ANALISIS PERBANDINGAN METODE MACHINE LEARNING UNTUK MEMPREDIKSI KASUS COVID-19 DI DUNIA

TESIS

Oleh :

Don Ardhito

1911601449

PROGRAM STUDI MAGISTER ILMU KOMPUTER

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS BUDI LUHUR

JAKARTA

GENAP 2020/2021

ANALISIS PERBANDINGAN METODE MACHINE LEARNING UNTUK MEMPREDIKSI KASUS COVID-19 DI DUNIA

TESIS

Diajukan untuk memenuhi salah satu persyaratan memperoleh gelar Magister Ilmu Komputer (M.Kom)

Oleh :

Don Ardhito

1911601449

PROGRAM STUDI MAGISTER ILMU KOMPUTER

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS BUDI LUHUR

JAKARTA

GENAP 2020/2021

PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama	: Don Ardhito
Nomor Induk Mahasiswa	: 1911601449
Program Studi	: Magister Ilmu Komputer
Bidang Peminatan	: Teknologi Sistem Informasi
Jenjang Studi	: Strata 2
Judul	: KOMPARASI MODEL UNTUK MEMPREDIKSI JUMLAH PENDERITA COVID-19 DI DUNIA

Laporan Proposal Tesis ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Sabtu 14 Agustus 2021

Tim Penguji:

Ketua	: Prof. Dr. Moedjiono, M.Sc
Anggota	: Dr. Achmad Solichin, S.Kom., M.T.I
Pembimbing	: Prof. Ir. Dana Indra Sensuse, Ph.D.
Ketua Program Studi	: Dr. Rusdah, S.Kom., M.Kom.

PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR

SURAT PERNYATAAN TIDAK PLAGIAT DAN PERSETUJUAN PUBLIKASI

Nama	: Don Ardhito
Nomor Induk Mahasiswa	: 1911601449
Konsentrasi	: Teknologi Sistem Informasi
Jenjang Studi	: Strata-2
Fakultas	: Teknologi Informasi

Menyatakan bahwa Tesis yang berjudul :

"ANALISIS PERBANDINGAN METODE MACHINE LEARNING UNTUK MEMPREDIKSI KASUS COVID-19 DI DUNIA"

Merupakan :

1. Karya tulis saya sebagai laporan tesis yang asli dan belum pernah diajukan untuk mendapatkan gelar akademik apapun, baik di Universitas Budi Luhur maupun di perguruan tinggi lainnya.
2. Karya tulis ini bukan saduran/terjemahan, dan murni gagasan, rumusan dan pelaksanaan penelitian/implementasi saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing akademik dan pembimbing di organisasi tempat riset.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan dalam naskah ini dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Saya menyerahkan hak milik atas karya tulis ini kepada Universitas Budi Luhur, dan oleh karenanya Universitas Budi Luhur berhak melakukan pengelolaan atas karya tulis ini sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh berdasarkan karya tulis ini, serta sanksi lainnya sesuai dengan norma di Universitas Budi Luhur dan Undang-Undang yang berlaku.

Jakarta, 14 Januari 2022

Don Ardhito

ABSTRAK

ANALISIS PERBANDINGAN METODE MACHINE LEARNING UNTUK MEMPREDIKSI KASUS COVID-19 DI DUNIA

Oleh :

Don Ardhito

1911601449

Berdasarkan data dari WHO, lonjakan kasus penderita Covid-19 telah terjadi beberapa kali sejak bulan Januari 2020 dan menyebabkan pemerintah mengalami kesulitan dalam menyediakan fasilitas kesehatan. Pada bulan Desember 2020 terjadi lonjakan kasus sebanyak 100% dari jumlah 400.000 penderita menjadi 800.000 dalam waktu 15 hari. Dan pada bulan April 2021 terjadi lonjakan kasus sebanyak 60%, dari 500.000 penderita menjadi 800.000 penderita. Munculnya varian baru (Delta) dan ketidaktaatan masyarakat terhadap prokes membuat penyebaran virus Covid-19 menjadi semakin luas, dan kurangnya antisipasi membuat pemerintah mengalami kesulitan dalam menyediakan fasilitas dan tenaga kesehatan untuk menangani pasien yang semakin banyak. Salah satu langkah antisipasi yang dapat dilakukan adalah dengan melakukan *forecasting* jumlah data penderita yang akan terjadi di masa depan sehingga pemerintah dapat mempersiapkan fasilitas berdasarkan data hasil dari *forecasting* sebagai gambaran untuk mengambil keputusan dalam menangani wabah ini. Beberapa penelitian untuk memprediksi jumlah lonjakan kasus di masa depan telah dilakukan oleh banyak pihak menggunakan model *forecasting* yang berbeda-beda, sehingga menghasilkan performa yang berbeda antara satu dengan lainnya. Perbedaan hasil penelitian tersebut dari sisi performa seperti akurasi dan mean error membuat hasil peramalan menjadi tidak konsisten antara satu dengan yang lain. Sehingga diperlukan suatu komparasi untuk mengetahui model yang paling tepat untuk digunakan dalam melakukan *forecasting* data penderita Covid-19. Hasil yang diharapkan dari penelitian ini adalah berupa rekomendasi model penelitian yang menghasilkan performa terbaik untuk melakukan *forecasting* data penderita Covid-19 di masa depan.

Kata kunci : Komparasi, Model Prediksi, Coronavirus, Covid-19, Model, Prediksi

ABSTRACT

ANALYSIS COMPARISON OF MACHINE LEARNING METHODS TO PREDICT COVID-19 CASES AROUND THE WORLD

By :

Don Ardhito

1911601449

According to the data from the WHO, spike in cases of Covid-19 was happened a few times since the early of 2020 causing the authorities to experience difficulties in providing health facilities. In December 2020, the first spike happened and the data showed 100% increment in cases from 400.000 cases to 800.000 cases in only 15 days. Then, in April 2021 the second spike happened and the data showed 60% increment in cases from 500.000 cases to 800.000 cases. The emergence of the new Covid-19 variant (Delta) and society disobedience towards government's health protocol regulation resulted the vast spread of the Covid-19 virus, and the lack of anticipation from the government and the society causes public to experience the difficulties to provide health facilities and workers to handle the fast-growing number of people infected by Covid-19. In order to overcome this problem, one of the anticipatory steps that can be taken is by conducting a forecast for the future data of the number of the Covid-19 cases so that the authorities will have a picture on how many resources needed to counter the outbreak. Several studies to predict the number of spikes in cases in the future have been carried out by many parties using different forecasting models. The differences in the results of these studies in terms of performance such as accuracy and mean error caused the forecasting results inconsistent with one another. Therefore, a comparison to find an appropriate model to use in forecasting is required. The expected results from this research are in the form of recommendations for research models that produce the best performance to forecast Covid-19 casualties in the future.

