

## **Analisis Efektivitas Social Distancing dan Pengaruh Suhu terhadap Tingkat Penyebaran Virus COVID-19**

Rahayuning Hardatin (G64170011), Rachmat Ghaly (G74170019), Efrad Galio (G74170052), Imam Muhajir (G74170064)<sup>1\*</sup>

Kelompok: 12, Kelas Paralel: 1

### **Abstrak**

Penyakit virus corona atau selanjutnya dikenal dengan nama COVID-19 merupakan penyakit yang memiliki kemampuan menular yang sangat cepat pada saat ini. Hal tersebut menyebabkan pemerintah setiap negara harus membuat kebijakan secara cepat dan tepat untuk mencegah penyebaran COVID-19. Salah satu kebijakan yang dilakukan adalah diberlakukannya *social distancing* (pembatasan sosial), di Indonesia diterapkan Pembatasan Sosial Berskala Besar (PSBB). Dalam penelitian kali ini ingin diketahui efektivitas dari pelaksanaan *social distancing* dan pengaruh suhu terhadap tingkat penyebaran virus COVID-19. Metode yang digunakan pada penelitian kali ini adalah *Decision Tree Regression* yang dapat digunakan untuk memprediksi angka *social distancing*. Didapatkan prediksi efektivitas *social distancing* dan hubungan antara suhu dengan tingkat penyebaran kasus COVID-19.

Kata Kunci: covid-19, *decision tree regression*, PSBB, *social distancing*, suhu

---

<sup>1</sup>Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor 16680

\*Mahasiswa Program Studi Ilmu Komputer, FMIPA-IPB; Surel: [rahayuning\\_hardatin@apps.ipb.ac.id](mailto:rahayuning_hardatin@apps.ipb.ac.id), [ghaly\\_cool97@apps.ipb.ac.id](mailto:ghaly_cool97@apps.ipb.ac.id), [efrad\\_fraqs@apps.ipb.ac.id](mailto:efrad_fraqs@apps.ipb.ac.id), [muhajir\\_29@apps.ipb.ac.id](mailto:muhajir_29@apps.ipb.ac.id)

# PENDAHULUAN

## Latar Belakang

Pada bulan Desember 2019, sejumlah pasien didiagnosa mengalami pneumonia yang disebabkan oleh makanan laut yang dijual di pasar grosir, Wuhan, China. Kejadian ini menuntun penemuan suatu virus baru yang berasal dari keluarga yang sama dengan SARS dan MERS. Virus baru ini dinamakan virus corona, kemudian penyakit yang disebabkan oleh virus tersebut dikenal dengan nama COVID-19. Gejala yang dapat dialami saat terkena virus tersebut yaitu demam, batuk, dan napas yang pendek. *The Center for Disease Control and Prevention* (CDC) percaya bahwa pasien Virus Corona dapat mengalami gejala-gejala ini dari 2 hari sampai 14 hari setelah terpapar virusnya.

Penyebaran COVID-19 dapat melalui droplet (tetesan kecil) yang dihasilkan saat orang yang terinfeksi batuk, bersin, atau menghembuskan napas. Sehingga orang dapat tertular saat menghirup udara yang mengandung virus atau saat menyentuh permukaan benda yang terkontaminasi yang kemudian diikuti tindakan menyentuh mata, hidung, atau mulut. Virus COVID-19 memiliki kemampuan menular yang sangat cepat dan telah menyebar ke hampir semua negara, termasuk Indonesia, hanya dalam waktu beberapa bulan. Hal tersebut menyebabkan pemerintah setiap negara harus membuat kebijakan secara cepat dan tepat. Beberapa negara menerapkan sistem *lockdown* atau pembatasan sosial untuk mencegah penyebaran COVID-19, sedangkan di Indonesia menerapkan sistem Pembatasan Sosial Berskala Besar (PSBB). Tingkat laju penyebaran virus COVID-19 ditentukan oleh banyak faktor, seperti penerapan sistem PSBB yang tidak disiplin atau mungkin faktor geografis suhu untuk setiap negara.

## Tujuan

Tujuan dari penelitian ini adalah untuk menganalisis Efektivitas Social Distancing dan Pengaruh Suhu terhadap Tingkat Penyebaran Virus COVID-19.

## Ruang Lingkup

Data yang digunakan adalah dataset kasus COVID-19 yang diperoleh dari *kaggle* dan *Imperial College* yang selanjutnya akan digunakan untuk pemodelan analisis efektivitas social distancing dengan menggunakan teknik *decision tree regression* serta akan digunakan untuk menganalisis hubungan tingkat penyebaran kasus COVID-19 dengan keadaan suhu wilayah. Informasi dari dataset yang digunakan total keseluruhan kasus sampai tanggal 25 Mei 2020. Kemudian dilakukan scraping untuk data suhu dengan menggunakan BeautifulSoup library. Data suhu yang digunakan diambil dari Wikipedia.

## Manfaat

Manfaat yang diharapkan dari penelitian ini adalah sebagai salah satu bahan pertimbangan dalam pengambilan tindakan untuk mencegah penyebaran COVID-19.

## TINJAUAN PUSTAKA

### Decision Tree Regression

Decision Tree adalah model pembelajaran yang digunakan untuk memprediksi target dengan mempelajari aturan keputusan dari fitur. Model ini dapat diartikan sebagai pembedahan data dengan membuat keputusan berdasarkan pada serangkaian pertanyaan. Algoritma decision-tree termasuk ke dalam kategori algoritma *supervised learning*. Decision-tree bekerja untuk variabel output kontinu maupun kategoris. Decision tree regression mengamati fitur dari suatu objek dan melatih suatu model dalam struktur pohon untuk memprediksi data di masa depan untuk menghasilkan output kontinu yang bermakna. Output kontinu berarti bahwa output / hasil tidak diskrit, yaitu, itu tidak diwakili hanya oleh seperangkat angka atau nilai yang dikenal.

