

UAS TI4141 Analitika Data

Analisis Data Penyebaran Virus COVID-19 di Italia

Oleh: Gilbert Tjia / 13417047

Saya mengerjakan UAS ini secara individu, tanpa menerima bantuan dari mahasiswa TI4141 ataupun individu lainnya. Saya memahami bahwa melakukan tindak plagiarisme dan kerja sama dalam pengerjaan UAS ini adalah terlarang, dan jika saya terbukti melakukan upaya plagiarisme dan kerja sama, saya bersedia menerima sanksi sesuai dengan kebijakan ITB.

Latar belakang

Sejak tanggal 12 Maret 2020, WHO telah menetapkan wabah COVID-19 sebagai pandemi. Pandemi sendiri dapat didefinisikan sebagai penyakit yang menyebar ke banyak orang di beberapa negara pada waktu yang bersamaan (WHO). Namun, beberapa negara sudah mengalami wabah COVID-19 dengan tingkat keparahan yang tinggi bahkan jauh sebelum penetapan yang dilakukan oleh WHO ini. Beberapa negara tersebut antara lain adalah China, Italia, dan Korea Selatan. Saat penulisan laporan ini dilakukan, ketiga negara tersebut sudah dapat mengontrol penyebaran virus COVID-19 di negaranya masing-masing, bahkan kegiatan perekonomian dan aktivitas sosial lainnya telah kembali berlangsung dengan “normal” di negara China dan Korea Selatan. Sedangkan, di Italia kegiatan perekonomian dan sosial masih belum dapat berlangsung seperti semula, walaupun kondisi wabah COVID-19 di negara tersebut sudah jauh lebih terkendali dibandingkan pada bulan Maret-April 2020.

Saat ini, negara Indonesia sedang berada pada tahap paling krusial dalam proses penanganan wabah COVID-19. Hal ini disebabkan karena pada saat ini, Indonesia hampir memasuki masa puncak dari wabah COVID-19. Langkah yang diambil pemerintah pada tahap krusial ini akan sangat berpengaruh pada tingkat keparahan serta laju penyebaran COVID-19 di Indonesia ke depannya. Keputusan yang diambil pemerintah terkait COVID-19 tentu harus mempertimbangkan banyak faktor, mulai dari faktor perekonomian, sosial, dan masih banyak faktor lainnya. Salah satu cara yang dapat dilakukan oleh pemerintah untuk membantu melakukan pengambilan keputusan yang tepat terkait penanganan COVID-19 di Indonesia, adalah dengan belajar dari negara-negara yang telah melalui tahap krusial ini, seperti China, Italia, dan Korea Selatan. Pada laporan ini, akan dilakukan analisis terhadap data penyebaran virus COVID-19 di Italia, agar selanjutnya dapat digunakan sebagai rekomendasi bagi pemerintah Indonesia dalam melakukan pengambilan keputusan terkait penanganan COVID-19 di negara ini.

Metodologi Penelitian

Framework utama yang akan digunakan pada penelitian ini adalah *framework* CRISP-DM (*cross industry standard process for data mining*). Berikut merupakan diagram yang menunjukkan alur *framework* CRISP-DM:



Gambar 1 Framework CRISP-DM

Berikut merupakan penjelasan singkat mengenai tiap bagian dari *framework* CRISP-DM tersebut:

1. Business Understanding

Merupakan tahap pertama dan terpenting dari *framework* CRISP-DM. Pada tahap ini dilakukan analisis dan pemahaman terhadap proses bisnis dari *stakeholder*, serta masalah yang sedang dihadapi oleh *stakeholder* tersebut. Masalah tersebut kemudian dirumuskan untuk selanjutnya dipecah menjadi sub-masalah yang lebih kecil untuk diselesaikan. Berdasarkan masalah dan sub-masalah tersebut, dapat ditentukan tujuan dari penelitian yang dilakukan

2. Data Understanding

Pada tahap ini, dilakukan pengecekan terhadap data yang akan digunakan untuk dianalisis. Pengecekan data ini biasanya dapat dilakukan dengan melakukan pengecekan terhadap kolom-kolom dari dataset, tipe data yang dimiliki, dan atribut-atribut lain yang berkaitan dengan data. Selain itu, pada tahap ini biasanya juga dilakukan proses *data preprocessing* untuk memastikan bahwa tidak ada data yang hilang / kurang, serta proses *data visualization* untuk mengetahui pola hubungan antar variabel dalam data, serta *insight* menarik yang mungkin belum terlihat sebelum dilakukannya proses visualisasi data

3. Data Preparation

Tahap selanjutnya adalah proses *data preparation*, dimana data mentah yang sebelumnya telah melalui proses *data preprocessing* selanjutnya diolah kembali agar dapat digunakan pada tahap *modelling*. Pada tahap ini, biasanya dilakukan penghapusan kolom-kolom yang tidak dibutuhkan, serta pengolahan data mentah menjadi data yang lebih terstruktur agar dapat sesuai dengan kebutuhan *modeling*

4. **Modeling**

Tahap *modeling* merupakan tahap pembuatan model *machine learning* yang akan digunakan untuk memecahkan masalah yang ada. Pada tahap ini, dilakukan pembagian dataset menjadi dataset *training* dan *testing*, yang bertujuan untuk melakukan pembentukan model, serta melakukan validasi model tersebut dengan membandingkannya dengan data aktual. Output dari tahap *modeling* ini dapat berupa prediksi, klasifikasi, ataupun *output* lain yang merupakan hasil perhitungan dari model yang telah dibuat

5. **Evaluation**

Setelah model telah dibuat dan *output* telah dihasilkan, selanjutnya dilakukan evaluasi terhadap *output* untuk mengetahui apakah *output* yang dihasilkan dari model sudah cukup untuk menjawab *business problem* yang telah dispesifikasikan pada tahap *business understanding*

6. **Deployment**

Tahap *deployment* merupakan tahap terakhir dari *framework* CRISP-DM, dimana pada tahap ini dilakukan implementasi model yang sebelumnya telah dibuat untuk membantu *stakeholder* dalam menyelesaikan permasalahan bisnis yang dihadapinya

Pengolahan Data dan Analisis

Pada bagian ini, akan dilakukan proses analisis terhadap dataset COVID-19 di Italia menggunakan *framework* CRISP-DM yang telah dijelaskan pada bagian sebelumnya. Berikut merupakan langkah-langkah analisis yang dilakukan:

Business Understanding

Sesuai dengan latar belakang permasalahan yang telah dibuat, diketahui bahwa pada saat ini pemerintah Indonesia sedang menghadapi masa krusial dalam proses penanganan wabah COVID-19 di Indonesia. Keputusan yang diambil oleh pemerintah pada masa ini akan sangat mempengaruhi nasib negara Indonesia ke depannya. Karena alasan inilah pemerintah Indonesia membutuhkan bantuan berupa rekomendasi strategi / langkah yang perlu diambil dalam menghadapi wabah COVID-19. Rekomendasi strategi tersebut dapat diperoleh dengan terlebih dahulu mempelajari negara lain yang telah melewati masa krusial penanganan wabah COVID-19. Dalam kasus ini, pemerintah Indonesia meminta bantuan untuk melakukan analisis terhadap strategi yang diambil oleh pemerintah Italia dalam menangani wabah tersebut di negaranya.

Berdasarkan skenario permasalahan tersebut, terlihat bahwa rumusan permasalahan utama yang perlu untuk diselesaikan pada penelitian ini adalah “Bagaimana strategi / langkah yang seharusnya diambil oleh pemerintah dalam menangani wabah COVID-19 di Indonesia?” Rumusan masalah ini kemudian dapat dipecah lagi menjadi beberapa sub-masalah yang lebih spesifik sebagai berikut:

- Bagaimana cara memprediksi laju kenaikan kasus COVID-19 Indonesia?
- Bagaimana cara menekan tingkat kematian dan persebaran COVID-19 di Indonesia?

Kesuksesan proyek ini dapat diukur melalui beberapa parameter keberhasilan berikut:

- Pemerintah Indonesia mampu mengetahui laju peningkatan kasus COVID-19 Indonesia
- Pemerintah Indonesia dapat mengambil keputusan yang mampu menekan angka kematian dan persebaran COVID-19 di Indonesia

Data Understanding dan Data Preparation

Dataset yang akan diolah dan dianalisis lebih lanjut adalah dataset COVID-19 di Italia yang dimulai dari tanggal 24 Februari 2020 hingga 23 April 2020. Terdapat 2 dataset berbeda yang diberikan oleh pemerintah Indonesia, yaitu dataset Regional dan dataset Provinsi. Pada penelitian ini, proses pengolahan dan analisis akan difokuskan pada dataset Regional. Alasan pemilihan dataset Regional menjadi dataset utama yang akan dianalisis adalah karena penelitian ini lebih memfokuskan pada data skala nasional untuk negara Italia, yang kemudian akan dianalisis untuk menentukan strategi yang tepat untuk diadopsi oleh pemerintah Indonesia. Berikut merupakan tabel penjelasan dari masing-masing variable dari dataset Regional.