Key Words : Comparison, Prediction Model, Coronavirus, Covid-19, Model, Prediction

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayahNya, sehingga penyusunan tesis ini dapat selesai dengan baik. Proses penyelesaian tesis ini berkat dukungan moril dan materiil dari banyak pihak, untuk itu penulis mengucapkan terimakasih dan penghargaan sebesar-besarnya kepada :

1. Dr. Ir. Wendi Usino, M.Sc., M.M., selaku Rektor dan para wakil Rektor di lingkungan Universitas Budi Luhur, yang telah memberikan kesempatan dan fasilitas kepada penulis untuk melakukan studi lanjut.
2. Dr. Deni Mahdiana, M.M., M.Kom., selaku Dekan Fakultas Teknologi Informasi Universitas Budi Luhur atas kesempatan dan dukungan penuh kepada penulis untuk menjadi bagian dari Program Magister Ilmu Komputer Universitas Budi Luhur.
3. Dr. Rusdah, M.Kom, selaku ketua Program Studi Magister Ilmu Komputer Universitas Budi Luhur beserta seluruh staf administrasi yang telah memberikan pelayanan dan dukungan akademik dengan penuh perhatian selama penulis menempuh studi.
4. Prof. Ir. Dana Indra Sensuse, M.LIS., Ph.D., selaku pembimbing dengan penuh perhatian dan kesabaran dalam memberikan bimbingan, arahan, saran, kritik dan solusi dalam penyempurnaan penulisan tesis ini.
5. Bapak dan Ibu dosen pada Program Magister Ilmu Komputer Universitas Budi Luhur yang telah memberikan bekal ilmu dan materi pembelajaran yang sangat bermanfaat bagi penulis selama menempuh studi.
6. Dr. Sri Rahayu SE, M.Si, selaku ibu penulis, Suryanto, S.Kom, M.M, selaku ayah penulis beserta Adrian Edenito dan Julio Evan Desnito selaku adik-adik penulis yang telah memberikan pengertian dan dukungannya dengan penuh kesabaran selama penulis menempuh studi.
7. Berbagai pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu dalam penyelesaian tesis ini.

Penulis menyadari bahwa tesis ini memiliki kekurangan dan ketidaksempurnaan, sehingga masih membutuhkan masukan, kritik dan saran untuk perbaikan. Semoga tesis ini dapat bermanfaat bagi pengembangan ilmu khususnya Ilmu Komputer dan menjadi referensi bagi pembaca dan pihak-pihak lain yang membutuhkan.

Jakarta, Agustus 2021

Penulis

(Don Ardhito)

DAFTAR ISI

LEMBAR PENGESAHAN iii
SURAT PERNYATAAN TIDAK PLAGIAT DAN PERSETUJUAN PUBLIKASI iv
ABSTRAK v
KATA PENGANTAR vii
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR GAMBAR xi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Masalah Penelitan 3
1.2.1 Identifikasi Masalah 3
1.2.2 Pembatasan Masalah 3
1.2.3 Rumusan Masalah 3
1.3 Tujuan dan Manfaat Penelitian 3
1.3.1 Tujuan Penelitian 3
1.3.2 Manfaat Penelitian 3
1.4 Tata-Urut Penulisan 4
BAB II LANDASAN TEORI DAN KERANGKA KONSEP/PEMIKIRAN 5
2.1 Tinjauan Pustaka 5
2.1.1 Coronavirus 5
2.1.2 Covid-19 6
2.1.3 Forecasting 7
2.1.4 *Time Series Analysis* 9
2.1.5 Regresi Linier Berganda 10
2.1.6 Support Vector Machine 10
2.1.7 ARIMA 11
2.1.8 Exponential Smoothing 13
2.2 Tinjauan Studi 18
2.3 Tinjauan Objek Penelitian 28
2.3.1 Profil Organisasi Kementerian Kesehatan 28
2.4 Kerangka Konsep/Pola Pikir Pemecahan Masalah 30

2.5 Hipotesis 32
BAB III METODOLOGI DAN RANCANGAN/DESAIN PENELITIAN 33
3.1 Metode Penelitian 33
3.1.1 Metode Data Mining 34
3.2 Sampling/Metode Pemilihan Sampel 37
3.3 Metode Pengumpulan Data 37
3.4 Langkah-langkah Penelitian 37
3.4.1 Identifikasi Masalah 38
3.4.2 Merumuskan dan Membatasi Masalah 38
3.4.3 Melakukan Studi Kepustakaan 38
3.4.4 Merumuskan Hipotesis atau Pertanyaan Penelitian 38
3.4.5 Menentukan Desain dan Metode Penelitian 39
3.4.6 Menyusun Instrumen dan Mengumpulkan Data 39
3.4.7 Menerapkan Metodologi untuk Memproses Data 39
3.4.8 Memberikan Kesimpulan 39
3.5 Jadwal Penelitian 40
BAB IV PEMBAHASAN 42
4.1 Business Understanding 42
4.2 Data Understanding 42
4.3 Data Preparation 43
4.4 Modelling 43
4.4.1 ARIMA 44
4.4.2 Exponential Smoothing 45
4.4.3 Linear Regression 46
4.4.4 Support Vector Regression 47
4.4.5 LSTM 48
4.5 Evaluation 49
4.6 Deployment 50
BAB V PENUTUP 66
5.1 Kesimpulan 66
5.2 Saran 66
DAFTAR PUSTAKA 67
LAMPIRAN 70