### R0 dan Rt

Reproduction Number (R0) adalah angka laju penyebaran penyakit, dapat juga diartikan sebagai angka reproduksi virus atau jumlah rata-rata orang bisa menularkan penyakit ke orang lain. Dari angka tersebut maka dapat diketahui tingkat dan kemampuan penyebaran suatu penyakit. Angka R berbeda-beda untuk setiap kasus. Sebagai contoh, angka R sama dengan 2 berarti satu orang dapat menularkan virus ke 2 orang lain. Hingga saat ini belum ada angka reproduksi pasti untuk Covid-19. Estimasi terbaiknya saat ini adalah R0 untuk Covid-19 di 2 sampai 2,5, lebih tinggi dua kali lipat dari flu biasa. Besar angka R0 di Indonesia adalah 2.5 (Sorta Tobing, 2020).

Sedangkan Rt adalah angka reproduksi setelah adanya intervensi pemerintah, intervensi dapat berupa PSBB (Pembatasan Sosial Berskala Besar). Angka Rt yang ideal adalah di bawah 1. Para ahli biasanya menghitung ke belakang, bukan ketika orang baru saja tertular. Data yang diambil adalah jumlah orang meninggal, masuk rumah sakit, atau positif Covid-19. R memberikan gambaran penularan sekitar dua sampai tiga pekan sebelumnya. R lebih dari satu berarti berbahaya karena penyebaran sebuah virus seperti efek bola salju, bisa membesar dengan cepat. Kalau angkanya di bawah satu, artinya penyakit tersebut akan menghilang dan orang tidak lagi tertular.

## METODE

### Data

Terdapat tiga dataset yang akan digunakan yaitu diperoleh dari *kaggle*, *Imperial College*, dan *Wikipedia*. Dataset yang diperoleh dari *kaggle* adalah *covid\_19\_clean\_complete.csv*. Dataset ini terdiri dari 27192 baris dan 8 variabel.

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
0	NaN	Afghanistan	33.0000	65.0000	2020-01-22	0	0	0
1	NaN	Albania	41.1533	20.1683	2020-01-22	0	0	0
2	NaN	Algeria	28.0339	1.6596	2020-01-22	0	0	0
3	NaN	Andorra	42.5063	1.5218	2020-01-22	0	0	0
4	NaN	Angola	-11.2027	17.8739	2020-01-22	0	0	0

Gambar 1 Tampilan data *covid\_19\_clean\_complete.csv* untuk 5 baris pertama

Dataset yang diperoleh dari *Imperial College* adalah *mitigation\_type.csv*. Dataset ini merupakan data simulasi untuk semua negara yang dibuat berdasarkan *profile China* ketika pandemi COVID-19 di China telah berakhir. Sehingga, dapat dikatakan dataset ini menggambarkan simulasi untuk setiap negara ketika virus COVID-19 telah berakhir. Dataset ini terdiri dari 2412 baris dan 9 variabel.

	Country	R0	Strategy	Social_distance	total_pop	total_infected	total_deaths	total_hospital	total_critical
1008	Indonesia	2.4	Enhanced social distancing of elderly	34%	273523621	145903545	419909	3635403	556681
1009	Indonesia	2.7	Enhanced social distancing of elderly	38%	273523621	158895807	478999	4020301	635202
1010	Indonesia	3.0	Enhanced social distancing of elderly	43%	273523621	167623842	653804	4659004	866911
1011	Indonesia	3.3	Enhanced social distancing of elderly	47%	273523621	175913029	711835	4959702	943646
1012	Indonesia	2.4	Social distancing whole population	37%	273523621	145288435	652399	4283278	864896

Gambar 2 Tampilan data mitigasi setiap negara dalam upaya penanggulangan wabah COVID-19 untuk 5 baris pertama.

Dataset yang diperoleh dari *Wikipedia* adalah *Avg\_world\_temp.csv*. Penulis melakukan *scrapping* data pada *Wikipedia* dengan menggunakan *library BeautifulSoup*. Dataset yang didapatkan terdiri dari 424 baris dan 16 variabel. Variabel bulan menggambarkan suhu rata-rata dari tahun ke tahun untuk setiap negara di setiap kota.