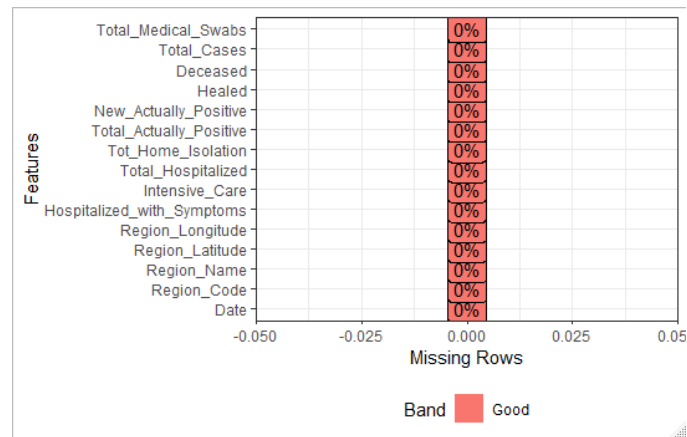
No	Variable	Description
1	Date	The Date of the Medical Detection
2	Region_Code	The ISTAT Regional Code provided by the Government
3	Region_Name	The Name of the Italian Region
4	Region_Latitude	The Region main Latitude
5	Region_Longitude	The Region main Longitude
6	Hospitalized_with_Symptoms	The Total Number of People Hospitalized with Syntoms
7	Intensive_Care	The Total Number of People in Intensive Care for that day in that Region
8	Total_Hospitalized	The Total Number of People Hospitalized for that day in that Region
9	Tot_Home_Isolation	The Total Number at Home Isolation for that day in that Region
10	Total_Actually_Positive	The Total Number of Positive People for that day in that Region
11	New_Actually_Positive	The Number of New People Actually Positive for that Day in that Region
12	Healed	The Total Number of Healed People
13	Deceased	The Total Number of Deceased People
14	Total_Cases	The Total Number of Cases
15	Total_Medical_Swabs	The Total Number of Medical Swabs done to the Population

Gambar 2 Penjelasan variabel dataset

Berdasarkan data yang diberikan, strategi analisis yang akan digunakan adalah melalui analisis terhadap tren wabah COVID-19 di Italia melalui analisis regresi dan *time series*, lalu menarik *insight* yang dapat bermanfaat bagi pemerintah Indonesia.

Langkah *data preprocessing* yang dilakukan terhadap dataset COVID-19 regional adalah sebagai berikut:

1. **Memasukkan dataset Regional pada IDE R dengan query read.csv()**
Query read.csv() digunakan untuk memasukkan dataset Regional ke dalam IDE R untuk selanjutnya dilakukan pengolahan
2. **Menangani *duplicate* data dengan query unique()**
Query unique() digunakan untuk menghilangkan data duplikat pada dataset Regional. Setelah dilakukan pengecekan lebih lanjut terhadap dataset, rupanya sudah tidak terdapat data duplikat pada dataset Regional
3. **Mengubah seluruh *blank cells* menjadi NA**
Proses ini dilakukan untuk melakukan standardisasi terhadap seluruh *missing values* yang terdapat pada dataset Regional
4. **Melakukan *plotting missing values* sebelum melalui proses *data cleaning***



Gambar 3 Hasil plotting terhadap missing values

Berdasarkan hasil *plotting missing values* pada dataset Regional, terlihat bahwa tidak terdapat *missing values* pada dataset tersebut

Proses *data preprocessing* untuk mengatasi *missing values* sudah selesai dilakukan. Selanjutnya, akan dilakukan proses visualisasi terhadap dataset untuk mendapatkan *insight*. Namun, sebelum melakukan proses visualisasi, dataset Regional perlu diagregasi terlebih dahulu untuk mendapatkan data nasional. Proses agregasi ini sebenarnya sudah termasuk ke dalam tahap *data preparation* dalam *framework* CRISP-DM. Namun, karena hasil proses agregasi dibutuhkan pada tahap *data understanding*, maka *framework* CRISP-DM yang sebelumnya telah ditunjukkan mengalami modifikasi menjadi alur berikut.



Gambar 4 Framework CRISP-DM yang telah dimodifikasi

Berikut merupakan kelanjutan dari proses *data preprocessing* yang telah dilakukan sebelumnya:

5. **Menghapus kolom Region_Code dari dataset Regional**
Kolom Region_Code tidak dibutuhkan dalam dataset tersebut karena analog dengan masing-masing Region yang terdapat di Italia
6. **Melakukan penghapusan waktu dari kolom Date**
Data waktu pada kolom Date dihapus karena dianggap kurang relevan dan tidak akan digunakan dalam proses pengolahan data selanjutnya
7. **Melakukan pengubahan tipe data Date dari *string* menjadi *date***

Proses ini dilakukan untuk menjadikan dataset Regional sebagai dataset *Time Series* yang akan memudahkan proses pemodelan dan analisis data

8. Melakukan agregasi data regional untuk mendapat data nasional

Tahap ini dilakukan untuk mendapatkan total data nasional negara Italia berdasarkan agregasi dari masing-masing Region di negara tersebut

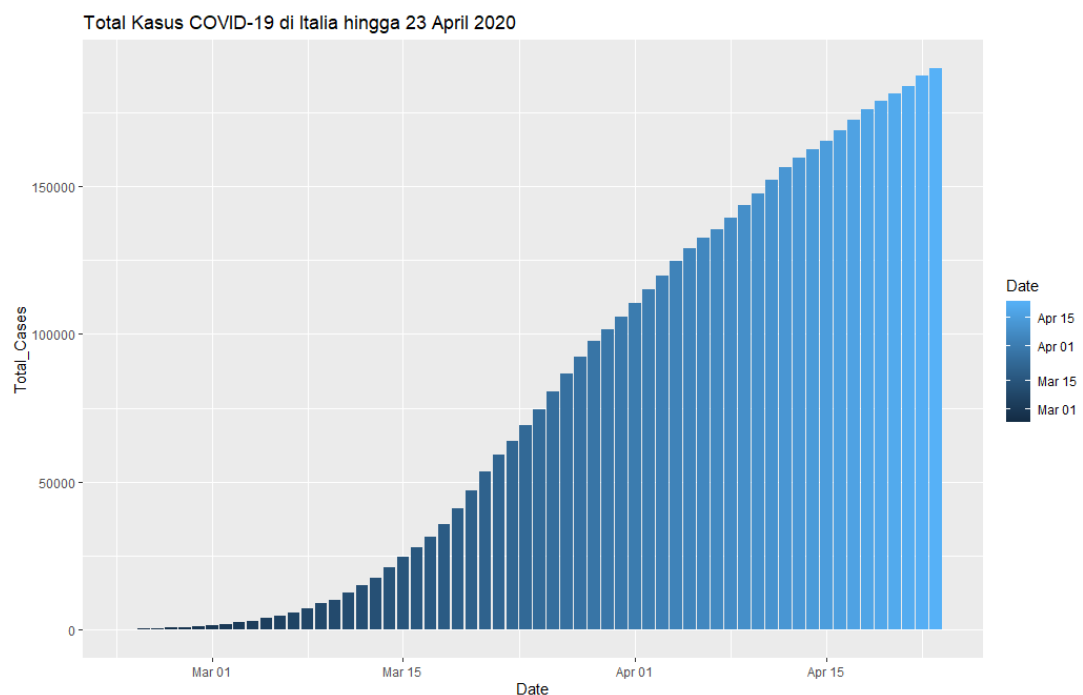
9. Melakukan perhitungan rasio kematian dan kesembuhan

Perhitungan kedua rasio ini dilakukan untuk kebutuhan visualisasi data nantinya

Exploratory Data Analysis

Tahap selanjutnya adalah tahap pencarian *insight* berdasarkan dataset yang telah dibersihkan dan diagregasi. Pada tahap ini akan dilakukan proses visualisasi data untuk membantu menemukan pola yang sebelumnya sulit untuk terlihat dari dataset.

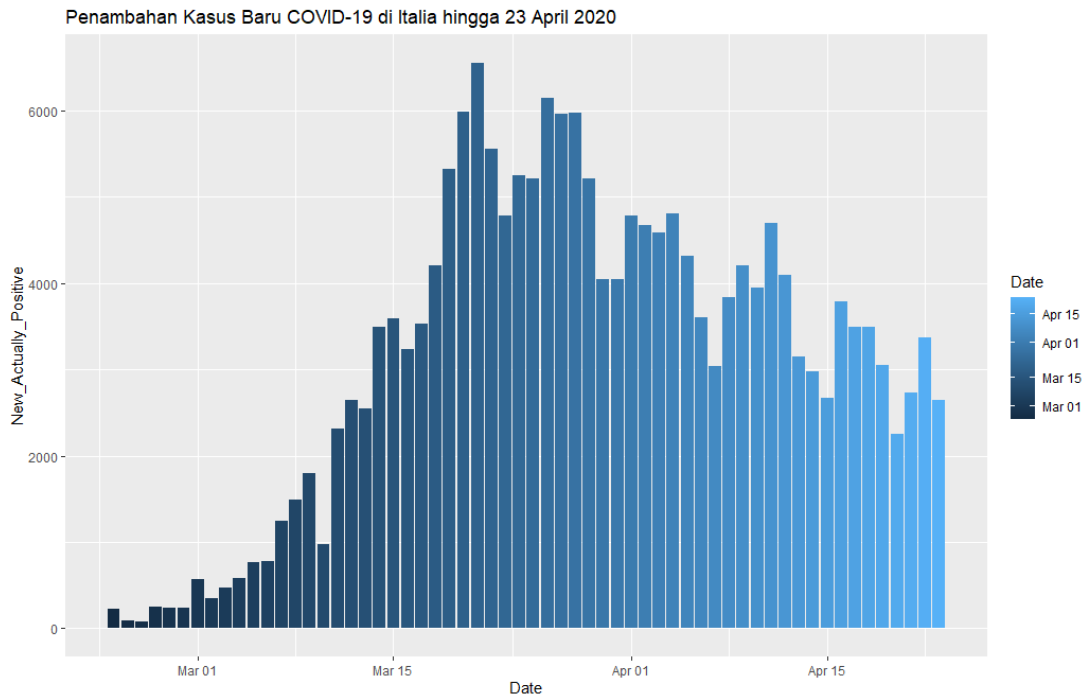
➤ **Tren total kasus COVID-19 di Italia hingga tanggal 23 April 2020**



Gambar 5 Total kasus COVID-19 di Italia hingga 23 April 2020

Berdasarkan grafik total kasus COVID-19 di Italia, terlihat bahwa peningkatan kasus COVID-19 secara eksponensial terjadi mulai sekitar awal bulan Maret 2020, kemudian mulai melambat saat memasuki awal bulan April 2020.