DAFTAR TABEL

Tabel 2.1 Varian Virus Corona dan Asalnya...……6

Tabel 2.2 Perbandingan Penelitian Terdahulu…….…...22

Tabel 3.1 Variabel pada Dataset yang Digunakan……..…….…..…….34

Tabel 3.2 Hasil Statistik Dataset Menggunakan Rapid Miner Studio……..…...35

Tabel 3.3 Timeline Pelaksanaan Penelitian…….…….…..40

Tabel 4.1 Hasil Percobaan Menggunakan ARIMA…….…….…..45

Tabel 4.2 Hasil Percobaan Menggunakan Exponential Smoothing…….…..46

Tabel 4.3 Hasil Percobaan Menggunakan Linear Regression……..…..47

Tabel 4.4 Hasil Percobaan Menggunakan Support Vector Regression...…….…..48

Tabel 4.5 Hasil Percobaan Menggunakan LSTM...…….…….…..49

Tabel 4.6 Peringkat Hasil Percobaan dari Semua Algoritma pada Training menggunakan RMSE.…….……..…..49

Tabel 4.7 Peringkat Hasil Percobaan dari Semua Algoritma pada Training menggunakan Absolute Error.…….…….…..50

Tabel 4.8 Hasil Pengolahan Data Menggunakan Aplikasi Prototipe.……63

DAFTAR GAMBAR

Gambar 1.1 Total Terkonfirmasi Positif (Kiri) dan Harian Terkonfirmasi Positif (Kanan)(WHO, 2021)……..…….……1

Gambar 1.2 Total Kematian (Kiri) dan Harian Kematian (Kanan) (WHO, 2021)...1

Gambar 2.1 Susunan Recurrent Neural Network…….……17

Gambar 2.2 Susunan Long Short-Term Memory……..18

Gambar 2.3 Struktur Organisasi Kementerian Kesehatan…...……..30

Gambar 2.4 Kerangka Konsep Penelitian……...……..31

Gambar 3.1 Metodologi Penelitian……...33

Gambar 3.2 Model untuk Mencari Korelasi……..35

Gambar 3.3 Hasil Evaluasi Korelasi Antar Variabel…….36

Gambar 3.4 Tahapan Penelitian……38

Gambar 4.1 Dataset Kasus Penderita Covid-19 dari Seluruh Dunia.……43

Gambar 4.2 Desain Model ARIMA.…….44

Gambar 4.3 Validation Data Training pada algoritma ARIMA.……44

Gambar 4.4 Desain Model Exponential Smoothing.…….45

Gambar 4.5 Validation Data Training pada algoritma Exponential Smoothing...45

Gambar 4.6 Desain Model Linear Regression…….…….46

Gambar 4.7 Validation Data Training pada algoritma Linear Regression…….46

Gambar 4.8 Desain Model Support Vector Regression…….47

Gambar 4.9 Validation Data Training pada algoritma Support Vector Regression……..47

Gambar 4.10 Desain Model LSTM (Neural Net)…….48

Gambar 4.11 Validation Data Training pada algoritma LSTM……48

Gambar 4.12 Hasil Pengolahan dengan Jumlah Training 80% (Variabel Konfirmasi Positif)…….53

Gambar 4.13 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Pertama)…….54

Gambar 4.14 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Kedua)..54

Gambar 4.15 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Ketiga)..55

Gambar 4.16 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Keempat)……56

Gambar 4.17 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Kelima).56

Gambar 4.18 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Keenam)…….57

Gambar 4.19 Hasil Pengolahan dengan Jumlah Training 80% (Variabel Sembuh)…….58

Gambar 4.20 Hasil Pengolahan dengan Jumlah Training 80% (Variabel Kematian)……...59

Gambar 4.21 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Pertama)…….59

Gambar 4.22 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Kedua)…….60

Gambar 4.23 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Ketiga)……...….61

Gambar 4.24 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Keempat)…….61

Gambar 4.25 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Kelima)……..…….62

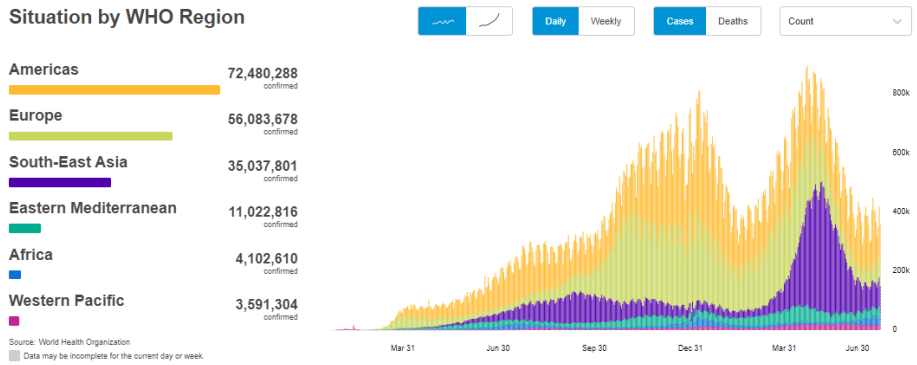
Gambar 4.26 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Keenam)…….…….63

BAB I
PENDAHULUAN

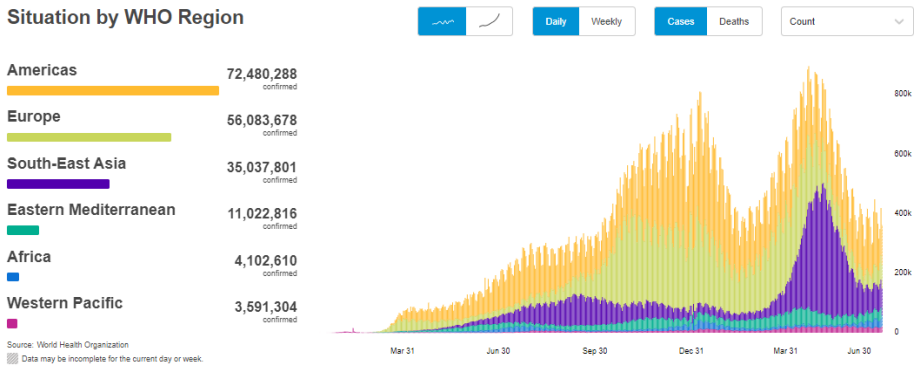
1.1 Latar Belakang

Lonjakan kasus penderita Covid-19 sampai saat ini telah beberapa kali mengalami peningkatan yang drastis semenjak pertama kali terdeteksi pada tahun 2020 lalu di Wuhan, China. Menurut data dari WHO, sampai 1 Juli 2021 terdapat setidaknya 181.521.067 orang yang telah terkonfirmasi positif Covid-19 secara keseluruhan dengan jumlah kasus terbaru selama 24 jam terakhir sebanyak 337.163 dan jumlah kematian sebanyak 3.937.437 orang secara keseluruhan dengan jumlah kematian terbaru selama 24 jam terakhir sebanyak 6.617 orang.