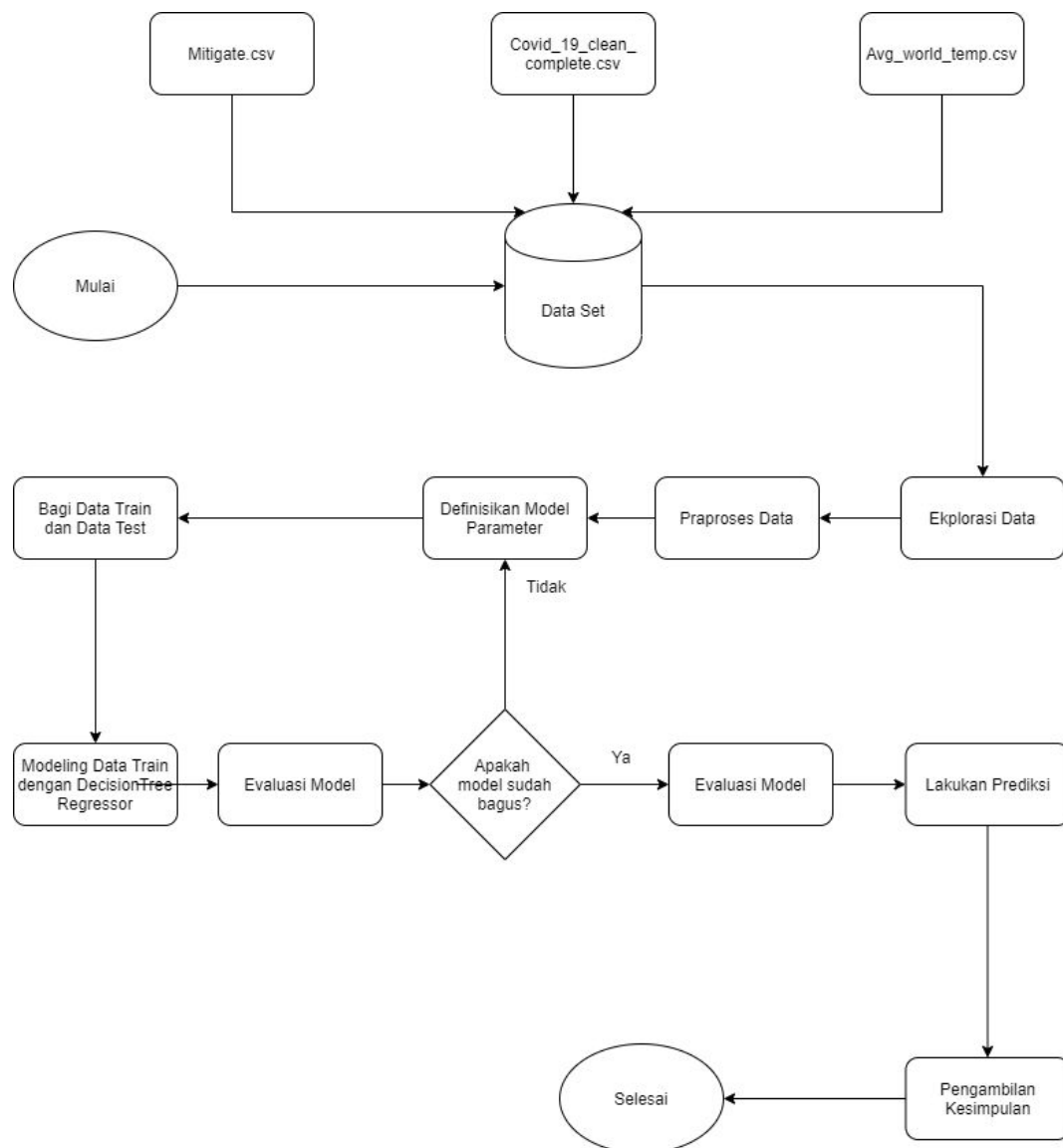
	Country	City	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg_Year	Continent
0	Algeria	Algiers	11.2	11.9	12.8	14.7	17.7	21.3	24.6	25.2	23.2	19.4	15.2	12.1	17.4	Africa
1	Algeria	Tamanrasset	12.8	15.0	18.1	22.2	26.1	28.9	28.7	28.2	26.5	22.4	17.3	13.9	21.7	Africa
2	Algeria	Reggane	16.0	18.2	23.1	27.9	32.2	36.4	39.8	38.4	35.5	29.2	22.0	17.8	28.3	Africa
3	Angola	Luanda	26.7	28.5	28.6	28.2	27.0	23.9	22.1	22.1	23.5	25.2	26.7	26.9	25.8	Africa
4	Benin	Cotonou	27.3	28.5	28.9	28.6	27.8	26.5	25.8	25.6	26.0	26.7	27.6	27.3	27.2	Africa

Gambar 3 Data suhu rata-rata setiap negara berdasarkan wikipedia setelah dilakukan *scrapping*

Data tambahan yang kami gunakan adalah data total penduduk, total rumah sakit, dan angka R0 di Indonesia. Total penduduk Indonesia adalah sebanyak 273,277,935 (Worldometers, 2020). Sedangkan total rumah sakit di Indonesia adalah sebanyak 2,831 (Wikipedia, 2020).

## Tahapan Kegiatan

Tahapan kegiatan yang dilakukan antara lain pengumpulan dataset, eksplorasi data, praproses data, modeling, evaluasi model, prediksi, kemudian pengambilan kesimpulan. Pada tahap pra-proses data dilakukan *encoding* pada beberapa tipe variabel, agregasi antara dataset serta dilakukan beberapa penambahan dan penghapusan variabel berdasarkan pemodelan yang dirancang. Alur tahapan penambangan data pada penelitian dalam bentuk desain implementasi dapat dilihat pada Gambar 4.



Gambar 4 Alur tahapan penambangan data pada penelitian

### Lingkungan Pengembangan

Perangkat keras yang digunakan dalam penelitian ini adalah komputer personal dengan spesifikasi sebagai berikut :

- a. Prosesor : Intel Core i3
- b. Memory : 4GB

Sedangkan perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut :

- a. Bahasa Pemrograman Python dan R
- b. Platform Kaggle sebagai media pengolah data tambahan, preprocessing, modelling, dan prediksi.

