

BAB II TINJAUAN PUSTAKA

2.1 COVID-19

Pada 2019, virus corona baru teridentifikasi sebagai pemicu wabah penyakit yang berasal dari Cina. Oleh sebab itu pada bulan Maret 2020 WHO (*World Health Organization*) secara resmi mengumumkan bahwa COVID-19 sebagai pandemi di seluruh dunia. COVID dapat menyebar karena mengandung virus yang dapat menginfeksi organ pernafasan dengan cepat yaitu *Severe Acute Respiratory syndrome Coronavirus-2* [10]. *Coronavirus* merupakan bagian virus yang dapat mengakibatkan penyakit semacam flu biasa, infeksi pada saluran *respiratory* (pernapasan) yang parah atau SARS (*Severe acute respiratory syndrome*), serta penyakit MERS (*Middle East respiratory syndrome*) [10] [11].

Gejala dan tanda-tanda penyakit *coronavirus* 2019 (COVID-19) mungkin muncul dua hingga 14 hari setelah terpapar. Setiap pasien yang terpapar COVID-19 memiliki gejala yang berbeda untuk gejala ringan seperti demam, sakit kepala, diare, nyeri dada, pilek, sakit tenggorokan, nyeri otot dan kelelahan [10]. Sedangkan untuk orang tua memiliki risiko yang sangat tinggi terkena COVID-19. Orang yang memiliki riwayat kesehatan yang tidak baik juga mungkin memiliki risiko lebih tinggi terkena penyakit serius akibat COVID-19 seperti penyakit jantung, kanker, diabetes, obesitas, perokok, penyakit ginjal kronis [10]. Kondisi lain dapat meningkatkan risiko penyakit serius, seperti asma, penyakit hati, kegemukan, penyakit paru-paru kronis seperti fibrosis kistik atau fibrosis paru, kondisi otak dan sistem saraf, sistem kekebalan yang melemah dari transplantasi sumsum tulang, HIV atau beberapa obat, diabetes tipe 1, dan tekanan darah tinggi [10].

Sejak pandemi COVID-19 masuk ke berbagai negara terutama di Indonesia, banyak aktivitas baru yang harus dilakukan khususnya untuk menjalankan aturan *physical distancing* dan *social distancing*. Seperti harus beraktivitas di rumah aja di mana berbagai hal dilakukan secara *online*, seperti *work from home*, belajar secara *online* antara lain menggunakan aplikasi *Zoom*, *Google Meet* dan *Ms.Teams*, hingga menjaga jarak fisik di tempat umum. Berbagai kebiasaan yang menjadi protokol pencegahan COVID-19 tersebut akan terus dilakukan hingga pandemi COVID-19

berkurang. Di kondisi sekarang harus tetap melakukan berbagai protokol kesehatan yang sebelumnya tidak pernah dilakukan sebelum COVID-19 muncul. Hal inilah yang disebut oleh banyak orang sebagai keadaan adaptasi '*new normal*'. Keadaan *new normal* atau kebiasaan baru ini belakangan sering disebut-sebut masyarakat dan WHO (*World Health Organization*) [12] [13].

New normal adalah kebiasaan perilaku hidup bersih tetap menjalankan aktivitas normal dibarengi penerapan protokol kesehatan untuk mencegah infeksi COVID-19 terhadap manusia [12]. Beberapa contoh situasi *new normal* yang nantinya akan diterapkan misalnya pusat perbelanjaan seperti *mall* dan kantor yang menyediakan bilik desinfektan, pengecekan suhu badan pengunjung, memonitor kepadatan pengunjung, dan penyediaan *hand sanitizer* di berbagai tempat. Kemudian untuk pembayaran lebih baik digunakan secara *cashless* melalui uang digital melalui *mobile banking*, *barcode scanner* ataupun *E-money*. Mengurangi kepadatan pengunjung dan menjaga jarak antarmanusia, contohnya antrian dengan jarak 1 sampai 1,5 meter di setiap lantai gedung, termasuk di *escalator* [13].

Ketika pandemi COVID-19 ini sudah ditemukan titik akhirnya, maka tetap harus menjalankan protokol kesehatan agar tidak ada penularan virus Corona lebih lanjut. Ini penting karena hingga saat ini vaksin COVID-19 masih dalam proses pengembangan dan belum ada vaksin resmi untuk COVID-19 ini. Sehingga tetap diperhatikan langkah-langkah pencegahan terhadap penyebaran virus di situasi *new normal* sebagai berikut [12] [13]:

1. Menghindari tempat ramai

Covid-19 dapat dengan mudah menyebar di beberapa tempat tertentu seperti tempat ramai yang penuh orang berdekatan, tempat yang sangat sempit di mana banyak orang ngobrol dalam jarak dekat dan ruangan yang tertutup dengan ventilasi yang buruk. Risikonya akan lebih tinggi di tempat-tempat yang memiliki beberapa faktor risiko ketiganya. Yang harus dilakukan adalah menghindari berada di tempat ramai, menjaga jarak 1 meter dengan orang lain, buka jendela agar terjadi pertukaran udara, dan jangan lupa memakai masker untuk selalu tetap berhati-hati.

2. Menjaga kondisi tubuh dengan memenuhi vitamin C

Lakukan berbagai tindakan yang memungkinkan untuk membantu menghentikan penyebaran Covid-19 yaitu dengan menjaga kondisi tubuh tetap sehat dan prima. Ini penting agar tubuh tetap memiliki imun dan antibodi yang bisa mencegah dari penularan virus Corona. Karena terkadang bisa saja tanpa sadar orang-orang melakukan aktivitas yang mengundang penularan Covid-19 seperti berkumpul di tempat ramai. Dalam menjaga kondisi tubuh bisa mengonsumsi vitamin C. Karena vitamin C memiliki banyak manfaat, seperti antioksidan untuk melindungi sel-sel tubuh yang rusak, dan dapat menjaga tubuh tetap fit.

3. Rajin mencuci tangan

Penularan virus Covid-19 dapat menyebar di kantor, pasar, *mall* dan di mana pun. Mencuci tangan dengan benar (setidaknya 20 detik) sangat efektif untuk mengurangi penyebaran virus dan bahkan membantu untuk tidak terinfeksi sejak dini. Mencuci tangan juga dapat mencegah infeksi melalui hidung, mata dan mulut. Disadari atau tidak, kita menyentuh hidung, mata dan mulut hingga empat kali dalam satu jam. Ini menambah banyak peluang bagi virus untuk memasuki tubuh dan menyebabkan infeksi. Mencuci tangan dengan baik dan menyeluruh dapat mengurangi kemungkinan bakteri dan virus jahat ini berada di dekat wajah. Jangan lupa untuk selalu rajin cuci tangan.

4. Memakai masker

Masker direkomendasikan sebagai penghalang sederhana untuk membantu mencegah penyebaran tetesan (*droplet*) ketika orang batuk, bersin, atau berbicara. Alasan memakai masker bukan untuk melindungi kulit wajah saja, namun dapat melindungi dari polusi udara yang tidak bersih dan virus Covid-19. Dalam memakai masker tentu ada hal penting yang perlu diperhatikan yaitu mengganti masker setiap 4 jam, jangan pakai masker secara berulang, jangan lupa untuk menggantung atau merobek masker sebelum membuang di tempat sampah, dan jangan tempatkan masker di dagu atau leher yang bisa membuat virus dan bakteri menyebar.

5. Rutin berolahraga

Olahraga dapat memberikan manfaat yang sangat baik bagi tubuh apabila dilakukan secara rutin yaitu minimal 3 kali dalam seminggu. Berolahraga dapat meningkatkan stamina, daya tahan tubuh, memperlancar sirkulasi darah dalam tubuh dan meningkatkan metabolisme. Ada beberapa hal-hal yang perlu diperhatikan dalam berolahraga di masa pandemi ini. Yang pertama perlu diperhatikan adalah intensitas dan durasi, olahraga tidak boleh dilakukan secara berlebihan. Lalu, ketika sedang berolahraga di luar rumah jangan lupa untuk menerapkan *physical distancing*. Dan yang terakhir pastikan ketika berolahraga kondisi badan harus tetap bugar dan fit.