Berdasarkan data yang diambil pada tangal 4 Juli 2021 dari situs WHO, jumlah kasus terkonfirmasi positif dan kematian akibat Covid-19 digambarkan sebagai berikut :



Gambar 1.1 Total Terkonfirmasi Positif (Kiri) dan Harian Terkonfirmasi Positif (Kanan) (WHO, 2021)



Gambar 1.2 Total Kematian (Kiri) dan Harian Kematian (Kanan) (WHO, 2021)

Penyebaran wabah Covid-19 telah mengalami beberapa peningkatan, dan 2 diantaranya adalah peningkatan yang terjadi secara masif. Di Indonesia sendiri, meningkatnya kasus Covid-19 secara drastis yang terjadi saat ini terjadi karena beberapa faktor, diantaranya adalah (Kemenkes, 2021) :

1. Faktor Eksternal
   Faktor eksternal dihubungkan dengan mobilitas dari penduduk, seperti besarnya arus mudik pada saat hari raya. Biasanya ini terjadi pasca lebaran dan diprediksi akan terus mengalami peningkatan sampai 10 hari ke depan.
2. Faktor Endogen
   Sedangkan dari faktor endogen, peningkatan disebabkan karena adanya mutasi virus ke beberapa varian baru yang lebih menular. Seperti mutasi varian Delta dari India, varian Alpha dari Inggris, dan varian Beta dari Afrika Selatan.

Berbagai macam upaya pemerintah telah dilakukan untuk mengantisipasi dan mengurangi dampak dari Covid-19, seperti penyederhanaan meja vaksinasi, mendorong vaksinasi dengan vaksin gotong royong, memberlakukan Pemberlakukan Pembatasan Kegiatan Masyarakat (PPKM) terutama di daerah ibukota, dan membatasi kegiatan perkantoran untuk sektor non-esensial dengan jumlah WFH sebanyak 100%, dan sektor esensial dengan jumlah WFH sebanyak 50%. Namun berdasarkan data yang didapat dari WHO menunjukkan bahwa kasus Covid-19 justru semakin melonjak tajam dibandingkan sebelumnya, hal ini menandakan bahwa upaya yang dilakukan pemerintah dan masyarakat masih kurang maksimal sehingga perlu dicari upaya lain yang dapat membantu menurunkan lonjakan tersebut.

Salah satu upaya yang dapat kita lakukan diantaranya adalah dengan melakukan prediksi data jumlah penderita Covid-19 di masa depan. Hal ini dilakukan dengan tujuan agar pemerintah dan masyarakat mendapatkan gambaran mengenai lonjakan kasus yang akan terjadi di masa depan sehingga pemerintah dapat mengambil keputusan yang tepat untuk melakukan antisipasi, seperti menyediakan fasilitas dan tenaga kesehatan yang memadai untuk menangani penderita Covid-19 dan masyarakat dapat mempersiapkan diri untuk menghadapi gelombang pandemi yang akan datang.

Menurut Sumayang (2003:24), f*orecasting*/peramalan adalah perhitungan bersifat objektif dan dilakukan menggunakan data masa lalu untuk menentukan hasil yang akan didapat di masa yang akan dating (Sumayang, 2003:24). Dalam hal ini, peramalan yang akan penulis lakukan adalah peramalan berapa jumlah penderita Covid-19 selama beberapa hari ke depan menggunakan dataset jumlah penderita Covid-19 di masa lalu. Variabel yang akan diramal adalah berapa banyak orang yang akan terkonfirmasi positif, berapa banyak orang yang akan sembuh, dan berapa banyak orang yang akan meninggal akibat Covid-19.

Ada beberapa algoritma yang dapat digunakan untuk melakukan peramalan, seperti ARIMA, Long-Short Term Memory, Support Vector Machine, dan Exponential Smoothing. Dari beberapa algoritma tersebut, penulis akan menguji performa terutama dari sisi akurasi dan persentase error menggunakan MAE, R2, dan RMSE untuk mengetahui algoritma mana yang memiliki performa terbaik untuk meramalkan data jumlah penderita Covid-19 di masa depan.

Banyak penelitian yang telah dilakukan menggunakan berbagai macam algoritma peramalan yang berbeda dan setiap penelitian tersebut memiliki hasil dan

akurasi peramalan yang berbeda-beda pula. Sehingga perlu dilakukan perbandingan untuk mencari algoritma terbaik yang dapat digunakan untuk melakukan prediksi dengan akurasi tertinggi.

1.2 Masalah Penelitan

1.2.1 Identifikasi Masalah

Berdasarkan penelitian untuk memprediksi data jumlah penderita Covid-19 yang telah banyak dilakukan sebelumnya, telah digunakan berbagai macam algoritma untuk melakukan *forecasting* data penderita Covid-19 di seluruh dunia. Namun hasil yang didapat pasti akan selalu berbeda karena algoritma yang digunakan juga berbeda antara satu dengan yang lainnya. Hasil yang berbeda ini tentunya akan memberikan akurasi yang berbeda pula. Sehingga perlu dilakukan komparasi antar algoritma tersebut agar didapatkan algoritma yang memiliki akurasi peramalan tertinggi sehingga hasil peramalan yang dilakukan menjadi lebih baik. Komparasi yang dilakukan adalah dari sisi akurasi dan selisih nilai antar data aktual dan data prediksi atau error.