## HASIL DAN PEMBAHASAN

### Praproses Data

Variabel yang akan digunakan pada dataset *covid\_19\_clean\_complete.csv* adalah Country/Region, Date, Confirmed, Deaths, dan Recovered. Untuk menggambarkan kasus yang sedang aktif, dibuat variabel baru yaitu Active. Rumus yang digunakan untuk pembuatan variabel Active dapat dilihat pada persamaan 1.

$$Active = Confirmed - Deaths - Recovered \quad (1)$$

	Country/Region	Date	Confirmed	Deaths	Recovered	Active
0	Afghanistan	2020-01-22	0	0	0	0
1	Albania	2020-01-22	0	0	0	0
2	Algeria	2020-01-22	0	0	0	0
3	Andorra	2020-01-22	0	0	0	0
4	Angola	2020-01-22	0	0	0	0

Gambar 5 Hasil Praproses dataset *covid\_19\_clean\_complete.csv*

Untuk dataset *mitigation\_type.csv* variabel yang tidak digunakan adalah *total\_critical*. Dikarenakan variabel Strategy merupakan tipe data ordinal maka dilakukan *encoding* terhadap variabel Strategy. Dimana nilai 'Unmitigated' akan berubah menjadi 0, nilai 'Enhanced social distancing of elderly' menjadi 1, dan nilai 'Social distancing whole population' menjadi 2. Lalu dilakukan perubahan nilai pada variabel Social\_Distance dari persentase menjadi integer agar dapat diolah.

	Country	R0	Strategy	total_pop	total_infected	total_deaths	total_hospital	Social_Distance
0	Afghanistan	2.4	1	38928341	22235810	42015	342532	34
1	Afghanistan	2.7	1	38928341	24001697	46456	371943	38
2	Afghanistan	3.0	1	38928341	25058690	57529	414970	44
3	Afghanistan	3.3	1	38928341	26214010	60640	434649	47
4	Afghanistan	2.4	2	38928341	21576989	55929	378738	37

Gambar 6 Hasil Praproses dataset *mitigation\_type.csv*

Dataset terakhir, yaitu *Avg\_world\_temp.csv* variabel yang akan digunakan adalah Country, City, Jan, Feb, Mar, Apr, May, dan Continent. Untuk menentukan suhu rata-rata di setiap negaranya, dilakukan pembuatan variabel baru, yaitu dengan melakukan agregasi rata-rata suhu untuk seluruh kota di setiap negaranya.

	Country	City	Jan	Feb	Mar	Apr	May	Avg_temp	Continent
0	Algeria	Algiers	11.2	11.9	12.8	14.7	17.7	18.660000	Africa
1	Algeria	Tamanrasset	12.8	15.0	18.1	22.2	26.1	18.660000	Africa
2	Algeria	Reggane	16.0	18.2	23.1	27.9	32.2	18.660000	Africa
3	Angola	Luanda	26.7	28.5	28.6	28.2	27.0	27.800000	Africa
4	Benin	Cotonou	27.3	28.5	28.9	28.6	27.8	28.566667	Africa

Gambar 7 Hasil Praproses dataset *Avg\_world\_temp.csv*

## Eksplorasi Data

Tahapan awal Eksplorasi Data dilakukan terhadap dataset *covid\_19\_clean\_complete.csv*. Sampai pada tanggal 25 Mei 2020, telah tercatat total kasus konfirmasi adalah sebanyak 5.495.055, total kematian sebanyak 346.232, total kesembuhan sebanyak 2.174.434, dan total yang aktif sebanyak 2.974.389.

Tabel 1 Jumlah total keseluruhan kasus COVID-19 sampai pada tanggal 25 Mei 2020

Date	Confirmed	Deaths	Recovered	Active
25 Mei 2020	5.495.055	346.232	2.174.434	2.974.389

Tingkat penyebaran kasus COVID-19 yang terjadi pada setiap negara berbeda-beda. Negara yang memiliki kasus terkonfirmasi terbanyak adalah US. Terdapat perbedaan yang besar antara US dan Brazil dibandingkan perbedaan antara Brazil dan Rusia seperti ditunjukkan pada Gambar 8.

	Country/Region	Confirmed	Deaths	Recovered	Active
0	US	1662302	98220	379157	1184925
1	Brazil	374898	23473	153833	197592
2	Russia	353427	3633	118798	230996
3	United Kingdom	262547	36996	1161	224390
4	Spain	235400	26834	150376	58190

Gambar 8 Lima negara yang memiliki kasus konfirmasi terbanyak.

Untuk melihat negara yang memiliki tingkat kematian dan tingkat kesembuhan yang paling tinggi dilakukan pembuatan variabel baru:

1. Ratio\_Deaths merupakan variabel yang menggambarkan persentase kematian akibat dari virus COVID-19.

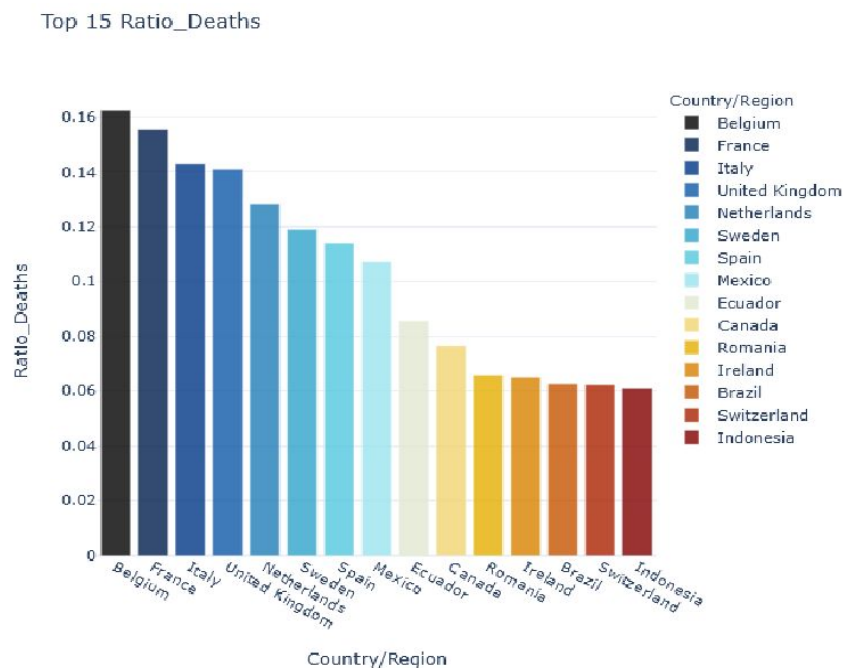
$$\text{Ratio\_Deaths} = \frac{\text{Deaths}}{\text{Confirmed}} \quad (2)$$