Penyebaran COVID-19 ini membuat adanya pembatasan aktivitas pada masyarakat yang berpengaruh pada aktivitas bisnis sehingga menyebabkan beberapa perekonomian berubah. Kinerja ekonomi di Indonesia turut melemah yang memberikan dampak pada situasi ketenagakerjaan Indonesia akibat penyebaran COVID-19. Penyebaran COVID-19 juga tentu memberikan dampak baik positif maupun negatif bagi semua lapisan masyarakat di Indonesia seperti [14] [15]:

1. Pengangguran meningkat

Terhambatnya kegiatan perekonomian di Indonesia secara drastis membuat beberapa pemilik perusahaan melakukan pemecatan karyawan untuk menekan kerugian.

2. Sistem pembelajaran yang berbeda

Sebagai upaya agar pandemi Covid-19 tidak semakin menyebar, pemerintah mengeluarkan kebijakan agar sekolah dan kampus meminta muridnya untuk belajar di rumah. Sistem pendidikan daring adalah sistem pembelajaran tanpa tatap muka antara mahasiswa dan dosen melainkan dilakukan secara *online* melalui aplikasi dan media yang menggunakan jaringan internet.

3. Pengeluaran berkurang

Salah satu dampak positif dari pandemi ini adalah pengeluaran berkurang. Meski banyak orang yang kehilangan pekerjaan dan pemasukannya, tetapi dengan di rumah saja biaya transportasi, makan di luar, nongkrong bersama teman, berbelanja di cafe sangat berkurang.

4. Gaya hidup lebih sehat dan bersih

Dampak yang satu ini sangat dirasakan masyarakat, ketika sehabis berpergian keluar rumah sebaiknya langsung mandi dan mencuci pakaian. Tidak hanya itu saja kepedulian masyarakat akan mencuci tangan dan memakai masker berhasil mengubah gaya hidup menjadi lebih higienis dan bersih.

2.2 Peramalan (*Forecasting*)

Dalam melakukan prediksi pandemi COVID-19 kapan akan berkurang atau berakhirnya diperlukan suatu peramalan. Peramalan (*forecasting*) adalah estimasi atau prediksi nilai variabel (sekumpulan variabel) pada suatu waktu di masa yang akan datang [16]. Istilah "perkiraan", "prediksi", "proyeksi", dan "prognosis" biasanya digunakan secara bergantian untuk menggambarkan perkiraan. Risiko dan ketidakpastian juga penting dalam peramalan dan prediksi. Tujuan untuk melaksanakan penelitian adalah agar pihak-pihak yang terkait dapat mempersiapkan strategi ekonomi, kesehatan, dan mengelola sumber dayanya untuk menghadapi akibat pandemi COVID-19 di masa yang akan datang.

Peramalan diklasifikasikan berdasarkan perbedaan waktu masa depan [16] [17]:

1. Peramalan jangka pendek

Mencakup waktu sampai 1 tahun tetapi biasanya kurang dari 3 bulan. Contoh: merencanakan pembelian saham dengan metode *trading*, jumlah karyawan, penugasan kerja, merencanakan pembelian stok obat-obatan di rumah sakit, merencanakan pembelian ATK di kantor dan kampus, serta peramalan curah hujan dan gempa.

2. Peramalan jangka menengah

Mencakup waktu bulanan antara 1 tahun sampai 3 tahun. Contoh: merencanakan penjualan barang, peramalan pasar dunia digital, pembuatan anggaran produksi dan alur kas.

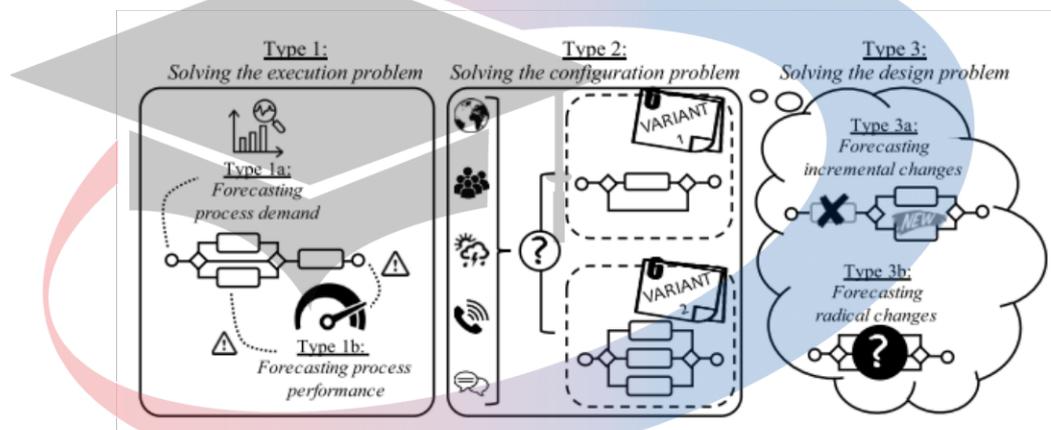
3. Peramalan jangka panjang

Mencakup perencanaan 3 tahun atau lebih. Proses peramalan jangka panjang yang terdiri dari pengelompokan dan prosedur klasifikasi didasarkan pada proses *ten-fold cross validation*. Proses ini memungkinkan pengoptimalan jumlah *cluster* untuk mencapai kesalahan perkiraan minimum dengan klasifikasi. Contoh:

membuat sebuah prototipe dari produk baru, pengembangan fasilitas dan prasarana, peramalan pandemi COVID-19, peramalan permintaan energi listrik, dan perencanaan transportasi udara di masa depan.

2.2.1 Proses Peramalan (*Forecasting*)

Ada 3 fase dalam proses peramalan yaitu memecahkan masalah eksekusi, memecahkan masalah konfigurasi dan memecahkan masalah desain [18]. Pada Gambar 2.1 adalah analogi proses peramalan.



Gambar 2. 1 Analogi proses peramalan

Dalam melakukan *Forecasting* terdapat beberapa proses yaitu [18]:

1. Memecahkan masalah yang akan digunakan (*Solving the execution problem*)
Proses yang pertama ini berkaitan dengan mengantisipasi masalah yang terkait dengan proses peramalan sehari-hari. Proses ini melibatkan 2 *task* peramalan yaitu permintaan proses peramalan dan melihat prediksi indikator seperti siklus. Inti dari proses ini adalah melihat masalah apa yang terjadi dan mengapa masalah itu harus dilakukan peramalan.
2. Memecahkan masalah konfigurasi (*Solving the configuration problem*)
Masalah konfigurasi ini mencakup memilih antara varian model alternatif seperti waktu, lokasi dan cuaca. Dalam hal ini, proses peramalan dapat membantu mengantisipasi varian model proses yang diperlukan pada titik waktu tertentu di masa mendatang.
3. Memecahkan masalah desain (*Solving the design problem*)

Hal ini mencakup informasi yang membangun tentang perubahan masa depan dalam karakteristik pelanggan, peramalan pandemi COVID-19, perkiraan proses dapat menemukan saluran mana yang harus atau tidak boleh ditawarkan di masa depan.

2.2.2 Metode Peramalan

Secara garis besar, terdapat 2 metode peramalan yaitu [16]:

1. Metode Kuantitatif

Metode kuantitatif adalah metode yang berdasarkan pada objektif dan sangat bergantung pada perhitungan matematika. Metode kuantitatif berisi model seperti *time series model* dan *associative models*. *Time series model* adalah model yang dapat digunakan dalam melihat prediksi masa depan berdasarkan pola yang terdapat di dalam data. Sedangkan, *Associative models* adalah model sebab akibat yang menggunakan faktor dan variabel yang mempengaruhi kuantitas yang akan diramalkan.

2. Metode Kualitatif

Metode kualitatif adalah metode yang berdasarkan pada penilaian, perilaku, tindakan, intuisi, emosi atau pengalaman pribadi baik secara deskripsi bahasa. Akan tetapi, metode ini tidak selalu bergantung pada perhitungan matematika. Metode ini sering digunakan dalam melihat peramalan perilaku konsumtif dan bentuk gaya hidup.

2.2.3 Langkah Sistem Peramalan

Adapun beberapa langkah yang perlu dilakukan dalam sistem peramalan yakni sebagai berikut [19]:

1. Menetapkan tujuan peramalan

Hal ini perlu dilakukan dengan mengidentifikasi masalah agar tujuan peramalan dapat dicapai.