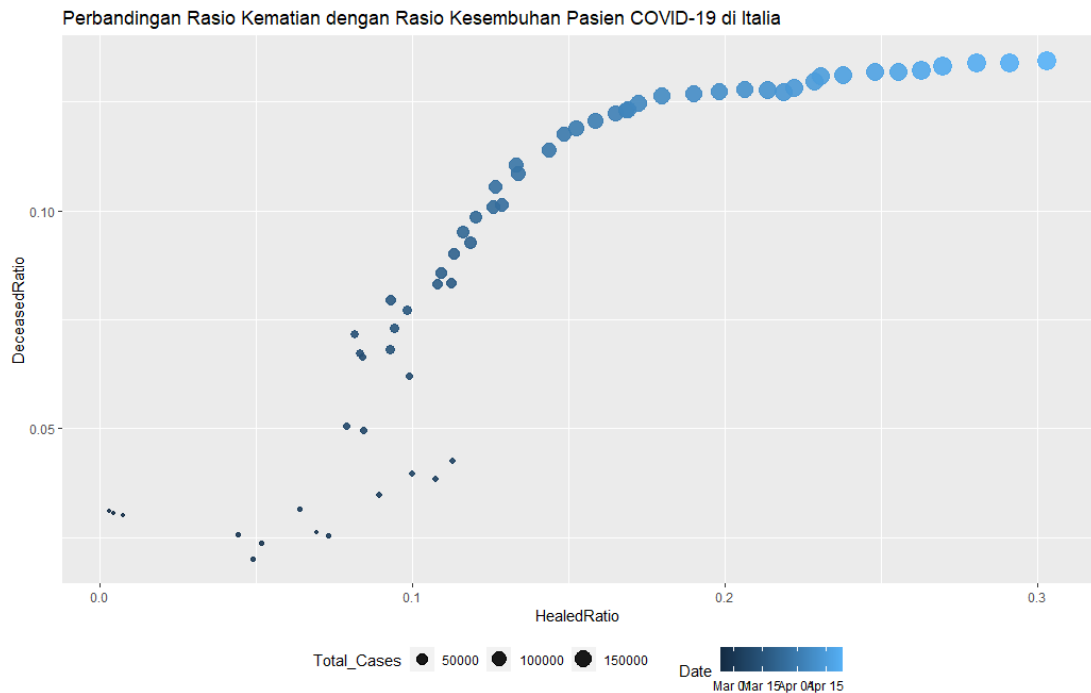
➤ **Tren penambahan kasus baru COVID-19 di Italia hingga tanggal 23 April 2020**



Gambar 6 Penambahan kasus baru COVID-19 di Italia hingga 23 April 2020

Berdasarkan grafik penambahan kasus baru COVID-19 di Italia hingga tanggal 23 April 2020, terlihat bahwa tren penambahan kasus baru COVID-19 di Italia telah melewati masa puncak yang terjadi pada pertengahan bulan Maret 2020. Setelah mencapai masa puncak tersebut, negara Italia sudah mampu mengontrol penambahan kasus baru COVID-19 dengan lebih baik. Hal ini ditunjukkan melalui tren penurunan jumlah kasus baru yang terjadi sejak akhir bulan Maret hingga akhir bulan April 2020.

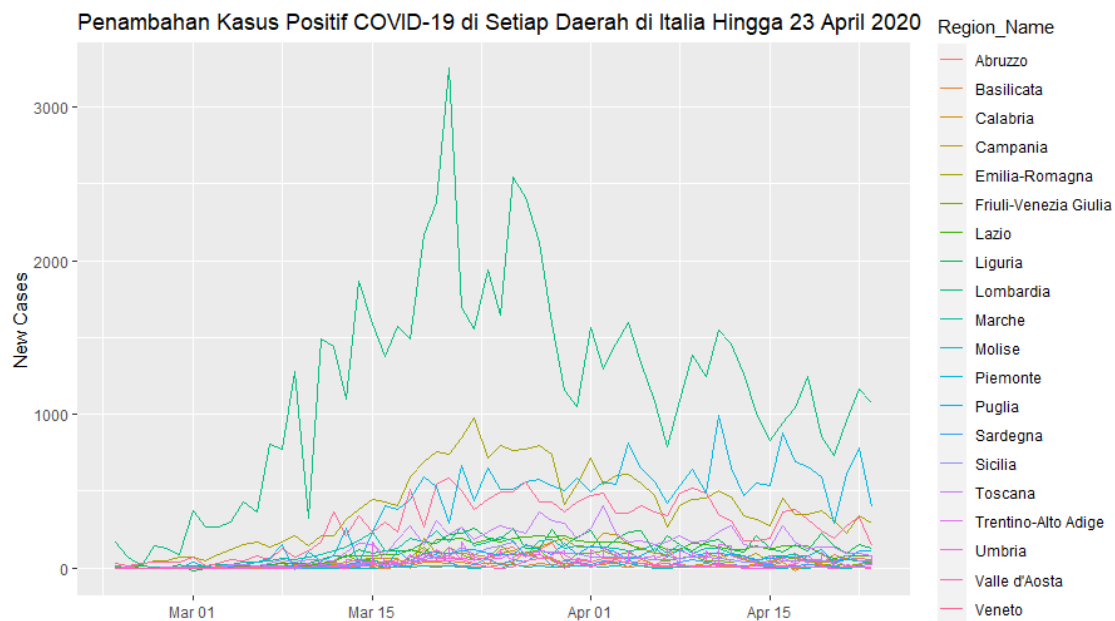
➤ **Grafik perbandingan rasio kematian dan rasio kesembuhan pasien COVID-19 di Italia**



Gambar 7 Perbandingan rasio kematian dan kesembuhan pasien

Grafik tersebut menunjukkan perbandingan antara rasio kematian dengan rasio kesembuhan pasien COVID-19 di Italia. Berdasarkan grafik tersebut, terlihat bahwa pada masa awal pandemi COVID-19, rasio kematian dan kesembuhan membentuk pola hubungan linear yang menunjukkan bahwa kedua rasio ini memiliki nilai yang serupa (jumlah kematian dan kesembuhan yang hampir sama). Namun, semenjak pertengahan bulan Maret 2020, grafik tersebut membentuk pola hubungan konstan yang menunjukkan bahwa rasio kesembuhan sudah semakin meningkat dibandingkan dengan rasio kematian.

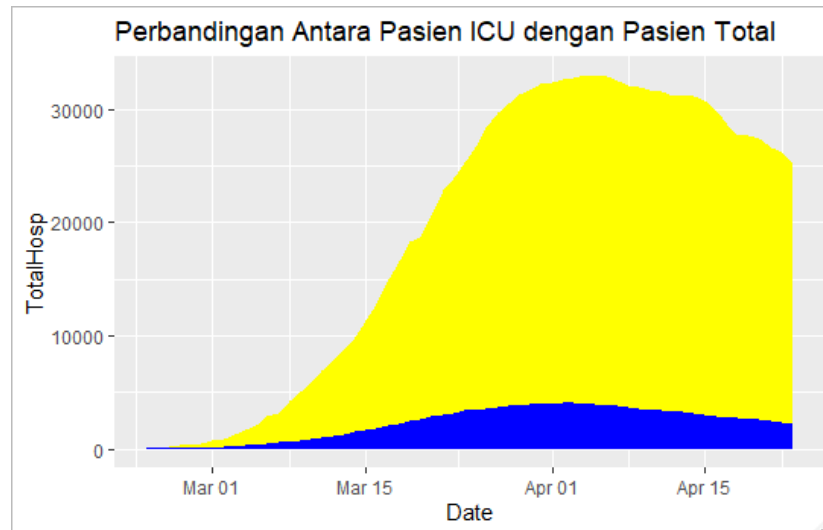
➤ Grafik penambahan kasus positif COVID-19 di setiap daerah di Italia



Gambar 8 COVID-19 pada skala regional

Grafik tersebut menunjukkan penambahan kasus positif COVID-19 setiap harinya di berbagai daerah di Italia hingga 23 April 2020. Terlihat bahwa terdapat beberapa daerah yang mengalami penambahan kasus positif yang jauh lebih cepat dan lebih tinggi dibandingkan dengan daerah-daerah lainnya. Daerah yang tergolong pada kategori ini contohnya adalah daerah Lombardy yang merupakan episentrum dari penyebaran COVID-19 di Italia.