1.2.2 Pembatasan Masalah

Data yang akan digunakan dalam penelitian ini adalah jumlah orang yang terkonfirmasi positif Covid-19, jumlah orang yang meninggal akibat Covid-19, dan jumlah orang yang sembuh dari Covid-19 dari bulan Februari 2020 sampai bulan Oktober 2021 di seluruh dunia.

1.2.3 Rumusan Masalah

Berdasarkan analisis uraian identifikasi masalah di atas, maka rumusan masalah yang dapat diambil adalah sebagai berikut :

1. Apa rekomendasi model peramalan data jumlah penderita Covid-19 yang dapat memberikan hasil dengan akurasi tertinggi dan rata-rata error terendah ?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk memberikan rekomendasi model *forecasting* data penderita Covid-19 berdasarkan data penderita Covid-19 di masa lalu dengan akurasi tertinggi dan rata-rata error terendah.

1.3.2 Manfaat Penelitian

1.3.2.1 Manfaat Teoritis

Penelitian ini diharapkan dapat menjadi referensi bagi penelitian-penelitian selanjutnya khususnya yang berhubungan dengan *forecasting* data penderita Covid-19.

1.3.2.2 Manfaat Praktis

Adapun manfaat praktis yang diharapkan dari penelitian ini adalah :

1. Bagi Organisasi
   a. Dapat menjadi salah satu acuan dalam memperkirakan jumlah lonjakan kasus Covid-19 di masa depan.
   b. Memberikan rekomendasi model prediksi dengan akurasi yang paling tinggi dan tingkat kesalahan yang paling rendah.

1.4 Tata-Urut Penulisan

Penulisan karya tulis ini terdiri dari 4 bab dengan rincian pembagian sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisi Latar Belakang, Masalah Penelitian, Tujuan dan Manfaat Penelitian, Tata-Urut Penulisan, dan Daftar Pengertian.

BAB II LANDASAN TEORI DAN KERANGKA KONSEP/PEMIKIRAN

Bab ini berisi Tinjauan Pustaka dan Studi yang terkait dengan bahasan penelitian, Tinjauan Objek Penelitian, Kerangka Konsep, dan Hipotesis.

BAB III METODOLOGI DAN RANCANGAN/DESAIN PENELITIAN

Bab ini berisi Metode Penelitian, Metode Sampling dan Pengumpulan Data, Instrumentasi, Teknik Analisis, Rancangan dan Pengujian Data/Sistem/Prototipe Model, Rencana Strategi, Langkah Penelitian, dan Jadwal Penelitian.

BAB IV PENUTUP

Bab ini berisi paparan kesimpulan dari pembahasan pada bab sebelumnya dan saran pengembangan penelitian selanjutnya.

BAB II
LANDASAN TEORI DAN KERANGKA KONSEP/PEMIKIRAN

Pada bab ini akan dibahas mengenai landasan teori yang berkaitan dengan pembahasan penelitian ini yaitu landasan teori mengenai Covid-19, landasan teori mengenai algoritma *forecasting*, tinjauan studi, tinjauan objek penelitian dan kerangka konsep/pemikiran. Selain itu juga akan dibahas mengenai hipotesis dari penelitian ini berdasarkan metodologi yang akan digunakan.

2.1 Tinjauan Pustaka

2.1.1 Coronavirus

Coronavirus adalah sekelompok virus yang memiliki strand RNA tunggal dan mudah untuk bermutasi. Coronavirus merupakan keluarga besar dari virus yang menyebabkan ISPA (Infeksi Saluran Pernafasan Atas) ringan hingga sedang baik untuk manusia maupun hewan (He, 2020). Penyakit yang umum disebabkan oleh virus ini adalah penyakit flu. Namun beberapa varian dari virus ini dapat menimbulkan penyakit flu yang lebih parah seperti MERS-CoV (Middle East Respiratory Syndrome), SARS-CoV (Severe Acute Respiratory Syndrome), dan Pneumonia. Hingga saat ini, terdapat 7 jenis virus dari keluarga coronavirus yang sudah teridentifikasi (US National Library of Medicine National Institutes of Health, 2021) :

1. HCoV-229E.
2. HCoV-OC43.
3. HCoV-NL63.
4. HCoV-HKU1.
5. SARS-CoV (Tiongkok).
6. MERS-CoV (Timur Tengah).
7. COVID-19 (Tiongkok).

Penyebaran wabah coronavirus bukan pertama kalinya terjadi di dunia. Menurut WHO, pada bulan November tahun 2002 di negara Tiongkok pernah terjadi wabah virus SARS-CoV. Virus ini kemudian menyebar ke berbagai negara seperti Hongkong, Vietnam, Indonesia, Amerika Serikat, Singapura dan banyak negara lainnya, hingga akhirnya selesai pada pertengahan 2003. Wabah tersebut menimbulkan korban sebanyak 8098 orang yang terinfeksi dan 774 orang meninggal karena wabah tersebut.

Selain itu ada juga wabah virus MERS (Middle East Respiratory Syndrome) yang terjadi pada tahun 2012 dan berasal dari Timur Tengah. MERS merupakan virus zoonotic, dimana virus tersebut ditransmisikan dari hewan ke manusia. Virus ini kemudian menyebar ke berbagai negara seperti Afrika dan Asia Selatan. Selama penyebaran wabahnya, virus ini menimbulkan korban jiwa sebanyak 858 orang (WHO, 2012).

2.1.2 Covid-19

Covid-19 sendiri merupakan varian dari Coronavirus yang sudah ada sejak dulu. Wabah yang disebabkan oleh keluarga virus corona telah terjadi beberapa kali sebelum wabah Covid-19 terjadi. Seperti wabah virus SARS pada tahun 2002, dan MERS pada tahun 2012. Namun untuk wabah Covid-19 yang terjadi kali ini, menimbulkan dampak yang jauh lebih besar dan resiko penularan lebih tinggi.

Pada akhir Desember 2019, wabah Covid-19 pertama kali muncul di kota Wuhan, provinsi Hubei, China. Hingga pada 31 Januari 2020 wabah tersebut telah menyebabkan 9270 orang terinfeksi dengan jumlah kematian sebanyak 213 orang, selain itu juga menginfeksi 106 orang dari negara lainnya pada saat itu (WHO, 2020).