2. Mengumpulkan informasi dan unsur yang akan diramal

Merupakan kegiatan pembelajaran dengan melihat sumber-sumber peramalan yang sudah dilakukan agar peramalan yang dihasilkan memiliki informasi yang valid.

3. Menentukan horizon waktu peramalan

Horizon adalah batas peramalan yang akan dilakukan. Terdapat peramalan jangka panjang, menengah, dan pendek.

4. Memilih model peramalan

Banyak model peramalan yang tersedia salah satunya adalah *time series*. *Time series* dapat memprediksi apa yang terjadi di waktu yang akan data dengan pola data.

5. Mengumpulkan data

Data peramalan bisa didapatkan melalui dataset *online* di *website Kaggle.com*. Situs *Kaggle.com* memiliki kumpulan dataset yang dapat digunakan untuk peramalan *time series*.

6. Membuat peramalan

Langkah yang ini yang perlu diperhatikan karena menyangkut dengan dataset yang digunakan. Ada beberapa *tools* yang dapat digunakan dalam membuat peramalan seperti *R Software*.

7. Menerapkan dan memvalidasi hasil *forecasting*

Langkah terakhir adalah melihat hasil *forecasting* dan memverifikasi model yang dihasilkan. Apakah hasil sudah akurat dengan data yang digunakan.

2.2.4 Jenis Peramalan

Adapun 3 jenis peramalan dengan fungsi dan kegunaan yang berbeda [16]:

1. Peramalan Ekonomi

Merupakan kegiatan memprediksi indikator ekonomi seperti jumlah yang beredar, suku bunga, harga saham, defisit anggaran, dan pertumbuhan ekonomi pada suatu negara.

2. Peramalan Teknologi

Peramalan teknologi adalah memprediksi tingkat kemajuan teknologi dan inovasi di masa depan seperti mesin teknik teknologi yang akan digunakan. Peneliti membuat prakiraan teknologi berdasarkan pengalaman masa lalu dan perkembangan teknologi saat ini. Tujuannya adalah mengevaluasi kemungkinan

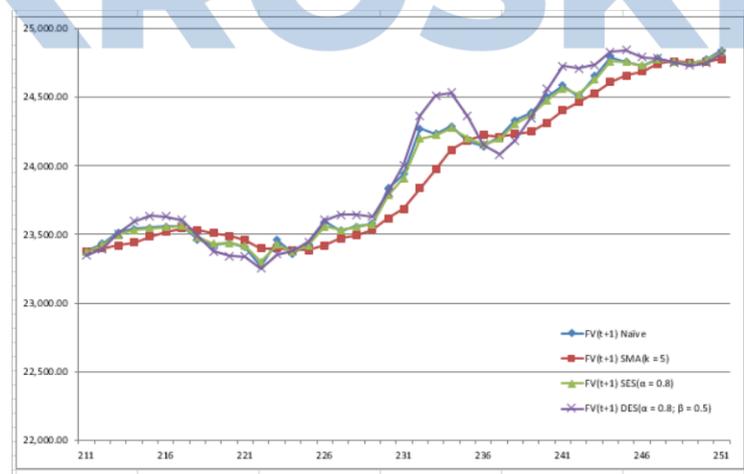
dan signifikansi dari berbagai kemungkinan perkembangan di masa depan sehingga manajer dapat membuat keputusan yang lebih baik.

3. Peramalan Permintaan

Peramalan permintaan atau *demand forecasting* merupakan membuat perkiraan tentang penggunaan permintaan pelanggan di masa depan dengan menggunakan data historis pelanggan. Peramalan permintaan yang akurat akan memberi informasi bisnis yang berharga tentang potensi mereka di pasar saat ini dan pasar lainnya. Tanpa peramalan permintaan, bisnis berisiko membuat keputusan kurang tepat tentang produk atau layanan di target pasar.

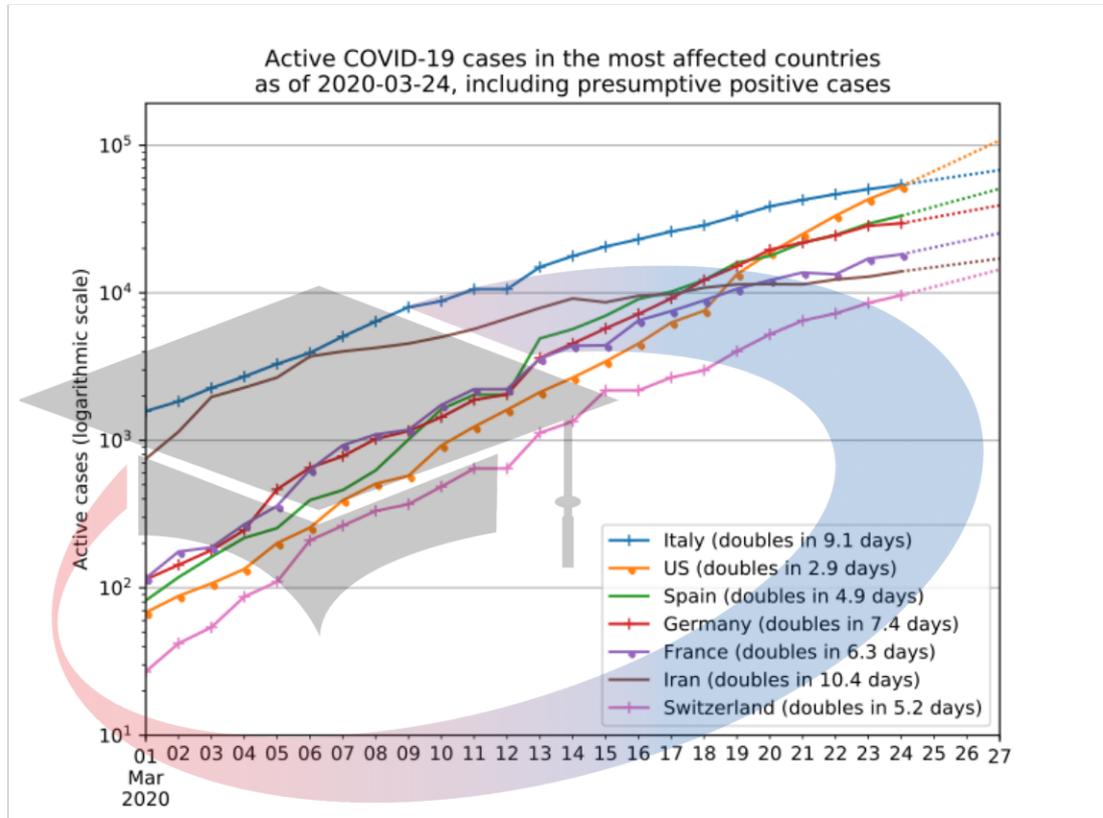
2.3 Time Series

Setiap hari semua manusia menghasilkan data *time series* baik dari lama waktu penggunaan *smartphone*, lama menonton televisi maupun lama tidur. Runtun waktu (*time series*) adalah urutan titik data numerik dalam data yang berurutan di mana, salah satu variabel berubah seiring frekuensi waktu dan yang nilai masa depannya terkait dengan nilai masa lalunya [20]. Frekuensi waktu ini bisa meliputi tahun, bulan, hari, jam maupun detik selama data di *save* dalam runtun waktu maka data itu adalah data *Time series* [20]. Runtun waktu berkaitan dengan analisis fluktuasi siklus, analisis tren, dan masalah musim seperti pandemi COVID-19. Seperti semua metode peramalan, kesuksesan dan keakuratan dari sebuah *time series* tidak dijamin. Tidak ada jumlah waktu minimum atau maksimum yang harus ditetapkan [21]. Model dari grafik data *time series* dapat dilihat pada Gambar 2.2 di bawah [22]:



Gambar 2. 2 Contoh grafik *Time series*

Adapun model grafik *time series* yang digunakan dalam kasus pandemi COVID-19 seperti Gambar 2.3 di bawah [2]:



Gambar 2. 3 Model *time series* pada COVID-19

2.3.1 Komponen *Time Series*

Adapun beberapa rangkaian *time series* terdiri dari 3 komponen sistematis yaitu *level*, *trend*, musim dan 1 komponen non-sistematis yang disebut *noise*, berikut penjelasannya [21]:

1. *Level* merupakan nilai rata-tata dalam suatu *series* atau rangkaian.
2. *Trend* yang melihat nilai naik atau turunnya dalam suatu rangkaian waktu.
3. Musim (*Seasonality*) adalah siklus jangka pendek yang berulang di rangkaian waktu.
4. *Noise* adalah perbedaan acak data dalam rangkaian atau *series*.

