

BAB 2

TINJAUAN PUSTAKA

Pada pengantar bab dua ini berisi teori penjelasan yang menjadi dasar dalam melaksanakan penelitian, teori tersebut terdiri dari pandemi COVID-19, prediksi, data, model deret waktu, facebook prophet, python, jupyter notebook dan evaluasi kinerja prediksi.

2.1 Pandemi Coronavirus Disease-2019 (COVID-19)

Coronavirus adalah suatu kelompok virus yang dapat menginfeksi manusia dan hewan sehingga menimbulkan penyakit. Coronavirus dapat mengakibatkan infeksi saluran pernapasan pada manusia yang diawali dengan gejala ringan seperti batuk hingga gejala serius seperti *Severe Acute Respiratory Syndrome* (SARS) dan *Middle East Respiratory Syndrome* (MERS). Novel coronavirus (2019-nCoV) adalah virus corona jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia dan belum diketahui hewan penularnya 2019-nCoV (Kementerian Kesehatan Republik Indonesia, 2020). Novel coronavirus pertama kali dilaporkan di China pada 31 Desember 2019 dan menjadi permasalahan kesehatan di seluruh dunia. Pada 30 Januari 2020, *World Health Organization* menyatakan bahwa COVID-19 sebagai *public health emergency of international concern (PHEIC)* atau yang disebut kedaruratan kesehatan masyarakat yang meresahkan dunia. Pandemi COVID-19 adalah suatu peristiwa menyebar luasnya virus corona yang dapat menyebabkan gejala yang ringan hingga berat seperti batuk, demam hingga orang yang berisiko tinggi (seperti jantung, diabetes, dan tekanan darah tinggi) dapat menyebabkan masalah kesehatan yang serius (Kementerian Dalam Negeri & Gugus Tugas COVID-19, 2020).

COVID-19 ditularkan dari manusia ke manusia melalui *droplet* yang keluar dari orang yang terinfeksi. Jika orang yang tidak tertular COVID-19 menghirup *droplet* yang terinfeksi maka orang tersebut akan terinfeksi virus. *Droplet* dapat menempel di benda seperti gagang pintu, kursi, meja, bahkan pada saat berjabat

tangan. Virus COVID-19 memiliki gejala yang ringan dari batuk dan demam hingga gejala yang berat seperti kesulitan bernapas hingga mati. Berdasarkan investigasi epidemiologi masa inkubasi COVID-19 adalah 1-14 hari dengan sumber utama infeksi terjadi pada pasien COVID-19 dan pembawa (*carier*) COVID-19. Karakteristik penyakit ini adalah mudah menular hingga menjangkiti banyak orang. Penyebaran COVID-19 di dunia terus meningkat dari waktu ke waktu hingga menyebar ke 215 negara, salah satunya di Indonesia (Belkacem, 2020). Di Indonesia kasus COVID-19 pertama kali ditemukan pada 2 Maret 2020 di Depok Jawa Barat oleh dua wanita dan telah dinyatakan sembuh namun pandemi ini terus menyebar ke seluruh Indonesia (Ardhiansyah & Ratih, 2020).

Peningkatan penyebaran virus corona di Indonesia yang signifikan hingga 20 Januari 2021 tercatat bahwa jumlah positif COVID-19 di Indonesia sebesar 939.948 jiwa dan meninggal sebesar 26.875 jiwa. Penyebaran COVID-19 di Indonesia menyeluruh hingga ke skala regional khususnya Kalimantan Timur yang menempati urutan ke-2 dengan insiden kasus COVID-19 tertinggi di Indonesia dan urutan ke-3 sebesar 36.2% dengan kenaikan kasus tertinggi di Indonesia (Satgas Penanganan COVID-19 Indonesia, 2021). Berdasarkan laporan analisis data COVID-19 incident rate di Kaltim sebesar 858.93 per 100.000 penduduk berarti hampir 860 jiwa per 100 ribu penduduk di Kaltim telah positif COVID-19. Penyumbang incident rate di Kaltim adalah Kota Bontang, Balikpapan, dan Kab. Kutai Timur. Lebih khususnya kota Bontang yang termasuk daerah dengan zona merah atau risiko tinggi berada di urutan pertama incident rate tertinggi di Kaltim dengan nilai sebesar 1161.78 kasus per 100 ribu penduduk. Tercatat pada 20 Januari 2021 persentasi kasus aktif di Bontang sebesar 24,2% dan jumlah kasus positif sebesar 2576 Jiwa. Meningkatnya jumlah kasus COVID-19 di Kota Bontang menjadi alasan peneliti untuk melakukan analisis prediksi penyebaran COVID-19.

Analisis prediksi COVID-19 telah menjadi area penelitian yang sangat penting untuk mendukung layanan kesehatan dalam merencanakan dan menangani penyebaran COVID-19 (Kumar & Susan, 2020). Salah satu negara yang menggunakan data untuk menangani penyebaran covid adalah Taiwan. Taiwan sukses menggunakan big data untuk analitik dalam menghasilkan peringatan gejala

www.itk.ac.id

klinis serta riwayat perjalanan dengan menghasilkan 124 item tindakan pencegahan yang terdiri dari kontrol perbatasan dari udara dan laut, identifikasi (kasus baru), karantina kasus yang mencurigakan, penemuan kasus proaktif, jaminan dan pendidikan publik, perumusan kebijakan sekolah, pengasuhan, dan kebijakan bantuan untuk bisnis (Wang, Ng, & Brook, 2020). Oleh karena itu analisis prediksi COVID-19 khususnya di kota Bontang di butuhkan dalam membantu mengambil keputusan di bidang sistem kesehatan khususnya Satgas COVID dan Rumah sakit untuk mengantisipasi kenaikan kebutuhan fasilitas kesehatan dan pengobatan dalam rangka antisipasi jumlah pasien yang akan datang.

Mempertimbangkan penyebaran COVID-19 yang terus meningkat dan menimbulkan korban jiwa, Presiden Joko Widodo menetapkan Keputusan Presiden Nomor 7 Tahun 2020 tentang Gugus Tugas Percepatan Penanganan COVID-19 dengan tujuan sebagai berikut.

1. Mempercepat penanganan COVID-19 melalui sinergi antara kementerian dan pemerintah daerah;
2. Meningkatkan ketahanan nasional khususnya di bidang kesehatan;
3. Meningkatkan pencegahan penyebaran COVID-19;
4. Meningkatkan sinergi pengambilan keputusan dan kebijakan operasional;
5. Meningkatkan kemampuan dalam mendeteksi, mencegah, dan merespon penyakit COVID-19 (Kementerian Dalam Negeri & Gugus Tugas COVID-19, 2020).