➤ **Grafik perbandingan total pasien dengan pasien ICU**

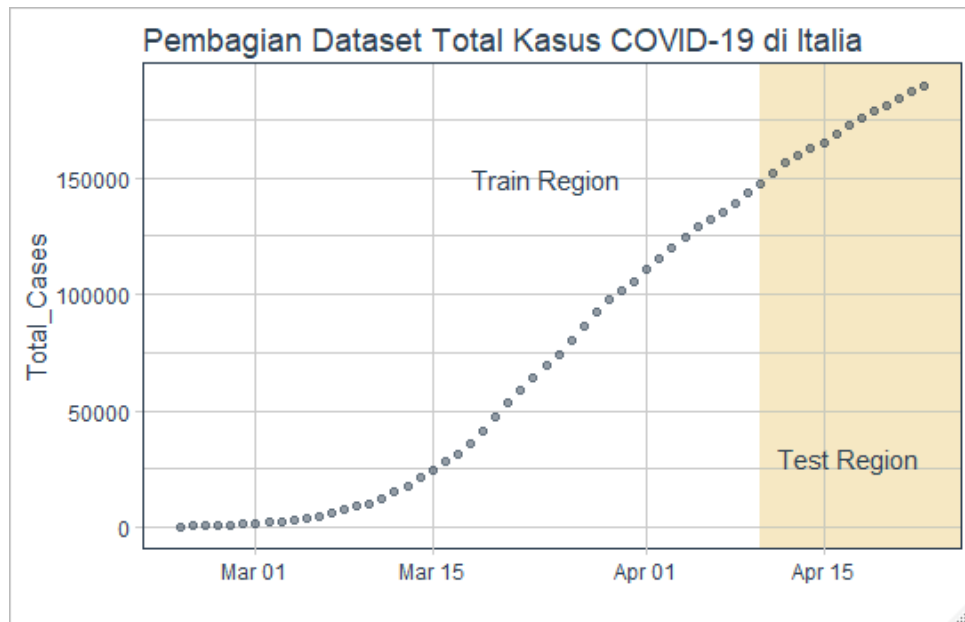


Gambar 9 Perbandingan pasien ICU dan total pasien

Grafik tersebut menunjukkan perbandingan pasien ICU (biru) terhadap total pasien (kuning). Dari grafik tersebut, terlihat bahwa peningkatan jumlah total pasien dan jumlah pasien ICU saling berkorelasi, dimana ketika terjadi peningkatan total pasien, pasien ICU pun ikut mengalami peningkatan, walaupun tidak signifikan peningkatan total pasien.

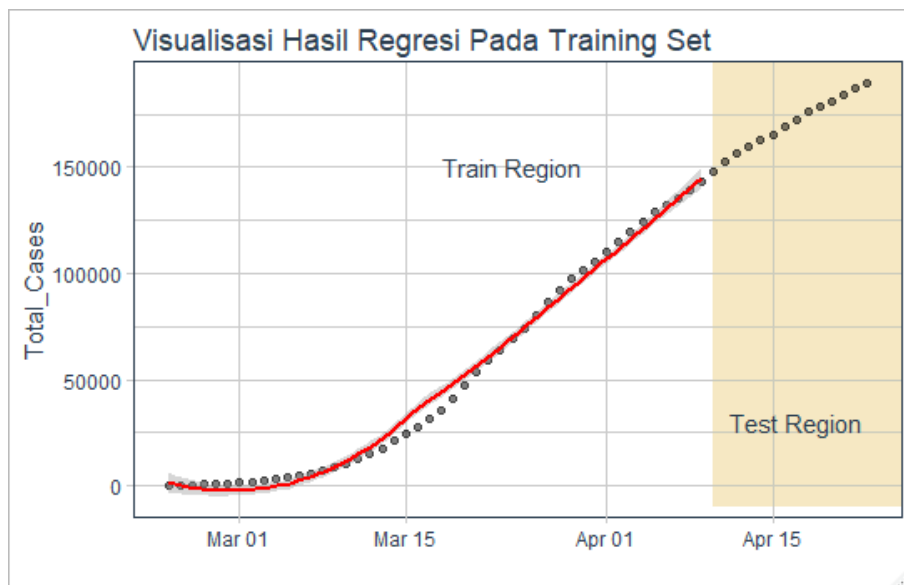
Modeling

Tahap selanjutnya dari *framework* CRISP-DM adalah tahap *modeling*. Pada tahap ini, akan dibentuk suatu model *machine learning* yang bertujuan untuk menjawab *business problem* yang telah ditetapkan sebelumnya. Sebelum melakukan pembuatan model *machine learning*, perlu terlebih dahulu dilakukan pembagian dataset menjadi dataset *training* dan *testing*. Pembagian dataset menjadi *train* dan *test set* dilakukan dengan proporsi 75% *training set* dan 25% *test set*. Berikut merupakan representasi grafis yang menunjukkan pembagian dataset tersebut.



Gambar 10 Pembagian dataset

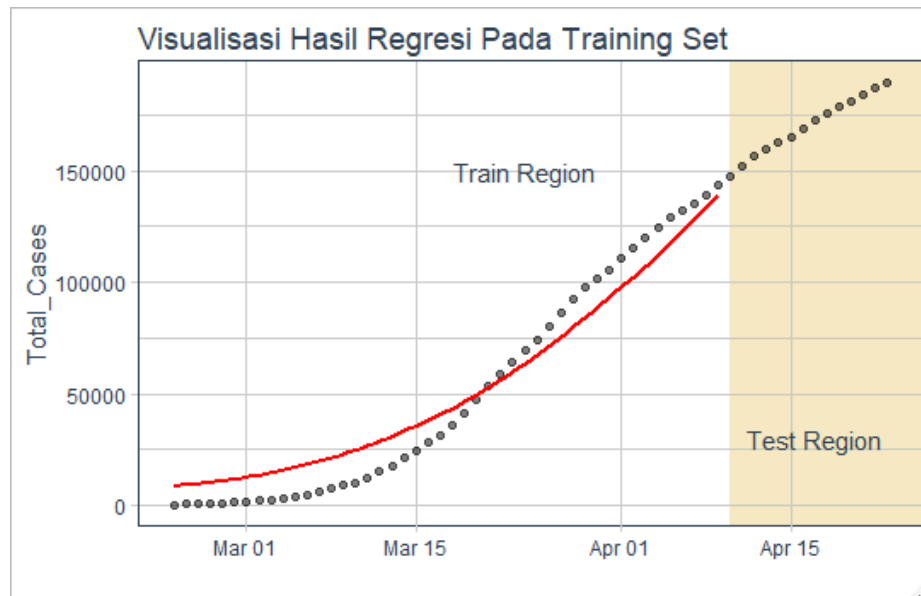
Selanjutnya, dataset *training* akan digunakan untuk melakukan pembentukan model *machine learning*. Model *machine learning* pertama yang akan digunakan pada penelitian ini adalah model *time series regression* yang merupakan suatu model regresi khusus untuk data *time series*. Proses *training* model dilakukan dengan melakukan *fitting* model *time series regression* terhadap dataset *training* yang telah dipersiapkan. Setelah model selesai dibuat, selanjutnya dilakukan proses validasi yang dilakukan dengan meminta model ML untuk melakukan prediksi terhadap dataset *test*. Hasil prediksi ini kemudian akan dibandingkan dengan data aktual untuk mengukur performansi dari model ML yang telah dibuat. Proses ini akan dijelaskan lebih lanjut pada tahap CRISP-DM berikutnya, yaitu *evaluation*.



Gambar 11 Visualisasi hasil regresi TS train set

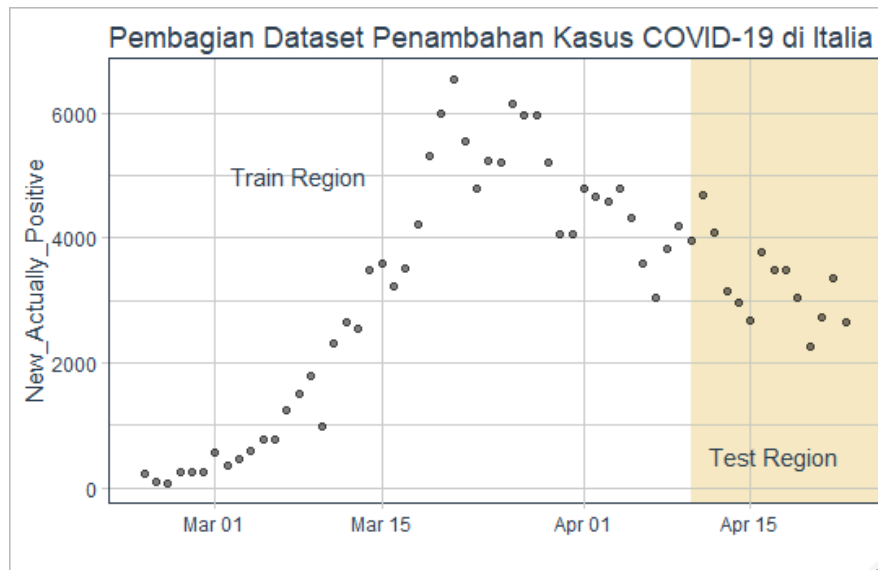
Garis merah yang terdapat pada grafik tersebut menunjukkan garis hasil *time series regression* yang dilakukan terhadap *training set*. Terlihat bahwa garis tersebut cukup mewakili / mendekati data *train set* sebenarnya. Hal tersebut juga sekaligus menunjukkan bahwa model telah selesai dibuat dan siap untuk divalidasi pada tahap selanjutnya. Selain analisis total kasus COVID-19 di Italia menggunakan

time series regression, pada penelitian ini digunakan pula model *logistic growth*. Alasan penggunaan model ini didasarkan pada hasil studi pustaka dan analisis terhadap berbagai kurva pandemi yang pernah terjadi sebelum pandemi COVID-19. Berdasarkan hasil analisis tersebut, didapatkan suatu *insight* menarik. Tren kurva dari setiap pandemi yang pernah terjadi, ternyata menunjukkan suatu pola yang sama. Pola ini adalah bahwa kenaikan total kasus pada masa awal pandemi pada umumnya akan mengikuti pola pertumbuhan eksponensial (*exponential growth model*), kemudian seiring berjalannya waktu, pola tersebut akan mengikuti pola pertumbuhan logistik (*logistic growth model*). Dengan mengikuti langkah *modeling* yang sama dengan sebelumnya, berikut merupakan performansi model *logistic growth* untuk memprediksi total kasus COVID-19 di Italia.