2.4 Analisis Runtun Waktu (*Time series analysis*)

Analisis runtun waktu adalah teknik analisis statistik yang berhubungan erat dengan runtun waktu untuk melihat bagaimana aset atau saham, angka kematian

maupun positif COVID-19 yang berubah secara fluktuasi seiring waktu [23]. Data *time series* berarti bahwa data ada dalam bentuk rangkaian runtun waktu atau interval tertentu. Analisis runtun waktu dapat digunakan juga untuk memeriksa bagaimana perubahan yang terkait dengan titik data yang dipilih dibandingkan dengan pergeseran di variabel lain selama periode waktu yang sama [24]. Dalam memilih metode di *time series* adalah mempertimbangkan jenis pola yang akan digunakan dalam penelitian.

2.4.1 Tipe pola data *Time Series*

Ada empat tipe umum dari pola data yaitu: *trend*, *seasonal*, *horizontal*, dan *cyclical* [8] [24]:

1. Pola data *Trend*

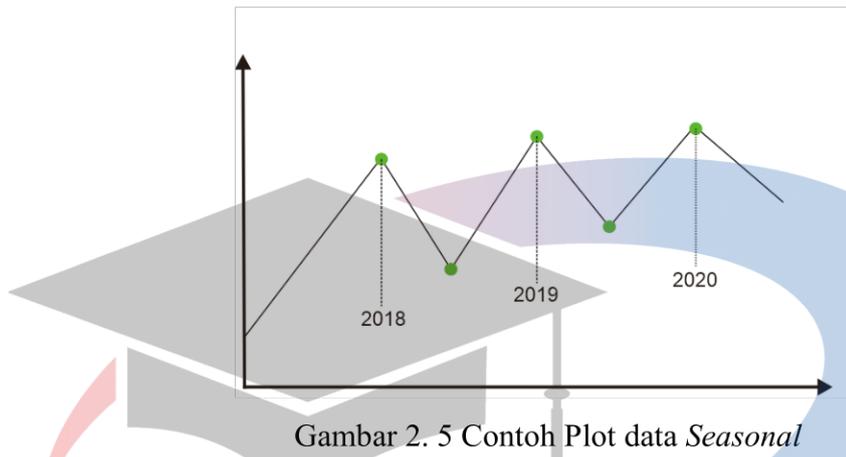
Pola data *trend* adalah ketika data naik atau menurun selama beberapa periode waktu tertentu, ada pola *trend*. *Trend* adalah komponen jangka panjang yang mewakili pertumbuhan atau penurunan dalam rangkaian waktu selama periode waktu yang diperpanjang. Gambar 2.4 berikut menunjukkan plot data tren.



Gambar 2. 4 Contoh Plot data *Trend*

2. Pola data *Seasonal*

Pola data *Seasonal* musiman adalah pola yang berulang tahun demi tahun. Pola ini dapat berulang setiap minggu, bulan, atau tahun tertentu. Gambar 2.5 berikut menunjukkan plot data *Seasonal*.



Gambar 2. 5 Contoh Plot data *Seasonal*

3. Pola data Horizontal

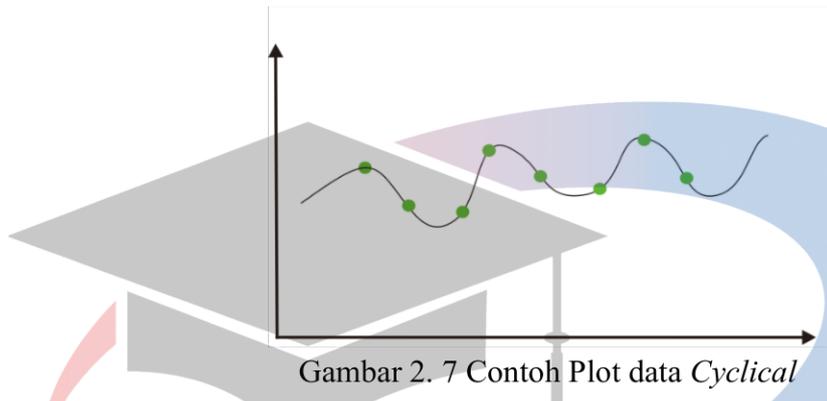
Pola data horizontal adalah data yang berfluktuasi di sekitar nilai rata-rata yang membentuk garis horizontal. Gambar 2.6 berikut menunjukkan plot data Horizontal.



Gambar 2. 6 Contoh Plot data Horizontal

4. Pola data *Cyclical*

Pola data *cyclical* adalah fluktuasi (naik dan turun) seperti gelombang di sekitar tren. Pola data ini berbeda dengan horizontal karena *Cyclical* dipengaruhi fluktuasi gelombang jangka panjang. Gambar 2.7 berikut menunjukkan plot data *Cyclical*.



Gambar 2. 7 Contoh Plot data *Cyclical*

2.4.2 Keuntungan Analisis Runtun Waktu

Adapun beberapa keuntungan dan manfaat yang dihasilkan menggunakan *time series analysis* adalah sebagai berikut [25] [26]:

1. Membersihkan data (*Cleaning data*)

Manfaat yang pertama dari analisis runtun waktu (*time series analysis*) adalah dapat membantu membersihkan data. Hal ini berguna untuk menemukan "sinyal atau tanda" yang sebenarnya ada di data, dengan memfilter *noise*. Artinya menerapkan berbagai rata-rata untuk mendapatkan hasil yang perspektif keseluruhan tentang makna data. Tentu saja, membersihkan data adalah bagian penting dari hampir semua jenis analisis data. Yang membuat untuk mendapatkan hasil analisis deret waktu memerlukan dengan sedikit usaha ekstra.

2. Digunakan sebagai peramalan

Hal inilah yang digunakan peneliti dalam *time series analysis* yaitu menggunakannya sebagai peramalan. Ini karena *time series analysis* pada dasarnya mengandung tentang pola dalam data, yang dapat digunakan untuk memprediksi titik data di masa mendatang. Menganalisis dan memahami data pada masa lampau itu dapat digunakan dalam peramalan, tetapi kemampuan memprediksi masa depan yang membantu membuat keputusan bisnis yang optimal.

2.5 Stasioneritas

Stasioneritas adalah ketentuan penting dalam model analisis runtun waktu menggunakan *Random Forest* untuk data *time series*. Data Stasioner merupakan data yang menghasilkan nilai akhir dari *mean*, *varians*, dan *autovarians* tetap sama dan tidak berubah ketika data akan dipakai, artinya dengan data stasioner *time series* akan terlihat lebih stabil [27] [28]. Deret waktu dengan tren dan musim, akan memengaruhi nilai deret waktu di waktu yang berbeda dan memengaruhi stasioneritas data [29]. Jadi, suatu runtun waktu dikatakan stasioner jika *mean*, variansi dan kovariansinya konstan terhadap waktu. Stasioneritas dapat dilihat pada plot *Time series*, jika hasil plot *Time series* pada grafiknya sejajar dengan sumbu waktu [25].

Suatu runtun waktu dikatakan tidak stasioner jika ia tidak memenuhi salah satu kriteria yang dimaksud seperti *mean*, variansi, dan kovariansinya konstan [27]. Jika memiliki data yang tidak stasioner dalam data *time series* dikarenakan *mean* dan varian yang tidak konstan, maka data tersebut dapat dibuat mendekati stasioner dengan cara melakukan *Differencing*. *Differencing* dilakukan dengan mengurangi suatu data dengan data sebelumnya untuk beberapa periode tertentu sampai data itu menjadi stasioner.

Secara umum, deret waktu stasioner tidak akan memiliki pola yang dapat diprediksi dalam jangka panjang. Plot waktu akan menunjukkan rangkaian secara kasar horizontal (meskipun beberapa perilaku siklik dimungkinkan), dengan varian konstan [29]. Data *Time series* dikatakan stasioner apabila terdapat kecenderungan baik dalam nilai rata-rata dan variasi.

