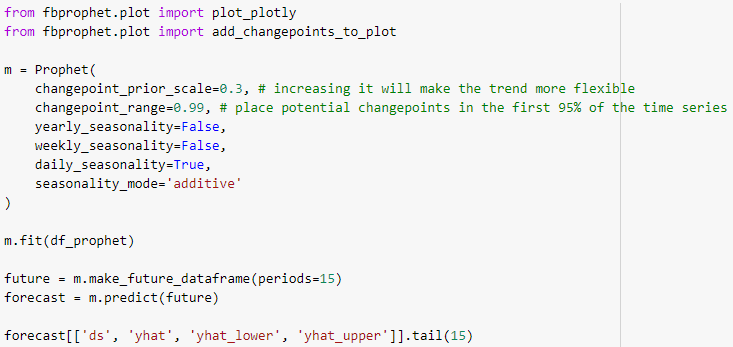
1. Melakukan plot dataset dengan plotly



1. Mengganti format dataframe menjadi format untuk FBProphet



1. Membangun model dengan FBProphet dan menampilkan hasil prediksi 15 hari ke depan



Untuk setiap metode pembuatan model prediksi, kami membuat dua buah model. Model pertama dibuat dari data jumlah kematian sampai tanggal 22 Maret 2020 dan model kedua dibuat dari data jumlah kematian sampai tanggal 30 Maret 2020. Hasil prediksi dari logistic model dan FBProphet untuk kedua model tersebut dapat dilihat pada tabel 1 dan tabel 2 berikut. Hasil prediksi jumlah maksimal kematian dengan logistic model untuk model pertama adalah 82 dan untuk model kedua adalah 205.

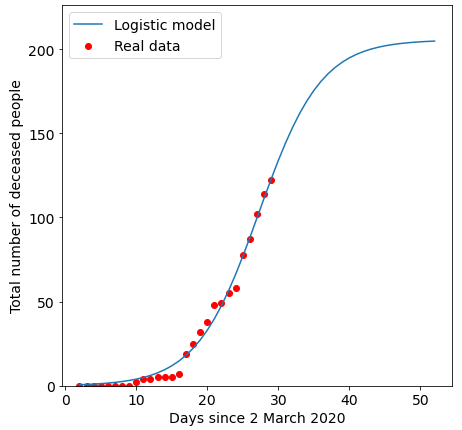
Tabel 1. Hasil prediksi dari model dengan menggunakan data sampai tanggal 22 Maret 2020

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Tanggal** | **Prediksi Jumlah Kematian**  **Dengan Logistic Model** | **Prediksi Jumlah Kematian**  **Dengan FBProphet** | **Data asli** |
| 1 | 23 Maret 2020 | 56 | 58 | 49 |
| 2 | 24 Maret 2020 | 63 | 64 | 55 |
| 3 | 25 Maret 2020 | 68 | 71 | 58 |
| 4 | 26 Maret 2020 | 72 | 76 | 78 |
| 5 | 27 Maret 2020 | 75 | 81 | 87 |
| 6 | 28 Maret 2020 | 77 | 85 | 102 |
| 7 | 29 Maret 2020 | 79 | 88 | 114 |
| 8 | 30 Maret 2020 | 80 | 92 | 122 |
| 9 | 31 Maret 2020 | 81 | 94 | 136 |

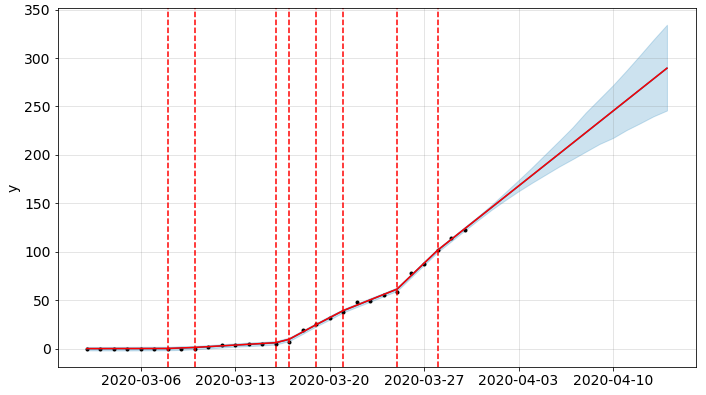
Tabel 2. Hasil prediksi dari model dengan menggunakan data sampai tanggal 30 Maret 2020