2.2 Prediksi Data Deret Waktu

Prediksi adalah perkiraan yang akan terjadi di masa depan yang dilakukan secara matematis yang didapatkan berdasarkan pengetahuan dan pengalaman di masa lalu. Model prediksi digunakan untuk memecahkan persoalan yang memerlukan prediksi dan estimasi (Kurniawan, 2010). Prediksi pada umumnya menggunakan data masa lampau untuk dianalisis menggunakan cara-cara tertentu. Ketepatan suatu prediksi berbeda untuk tiap persoalan bergantung pada berbagai faktor sehingga hasil prediksi yang diperoleh tidak akan selalu didapatkan dengan seratus persen. Prediksi dapat dilihat sebagai jenis klasifikasi, perbedaannya adalah

www.itk.ac.id

prediksi itu memprediksi keadaan masa depan sedangkan klasifikasi keadaan saat ini. pengaplikasian dari prediksi seperti *flooding*, *speech recognition*, *machine learning*, dan pengenalan pola. Dalam memprediksi masa depan dapat menggunakan analisis deret waktu atau teknik regresi, serta pendekatan lainnya (Dunham, 2003). Jadi prediksi data deret waktu adalah perkiraan yang akan terjadi di waktu mendatang yang dilakukan secara matematis menggunakan himpunan data terdahulu berdasarkan waktu kejadian. Berikut adalah rincian penjelasan prediksi, data dan model deret waktu.

2.2.1 Jenis- Jenis Prediksi

Prediksi atau peramalan dapat dibedakan dari berbagai pendekatan berikut jenis prediksi dari sudut horizon waktu yaitu.

1. Peramalan jangka pendek, yaitu peramalan dengan kurun waktu mulai dari satu hari hingga satu musim atau dapat sampai satu tahun kedepan. Peramalan ini biasanya digunakan untuk operasional.
2. Peramalan jangka menengah, yaitu peramalan dengan kurun waktu dari satu musim hingga dua tahun.
3. Peramalan jangka panjang, yaitu peramalan untuk kurun waktu minimal lima tahun. Peramalan ini biasanya digunakan untuk membantu organisasi atau perusahaan dalam menentukan strategi dan kompetitif organisasi atau perusahaan (Santoso, 2009).

2.2.2 Metode Prediksi

Metode prediksi atau peramalan dapat dikelompokkan menjadi dua metode yaitu.

1. Metode Kualitatif

Metode prediksi kualitatif mengutamakan pada data pendapat dan intuisi manusia dari pada penggunaan data historis yang dimiliki. Metode ini biasa digunakan dalam pengambilan keputusan sehari-hari.

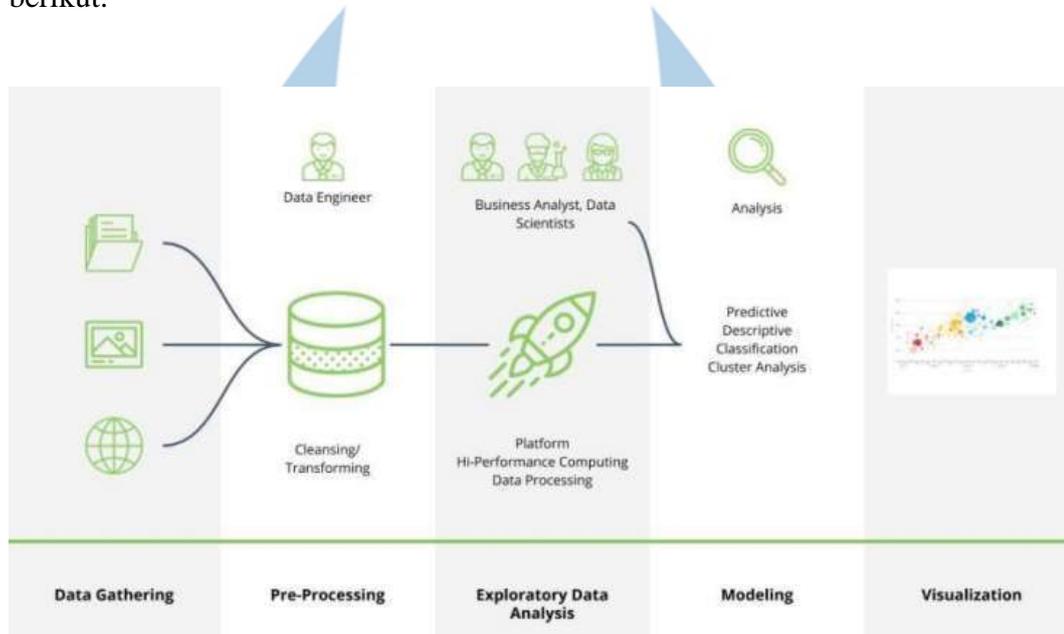
2. Metode Kuantitatif

Metode kuantitatif dapat digunakan jika memenuhi kondisi yaitu pertama tersedia informasi tentang masa lalu, kedua informasi dapat dikuantitatifkan dalam bentuk data numerik, ketiga dapat diasumsikan bahwa beberapa aspek pola masa lalu akan terus berlanjut dimasa yang akan datang. Model prediksi kuantitatif dapat dibagi menjadi dua yaitu model time series dan model kausal. Model time series adalah model yang berdasarkan pada input data basis waktu, model ini dilakukan dengan mengamati pola, menggunakan model yang tepat dan kemudian dapat dilakukan prediksi. Sedangkan model kausal adalah model yang berdasarkan hubungan sebab akibat sehingga variabel model lebih dari satu (Efendi, 2010).

2.2.3 Data

Data adalah kumpulan fakta atau gambaran kejadian yang terekam dan tidak memiliki arti. Data dapat diartikan sebagai kumpulan fakta yang direpresentasikan kedalam beberapa bentuk baik angka, huruf, simbol yang diproses sehingga dapat menghasilkan sebuah informasi (Nofriansyah, 2015). Data adalah fakta dan gambaran secara umum tentang sesuatu kejadian, kegiatan, dan transaksi yang direkam, disimpan, dan diklasifikasikan yang tidak dapat digunakan oleh penggunanya sebelum data tersebut diolah sehingga menghasilkan suatu arti yang khusus (Turban, Aronson, & Liang, 2005). Ilmu data atau data science adalah ilmu yang mempelajari bagaimana mengungkapkan informasi dan pengetahuan dari kumpulan data dengan pendekatan sains seperti programming, statistika, database engineering, dan matematika. Ilmu data menggunakan teknik seperti klasifikasi,

regresi, dan cluster analysis seperti halnya data mining. Untuk mengimplementasikan data science maka diperlukan siklus pemrosesan data yang ditentukan oleh seorang data analis. Siklus pemrosesan data dapat dilihat berdasarkan gambar sebagai berikut.