Berikut adalah rumus dan variabel dasar dari stasioner [30]:

- a) Mean : $E(Y_t) = \mu$ (konstan untuk semua t)
- b) Varians : $Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$ (konstan untuk semua t)
- c) Kovarians : $Cov(Y_t, Y_{t+k}) = E[Y_t - \mu)(Y_{t+k} - \mu) = Y_k$
(konstan untuk semua $t, k \neq 0$)

Fungsi dari perbedaan waktu $|t - s|$ dapat ditulis sebagai berikut [31]:

$$\text{Cov}(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)] = \mu_k \quad (1)$$

Di mana:

- a) μ = nilai rata-rata (*mean*)
- b) Y_k = kovarians
- c) σ^2 = varians
- d) Y_t = data ke- t

2.5.1 Konsep Differencing

Suatu deret waktu yang tidak stasioner (terdapat nilai yang berubah) harus diubah menjadi data stasioner dengan melakukan *differencing*. Makna dari *differencing* adalah menghitung nilai perubahan atau perbedaan nilai selisih observasi [25]. Nilai selisih yang diperoleh dicek lagi apakah stasioner atau tidak. Jika data yang dihasilkan belum stasioner maka dilakukan *differencing* lagi. Dasar dari *differencing* adalah mengurangi antara peramalan pengamatan Z_t dengan hasil peramalan pengamatan sebelumnya Z_{t-1} .

Secara umum rumus *differencing* sebagai berikut:

$$\Delta Z_t = Z_t - Z_{t-1} \text{ dan } \Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \quad (2)$$

Di mana:

- a) Z_t = hasil peramalan yang diamati
- b) Z_{t-1} = hasil peramalan sebelumnya

2.5.2 Konsep Augmented Dickey-Fuller (ADF)

Salah satu konsep penting yang dipakai untuk mengetahui stasioneritas data yaitu melalui uji akar unit (*unit root test*). Teknik ini dapat melihat apakah data stasioner atau tidak stasioner. Uji *unit root test* ialah pengujian yang terkenal dan umum digunakan, dikembangkan oleh David Dickey dan Wayne Fuller dengan sebutan *Augmented Dickey-Fuller (ADF) Test*. Tes *Augmented Dickey-Fuller (ADF)* adalah bentuk analisis *unit root* yang menggabungkan kelambatan untuk memperhitungkan korelasi serial [28]. *Augmented Dickey-Fuller* secara statistik didefinisikan sebagai pengujian hipotesis nol dari "proses ARIMA (p, 1,0) yang

(*autoregressive integrated moving average*) terhadap alternatif stasioner ARIMA ($p + 1, 0, 0$)" [9]. Jika x_t adalah deret waktu, maka regresi ADF mengambil bentuknya.

$$\Delta x_t = \mu + \gamma t + \alpha x_{t-1} + \sum_{j=1}^{k-1} \beta_j \Delta x_{t-j} + \mu_t \quad (3)$$

Di mana:

- a) Δ : operator pembeda
- b) x_t : nilai x pada orde ke t (*time series*)
- c) x_{t-1} : nilai x pada orde ke $t-1$ (*time series*)
- d) μ : nilai harapan dari variabel random x_t
- e) μ_t : *white noise*
- f) γ : parameter bebas
- g) β : Beta hasil kovarian dan varian

Jumlah lag yang dimasukkan dalam analisis secara otomatis ditentukan menggunakan Akaike Information Criteria (AIC), dan jumlah maksimum lag yang dipertimbangkan p_{max} didasarkan pada paradigma yang diuraikan dalam persamaan di bawah ini [28].

$$P_{max} = 12 \left(\frac{T}{100} \right)^{\frac{1}{4}} \quad (4)$$

T mewakili ukuran total sampel (dalam hal ini, jumlah hari). Tes ADF mengembalikan statistik uji t yang terletak pada analisis menentukan apakah akan menolak hipotesis nol dalam distribusi f_a dengan $df = T - 1$. Nilai p yang dihasilkan dari nilai alternatif seperti persamaan di bawah ini [28].

$$H_0: \rho = 1 \qquad H_a: \rho < 1 \quad (5)$$

Di mana ρ merepresentasikan *root*. Seperti yang tertulis dari namanya, $H_0 \rho = 1$ menandakan unit *root*, sedangkan $H_a \rho < 1$ menandakan runtun waktu stasioner.

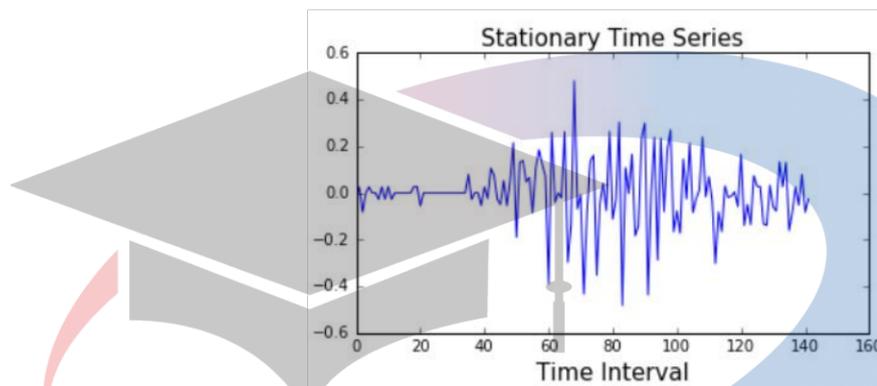
Maka pengambilan keputusannya adalah [28]:

1. Jika $\rho > 0$, maka hasilnya stasioner.
2. Jika $\rho = 0$, hipotesis nol ditolak, maka hasilnya tidak stasioner.

2.5.3 Pola grafik data Stasioner

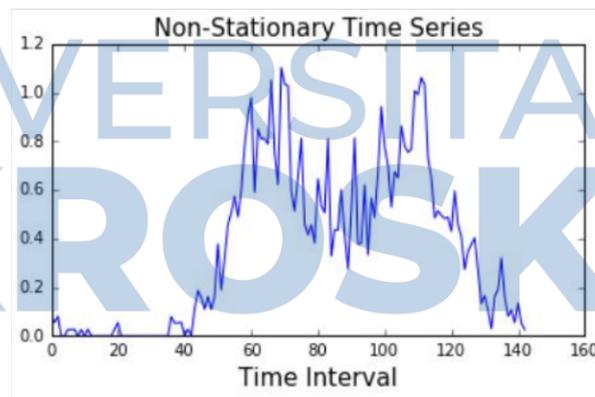
Pola data grafik antara stasioner dan non-stasioner berbeda, berikut perbedaan antara kedua grafik [27]:

1. Plot runtun waktu stasioner, pergerakan grafik lebih condong ke arah tengah dan menunjukkan pergerakan yang fluktuaktif di tengah grafik saja seperti pada Gambar 2.8.



Gambar 2. 8 Plot runtun waktu stasioner

2. Plot runtun waktu non-stasioner, menunjukkan adanya *downtrend* secara tiba-tiba kemudian terjadi *uptrend* pada plot grafik secara fluktuaktif seperti pada Gambar 2.9 di bawah.



Gambar 2. 9 Plot runtun waktu non-stasioner

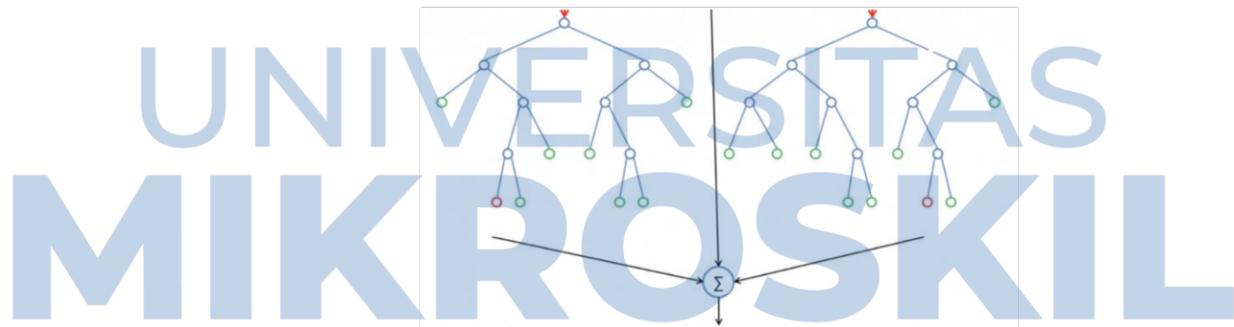
2.6 Metode *Random Forest*

Metode *Random Forest* merupakan salah satu metode yang digunakan untuk klasifikasi dan regresi. *Random Forest* dapat digunakan untuk *time series* dengan cara meningkatkan keakuratan metode klasifikasi dengan menggabungkan (*ensemble*) metode klasifikasi [32] [33]. *Random Forest* dilakukan dengan teknik *data mining* (penggalian data) yang dasar yaitu pohon keputusan (*decision tree*) [34]. *Random*

Forest berbeda dengan *decision tree* meskipun kedua metode ini hampir sama. Dalam garis besar *Random Forest* terdiri atas beberapa pohon keputusan (*decision tree*), di mana gabungan dari *decision tree* digunakan untuk *classification* data ke suatu kelas. *Random Forest* adalah kombinasi pengklasifikasi yang terdiri dari kumpulan pengklasifikasi pohon, di mana setiap pohon sangat bergantung kepada nilai nilai acak untuk menghasilkan ke pohon yang paling populer [34] [32].