2. Ratio\_Recovered merupakan variabel yang menggambarkan persentase kesembuhan dari virus COVID-19.

$$\text{Ratio\_Recovered} = \frac{\text{Recovered}}{\text{Confirmed}} \quad (3)$$

3. Best\_Ratio merupakan variabel yang mengukur negara manakah yang memiliki Ratio\_Deaths dan Ratio\_Recovered terbaik.

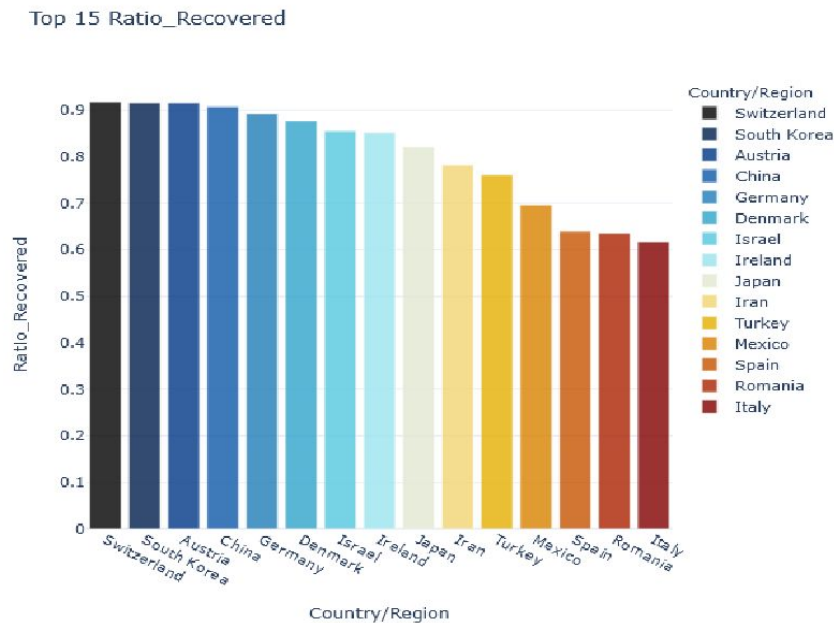
$$\text{Best\_Ratio} = \text{Ratio\_Deaths} - \text{Ratio\_Recovered} \quad (4)$$



Gambar 9 Lima belas negara dengan tingkat Kematian tertinggi.

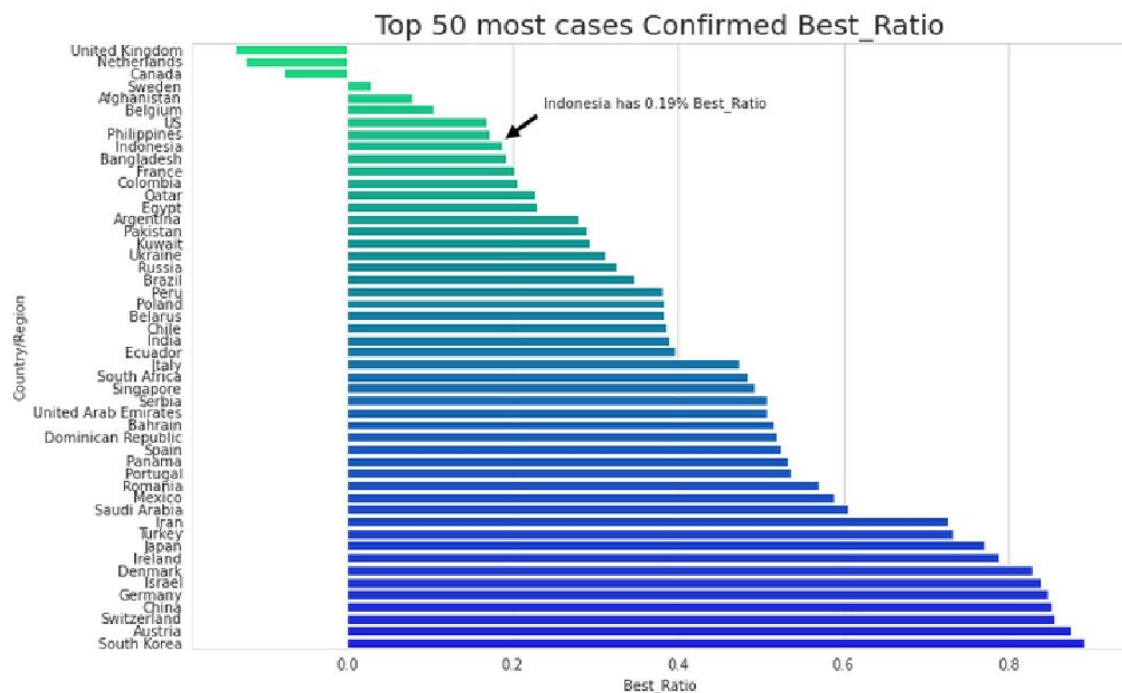


Dari Bar plot diatas, pada Gambar 9, Belgium merupakan negara yang memiliki rasio kematian tertinggi yaitu sekitar 0.16, disusul oleh France dengan ratio kematian sekitar 0.15. Indonesia menempati urutan ke-15 dengan nilai rasio kematian sebesar 0.06. Hal ini menandakan bahwa Indonesia merupakan negara yang memiliki tingkat kematian yang paling tinggi di ASEAN bahkan ASIA.