Sehingga beberapa hari kemudian, WHO menamakan virus ini sebagai Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Berdasarkan laporan dari WHO, sampai tanggal 27 Februari 2020 tercatat sebanyak 78.630 kasus dan 2747 kematian di China, kemudian menyebar ke 46 negara lain dan menyebabkan 3664 orang terinfeksi virus ini.

Patogen yang menyebabkan Covid-19 diidentifikasi sebagai 2019-nCoV pada tahun 2019. SARS-CoV-2 merupakan anggota dari keluarga coronavirus yang memiliki strand tunggal. Beberapa laporan mengatakan bahwa virus ini berasal dari kelelawar, hal ini didasarkan pada kesamaan urutan genetik dengan virus CoV lain. Sedangkan untuk hewan yang menjadi inang utama yang mentransmisikan virus ini kepada kelelawar masih belum diketahui.

Menurut WHO, sampai saat ini terdapat setidaknya 10 varian virus corona yang dilaporkan dari beberapa negara di dunia, diantaranya adalah :

Tabel 2.1 Varian Virus Corona dan Asalnya

Varian	**Asal Negara**
B117 – Alpha Variant	Inggris
B1351 – Beta Variant	Afrika Selatan
P1 – Gamma Variant	Brazil
B16172 – Delta Variant	India
B1427/B1429 – Epsilon Variant	Amerika Serikat
P2 – Zeta Variant	Brazil
B1525 – Eta Variant	Inggris
P3 – Theta Variant	Filipina
B1526 – Iota Variant	Amerika Serikat
B16171 – Kappa Variant	India

2.1.3 Forecasting

Forecasting/peramalan adalah perhitungan bersifat objektif dan dilakukan menggunakan data masa lalu untuk menentukan hasil yang akan didapat di masa yang akan datang (Sumayang, 2003:24).

Sedangkan menurut Render dan Heizer (2001:47) peramalan adalah ilmu pengetahuan dan seni untuk memperkirakan data pada masa yang akan dating. Peramalan dilakukan dengan mengambil data di masa lalu dan memproyeksikan data tersebut ke masa depan menggunakan model matematika.

Tujuan dari dilakukannya peramalan adalah untuk memperkirakan nilai data di masa depan, sehingga didapatkan perkiraan yang nilainya mendekati nilai actual. Hasil dari peramalan tidak akan sempurna, namun hasil permalan dengan nilai akurasi tinggi akan memberikan gambaran bagi kita untuk mengambil keputusan dalam melakukan perencanaan (Sofyan, 2013:15).

Proses *forecasting* termasuk ke dalam metode penelitian kuantitatif dimana penggunaan metode ini adalah berdasarkan adanya ketersediaan raw dataset dengan serangkaian metode matematis yang digunakan untuk melakukan prediksi nilai data di masa depan. Beberapa model penelitian yang termasuk ke dalam metode kuantitatif adalah (Arna, 2010) :

1. Model Regresi
   Merupakan turunan dari metode Regresi Linier untuk melakukan prediksi sebuah variabel yang memiliki hubungan secara linier dengan variabel independen yang diketahui.

2. Model Ekonometrik
   Sekumpulan persamaan regresi dimana tersedia variabel dependen yang mempengaruhi segmen-segmen ekonomi seperti harga dan lain sebagainya.

3. Model Deret Waktu (*Time Series Analysis*)
   Dengan menggunakan garis tren yang mewakili data pada masa lalu berdasarkan kecenderungan datanya dan menampilkan data tersebut ke masa depan.

Kemudian menurut Sofyan (2013:14), peramalan memiliki prinsip-prinsip sebagai berikut :

1. Tidak ada peramalan yang sempurna, maksudnya adalah tidak ada peramalan yang memiliki akurasi sampai 100% dan sesuai dengan kenyataan seluruhnya. Peramalan data hanya mengurangi ketidakpastian yang terjadi.

2. Peramalan selalu menghasilkan informasi mengenai ukuran *error* yang biasa dilakukan dengan menghitung MAE, MAPE, R2 dan sebagainya, sehingga peneliti perlu menyampaikan tingkat kesalahan yang terjadi pada data yang diramalnya.

3. Peramalan untuk waktu berjangka pendek umumnya lebih akurat dibanding peramalan untuk waktu berjangka panjang. Hal ini disebabkan karena pada peramalan berjangka pendek, faktor yang mempengaruhi lebih sedikit dibanding peramalan jangka panjang.

Adapun faktor-faktor yang mempengaruhi *forecasting* adalah (Sofyan, 2013:15) :

1. *Time Horizon*
   Ada dua aspek horizon waktu yang perlu diperhatikan, pertama adalah range waktu untuk target peramalan sebaiknya disesuaikan dengan model yang digunakan, kedua adalah periode waktu peramalan.

   Menurut Heizer dan Render (2014:140), *time horizon* itu sendiri dibagi menjadi 3, yaitu :
   a. Jangka Pendek
      Jangka pendek meliputi rentang waktu antara 3 bulan sampai 1 tahun. Biasanya rentang waktu ini digunakan untuk meramalkan tingkat produksi, jumlah tenaga kerja, dsb.
   b. Jangka Menengah
      Jangka menengah meliputi rentang waktu antara 1 hingga 3 tahun. Biasanya rentang waktu ini digunakan untuk meramalkan penjualan, anggaran produksi, anggaran kas, dsb.
   c. Jangka Panjang
      Jangka Panjang meliputi rentang waktu di atas 3 tahun. Biasanya rentang waktu ini digunakan untuk meramalkan produk baru, pengembangan fasilitas, dan penelitian dan pengembangan.

2. Pola Data
   Tipe dari pola data yang dihasilkan dari peramalan akan bersifat kontinyu.

3. Jenis Model
   Model yang dimaksud adalah sebuat deret yang menggambarkan waktu sebagai parameter yang penting untuk mengetahui perubahan di dalam pola yang dapat dijelaskan melalui analisis. Model lainnya adalah sebab-akibat, dimana hasil prediksi yang dihasilkan akan bergantung dari peristiwa yang terjadi di masa lalu.