Proses penggabungan nilai taksiran banyak pohon serupa dengan yang dilakukan dalam metode *bagging* (dikumpulkan). Proses ini menghasilkan satu pohon dengan ukuran dan bentuk berbeda. Hasil yang diharapkan adalah satu kumpulan pohon memiliki korelasi yang kecil antar pohon. Dapat dikatakan, semakin banyak pohon (*tree*) dalam sebuah hutan (*forest*) maka semakin kuat dan akurat hasil yang didapatkan [35].

Satu keuntungan besar dari *Random Forest* adalah dapat digunakan untuk masalah klasifikasi dan regresi, yang merupakan sebagian besar sistem *machine learning* saat ini. *Random Forest* dalam klasifikasi, karena klasifikasi terkadang dianggap sebagai elemen penyusun *machine learning*. Pada Gambar 2.10 di bawah terdapat bagaimana *Random Forest* akan terlihat dengan beberapa pohon [35].



Gambar 2. 10 Analogi *Random Forest* seperti layaknya 2 pohon

Berikut merupakan algoritma *Random Forest*. Khususnya, model algoritma yang digunakan untuk regresi. Model khusus dari algoritma sudah diimplementasikan dalam paket *Random Forest R* [33].

\mathbf{u} merupakan vektor acak dengan elemen k . Tujuannya yakni untuk memprediksi v dengan mengestimasi \mathbf{u} .

Fungsi regresi:

$$m(\mathbf{u}) = \mathbf{u}E[\vartheta|\mathbf{u} = \mathbf{u}] \quad (6)$$

Di mana:

\mathbf{u} = vektor acak dengan elemen k ,

$m(\mathbf{u})$ = fungsi regresi

Dan untuk mencari sampel, yakni dengan rumus

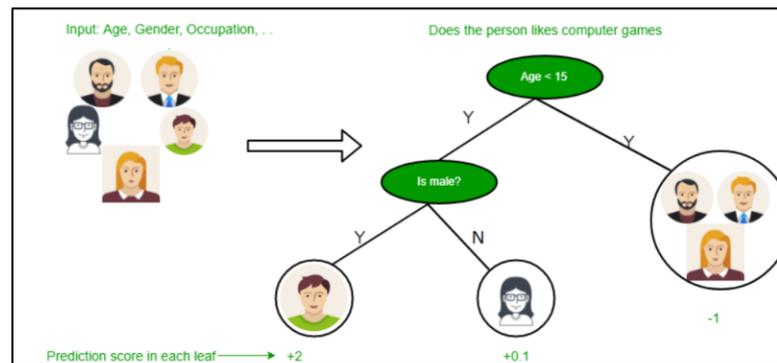
$$S_s = ((\mathbf{u}_1, \vartheta_1), \dots, (\mathbf{u}_2, \vartheta_2)) \quad (7)$$

Di atas merupakan realisasi independen dari variabel acak (u, v) . Oleh karena itu, tujuannya adalah untuk membuat perkiraan m_s dari fungsi m . *Random Forest* adalah prediktor yang dibangun oleh pohon regresi acak yang tumbuh. Untuk j -th pohon dalam bagiannya, nilai prediksi pada u dilambangkan dengan $m_s(u; \theta_j, S_s)$, di mana $\theta_1, \dots, \theta_M$ adalah variabel acak independen, didistribusikan sebagai θ dan tidak bergantung pada S_s . Modifikasi kelengkungan lokal diimplementasikan dengan menggunakan "koefisien Theta" disimbolkan dengan θ [34]. Variabel acak θ digunakan untuk mengambil sampel kembali set pas sebelum penanaman pohon individu dan untuk memilih arah pemisahan yang berurutan. Prediksi kemudian diberikan oleh rata-rata nilai prediksi dari semua pohon [34].

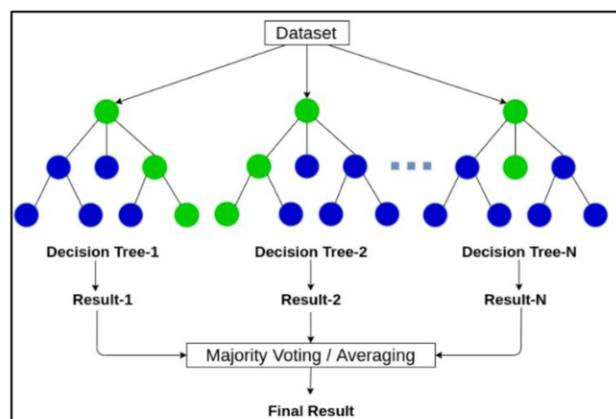
2.6.1 Perbedaan *Decision tree* dan *Random Forest*

Antara *Decision tree* dan *Random Forest* memiliki kesamaan namun berbeda, berikut perbedaan antara *Decision tree* dan *Random Forest* [36]:

1. Pohon keputusan (*Decision tree*) adalah algoritma pembelajaran mesin (*machine learning*) yang diawasi yang dapat digunakan untuk masalah klasifikasi dan regresi [36]. Pohon keputusan menggunakan representasi pohon untuk memecahkan masalah di mana setiap simpul daun berhubungan dengan label kelas dan atribut direpresentasikan pada simpul internal pohon. Merepresentasikan fungsi boolean apapun pada atribut diskrit menggunakan pohon keputusan. Gambar 2.11 di bawah merupakan ilustrasi pohon keputusan (*Decision tree*).

Gambar 2. 11 Ilustrasi *Decision tree*

2. *Random Forest* berbeda dengan algoritma *decision tree*. Karena *decision tree* cukup mudah untuk dipahami dan diinterpretasikan. Namun seringkali, satu pohon tidak cukup untuk menghasilkan hasil yang efektif. Di sinilah algoritma *Random Forest* memiliki kelebihan. *Random Forest* adalah algoritma pembelajaran mesin berbasis pohon yang memanfaatkan kekuatan beberapa pohon keputusan untuk membuat keputusan. Seperti namanya, ini adalah “hutan” pepohonan. Tetapi mengapa disebut sebagai hutan "acak"? itu karena hutan pohon keputusan yang dibuat secara acak. Setiap node di pohon keputusan bekerja pada subset fitur acak untuk menghitung output. *Random Forest* kemudian menggabungkan keluaran dari pohon keputusan individu untuk menghasilkan keluaran akhir. Proses menggabungkan *output* dari beberapa model individu disebut *Ensemble Learning*. Sehingga dapat ditarik kesimpulan secara garis besar bahwa Algoritma *Random Forest* menggabungkan *output* dari beberapa Pohon Keputusan (yang dibuat secara acak) untuk menghasilkan *output* akhir. Berikut adalah ilustrasi *random forest*.