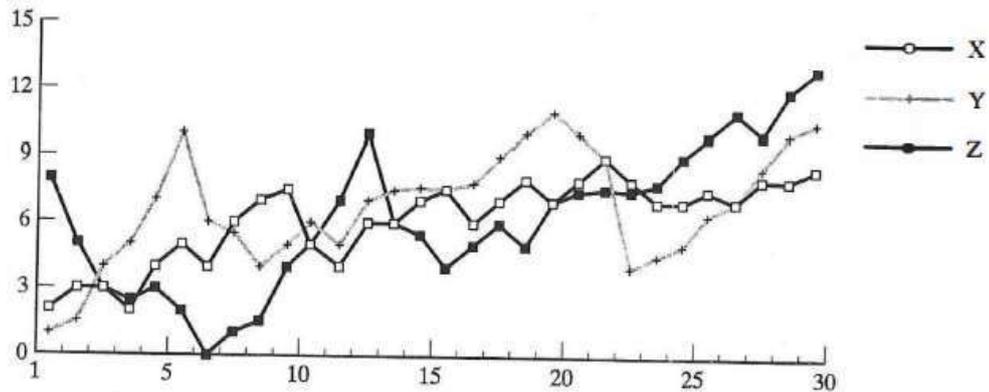
Gambar 2.1 Siklus Pemrosesan Data

Pada gambar 2.1 pertama data dikumpulkan dari berbagai sumber dan berbagai format baik gambar, teks, dan video. Proses ini dilakukan oleh seorang data engineer. Kedua setelah data dikumpulkan maka dilakukan pre-processing data dengan proses pembersihan data dan data transformation. pembersihan data atau cleansing data adalah proses mengoreksi data yang bias dan kurang lengkap yang tidak mempresentasikan data yang akan dianalisa. Sedangkan data transformation adalah proses mengubah data menjadi format yang berbeda seperti data linguistik menjadi data numerik, data string menjadi data datetime, dan data nilai kontinu menjadi diskrit. Proses pre-processing biasanya dilakukan oleh seorang data scientist. Ketiga dilakukan tahap exploratory data analysis yang digunakan untuk mendapatkan hasil analisis secara cepat tanpa mengimplementasikan *machine learning*. keempat dilakukan tahap modelling, tahap ini digunakan untuk mengimplementasikan *machine learning* dengan melakukan prediksi, deskriptif classifikasi, dan cluster analysis. Di tahap ini data yang digunakan dibagi menjadi data latih dan data uji untuk dapat mencari tingkat akurasi dari sistem yang

digunakan. Setelah mendapatkan hasil analisis maka dilakukan interpretasi data dengan visualisasi data kepada stakeholder (Wahyudin, Tosida, & Andria, 2019). Dalam penelitian ini penulis menggunakan data COVID-19 di Kota Bontang yang berbentuk numerik dan gambar infografis yang didapatkan di Satgas COVID-19 Kota Bontang. Data yang digunakan dalam penelitian ini adalah data jumlah terkonfirmasi positif, terkonfirmasi sembuh, dan terkonfirmasi meninggal akibat COVID-19 di Kota Bontang.

2.2.4 Model Deret Waktu

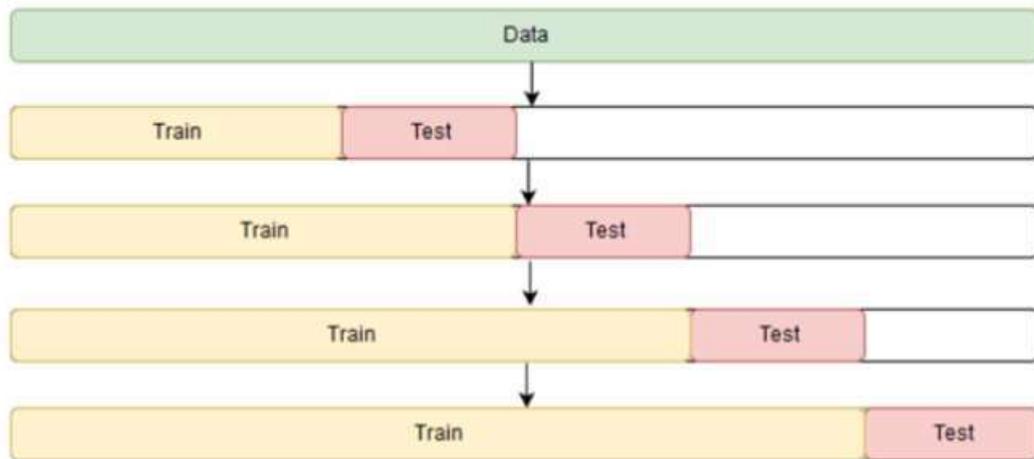
Deret waktu atau *time series* adalah himpunan observasi data terurut berdasarkan waktu. Metode *time series* adalah metode prediksi yang menggunakan analisa pola hubungan antara variabel waktu dengan variabel perkiraan, Umumnya pola data deret waktu terbagi menjadi empat yaitu horizontal, trend, musiman dan siklis. Pola horizontal adalah suatu pola kejadian yang tidak terduga dan bersifat acak tetapi kemunculannya dapat mempengaruhi fluktuasi data deret waktu. Pola *trend* adalah pola kecenderungan *time series* untuk meningkat, menurun atau tetap dalam jangka waktu yang panjang. Pola musiman adalah pola kejadian yang fluktuatif dari suatu data yang terjadi secara periodik dalam kurun waktu satu tahun seperti harian, mingguan, bulanan, triwulan, dan kuartalan. Pola siklis adalah pola perubahan data *time series* dalam jangka menengah yang disebabkan oleh keadaan yang berulang dalam siklus. Pola ini berlanjut dalam jangka waktu lebih dari setahun (Hanke & Wichers, 2005). Analisis *time series* adalah analisis sekumpulan himpunan data dalam suatu periode waktu yang lampau yang berguna untuk memprediksi bagaimana kondisi di masa depan. Plot dari deret waktu digunakan untuk memvisualisasi berdasarkan waktu kejadian. Berikut contoh dari deret waktu.