Gambar 10 Lima belas negara dengan tingkat Kesembuhan tertinggi.

Sedangkan untuk rasio kesembuhan ditunjukkan pada Gambar 10, posisi pertama diduduki oleh Switzerland dan disusul oleh Korea Selatan dan Austria. Ketiga negara tersebut memiliki rasio kesembuhan sekitar 0.9. Sedangkan China sendiri memiliki rasio kesembuhan sekitar 0.89.



Gambar 11 Ratio terbaik untuk 50 Negara dengan kasus konfirmasi terbanyak.

Dapat dilihat pada Bar Plot diatas, pada Gamabr 11, bahwa Best\_Ratio paling terbaik diduduki oleh Korea Selatan dan disusul oleh Austria dan Switzerland. Hal ini berarti tingkat kematian Switzerland masih lebih tinggi dibandingkan Korea Selatan dan Austria. Korea Selatan merupakan negara yang memiliki tingkat kesembuhan yang paling tinggi dan tingkat kematian yang paling rendah.

Terdapat tiga negara yang memiliki nilai Best\_Ratio negatif, yaitu United Kingdom, Netherlands, dan Canada. Hal ini berarti tingkat kematian untuk ketiga negara tersebut masih lebih tinggi dibandingkan tingkat kesembuhan. Indonesia sendiri berada pada peringkat ke-9 dengan nilai Best\_Ratio sebesar 0.19.

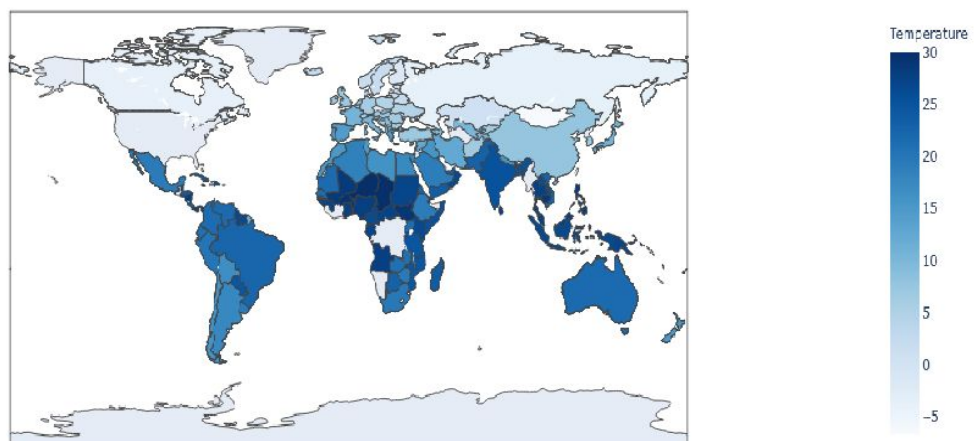
Tahapan selanjutnya adalah analisis suhu dengan tingkat penyebaran virus COVID-19. Sebelum dilakukan analisis diperlukan *merging* data terhadap dataset *covid\_19\_clean\_complete.csv* dan *Avg\_world\_temp.csv*.

	Confirmed	Deaths	Recovered	Active	Country	Avg_temp	Continent
0	17267	294	1522	15451	Afghanistan	6.68	Asia
1	1184	33	898	253	Albania	11.18	Europe
2	9733	673	6218	2842	Algeria	18.66	Africa
3	9733	673	6218	2842	Algeria	18.66	Africa
4	9733	673	6218	2842	Algeria	18.66	Africa

Gambar 12 Hasil Merging dataset.

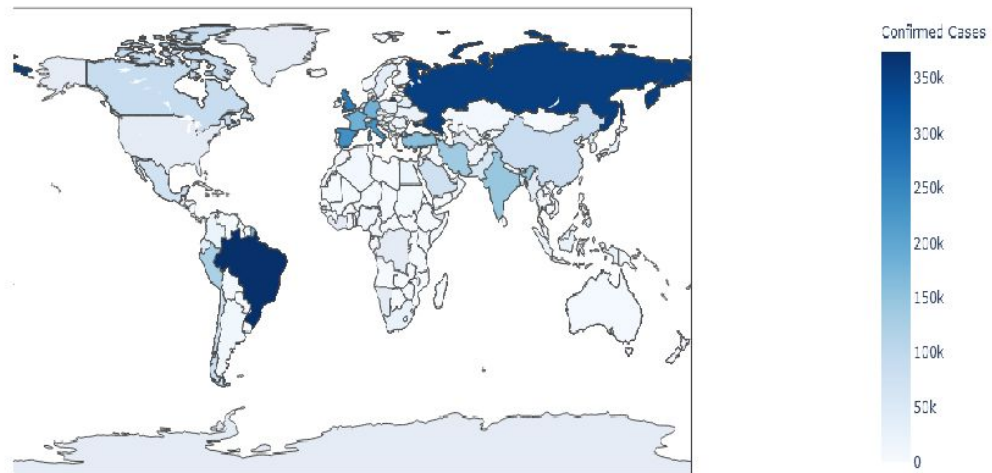
Setelah dilakukan pembuatan dataset *merging*, dilakukan visualisasi spasial untuk variabel Avg\_temp dan Confirmed. Hasil visualisasi suhu ditunjukkan pada Gambar 13 dan visualisasi konfirmasi kasus ditunjukkan pada Gambar 14.