4. Biaya
   Dalam hal ini, unsur biaya yang dimaksud adalah *development cost, error*, operasi, dan kesempatan untuk menggunakan metode lainnya.

5. Akurasi
   Tingkat akurasi yang dibutuhkan akan bergantung dari rincian yang dilakukan saat peramalan.

6. Kemudahan Penggunaan
   Dalam melakukan peramalan, kita sebaiknya menggunakan metode-metode yang mudah dimengerti dan diaplikasikan dalam data yang akan diramal.

2.1.4 *Time Series Analysis*

Data *time series* merupakan data yang ditampilkan bersamaan dengan deret waktu pada satu periode waktu tertentu. Sedangkan peramalan time series merupakan prediksi yang menyatakan asumsi bahwa masa yang akan datang merupakan fungsi dari histori (masa lalu) dengan tujuan untuk mencari pola dalam deret data masa lalu dan menerjemahkan pola tersebut ke masa yang akan datang (Heizer dan Render, 2009:169).

Menurut Arsyad (2001), *data time series* terbagi menjadi beberapa pola, yaitu sebagai berikut :

1. Pola Siklus
   Pola siklus merupakan perubahan data menuju naik atau turun yang menyebabkan pola ini menjadi bervariasi dan berubah dari suatu siklus ke siklus berikutnya.

2. Pola Random
   Pola random merupakan pola yang berisi data secara acak dan tidak beraturan sehingga tidak dapat digambarkan. Beberapa hal yang menyebabkan terjadinya pola acak ini biasanya adalah karena sesuatu hal yang tidak terduga sebelumnya, seperti kerusuhan, perang, bencana alam, dan lain sebagainya. Untuk menganalisa pola acak, maka digunakan indeks dengan nilai 100% atau 1 dikarenakan bentuknya yang tidak beraturan dan tidak dapat diramalkan.

3. Tren
   Tren/kecenderungan merupakan sebuah komponen yang bersifat jangka panjang dimana komponen tersebut memiliki kecenderungan tertentu dalam membentuk suatu pola data, baik yang arahnya naik ataupun turun secara *time series*, sehingga pola tren jangka panjang jarang menunjukkan pola yang tetap.

4. Season/Musiman
   Pola musiman akan memperlihatkan sebuah pergerakan yang selalu berulang-ulang dari satu periode ke periode berikutnya dengan cara teratur. Untuk mendapatkan pola musiman, kita dapat mengelompokkan data-data baik secara mingguan, bulan, atau kuartal.

2.1.5 Regresi Linier Berganda

Regresi Linier merupakan salah satu model yang sering digunakan dalam penelitian untuk menemukan nilai variabel di masa yang akan datang menggunakan nilai variabel di masa lalu.

Regresi Linier Memiliki turunan yang disebut sebagai Regresi Linier Berganda. Regresi Linier Berganda adalah analisa asosiasi yang bersamaan untuk mencari pengaruh multi variabel independen terhadap satu variabel dependen dengan skala interval (Narimawati, 2008). Menurut Sugiyono (2016:192), Regresi Linier Berganda adalah regresi dengan susunan variabel satu dependen dan dua atau lebih independen.

Secara umum, untuk menghitung regresi linier berganda kita dapat menggunakan model seperti dibawah ini :

$$Y = a+b_1X_1+b_2X_2+b_3X_3+\ldots+b_nX_n \quad (3.1)$$

Keterangan :

Y	: Variabel Dependen
X (1,2,3,…)	: Variabel Independen
a	: Nilai Konstanta
b (1,2,3,…)	: Nilai Koefisien Regresi

Model diatas dapat digunakan apabila kita ingin mengetahui :

1. Kekuatan hubungan antara multi (dua atau lebih) variabel independent dengan satu variabel dependen (contohnya bagaimana curah hujan, suhu dan jumlah pupuk yang ditambahkan mempengaruhi pertumbuhan tanaman).
2. Nilai variabel dependen pada nilai variabel independent tertentu (contohnya adalah mencari hasil yang diharapkan dari tanaman pada curah hujan, suhu dan penambahan pupuk).

Adapun tujuan kita menggunakan analisis Regresi Linier Berganda adalah :

1. Melakukan prediksi dan peramalan dimana penggunaannya secara substansial dan tanpa tumpeng tindih dengan Machine Learning.
2. Di beberapa situasi, analisis Regresi Linier Berganda dapat menyimpulkan hubungan kasual antara variabel independen dan variabel dependen. Karena regresi itu sendiri hanya mencari hubungan antara variabel dependen dan kumpulan variabel independen dalam populasi data tetap.

2.1.6 Support Vector Machine

SVM (*Support Vector Machine*) merupakan salah satu jenis algoritma *Machine Learning* (ML) yang digunakan untuk melakukan regresi dan klasifikasi (Du, 2013). Sedangkan menurut Fachrurazi (2011), SVM adalah teknik untuk

memprediksi suatu data yang dapat dilakukan dalam bentuk regresi atau klasifikasi. SVM menyelesaikan masalah regresi menggunakan fungsi linier. Sedangkan untuk menyelesaikan masalah non-regresi linier, SVM memetakan vektor masukan (x) ke n-*dimensional space* yang disebut dengan *feature space* (z). Pemetaan ini dilakukan oleh teknik non-linear setelah regresi linier diaplikasikan ke dalam *space*. Selain itu, teknik SVM juga berfungsi sebagai pemisah untuk membedakan observasi yang mengandung nilai variabel target yang berbeda (William, 2011).

Karakteristik SVM secara keseluruhan adalah sebagai berikut (Nugroho, 2003) :

1. SVM pada prinsipnya adalah linear *classifier*.
2. Melakukan *pattern recognition* dengan cara mentransformasikan data pada *input space* ke dalam *feature space* dan dilakukan optimasi di ruang tersebut. Ini adalah salah satu fitur SVM yang membedakannya dengan *pattern recognition* lainnya, dimana yang lain melakukan optimasi hasil transformasi pada dimensi yang lebih rendah daripada *input space*.
3. Memiliki SRM (*Structural Risk Minimization*).
4. SVM umumnya hanya mampu menangani dua kelas klasifikasi, namun dengan adanya pengembangan *pattern recognition* membuat SVM dapat menangani lebih dari dua kelas.