Gambar 2.2 Plot Deret Waktu

2.2.5 Cross Validation Pada Time Series

Dalam analisis *machine learning* prediksi deret waktu diperlukan pembagian untuk data latih dan data uji. Namun dalam kasus data deret waktu, tidak dapat dilakukan pembagian data secara acak karena kumpulan data deret waktu memiliki sifat autoregresif, tren, musiman, dan siklus yang tidak memungkinkan pemisahan data secara acak. Pemisahan data secara acak pada data runtun waktu mengakibatkan terjadi kebocoran data latih dan data uji sehingga angka hasil validasi bersifat baik namun ketepatan model buruk yang biasanya disebut autokorelasi. Oleh karena itu diperlukan pemisahan khusus untuk data deret waktu menggunakan split deret waktu atau *time series split*. *Time series split* didukung dengan menggunakan library *sklearn* pada *python* namun perlu ditentukan terlebih dahulu split yang digunakan. *Time series split* adalah fungsi untuk membagi data pelatihan dan data uji sebagai bagian tertentu dari akhir kumpulan data contohnya jika memiliki data observasi selama 10 tahun maka data latih yang digunakan adalah 7 tahun pertama dan data uji yang digunakan adalah 3 tahun terakhir (Miyaki, 2019). *Split validation* merupakan teknik validasi yang membagi data menjadi dua bagian menjadi data training dan data testing. Metode yang dapat digunakan untuk memvalidasi silang model deret waktu adalah validasi silang secara bergulir, jadi dimulai dengan sebagian kecil data untuk pelatihan, dan dilanjutkan dengan data uji. Berikut gambaran mengenai *cross validation* pada deret waktu.



Gambar 2.2.3 Cross Validation Time Series

2.3 Facebook Prophet

Facebook Prophet adalah prosedur untuk meramalkan data deret waktu berdasarkan model aditif yang dikembangkan oleh Facebook. *Facebook Prophet* awalnya digunakan untuk menangani masalah deret waktu di bidang bisnis namun saat ini banyak digunakan untuk analisis prediksi pada data covid karena karakteristik dari data covid yang bersifat time series dengan tingkat penyebaran yang eksponensial serta dipengaruhi oleh tindakan pencegahan karantina dan non karantina, maka tren nonlinear dengan sifat exponential smoothing di Facebook Prophet dinilai sangat cocok untuk data historis COVID-19. Selain itu berdasarkan penelitian dari (Battineni, Chintalapudi, & Amenta, 2020) model facebook prophet dapat berjalan dengan baik untuk populasi karantina dan non karantina serta dapat berjalan efektif untuk data di situasi pandemi saat ini. Facebook Prophet menggunakan model deret waktu yang mudah digunakan terdiri dari tiga komponen utama yaitu tren, musim, dan hari libur. Fungsi tren yaitu memodelkan perubahan non-periodik dalam deret waktu. Fungsi musim yaitu mewakili perubahan periodik (misalnya harian, mingguan, bulanan, dan tahunan). Fungsi libur yaitu mempresentasikan efek liburan yang terjadi pada jadwal yang tidak teratur selama satu atau beberapa hari. Dalam penelitian ini, penulis memfokuskan pada fungsi tren non-periodik dan mengecualikan fungsi musiman dan libur, karena pada kasus COVID-19 fungsi musiman tidak dapat digunakan karena membutuhkan minimal setahun untuk memanfaatkan musim tahunan juga sifat dari virus COVID-19 ini

bukanlah ada di setiap tahun sedangkan fungsi libur tidak dapat digunakan karena virus corona tidak terpengaruh oleh hari libur, hari kerja atau bulan lainnya. Bentuk komponen Prophet dalam bentuk rumus sebagai berikut.

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad 2.1$$

Dimana :

$y(t)$ adalah nilai perkiraan pada saat t waktu

$g(t)$ adalah tren non-periodik pada saat t waktu

$s(t)$ adalah musiman periodik (mingguan, bulanan, tahunan) pada saat t waktu

$h(t)$ adalah efek liburan tidak teratur pada t waktu

$e(t)$ adalah perubahan istimewa yang tidak diakomodasi oleh model (Belkacem, 2020)

Pada intinya, Facebook Prophet adalah model regresi dengan parameter yang dapat diinterpretasikan dan telah banyak digunakan (Taylor & Letham, 2018). Model ini berfungsi paling baik dengan deret waktu yang memiliki efek tren seperti COVID-19 selain itu prophet sangat kuat dalam menangani data yang hilang, pergeseran tren dan menangani data yang menyimpang terlalu jauh dari rangkaian data lainnya (outliers) dengan baik (Facebook Open Source, 2020). Penggunaan Facebook Prophet didasarkan atas kelebihanya yaitu :

1. Akurat, cepat, dan fleksibel

Prophet digunakan di banyak aplikasi di Facebook untuk menghasilkan perkiraan yang andal dalam rangka perencanaan dan penetapan tujuan. Menurut Facebook pendekatan prophet berperforma lebih baik dibandingkan pendekatan lainnya pada sebagian besar kasus yang telah dicoba. Selain itu kinerja dari Prophet sangat akurat dalam berbagai masalah termasuk penyakit epidemik (Tseng & Shih., 2019).

2. Penggunaan yang sepenuhnya otomatis

Fb Prophet dapat melakukan prakiraan yang otomatis pada data yang berantakan tanpa upaya manual. Prophet sangat mudah menangani outliers, data yang hilang, dan perubahan dramatis dalam deret waktu.

3. Dapat digunakan di bahasa Python dan R

Facebook telah menyediakan prosedur prophet di R dan Python yang dapat digunakan sesuai dengan bahasa yang disukai oleh peneliti dan data science (Facebook Open Source, 2020).

2.4 Python

Python adalah bahasa pemrograman tingkat tinggi dan termasuk bahasa berorientasi objek yang dinamis dan dikembangkan oleh Guido Van Rossum pada tahun 1990 di Amsterdam. Python dikenal sebagai bahasa pemrograman yang mudah dipelajari karena struktur sintaknya rapi dan mudah dipahami (Muahradian, 2018). Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbatasan kode. Python banyak digunakan di bidang sains, data science, robotika, antariksa, pemrograman, machine learning, ekonomi, dan developing game. Python memiliki sebuah library yang populer seperti Scipy dan Sckit package yang berguna untuk membuat aplikasi machine learning dan artificial intelligence. Selain itu terdapat matplotlib untuk membuat grafik dalam keperluan saintifik (Supardi, 2017).