Country vs Temperature



Gambar 13 Visualisasi Spasial Suhu untuk seluruh negara.

Country vs Confirmed

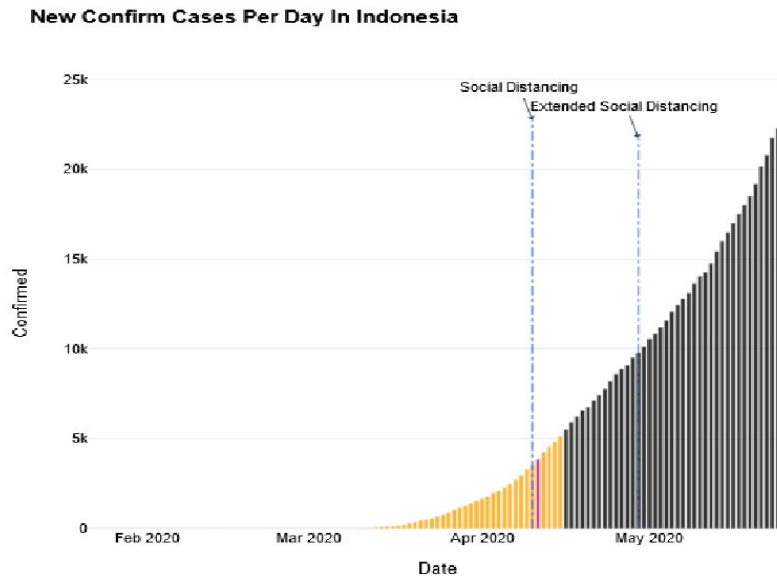


Gambar 14 Visualisasi Spasial Jumlah Konfirmasi untuk seluruh negara.

Kedua Gambar diatas merupakan visualisasi spasial untuk variabel *Avg\_temp* dan *Confirmed*. Pada visualisasi spasial, yaitu Country vs Temperature negara yang memiliki warna biru muda memiliki suhu rata-rata yang rendah, sedangkan untuk negara yang memiliki warna biru tua memiliki suhu rata-rata yang tinggi. Sedangkan pada visualisasi spasial Country vs Confirmed negara yang memiliki warna biru muda memiliki jumlah konfirmasi yang rendah, sedangkan untuk negara yang memiliki warna biru tua memiliki jumlah konfirmasi yang tinggi.

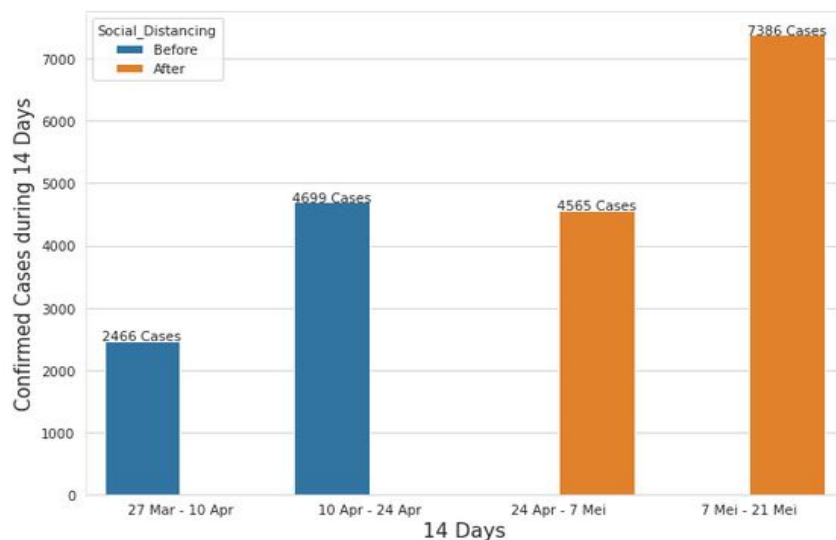
Jika kedua gambar diatas dibandingkan dapat dilihat dibagian Rusia untuk visualisasi spasial Country vs Temperature memiliki warna biru muda sedangkan untuk visualisasi spasial Country vs Confirmed memiliki warna biru tua. Hal ini juga berlaku untuk negara-negara yang berada di kawasan Asia, Afrika, dan Australia. Dengan demikian, dapat diambil kesimpulan bahwa negara yang memiliki suhu yang tinggi cenderung memiliki jumlah konfirmasi yang rendah dan negara yang memiliki suhu yang rendah cenderung memiliki jumlah konfirmasi yang tinggi.

Pembatasan Sosial Berskala Besar (PSBB) di Indonesia merupakan salah satu tindakan pemerintah untuk mengurangi laju persebaran Covid-19. Kebijakan tersebut dilakukan dari tanggal 10 April 2020 dan diperpanjang pada tanggal 23 April 2020.



Gambar 15 Bar plot untuk Jumlah Konfirmasi di Indonesia

Pada grafik yang ditunjukkan pada Gambar 15, terlihat tidak ada penurunan laju peningkatan Covid-19 setelah diterapkannya PSBB. Jumlah konfirmasi terus meningkat secara eksponensial, namun hal tersebut belum dapat digunakan untuk menyimpulkan bahwa PSBB yang dilakukan tidak efektif sama sekali. Diperlukan analisis lebih mendalam untuk mengetahui seberapa efektif PSBB yang telah diterapkan.



Gambar 16 Sebelum dan Sesudah PSBB

Pada Gambar 16 ditampilkan perbandingan jumlah konfirmasi kasus COVID-19 untuk sebelum dan sesudah diberlakukannya PSBB. Dikarenakan penerapan PSBB diterapkan dari tanggal 10 April dan klarifikasi virus COVID-19 pada seseorang membutuhkan waktu 14 hari, maka dari tanggal 10 April sampai dengan 24 April ditetapkan sebagai selang waktu untuk kasus yang terinfeksi sebelum penerapan PSBB

dilakukan. Sedangkan 24 April sampai dengan 7 Mei merupakan selang waktu untuk kasus yang terinfeksi setelah penerapan PSBB.