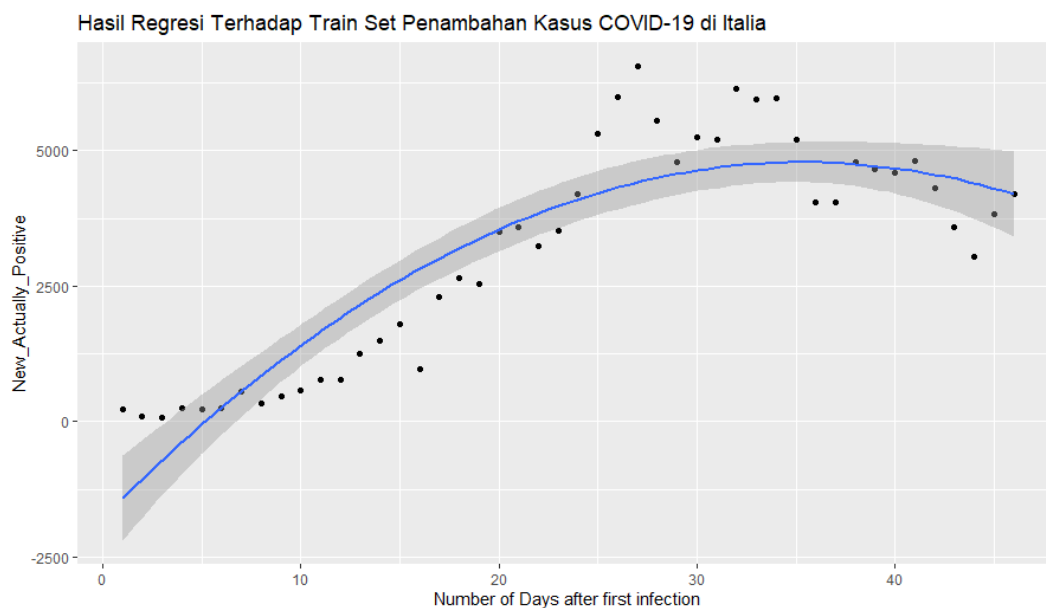
Gambar 12 Visualisasi hasil regresi LG pada train set

Serupa dengan model *time series regression*, model *logistic growth* juga mampu melakukan regresi terhadap *train set* dengan cukup baik. Bentuk "*logistic*" dari model ini belum terlihat pada tahap awal regresi. Bentuk ini nantinya akan terlihat ketika dilakukan *forecast* untuk beberapa periode ke depan. Setelah dilakukan analisis terhadap total kasus COVID-19 di Italia, selanjutnya akan dilakukan analisis terhadap penambahan kasus baru COVID-19 di Italia setiap harinya. Berikut merupakan plot penambahan kasus baru COVID-19 di Italia hingga tanggal 23 April 2020 beserta pembagian dataset tersebut menjadi *train* dan *test set*.



Gambar 13 Pembagian dataset kenaikan kasus

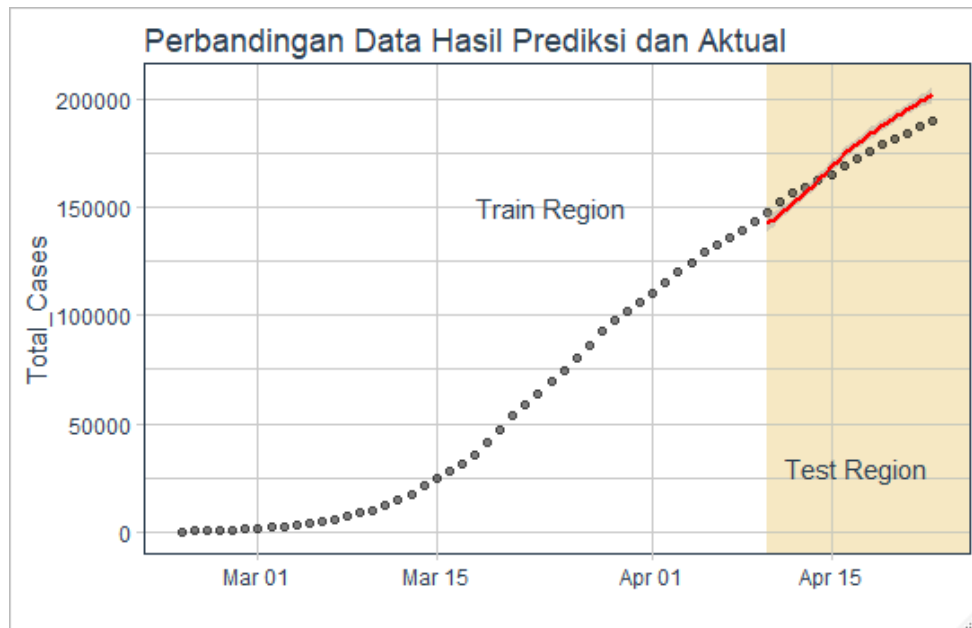
Selanjutnya, akan dilakukan pembuatan model *machine learning* untuk dataset penambahan kasus COVID-19 di Italia. Model *machine learning* yang akan digunakan untuk menganalisis dataset tersebut adalah model *polynomial regression*. Berikut merupakan hasil regresi menggunakan *polynomial regression* dengan $n=2$.



Gambar 14 Hasil regresi terhadap train set penambahan kasus

Evaluation

Seperti yang telah dijelaskan sebelumnya pada tahap *modeling*, selanjutnya akan dilakukan proses evaluasi terhadap performansi model yang telah dibuat. Proses evaluasi ini dilakukan melalui proses perbandingan antara hasil prediksi yang dihasilkan oleh model dengan data *real*. Data yang akan dijadikan sebagai pembandingan pada penelitian ini adalah *test set* yang sebelumnya telah dipilih pada tahap *modeling*. *Test set* sendiri terdiri atas 25% data terbaru dari dataset total kasus COVID-19 di Italia, yaitu data dari tanggal 10 April 2020 hingga 23 April 2020. Berikut merupakan hasil validasi yang telah dilakukan terhadap model *time series regression* maupun *logistic growth*.



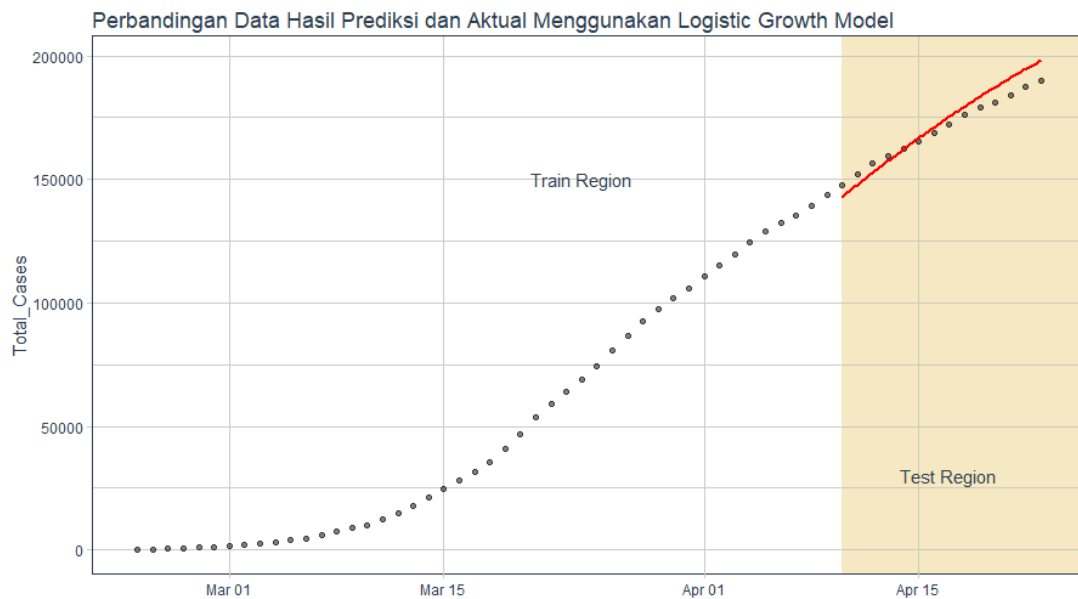
Gambar 15 Perbandingan data hasil prediksi dan aktual model TS

Secara kualitatif, terlihat bahwa model *machine learning* yang telah dibuat mampu melakukan prediksi terhadap data aktual dengan cukup baik. Tetapi, untuk memastikan hal ini perlu dilakukan pengukuran secara kuantitatif dengan melakukan perhitungan terhadap *error* dan R^2 dari model tersebut. Parameter yang dijadikan sebagai dasar pengukuran *error* model pada penelitian ini adalah RMSE (*root mean squared error*) dan MAPE (*mean absolute percentage error*).

Tabel 1 Performansi model TS regression

RMSE	7760.987
R^2	0.9859
MAPE	0.0404

Berdasarkan perhitungan terhadap *error* model, terlihat bahwa performansi model tersebut cukup baik, melihat nilai MAPE yang lebih kecil dari 5% dan nilai R^2 yang sudah sangat tinggi. Selanjutnya, hal yang sama akan dilakukan terhadap model *logistic growth*.