Dengan berbagai library python memungkinkan programmer membuat aplikasi dengan menggunakan sintaks yang tampak sederhana. Sintaks atau source code dalam python akan dikompilasi menjadi format perantara yang dikenal dengan bytecode yang selanjutnya di eksekusi. Kelemahan bahasa python dibandingkan bahasa pemrograman lainnya terletak pada kecepatan eksekusinya selain itu python bukan perangkat bantu terbagus untuk pengantar komponen berperforma kritis (Lutz, 2013). Namun kelebihan bahasa ini adalah pertama source code yang mudah dibaca dan koheren sehingga memberikan kenyamanan pada programmer selain itu python bersifat ekspresif dengan penulisan koding yang to the point sehingga mudah di pelajari, kedua meningkatkan produktivitas programmer karena program dapat dijalankan secara langsung tanpa proses compiling dan linking yang biasanya dibutuhkan pada bahasa pemrograman lainnya selain itu python mudah dalam instalasi. Ketiga python dapat berjalan diberbagai perangkat sehingga tidak perlu penyesuaian, selain itu python menyediakan Graphical User Interface yang dapat digunakan di berbagai platform. Keempat python dilengkapi dengan library yang dapat digunakan programmer dan berorientasi obyek. Kelima python dapat

berintegrasi dengan berbagai komponen dan memanggil library C++ dan C (Lutz, 2013).

www.itk.ac.id

2.4.1 Library Python

Bahasa pemrograman python memiliki lebih dari 100.000 library yang bukan hanya membantu mengembangkan program yang terintegrasi namun dapat menyelesaikan masalah berupa analisis data dengan membuat visualisasi data dari berbagai tampilan library. Kemudian menghubungkannya ke web dengan python sebagai back-end servernya. Dalam penelitian Tugas Akhir ini, penulis menggunakan library sebagai berikut.

1. Pandas

Pandas adalah kependekan dari python for data analysis yang berfokus pada manipulasi data, persiapan data, dan cleaning data. Nama pandas berasal dari Word Panel Data-Ekonometrika dari Data Multidimensi pada tahun 2008 oleh Wes Mckinney yang mengembangkan pandas untuk analisis data. Dengan menggunakan pandas, developer dapat menyelesaikan lima langkah umum dalam pemrosesan dan analisis data (Tutorials Point, 2017). Pandas adalah tools analisa data untuk python yang bersifat open source yang menyediakan struktur dan operasi data berperforma tinggi sehingga mudah digunakan dalam memanipulasi data numerik dan runtun waktu (Ardhiansyah & Ratih, 2020). Pandas banyak digunakan dilingkungan akademik, komersial, finansial, ekonomi, statistika, dan analitik. Berikut fitur utama pandas yaitu.

- a. Objek dataframe yang efisien dan cepat dengan pengindeksan yang otomatis dan disesuaikan.
- b. Tools untuk memuat data ke dalam objek data dengan berbagai format file
- c. Menyelaraskan data yang hilang
- d. Fungsionalitas time series
- e. Penanganan indeks tanggal-waktu secara eksplisit dalam data dan rentang waktu
- f. Memudahkan metode reshape dan resampling seperti down sampling, up sampling, dan agregasi

www.itk.ac.id

- g. Kolom dari struktur data dapat dihapus atau disisipkan (Tutorials Point, 2017).

www.itk.ac.id

Pandas menggabungkan fitur komputasi array berkinerja tinggi dari numpy dengan kemampuan manipulasi data yang fleksibel dari spreadsheet dan database relasional seperti SQL (Mckinney, 2013).

2. Matplotlib

Matplotlib adalah library python yang sangat populer digunakan untuk menampilkan hasil analisis berupa gambar ilmiah baik dua dimensi maupun tiga dimensi. Matplotlib awalnya dibuat oleh John D Hunter pada tahun 2002 dan sekarang dikembangkan oleh tim developer yang besar. Matplotlib adalah library visualisasi data multiplatform yang dibangun di atas array numpy dan dirancang untuk bekerja dengan tumpukan SciPy yang lebih luas. Salah satu fitur matplotlib yang paling penting adalah kemampuan untuk menjalankan banyak operator sistem dan backend grafis (VanderPlas, 2017). Matplotlib memiliki sejumlah perangkat tambahan seperti mplot3d untuk plot 3D dan peta dasar untuk pemetaan dan proyeksi.

3. Numpy

Numpy adalah kependekan dari Numerical Python yang digunakan untuk memanipulasi array dan matriks untuk memproses data numerik dalam skala besar. Numpy memuat fungsi built-in seperti melakukan transformasi fourier diskrit, melakukan operasi aljabar, membuat, menggabungkan dan memisahkan array N-dimensional, membuat matriks khusus, membuat histogram, membentuk dasar library lain seperti scikit-learn dan operasi himpunan matematika lainnya. Numpy banyak digunakan dalam analisis data dan menggantikan MATLAB saat penggunaan matplotlib (Herho, 2017). Numpy menyediakan fitur antara lain.

- a. N- dimensional objek array yang cepat dan efisien
- b. Fungsi untuk melakukan perhitungan element-wise dengan array
- c. Tools untuk membaca dan menulis kumpulan data berbasis array ke disk
- d. Alat untuk mengintegrasikan kode pemrograman bahasa C, C++ dan Fortran ke Python (Mckinney, 2013).

4. Scikit-Learn

www.itk.ac.id

Scikit-Learn adalah sebuah library open source serbaguna untuk analisis data yang ditulis dengan bahasa pemrograman python. Scikit-learn berisi sejumlah implementasi untuk berbagai algoritma machine learning yang populer. Untuk menginstal scikit-learn dibutuhkan python dengan versi diatas 2.6, numpy versi diatas 1.6.1, dan SciPy dengan versi lebih dari 0.9 (Alleo, et al.). scikit-learn menampilkan berbagai algoritma machine learning baik supervised dan unsupervised learning yang menggunakan interface dan berorientasi task yang konsisten, sehingga memungkinkan perbandingan metode yang mudah untuk aplikasi tertentu. Karena scikit learn bergantung pada ekosistem python maka, ia dengan mudah diintegrasikan ke dalam aplikasi di luar range tradisional analisis data statistik (Michel, Gramfort, Varoquaux, Eger, & Thirion, 2011).