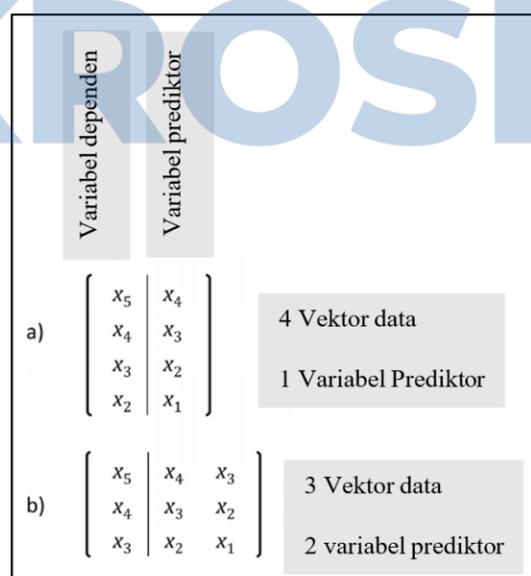
Gambar 2. 12 Ilustrasi *Random Forest*

2.6.2 *Random Forest* untuk peramalan *time series*

Menggunakan *Random Forest* untuk peramalan runtun waktu merupakan langkah yang mudah dan mirip dengan cara *Random Forest* yang digunakan untuk regresi [34]. Misalkan g adalah fungsi yang diperoleh dari *Random Forest*, yang akan digunakan untuk peramalan x_{n+1} , diberikan x_1, \dots, x_n . Jika menggunakan variabel tertinggal yaitu k , maka x_{n+1} yang diramalkan diberikan oleh persamaan berikut untuk $t = n + 1$:

$$x_t = g(x_{t-1}, \dots, x_{t-k}), t = k + 1, \dots, n + 1 \quad (8)$$

Fungsi g tidak dalam bentuk tertutup, tetapi bisa diperoleh dengan melatih algoritma *Random Forest* menggunakan set pelatihan ukuran $n - k$. Dalam setiap sampel set fitting variabel dependennya adalah x_t , untuk $t = k + 1, \dots, n + 1$ sedangkan variabel prediktornya adalah x_{t-1}, \dots, x_{t-k} . Ketika jumlah variabel prediktor k meningkat, ukuran set pelatihan $n - k$ berkurang. Set pelatihan, yang mencakup sampel $n - k$, dibuat menggunakan fungsi *Cases Series* dari paket *rminer R Software* [35]. Akhirnya, pemasangan dilakukan dengan menggunakan fungsi *training* dari paket R CARET dan nilai ramalan x_{n+1} , diperoleh menggunakan fungsi *prediksi* dari paket R CARET. CARET sendiri merupakan kepanjangan dari *Classification And Regression Training* [35]. Gambar 2.13 di bawah merupakan Penjelasan bagaimana sampel *test* berubah dengan jumlah variabel prediktor untuk deret waktu dengan $n = 5$ dan (a) $k = 1$; (b) $k = 2$.



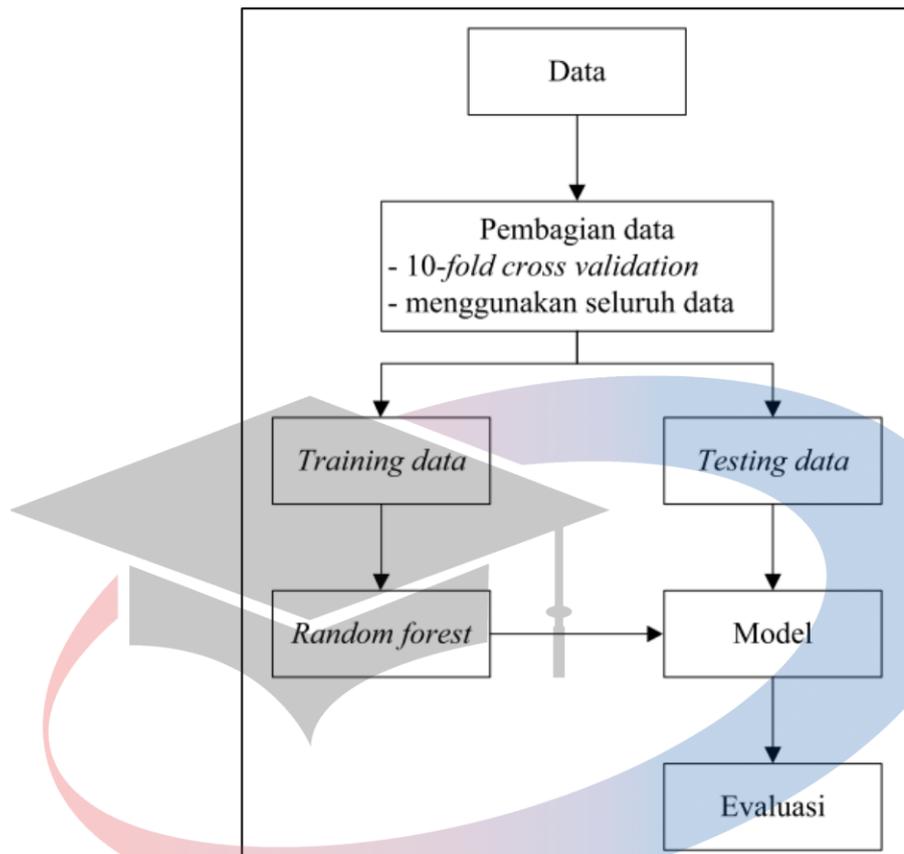
Gambar 2. 13 Sampel *test* berubah dengan jumlah variabel prediktor

2.6.3 Implementasi *Random Forest*

Secara sederhana, algoritma pembentukan *Random Forest* dapat digambarkan sebagai berikut. Bayangkan dataset yang dimiliki berukuran n dan terdiri atas d variabel penjelas (*predictor*). Maka tahap-tahap penyusunan dan peramalan menggunakan *Random Forest* sebagai berikut [37] [38]:

1. *Bootstrap sampling*, yaitu tahap di mana menarik contoh acak untuk membangun pohon prediksi dari dataset (teknik ini juga dinamakan *replacement*).
2. *Random sub-setting*, pada tahap ini dilakukan penyusunan pohon berdasarkan dataset yang digunakan, tetapi pada setiap proses pemilahan pilihan secara acak jumlah variabel *predictor* $m < d$ variabel penjelas, dan melakukan pemisahan (masing-masing pohon memprediksi dengan *predictor* yang *random*).
3. Ulangi langkah pertama sampai kedua (1-2) sebanyak k kali sehingga didapatkan hasil k buah pohon acak (*Random Forest*).
4. Melakukan pendugaan gabungan berdasarkan k buah pohon tersebut dengan cara *majority vote* untuk mengklasifikasi pada regresi (dalam proses penggabungan nilai dugaan dari beberapa pohon yang dihasilkan menggunakan metode *bagging*).

Pada Gambar 2.14 di bawah ini merupakan langkah mengimplementasi algoritma *random forest* untuk prediksi pandemi COVID-19. Yang pertama adalah melakukan input data dari hasil transformasi data (terdiri dari atribut penjelas dan target). Lalu dibagi menjadi 2 (dua) jenis data yaitu *training data* dan *testing data* dengan metode *K-Fold Cross Validation* [39]. Di mana, nilai k dipilih sebanyak 10. *Training data* digunakan hanya sebagai data masukan untuk algoritma *random forest*. Sedangkan, *testing data* dipakai sebagai mengevaluasi model keluaran (*output*) yang dihasilkan algoritma *random forest* [39].



Gambar 2. 14 Implementasi *Random Forest*

Evaluasi pada *random forest* dilakukan dengan beberapa parameter yaitu akurasi, *recall*, *f-measure* dan presisi[39]. Akurasi adalah parameter yang sederhana untuk melihat kinerja algoritma klasifikasi *random forest*, dengan melihat berapa persentase keakuratan dan kebenaran peramalan (*forecasting*).

2.7 Ketepatan Penggunaan Metode Peramalan

Dalam memakai tata cara peramalan sangat berarti dicermati untuk menciptakan hasil akhir yang kokoh serta akurat. *Random Forest* diseleksi sebab salah satu algoritma *machine learning* yang sangat populer baik dalam membongkar permasalahan klasifikasi dan regresi dibandingkan *decision tree* [36]. Model *Random Forest* sangat cocok digunakan dalam penelitian pandemi COVID-19 ini karena tidak teralalu memiliki banyak pohon di hutan. Hasil peramalan ini dengan jelas menampilkan jika memakai sebagian variabel terkini sebagai prediktor sepanjang proses pemasangan yang menimbulkan akurasi prediksi yang lebih besar pada algoritma *Random Forest* yang digunakan dalam riset ini.