Dapat dilihat sebelum penerapan PSBB dilakukan, yaitu pada selang waktu 10 April sampai dengan 24 April kasus yang terkonfirmasi adalah sebanyak 4699 kasus. Sedangkan ketika penerapan PSBB dilakukan, yaitu pada selang waktu 24 April sampai dengan 7 Mei kasus yang terkonfirmasi adalah sebanyak 4565 kasus. Terlihat terjadi penurunan sebelum dan sesudah penerapan PSBB sebesar 134 kasus, namun empat belas hari kemudian meningkat secara eksponensial, yaitu sebesar 7386 kasus.

## Modeling

Cara lain yang digunakan untuk mengetahui seberapa efektivitas PSBB yang dilakukan adalah dengan melakukan prediksi. Data yang akan digunakan untuk menjadi data training adalah dataset *mitigation\_type.csv*

	Country	R0	Strategy	total_pop	total_infected	total_deaths	total_hospital	Social_Distance
0	Afghanistan	2.4	1	38928341	22235810	42015	342532	34
1	Afghanistan	2.7	1	38928341	24001697	46456	371943	38
2	Afghanistan	3.0	1	38928341	25058690	57529	414970	44
3	Afghanistan	3.3	1	38928341	26214010	60640	434649	47
4	Afghanistan	2.4	2	38928341	21576989	55929	378738	37

Gambar 17 Data training: *mitigation\_type.csv*

Variabel X adalah R0, Strategy, total\_pop, total\_infected, total\_deaths, dan total\_hospital. Sedangkan untuk variabel y adalah Social\_Distance.

Sebelum melakukan prediksi dilakukan pengecekan performansi model. Cross-Validation yang dilakukan menggunakan KFold, dimana fold yang dipakai sebesar 2 fold. Metric yang digunakan untuk mengukur performansi model adalah RMSE dan model yang digunakan adalah DecisionTreeRegressor atau Decision Tree Regression. Hasil RMSE didapatkan sebesar 28.57.

Setelah pengukuran performansi model, dilakukan pembuatan data test yang akan di prediksi dengan kondisi Indonesia saat ini, yaitu 25 Mei 2020.

	Country	R0	Strategy	total_pop	total_infected	total_deaths	total_hospital
0	Indonesia	2.5	Social distancing whole population	273277935	28233	1698	2831

Gambar 18 Data test

Hasil prediksi yang diperoleh ada 33%. Sehingga dapat dikatakan bahwa efektivitas PSBB yang dilakukan di Indonesia untuk kondisi saat ini adalah sebesar 33%.

## KESIMPULAN

Suhu dan tingkat perkembangan virus memiliki hubungan yang berbanding terbalik. Negara yang memiliki tingkat kesembuhan paling baik dan tingkat kematian yang paling rendah adalah Korea Utara yaitu sebesar 90%. Terdapat tiga negara yang memiliki tingkat kematian yang lebih besar dari tingkat kesembuhan, yaitu United Kingdom, Canada, dan Netherland. Tingkat efektifitas dari Social Distancing di Indonesia masih sangat kecil.

## SARAN

Nilai RMSE yang didapatkan masih sangat besar, sehingga masih bisa ditingkatkan dengan memilih model lain atau membuat feature engineering. Suhu rata-rata untuk setiap negara masih direpresentasikan oleh beberapa kota saja, diperlukan data suhu untuk seluruh kota di setiap negaranya dan diperlukan data jumlah konfirmasi untuk setiap provinsi di setiap negara sehingga dapat dilakukan analisis hubungan antara jumlah konfirmasi dengan suhu secara lebih spesifik.

## DAFTAR PUSTAKA

- Devakumar. 2020. *COVID-19 Dataset: Number of Confirmed, Death and Recovered cases every day across the glob* [Internet]. [diunduh 2020 Mar 12] . Tersedia pada:<https://www.kaggle.com/imdevskp/corona-virus-report> .
- Patrick GT Walker et al. 2020. *The Global Impact of COVID-19 and Strategies for Mitigation and Suppression* [Internet]. [diunduh 2020 Mar 12]. Tersedia pada:<https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-Global-Impact-26-03-2020v2.pdf> .
- Tobing, Sorta. 2020. Rumus R0 dan Rt yang Jadi Acuan Jokowi untuk Masuki Fase New Normal [Internet]. [diakses 2020 Mei 20]. Tersedia pada:<https://katadata.co.id/berita/2020/05/28/rumus-r0-dan-rt-yang-jadi-acuan-jokowi-untuk-masuki-fase-new-normal> .
- Wikipedia. 2020. *List of cities by average temperature* [Internet]. [diakses 2020 Mei 20]. Tersedia pada:[https://en.wikipedia.org/wiki/List\\_of\\_cities\\_by\\_average\\_temperature](https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature).
- Wikipedia. 2020. Daftar rumah sakit di Indonesia [Internet]. [diakses 2020 Mei 20]. Tersedia pada:[https://id.m.wikipedia.org/wiki/Daftar\\_rumah\\_sakit\\_di\\_Indonesia](https://id.m.wikipedia.org/wiki/Daftar_rumah_sakit_di_Indonesia).
- Worldometers. 2020. Indonesia Population [Internet]. [diakses 2020 Mei 20]. Tersedia pada:<https://www.worldometers.info/world-population/indonesia-population/>.