2.5 Jupyter Notebook

Untuk melakukan analisis data, peneliti menggunakan bahasa pemrograman python dengan sebuah interpreter yang interaktif. Interpreter adalah suatu kode python yang dieksekusi baris perbaris. Salah satu interpreter python yang terkenal dan interaktif adalah jupyter notebook. Jupyter Notebook adalah tools antarmuka grafis berbasis browser ke shell Ipython, yang harus terhubung ke proses python yang sedang berjalan untuk mengeksekusi kodenya. Sebelum mengeksekusi kode maka browser akan secara default terbuka dengan URL lokal (<http://localhost:8888/>) (VanderPlas, 2017). Selain itu jupyter notebook adalah aplikasi antar muka berbasis website yang digunakan untuk membuat program analisis yang menggunakan metode machine learning sehingga dapat memvisualisasi hasil analisisnya (Budiarti, 2017). Untuk menginstall jupyter notebook maka melalui aplikasi Anaconda Data Science Platform yang dapat diunduh melalui www.anaconda.com. Pada penelitian ini penulis menggunakan bahasa python dengan versi 3.7 yang akan secara default disarankan oleh Anaconda. Penggunaan aplikasi Anaconda memberikan kemudahan seperti melakukan konfigurasi semua library yang berhubungan dengan metode yang digunakan selain itu dapat menggunakan environment sesuai kebutuhan project tanpa mengganggu environment project lainnya. Kemudian Anaconda menyediakan akses tools data science dan developer software seperti Microsoft Visual Studio Code, R Studio, Spyder dan software lainnya

2.6 Evaluasi Performa Model Prediksi

Untuk mengetahui keakuratan dari hasil prediksi yang telah dilakukan maka perlu diuji kinerja modelnya, hal ini biasanya disebut dengan proses evaluasi atau validasi model. Terdapat beberapa metode yang dapat digunakan untuk mengevaluasi model hasil prediksi. Dalam penelitian ini penulis menggunakan mean absolute error (MAE) dan confusion matrix dengan penjelasan sebagai berikut

2.6.1 Mean Absolute Error (MAE)

MAE adalah rata-rata nilai absolut dari kesalahan meramal. Mean absolute error adalah perbedaan rata-rata antara kedua metode yang dihitung. Mean Absolute Error (MAE) merupakan pengukuran yang umum digunakan untuk prediksi error pada analisis runtun waktu. Dalam ukuran MAE, model prediksi dapat dikatakan paling baik apabila nilai MAE mendekati angka 0 (nol). Nilai MAE ini menyatakan rata-rata besar kesalahan atau error pada model yang digunakan. (Mahmoud, Abdel-Nasser, & Kashef, 2020). Berikut rumus dari MAE.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Aktual - Forecast| \quad 2.2$$

Dimana

Aktual-Forecast : hasil pengurangan antara nilai asli dengan nilai hasil peramalan

n : jumlah periode yang digunakan dalam perhitungan (L. Y. Wei, 2016).

2.6.2 Mean Absolute Percentage Error (MAPE)

MAPE atau *mean absolute percentage error* adalah rata-rata nilai persentase absolut positif dari kesalahan meramal. MAPE merupakan perhitungan yang menunjukkan nilai mutlak dari nilai aktual dan nilai prediksi (Kharista, 2018). Berikut rumus dari MAPE.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Aktual - Forecast|}{Aktual} \times 100\% \quad 2.3$$

Dimana

Aktual-Forecast : hasil pengurangan antara nilai asli dengan nilai hasil peramalan

n : jumlah periode yang digunakan dalam perhitungan (L. Y. Wei, 2016).

Hasil peramalan dengan nilai MAPE menunjukkan semakin kecil nilai persentasenya maka hasilnya akan semakin akurat. Berikut tabel kriteria penilaian akurasi dari nilai MAPE menurut Lewis (Lewis, 1982).

Tabel 2.1 Kriteria Akurasi Prediksi

MAPE	Interpretasi
<10%	Kemampuan akurasi peramalan sangat baik
10-20%	Kemampuan akurasi peramalan baik
20-50%	Kemampuan akurasi peramalan layak
>50%	Kemampuan akurasi peramalan buruk

2.7 Dashboard

Dashboard adalah antar muka pengguna yang berada diantara data dan desain yang menampilkan berbagai indikator, grafik, dan visualisasi data yang bertujuan untuk membantu pengguna membuat keputusan yang tepat, akurat, dan cepat berdasarkan data yang tersedia. Dashboard adalah platform berbasis GUI yang menyediakan bagan, laporan, dan grafis yang saling terkait antara data. Sumber data dari dashboard terdiri data internal berupa laporan, database, dan dokumen, serta data external organisasi (Maharani, 2020).

Dashboard adalah sebuah platform yang digunakan untuk memvisualisasi data agar informasi yang didapatkan mudah untuk dipahami. Dashboard banyak digunakan di perusahaan atau bisnis agar mempermudah visualisasi performance perusahaan ke stakeholder. Dashboard dapat digunakan untuk menampilkan analisa hasil perkiraan berdasarkan data terdahulu menggunakan gambar grafik (Malik, 2005).

2.8 Google Data Studio

Google data studio adalah program untuk visualisasi data yang diluncurkan oleh Google pada Maret 2016 sebagai bagian dari Google Analytics 360 Suite. Google data studio dapat digunakan secara gratis secara global pada Maret 2017. Google data studio merupakan alat untuk memvisualisasi data yang digunakan

untuk membantu orang memahami *insight* data dengan menempatkan data dalam bentuk visual selain itu, membuat laporan yang lebih menarik dan dinamis sehingga mudah dipahami. Google data studio menggunakan skema warna dan template yang konsisten sehingga mudah digunakan. Google data studio mengandalkan berbagai tools visualisasi seperti grafik dan chart dengan menambahkan fitur seperti filter, maximum, minimum, location, dan aggregate. Selain itu dapat mengintegrasikan berbagai sumber data external dalam suatu dashboard baru. Google data studio dapat diakses melalui link <https://datastudio.google.com> (Putra, 2020).

2.9 Penelitian Terdahulu

Berikut adalah rangkuman hasil penelitian terdahulu yang memiliki keterkaitan dengan penelitian prediksi data COVID 19.