Random Forest mempunyai beberapa keuntungan antara lain[36]:

1. Algoritma *Random Forest* bisa digunakan buat peramalan *time series*. Artinya algoritma ini tidak menghadapi penumpukan kesalahan yang ditaksir, serta kedua prediktornya gampang untuk memasukkan prediktor eksogen.
2. Permasalahan *overfitting* tidak hendak sempat timbul kala kita memakai algoritma *Random Forest* dalam permasalahan klasifikasi. *Overfitting* merupakan penyimpangan informasi yang sangat dekat dengan informasi aslinya.
3. *Random Forest* mempunyai kekuatan buat menanggulangi kumpulan data dengan ukuran yang lebih besar semacam data COVID-19 yang mencapai ratusan ribu. Ini dapat secara otomatis menyeimbangkan kumpulan data saat kelas lebih jarang daripada kelas lain dalam data. Metode *random forest* ini juga menangani variabel dengan cepat, sehingga cocok untuk data yang rumit seperti COVID-19 di Indonesia.

2.7.1 MAPE (*Mean Absolute Percent Error*)

Adapun salah satu parameter penting untuk melihat persentase kesalahan hasil peramalan yaitu MAPE (*Mean Absolute Percent Error*). *Mean Absolute Percent Error* (MAPE mengukur akurasi sebagai persentase, dan dapat dihitung sebagai kesalahan persentase sebuah prediksi [40]. Terdapat 4 kategori nilai MAPE yaitu [41]:

- a. <10 % = Hasil prediksi sangat akurat
- b. 10-20 % = Hasil prediksi baik
- c. 20-50% = Hasil prediksi layak
- d. >50% = Hasil prediksi kurang akurat

Fungsi MAPE (*Mean Absolute Percent Error*):

$$MAPE = \frac{1}{2} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (9)$$

Di mana:

Y_t : nilai aktual pada periode t

\hat{Y}_t : nilai *forecast* pada periode t

2.8 Penelitian Terdahulu

Penelitian terdahulu adalah dasar penulis dalam melakukan penelitian agar penulis dapat melakukan eksplorasi teori dan model yang digunakan. Demikian juga, penelitian terdahulu digunakan sebagai referensi dan aspek penting yang membuat penelitian tidak memiliki unsur plagiat penelitian. Tabel 2.1 di bawah merupakan beberapa penelitian terdahulu terkait penelitian yang dilakukan penulis.

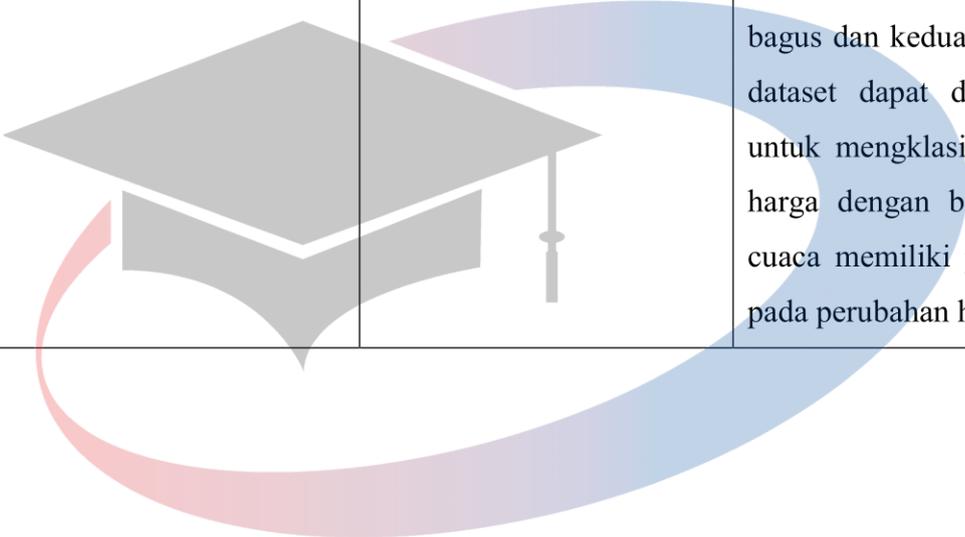
Tabel 2. 1 Penelitian terdahulu

Nama peneliti	Judul penelitian	Hasil penelitian
Syauqi Amri Yahya, 2019 [37].	Klasifikasi ketepatan lama studi mahasiswa menggunakan metode <i>support vector machine</i> dan <i>random forest</i>	Metode <i>Random Forest</i> dapat digunakan dengan sangat bagus dengan hasil nilai optimum $m=2$ dan $k=500$ sebesar 80%. Di mana jurusan dengan mahasiswa kelulusan tepat waktu terbesar adalah Ekonomi Islam sebesar 61,5%.
Bagus Sartono dan Utami Dyah Syafitri, 2015 [38].	Metode <i>decision tree</i> : solusi pilihan untuk mengatasi kelemahan pohon regresi dan klasifikasi tunggal	Akurasi dugaan yang tinggi pada pohon gabungan relatif dapat diperoleh dengan mudah, dibanding menerapkan metode klasifikasi yang lebih rumit seperti <i>support vector machine (SVM)</i> atau <i>kernel classification</i> .

<p>I Made Budi Adnyana, 2016 [42].</p>	<p>Prediksi lama studi mahasiswa dengan Metode <i>Random Forest</i></p>	<p>Hasil pengujian menunjukkan nilai <i>producer accuracy</i> untuk kelas “Lulus Tepat Waktu” diperoleh nilai sebesar 81,81 % dan kelas “Lulus Lewat Waktu Studi” diperoleh nilai <i>producer accuracy</i> sebesar 84,98%</p>
<p>P.J.Moore1, J. Gallacher, T.J. Lyons, 2019 [43].</p>	<p><i>Random forest prediction of Alzheimer’s disease using pairwise selection from time series data</i></p>	<p>Metode <i>Random Forest</i> perkiraan dapat dibandingkan menggunakan kumpulan data papan peringkat yang dipilih dari <i>Alzheimer’s Disease Neuroimaging Initiative</i> (ADNI). Untuk diagnosis, ditemukan mAUC 0,82, dan akurasi klasifikasi 0,73. Hasilnya menunjukkan bahwa metode tersebut efektif dan dapat dibandingkan dengan metode lain.</p>
<p>Michael J Kane, Natalie Price, Matthew Scotch dan Peter Rabinowitz, 2014 [44].</p>	<p><i>Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks</i></p>	<p>Telah diterapkan model deret waktu ARIMA dan <i>Random Forest</i> pada data kejadian wabah flu burung yang sangat patogen (H5N1) di Mesir, tersedia</p>

		<p>melalui sistem online EMPRES-I. Bahwa model <i>Random Forest</i> mengungguli model ARIMA dalam kemampuan prediksi. Artinya, bahwa model <i>Random Forest</i> sangat efektif untuk memprediksi wabah Flu burung (H5N1) di Mesir</p>
<p>Rahmat Robi Waliyansyaha dan Nugroho Dwi Saputro, 2020 [32].</p>	<p>Peramalan Calon Mahasiswa Baru Menggunakan Metode <i>Random Forest</i></p>	<p>Hasil pengujian menggunakan <i>Random Forest</i> didapatkan 5 program studi dengan peningkatan jumlah mahasiswa baru yang signifikan dan 5 program studi dengan jumlah mahasiswa baru terendah. Program Studi dengan peningkatan jumlah mahasiswa yaitu: Program Studi Manajemen sebesar 75% sedangkan untuk prodi dengan jumlah mahasiswa baru terendah yaitu: Magister Pendidikan dan Bahasa Indonesia sebesar 2,6%.</p>

<p>Yunita Ayu Lestari Sudarjat, Fhira Nhita, Deni Saepudin, 2017 [45].</p>	<p>Peramalan Harga Komoditas Pertanian Menggunakan Algoritma <i>Random Forest</i></p>	<p>Berdasarkan hasil prediksi, rata-rata akurasi pada data training sebesar 78.25% dan pada data testing sebesar 95.75% secara keseluruhan sistem mempunyai kinerja yang bagus dan kedua scenario dataset dapat digunakan untuk mengklasifikasikan harga dengan baik. Dan cuaca memiliki pengaruh pada perubahan harga.</p>
--	---	---



UNIVERSITAS
MIKROSKIL