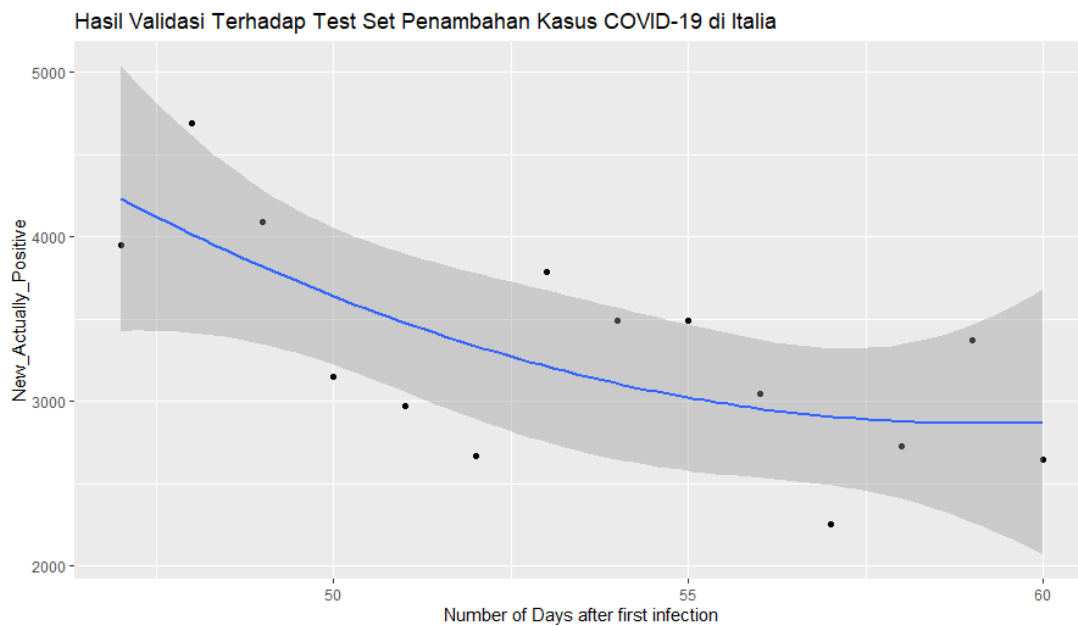


Gambar 16 Perbandingan data prediksi dan aktual model LG

Tabel 2 Performansi model LG

RMSE	4779.57
R^2	0.9987
MAPE	0.0243

Dapat terlihat bahwa nilai RMSE dan MAPE yang didapatkan melalui model *logistic growth* ini lebih kecil dibandingkan nilai RMSE dan MAPE yang diperoleh melalui model *time series regression*. Nilai R^2 model ini pun sedikit lebih tinggi dibanding nilai R^2 pada model *time series regression*. Hal ini secara sekilas menunjukkan bahwa model *logistic growth* merupakan model yang memiliki performansi lebih baik (lebih mendekati kondisi data *real*) dibandingkan model *time series regression*. Proses validasi terakhir yang perlu untuk dilakukan adalah proses validasi terhadap model *polynomial regression* yang digunakan untuk memprediksi jumlah penambahan kasus baru COVID-19 setiap harinya.



Gambar 17 Hasil validasi test set penambahan kasus

Tabel 3 Performansi model penambahan kasus

RMSE	690.71
R^2	0.4011
MAPE	0.1874

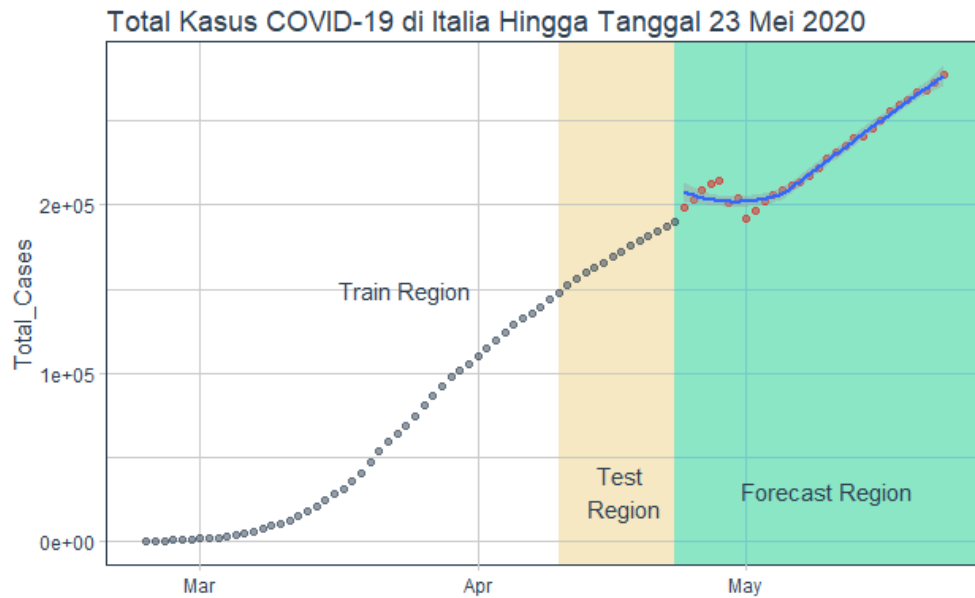
Berdasarkan hasil validasi yang telah dilakukan, terlihat bahwa RMSE yang dihasilkan dari model ini jauh lebih kecil dibandingkan RMSE yang dihasilkan oleh 2 model sebelumnya. Namun, nilai R^2 dan MAPE dari model ini lebih buruk dibandingkan 2 model sebelumnya.

Deployment

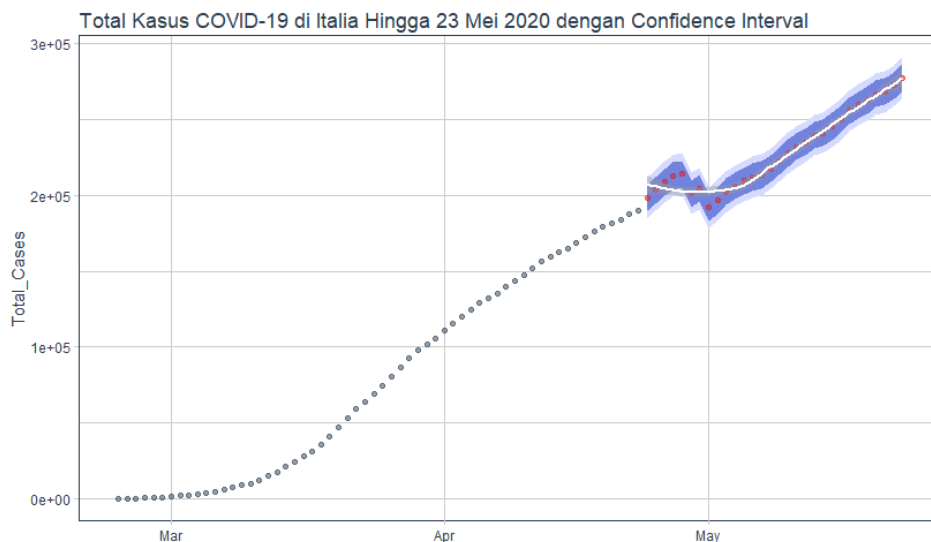
Tahap terakhir dari *framework* CRISP-DM adalah tahap *deployment*. Khusus pada penelitian ini, tahap *deployment* dibagi menjadi 2, yaitu *deployment* model untuk dataset COVID-19 di Italia untuk melakukan *forecasting*, serta *deployment* model untuk digunakan oleh pemerintah Indonesia.

Deployment model untuk Dataset COVID-19 di Italia

Setelah melakukan pengujian terhadap validitas model, langkah selanjutnya adalah melakukan prediksi menggunakan model yang telah dibuat. Prediksi yang akan dilakukan pada penelitian ini adalah prediksi terhadap jumlah total kasus di Italia selama 30 hari setelah hari setelah tanggal 23 April 2020 (hari terakhir pencatatan pada dataset).