Tabel 2.2 Penelitian terdahulu

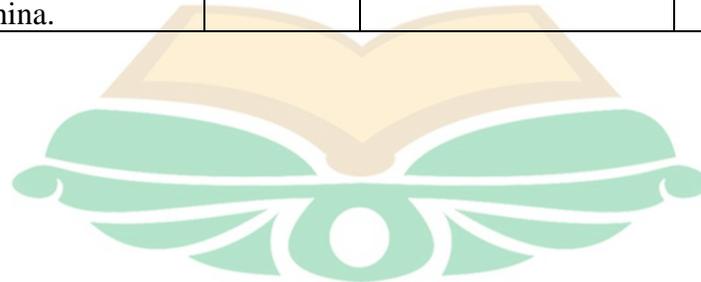


No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
1	(Wiguna, et al., 2020)	2020	Data runtun waktu COVID DKI Jakarta	Maret-awal juli 2020	ARIMA, Analisis dan visualisasi data dilakukan menggunakan program Python dan Tableau.	MAPE dengan nilai eror 20,97% dan Root Mean Squared Error (RMSE) sebesar 42,415.	Hasil berupa analisis data COVID-19 berdasarkan jenis kelamin, kelompok umur, hasil uji lab, dan kenaikan kasus harian beserta prediksinya. Dimana hasil analisis prediktif memperlihatkan tren kasus positif harian yang cenderung naik di kurun waktu 14 hari ke depan dari data yang digunakan
2	(Watratana, B., & Moeis, 2020)	2020	Data COVID-19 Indonesia	15 Mei 2020	Naive bayes dengan tools WEKA.	Correctly Classified Instance adalah sebesar 48,4848%	Hasilnya metode Naive Bayes menunjukkan bahwa 16 data dari 33 data yang di uji dalam kasus Covid-19 per provinsi dengan keakuratan sebesar 48,4848%, di mana dari 33 data yang di uji dalam kasus Covid-19 per provinsi yang diuji terdapat 16 data yang berhasil diklasifikasikan dengan benar.

No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
3	(Ardhiansyah & Ratih, 2020)	2020	Data COVID-19 Indonesia	Maret-Juli	SDLC, penggalian data runtun waktu dengan prosedur Facebook Prophet	MAE dan MAPE	Hasil prediksi jumlah pasien terjangkit dan pasien meninggal berturut turut pada percobaan 1 sebanyak 262 orang dan 18 orang, sedangkan pada percobaan 2 sebanyak 287 orang dan 10 orang, serta percobaan 3 sebanyak 656 orang dan 98 orang
4	(Chandra & Budi, 2020)	2020	data registrasi mahasiswa UK. Maranatha	1993-2019	Compare metode ARIMA dan Prophet	RMSE	Model Prophet mampu berkinerja lebih unggul dari ARIMA berdasarkan dari nilai RMSE
5	(Albana & Azhari, 2020)	2020	Data corona di Surabaya	Maret hingga 23 April 2020.	Metode yang digunakan adalah Simulasi Monte Carlo.	Galat under prediction dan over-prediction.	Nilai galat terbesar dari under prediction adalah 11% sedangkan untuk over-prediction sebesar 23%.



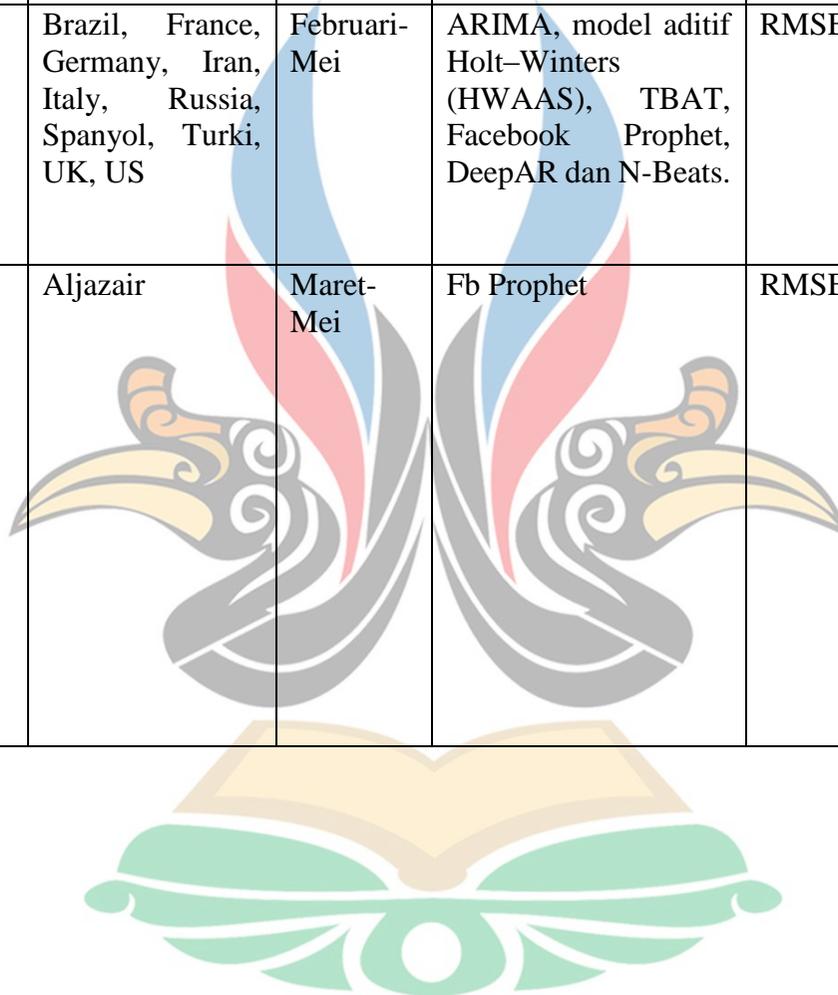
No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
6	(Abdulmajeed, Adeleke, & Popoola, 2020)	2020	Data Nigeria	27 Februari – 5 April 2020	Gabungan ARIMA, FB Prophet, dan Exponential Holt-Winters yang dikombinasikan dengan GARCH dan diimplementasikan menggunakan brute search dengan bahasa pemrograman python.	-	visualisasi data perbandingan kasus COVID-19 di Nigeria dengan Afrika selatan dan dampak kebijakan.
7	(Hamzah, et al., 2020)	2020	Data COVID-19 di dunia berasal dari Universitas John Hopkins, WHO dan DingXiangYuan, situs web yang disahkan oleh pemerintah China.	240 hari dari tanggal 20 Januari 2020	Pemodelan prediktif Susceptible-Exposed-Infectious-Recovered (SEIR).	-	Jumlah kasus yang dikonfirmasi diperkirakan melebihi 76.000 kasus dan mencapai puncak wabah sebelum 20 Februari 2020.