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Tanggal** | **Prediksi Jumlah Kematian**  **Dengan Logistic Model** | **Prediksi Jumlah Kematian**  **Dengan FBProphet** | **Data asli** |
| 1 | 31 Maret 2020 | 134 | 135 | 136 |
| 2 | 1 April 2020 | 144 | 146 | 157 |
| 3 | 2 April 2020 | 153 | 157 | 170 |
| 4 | 3 April 2020 | 162 | 168 | 181 |
| 5 | 4 April 2020 | 169 | 179 | 191 |
| 6 | 5 April 2020 | 175 | 190 | 198 |
| 7 | 6 April 2020 | 181 | 201 | 209 |
| 8 | 7 April 2020 | 185 | 212 |  |
| 9 | 8 April 2020 | 189 | 223 |  |

Grafik hasil plotting untuk fungsi regresi dan data asli untuk model hasil dari logistic model dan FBProphet menggunakan data sampai tanggal 30 Maret 2020 dapat dilihat pada gambar 1 dan gambar 2 berikut.



Gambar 1. Hasil plotting fungsi regresi dan data asli dari logistic model



Gambar 2. Hasil plotting fungsi regresi dan data asli dari FBProphet

Teknik analisis yang digunakan adalah dengan membandingkan hasil prediksi yang dihasilkan oleh model dengan kenyataan yang ada di dunia nyata. Kita ingin melihat seberapa baik dan akurat hasil prediksi model jika dibandingkan dengan data asli di lapangan. Data asli yang digunakan sebagai pembanding didapatkan dari situs resmi pemerintah Indonesia untuk Covid-19[[4]](#footnote-4).

**Hasil Analisis**

Jika melihat hasil prediksi untuk model yang menggunakan data sampai tanggal 22 Maret 2020, kedua model mampu memprediksi dengan baik sampai tanggal 27 Maret 2020. Pada tanggal 28 Maret 2020, terjadi peningkatan jumlah kematian yang cukup signifikan yang gagal diprediksi oleh kedua model. Jika melihat ke lapangan, hal ini mungkin terjadi akibat peningkatan jumlah kasus positif yang cukup signifikan akibat diadakan rapid test di beberapa daerah.

Jika melihat hasil prediksi untuk model yang menggunakan data sampai tanggal 30 Maret 2020, kedua model mampu memprediksi dengan cukup baik sampai tanggal 6 April 2020. Model yang dihasilkan oleh FBProphet mampu memprediksi lebih baik dibandingkan logistic model dengan error yang lebih kecil.

Jika membandingkan hasil prediksi dari logistic model dan FBProphet, dapat dilihat bahwa model yang dihasilkan oleh FBProphet cenderung menghasilkan nilai prediksi dengan pertumbuhan nilai yang lebih besar dari hari ke hari. Model ini dirasa lebih cocok untuk memprediksi pertumbuhan jumlah kematian akibat Covid-19 di Indonesia yang memang pertumbuhannya cenderung cepat.

**Diskusi**

Model prediksi yang dibuat dalam projek ini hanya menggunakan data jumlah kematian sebelumnya. Kita tahu bahwa banyak faktor yang mempengaruhi pertumbuhan jumlah kematian akibat Covid-19, seperti jumlah kasus positif, ketersediaan rumah sakit, dokter dan alat kesehatan, dan lain-lain. Penambahan faktor variabel lain untuk pembuatan model prediksi seharusnya membuat hasil prediksi model menjadi lebih presisi.

Model prediksi yang dibuat dalam projek ini menggunakan data jumlah kematian akibat Covid-19 untuk seluruh Indonesia sehingga prediksi yang dihasilkan pun merupakan jumlah kumulatif dari seluruh Indonesia. Permasalahan yang ingin coba diselesaikan dengan model prediksi ini adalah antisipasi jumlah korban meninggal akibat Covid-19 kaitannya dengan persiapan penanganan khusus untuk jenazah korban meninggal. Jika tersedia data jumlah kematian untuk setiap provinsi atau daerah, akan lebih baik jika dibuat model prediksi untuk setiap provinsi atau daerah tersebut supaya dapat menjawab permasalahan dengan lebih baik.

1. https://www.kaggle.com/ardisragen/indonesia-coronavirus-cases [↑](#footnote-ref-1)
2. https://www.kaggle.com/rizkyalifr/logistic-model-for-indonesia-covid-19 [↑](#footnote-ref-2)
3. https://www.kaggle.com/ardisragen/predicting-coronavirus-positive-cases-in-indonesia [↑](#footnote-ref-3)
4. https://www.covid19.go.id/situasi-virus-corona/ [↑](#footnote-ref-4)