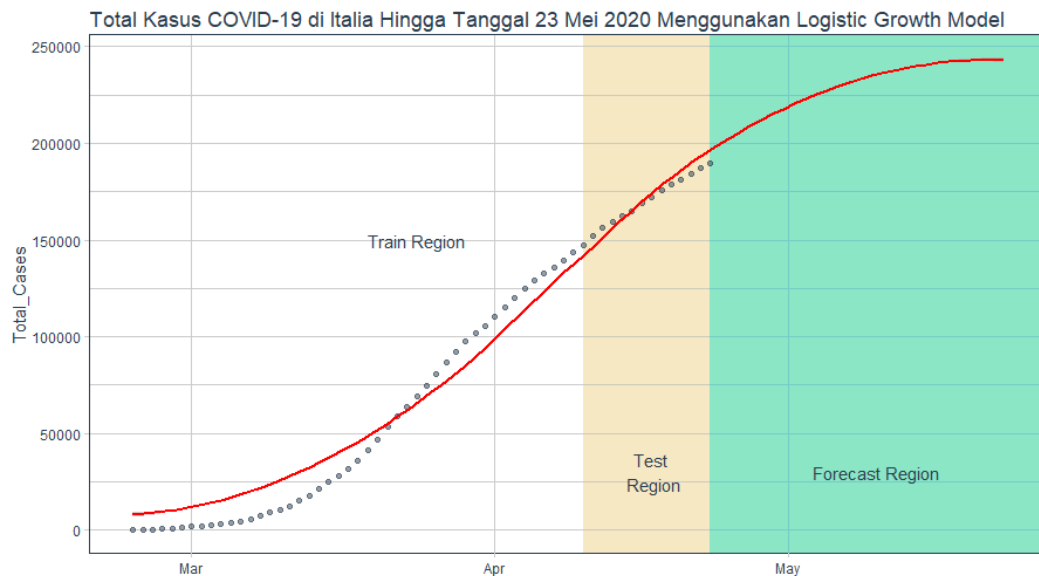


Gambar 18 Hasil forecast menggunakan TS regression



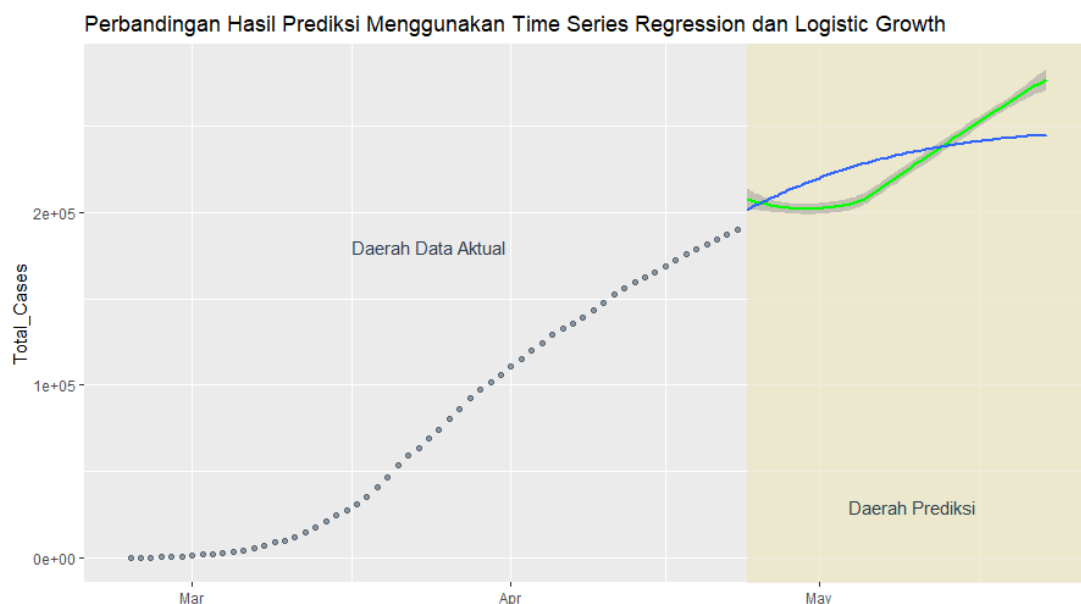
Gambar 19 Hasil forecast TS regression dengan confidence interval

Pada kedua grafik tersebut, terlihat bahwa model *machine learning* yang telah dibuat memprediksi bahwa total kasus COVID-19 di Italia akan terus mengalami peningkatan secara eksponensial mengikuti tren yang telah terjadi pada bulan Maret-April 2020. Sesuatu hal yang menarik pada grafik tersebut adalah pada awal masa *forecast* (menjelang awal bulan Mei), terlihat bahwa model tersebut memprediksi terjadinya penurunan jumlah total kasus COVID-19 di Italia sebelum terjadi kenaikan kembali. Hal ini sebenarnya tidak mungkin terjadi di dunia nyata, karena total kasus COVID-19 merupakan akumulasi dari total kasus di hari sebelumnya, sehingga tidak mungkin terjadi penurunan seperti yang diprediksi pada model tersebut. Hasil prediksi tersebut kemungkinan besar disebabkan karena model *time series regression* ini memperhitungkan *seasonality* yang terjadi pada dataset *training*.



Gambar 20 Hasil forecast menggunakan LG

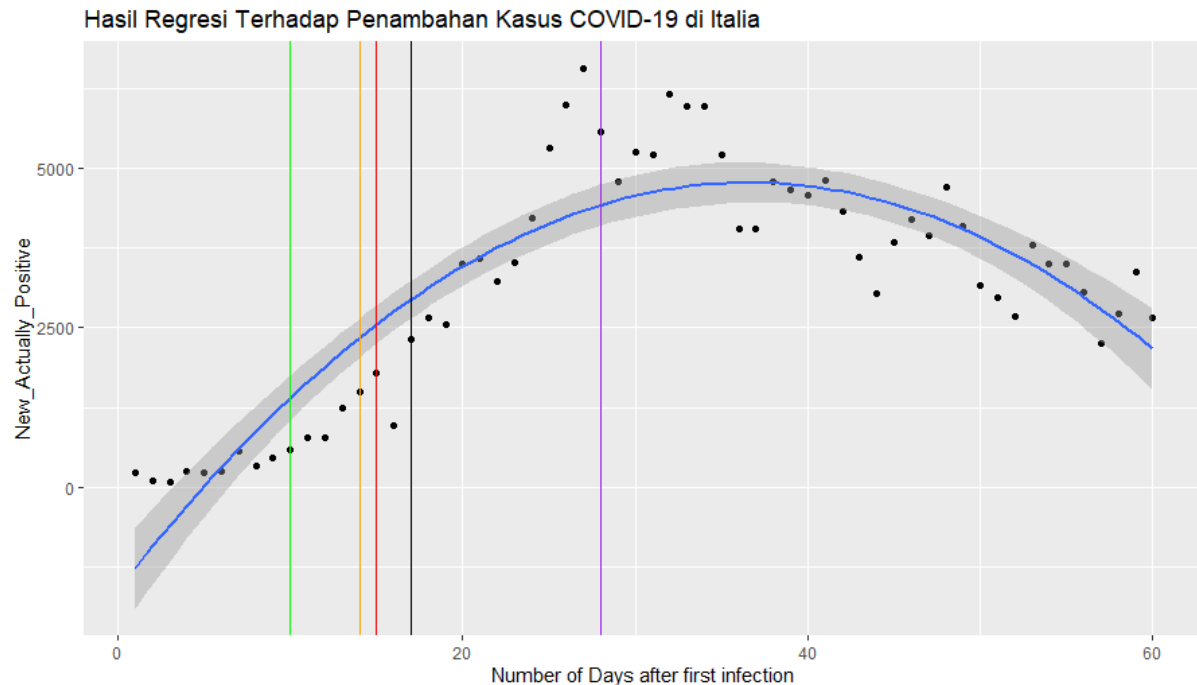
Berdasarkan model *logistic growth*, terlihat bahwa peningkatan total kasus COVID-19 hanya akan meningkat sampai suatu titik tertentu, kemudian laju peningkatan akan menurun secara drastis. Hal ini tentu menunjukkan pola yang lebih sesuai dengan sifat pandemi, dimana dalam jangka waktu yang lebih panjang, setiap kasus pandemi akan sampai pada salah satu diantara dua skenario berikut: 1) Tidak terdapat lagi orang yang dapat terinfeksi karena seluruh orang sudah pernah terinfeksi, atau 2) Ditemukan suatu vaksin / obat yang mampu mengobati penyakit tersebut. Terakhir, akan dilakukan perbandingan terhadap hasil prediksi menggunakan *time series regression* dan *logistic growth* seperti yang terlihat pada gambar berikut:



Gambar 21 Perbandingan hasil prediksi

Ketika membandingkan kedua hasil prediksi tersebut dalam satu grafik yang sama, terlihat bahwa baik hasil prediksi menggunakan *time series regression* maupun *logistic growth* sangat mungkin untuk terjadi. Hal ini tentu sangat tergantung pada beberapa faktor, antara lain populasi penduduk dari suatu negara, lama waktu sejak negara tersebut mengalami kasus infeksi pertama, regulasi dan

strategi yang diambil oleh pemerintah di negara tersebut, serta masih banyak faktor lainnya. *Deployment* model terakhir adalah model *polynomial regression* untuk memprediksi tren penambahan jumlah kasus positif COVID-19 setiap harinya. Penggunaan model *polynomial regression* ini sebenarnya hampir sama dengan kedua model sebelumnya, yaitu dapat digunakan sebagai alat untuk memprediksi tren penambahan jumlah kasus COVID-19 di Indonesia.



Gambar 22 Hasil regresi ditambah tanggal penting

Berdasarkan model regresi tersebut, terlihat bahwa jumlah penambahan kasus baru COVID-19 di Italia saat ini sedang mengalami *downtrend* yang artinya ada harapan bahwa penyebaran pandemi COVID-19 ini akan berhenti dalam waktu dekat.

Deployment Model untuk Pemerintah Indonesia

Model *machine learning* yang telah dibuat pada penelitian ini tentu dapat digunakan oleh pemerintah Indonesia untuk melakukan estimasi laju peningkatan penyebaran COVID-19 di Indonesia serta estimasi tanggal berakhirnya kasus COVID-19 di Indonesia. Proses *deployment* model untuk Indonesia tentu akan sangat mirip dengan proses *deployment* model untuk negara Italia. Hanya saja perlu dilakukan beberapa penyesuaian parameter agar *output* dari model tersebut dapat merepresentasikan data nyata laju peningkatan COVID-19 di Indonesia dengan lebih baik. Hasil prediksi ini kemudian dapat digunakan oleh pemerintah Indonesia, misalnya untuk melakukan proses pengambilan keputusan mengenai saat yang tepat untuk memperketat peraturan pembatasan sosial ataupun pengambilan keputusan *lockdown*. Pada akhirnya, seluruh model yang dibuat pada penelitian ini hanya merupakan suatu sistem yang mampu memberikan hasil prediksi berdasarkan data, untuk selanjutnya dianalisis dan dipertimbangkan lebih lanjut oleh *stakeholder* terkait.