No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
8	(Gaur, 2020)	2020	AS, China, Spanyol dan Italia	22 Januari 2020 hingga 19 Mei 2020	model ARIMA by FB-Prophet	RMSE	hasil prediksi menggunakan grafik dan perbandingan tiap negara yaitu India negara yang bagus dalam pengendalian COVID-19 dibandingkan dengan Amerika, China, Italia, dan Spanyol
9	(Narin, Kaya, & Pamuk, 2020)	2020	Data 341 pasien yang terinfeksi virus corona pneumonia menggunakan radiografi rontgen dada	-	jaringan saraf konvolusional (CNN)	Confusion matrix	model ResNet50 pra-terlatih memberikan kinerja klasifikasi tertinggi (akurasi 96,1% untuk Dataset-1, akurasi 99,5% untuk Dataset2 dan akurasi 99,7% untuk Dataset-3) di antara empat model yang digunakan.
10	(Pandey, Chaudhary, Gupta, & Pal, 2020)	2020	Data COVID-19 di India	Januari-Maret	Model SEIR dan model Regresi	RMSLE. Dengan nilai akurasi untuk model SEIR dicapai 1,52 dan model regresi dicapai 2.01	Dalam dua minggu ke depan prediksi kasus naik antara 5000-6000 kasus

No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
11	(Ardabili, et al., 2020)	2020	Italia, Jerman, Iran, AS, dan China	Januari-Maret	multi-layered perceptron (MLP), dan sistem inferensi fuzzy berbasis jaringan adaptif (ANFIS)	RMSE	MLP lebih akurat dibandingkan ANFIS
12	(Siemuri, Alabi, & Elmusrati, 2020)	2020	Finlandia, US, China, UK, Perancis, Jerman, Itali, Kanada, Iran, Swedia, Mesir, dan Nigeria	7 April-3 Mei 2020	Facebook Prophet	Validate forecasted manual	Grafik peramalan tren virus untuk tanggal 8 sampai 14 Mei 2020. Model kami menunjukkan kinerja yang baik jika dibandingkan dengan laporan resmi dari Organisasi Kesehatan Dunia. Akurasi perkiraan rata-rata modelnya adalah 79,6%.
13	(Ghamizi, et al., 2020)	2020	Belgium, Perancis, Jerman Yunani, Italia Luksemburg Spanyol, Brazil Kamerun, Kanada, Jepang Britania Raya	Mei-April	Regresi waktu SEIR dengan model DN-SEIR	RMSE	menghasilkan prediksi dengan tingkat kesalahan yang jauh lebih rendah daripada model epidemiologi murni dalam 75% kasus dan mencapai skor R2 95% ketika pembelajaran ditransfer dan diuji di negara-negara yang tidak terlihat.

No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
14	(Papastefanopoulos, Linardatos, & Kotsiantis, 2020)	2020	Brazil, France, Germany, Iran, Italy, Russia, Spanyol, Turki, UK, US	Februari-Mei	ARIMA, model aditif Holt-Winters (HWAAS), TBAT, Facebook Prophet, DeepAR dan N-Beats.	RMSE	Hasil berupa Prediksi persentase kasus aktif dan melihat tujuh hari ke depan, untuk sepuluh negara dengan jumlah kasus terkonfirmasi tertinggi pada 4 negara di bulan Mei 2020.
15	(Belkacem, 2020)	2020	Aljazair	Maret-Mei	Fb Prophet	RMSE	Hasil evaluasi menunjukkan bahwa deret waktu Prophet akurat memprediksi jumlah kasus dan kematian di Aljazair dengan skor RMSE rendah masing-masing 218,87 dan 4,79, sedangkan perkiraan menunjukkan bahwa jumlah kasus dan kematian diperkirakan akan meningkat dalam beberapa minggu mendatang.



No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
16	(Erraissi, Azouazi, Belangour, & Banane, 2020)	2020	Maroko	Maret-April	menggunakan machine learning spark ML dengan algoritma decision tree, gaussian naive bayes, support vector machine, logistic regression, dan random forest.	Confusion Metrix	Hasilnya menyatakan masyarakat maroko pada bulan suci ramadhan tidak mengikuti instruksi untuk melawan covid (tidak patuh).
17	(Topping, et al., 2020)	2020	Data kualitas udara pada masa COVID-19 di Manchester Inggris	2015-2020	Facebook Prophet dan model regional EMEP	Evaluasi model validasi silang dan penyesuaian data historis	metode prophet lebih baik dibandingkan The European Monitoring and Evaluation Programme for Transboundary Long-Range Transported Air Pollutants (EMEP) model



No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
18	(Agbehadji, Awuzie, & Ngowi, 2020)	2020	Data COVID-19	2020	menggunakan big data analisis, machine learning dengan algoritma CNN, RNN, LSTM	-	Penerapan model kecerdasan buatan dan Big data analisis yang digabungkan dapat membantu identifikasi dan tracing COVID-19 selain itu dengan penggunaan model CNN berbasis kecerdasan buatan dapat meningkatkan akurasi identifikasi fitur COVID-19
19	(Pinter, Felde, Mosavi, Ghamisi, & Gloaguen, 2020)	2020	data dari Hungary	4 Maret-19 April	pendekatan Machine learning hybrid untuk memprediksi COVID-19, menggunakan sistem inferensi fuzzy berbasis jaringan adaptif (ANFIS) dan algoritma persaingan perceptron-imperialis berlapis (MLP-ICA).	MAPE dan MSE	total kasus di prediksi dari 20-28 april menggunakan MPL-ICA berjumlah 167.88 dan menggunakan ANFIS 194.10 sedangkan untuk total mortality rate pada MLP-ICA berjumlah 8,32 dan ANFIS berjumlah 15,25.



No	Peneliti	Tahun	Data	Periode Data	Metode	Evaluasi	Hasil
20	(Ayyoubzadeh, Ayyoubzadeh, Zahedi, Ahmadi, & Kalhori, 2020)	2020	Data Covid-19 di Iran	15 Februari-18 Maret 2020	Model regresi linier dan memori jangka pendek panjang (LSTM)	Validasi silang dan RMSE	Model regresi linier memprediksi kejadian dengan RMSE 7,562 (SD 6,492).RMSE model LSTM adalah 27.187 (SD 20.705