Analisis dan Rekomendasi Keputusan untuk Pemerintah Indonesia

Berdasarkan hasil analisis terhadap data dan model yang telah dilaksanakan pada bagian sebelumnya, dibuat suatu rangkuman rekomendasi keputusan untuk pemerintah Indonesia dalam menghadapi wabah COVID-19 di Indonesia. Sebelum memberikan rekomendasi keputusan, berikut merupakan sedikit gambaran mengenai langkah yang diambil pemerintah Italia dalam menghadapi kasus COVID-19 ini, beserta dengan kode warna yang akan digunakan pada grafik visualisasi data:

4 Maret 2020 → Pemerintah menutup sekolah dan universitas (Hijau)

8 Maret 2020 → Pemerintah memberlakukan *partial lockdown* untuk beberapa *region*, salah satunya adalah Lombardy, *region* dengan kasus COVID-19 terbanyak di Italia (Oranye)

9 Maret 2020 → Pemerintah memberlakukan *lockdown national* (Merah)

11 Maret 2020 → Pemerintah menutup seluruh restoran dan bar (Hitam)

22 Maret 2020 → Pemerintah menutup pabrik dan pekerjaan non-esensial (Ungu)

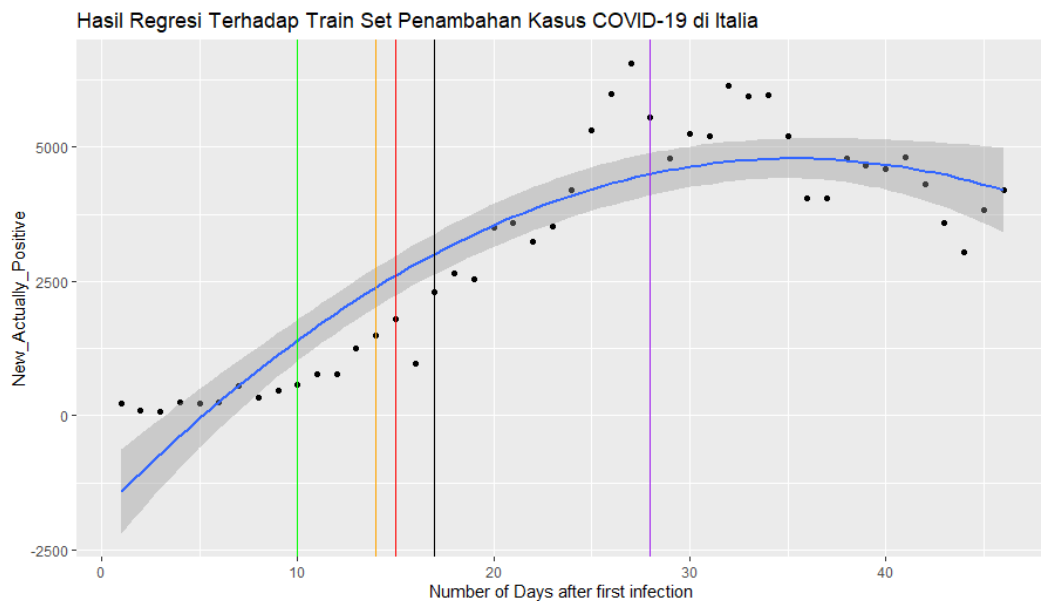
Berikut merupakan analisis terhadap strategi yang diambil oleh pemerintah Italia disertai dengan visualisasi data hasil analisis pada bagian sebelumnya, serta rekomendasi bagi pemerintah Indonesia:

1. Pemberlakuan *lockdown parsial* terbukti cukup efektif menekan penambahan kasus baru di Lombardy

Berdasarkan hasil visualisasi terhadap penambahan kasus baru COVID-19 di Italia pada skala regional, terlihat bahwa Lombardy merupakan daerah yang terkena dampak terparah dari wabah COVID-19 ini. Daerah Lombardy sendiri telah memberlakukan pembatasan sosial berskala besar jauh sebelum daerah ini melaksanakan *partial lockdown*. Daerah ini juga merupakan salah satu daerah pertama yang memberlakukan *partial lockdown*. Efek dari langkah cepat ini mungkin tidak terlihat pada pertengahan Maret hingga pertengahan April 2020, dimana terjadi puncak penambahan kasus baru penderita COVID-19. Namun, setelah daerah ini melewati masa puncak tersebut, penurunan jumlah penambahan kasus baru menurun dengan cukup cepat pada pertengahan hingga akhir April 2020. Berbeda dengan Lombardy, daerah lain seperti Piemonte baru mengalami peningkatan total kasus baru COVID-19 menjelang akhir bulan April 2020.

Pemerintah Indonesia juga dapat mempelajari strategi yang diambil pemerintah Italia ini. Lombardy dapat dianalogikan sebagai episentrum penyebaran wabah COVID-19, sama seperti kota Jakarta di Indonesia. Pemerintah Indonesia dapat melakukan analisis terhadap tren penambahan kasus baru COVID-19 di daerah Jakarta setiap harinya dan menggunakan data tersebut sebagai pertimbangan untuk melaksanakan *partial lockdown*. Proses ini tentu harus dibarengi dengan pertimbangan sosial, ekonomi, serta kesiapan pemerintah untuk memberikan bantuan bagi penduduknya.

2. Pemberlakuan *lockdown* oleh pemerintah Italia baru menghasilkan efek yang signifikan setelah kurang lebih 1 bulan



Pemberlakuan *lockdown* nasional yang dilakukan oleh pemerintah Italia pada awalnya terlihat tidak memperbaiki situasi karena penambahan kasus baru COVID-19 di Italia setiap harinya masih terus mengalami peningkatan. Tetapi, setelah *lockdown* nasional berjalan selama kurang lebih 1 bulan, barulah terlihat efek dari pemberlakuan *lockdown* tersebut, dimana terjadi penurunan jumlah penambahan kasus baru COVID-19 di Italia.

Pelajaran penting yang dapat diambil oleh pemerintah Indonesia adalah bahwa pengambilan keputusan untuk melakukan *lockdown* harus dilakukan sebelum tren penambahan jumlah kasus dan tren total kasus menunjukkan pola eksponensial. Hal ini dikarenakan pemerintah Italia yang memberlakukan *lockdown* di saat tren penambahan dan total kasus sudah menunjukkan pola eksponensial. Hal ini mengakibatkan memuncaknya jumlah kasus COVID-19 di Italia bahkan setelah diberlakukan *lockdown*. Apabila pemerintah Indonesia ingin mengikuti langkah pemerintah Italia untuk memberlakukan *lockdown*, keputusan tersebut harus diambil dengan cepat sebelum tren kenaikan kasus COVID-19 sudah menjadi tidak terkendali.

3. Negara Italia tidak melakukan pembatasan sosial berskala besar sebelum memberlakukan *lockdown*

Salah satu keunikan dari negara Italia dalam menghadapi wabah COVID-19 ini adalah bahwa pemerintah Italia tidak memberlakukan PSBB sebelum memberlakukan *lockdown*. Hal ini terbilang cukup unik karena pada umumnya, negara lain akan memilih untuk melakukan pembatasan sosial berskala besar, yang kemudian diikuti dengan *lockdown* apabila strategi pertama tersebut tidak berhasil. Hal ini kemungkinan juga menjadi salah satu alasan tingginya kasus penyebaran COVID-19 di Italia bahkan setelah diberlakukannya *lockdown* nasional.

Saat ini, pemerintah Indonesia sedang melaksanakan PSBB. Hal ini tentu merupakan suatu langkah yang cukup baik mengingat hal ini tidak dilakukan oleh negara Italia pada tahap awal penyebaran wabah COVID-19 di negara tersebut. Namun, pemerintah Indonesia juga harus terus mengamati efektivitas dari PSBB ini. Apabila setelah melaksanakan PSBB tren penambahan dan total kasus COVID-19 masih menunjukkan tren eksponensial, di saat itulah

pemerintah Indonesia perlu bersiap untuk melakukan *lockdown*, baik *lockdown* parsial maupun nasional.

4. Model prediksi total kasus COVID-19 dapat digunakan sesuai dengan kondisi Indonesia saat ini

Pada penelitian ini, telah dibentuk sebanyak 3 model *machine learning* yang dapat digunakan oleh pemerintah, sesuai dengan kebutuhan. Penggunaan model prediksi total kasus COVID-19 dapat dilakukan baik menggunakan *time series regression* maupun *logistic growth model*. Pemilihan model prediksi harus didasarkan pada tren dan pola data total kasus COVID-19 di Indonesia. Apabila pola data total kasus COVID-19 di Indonesia masih menunjukkan pola hubungan eksponensial, serta pelaksanaan prediksi dilakukan pada tahap awal penyebaran COVID-19, maka model prediksi *time series regression* lebih cocok untuk digunakan. Sebaliknya, apabila total kasus telah mendekati tahap akhir pandemi, yang juga dicirikan melalui penurunan jumlah penambahan kasus baru COVID-19 di Indonesia, maka dapat digunakan *logistic growth model*. Kedua model ini juga dapat diibaratkan sebagai prediksi "optimis" (*logistic growth*) dan "pesimis" (*time series regression*).

5. Model prediksi penambahan kasus baru COVID-19 setiap harinya dapat digunakan untuk memprediksi akhir dari pandemi

Model *polynomial regression* yang digunakan pada dataset penambahan kasus baru COVID-19 di Italia setiap harinya juga dapat digunakan pada dataset penambahan kasus baru COVID-19 di Indonesia untuk memprediksi tanggal wabah COVID-19 ini akan berakhir. Data ini dapat digunakan oleh pemerintah untuk menentukan strategi ekonomi dan sosial yang tepat untuk dapat menyokong negara hingga pandemi ini berakhir.

6. Pentingnya persiapan rumah sakit dan IGD

Peningkatan jumlah kasus positif COVID-19 tentu akan berpengaruh pada jumlah pasien yang membutuhkan fasilitas rumah sakit dan IGD. Pemerintah Indonesia harus bersiap untuk menghadapi hal tersebut terutama pada masa puncak penyebaran COVID-19 (saat peningkatan total kasus menunjukkan pola eksponensial).

Daftar Pustaka

Overgoor, G. (2019). *Letting The Computers Take Over: Using AI to Solve Marketing Problems*. California: California Management Review.

Fleming, S. (n.d.). What's a logarithmic graph and how does it help explain the spread of COVID-19? Retrieved from <https://www.weforum.org/agenda/2020/04/covid-19-spread-logarithmic-graph/>

Mansoor, R. (n.d.). Growth Curves and an Exponential Pandemic. Retrieved from <https://www.hadean.com/blog/growth-curves-and-an-exponential-pandemic>