Selain itu, juga terdapat kelebihan dan kekurangan dari model SVM, kelebihan SVM adalah :

1. Generalisasi
   Definisi dari generalisasi adalah suatu kemampuan dari metode tertentu untuk melakukan klasifikasi dari pattern atau pola yang berada di luar data yang digunakan dalam fase *learning* pada metode itu.
2. *Curse of Dimensionality*
   Definisi dari *Curse of Dimensionality* adalah masalah yang dialami suatu metode pengenalan pola (*pattern recognition*) dalam memprediksi parameter yang disebabkan oleh jumlah sampel data yang berjumlah lebih sedikit dibandingkan dimensional ruang dari vektor.
3. Feasibility
   Pengimplementasian SVM relatif lebih mudah dikarenakan adanya proses penentuan support vector yang dirumuskan dalam *Quadratic Programming* (QP) *problem* (Nugroho, 2003).

Sedangkan kelemahan dari SVM adalah :

1. Tidak cocok diterapkan pada masalah berskala besar, maksudnya adalah besarnya jumlah sampel yang diolah.
2. Secara teoritik, SVM dikembangkan untuk menangani masalah dengan dua kelas. Namun adanya pengembangan *pattern recognition* membuat SVM dapat menangani lebih dari dua kelas.

2.1.7 ARIMA

Model ARIMA biasa digunakan dalam perhitungan statistik dan ekonometrik dalam analisa *time series*. ARIMA merupakan pengembangan dari

model ARMA (Autoregressive Moving Average) yang digabungkan dengan metode Box Jenkins yang biasa digunakan untuk melakukan prediksi nilai yang akan datang dalam bentuk rentang waktu nilai sekarang dan nilai sebelumnya (Pankratz, 2009).

ARIMA dikembangkan pertama kali oleh George Box dan Gwilym Jenkins untuk memodelkan analisa *time series* sehingga model ini sering juga disebut dengan Box-Jenkins. ARIMA merupakan gabungan dari tiga model lain yaitu AR (Autoregressive Model), MA (Moving Average), dan ARMA (Autoregressive Moving Average) (Whitten, 2007).

Model ARIMA sendiri telah banyak digunakan dalam penelitian sebelumnya untuk melakukan *forecasting* wabah penyakit seperti Hepatitis, Demam Berdarah, Malaria, TBC, Influenza, dan Zoonotic Cutaneous Leishmaniasis (Guleryuz, 2021).

Adapun tahapan untuk melakukan pencarian model adalah :

1. Identifikasi Model
   Suatu hal yang perlu digaris bawahi adalah umumnya *time series* bersifat nonstasioner dan model-model dalam ARIMA seperti AR dan MA hanya bersinggungan dengan deret yang bersifat stasioner (stasioner berarti tidak adanya penambahan atau pengurangan pada nilai data). Sehingga perlu dilakukan perubahan dari data non stasioner menjadi data stasioner dengan cara *differencing* (menghitung nilai perubahan atau selisih dari nilai observasi). Adapun persamaan *differencing* menurut Hanke (2009) adalah sebagai berikut :

   $$X't = Xt - BXt \qquad (3.2)$$

   Keterangan :
   X't : Nilai deret setelah differencing
   Xt : Nilai deret pada waktu t
   B : Orde differencing

2. Fungsi Autokorelasi
   Fungsi Autokorelasi (ACF) merupakan hubungan antara data pada periode waktu t dengan periode waktu sebelumnya t-1. Median dan varian dari sebuah data *time series* mungkin tidak akan memiliki manfaat apabila deret tersebut bersifat non stasioner, hanya saja kita masih bisa menggunakan nilai maksimum dan minimum untuk melakukan *plotting*.

3. Fungsi Autokorelasi Parsial
   Fungsi Autokorelasi Parsial (PAFC) memiliki fungsi untuk mengukur tingkat kecerdasan antara Xt dan Xt-k, jika pengaruh dari lag time terpisah. Tujuan utama dalam analisa *time series* adalah untuk mencari model ARIMA yang paling tepat.

4. Model Autoregresif

Pada tahap sebelumnya yaitu PAFC berfungsi untuk mencari tingkat keeratan antara Yt dan Yt-k jika pengaruh dari time lag adalah 1,2,3…,k. Tujuan penggunaan PAFC dalam analisa *time series* adalah untuk membantu mencari model ARIMA yang tepat untuk prediksi, lebih tepatnya untuk menentukan ordo p dari model AR.

2.1.8 Exponential Smoothing

Exponential Smoothing pertama kali diperkenalkan pada akhir tahun 1950. Algoritma ini digunakan untuk memperhalus (*smoothing*) data berbasis rentang waktu/*timeseries* berdasarkan fungsi eksponensial. Fungsi eksponensial digunakan untuk menemukan berat yang terus menerus turun secara eksponensial, sehingga semakin kecil jarak ke waktu data terakhir, maka nilai beratnya akan semakin tinggi (Brownlee, 2013:171). Algoritma ini cocok untuk memprediksi data yang bersifat univariat (data yang hanya memiliki 1 variabel dependen).

Exponential Smoothing sendiri memiliki 3 metode, yaitu Single Exponential Smoothing (SES), Double Exponential Smoothing (Holt Linear Trend Model), dan Triple Exponential Smoothing (Holt-Winters Exponential Smoothing).

1. Single Exponential Smoothing
   Single Exponential Smoothing (SES) atau disebut juga dengan Simple Exponential Smoothing merupakan metode *forecasting* yang digunakan untuk memprediksi data yang bersifat univariat dimana data tersebut tidak memiliki tren.

   Metode ini menggunakan sebuah parameter yang disebut dengan Alpha, yang berfungsi sebagai *smoothing factor*. Parameter Alpha akan mengontrol rata-rata nilai prediksi yang dipengaruhi oleh rentang waktu dari observasi data sebelumnya yang akan berkurang secara eksponensial. Alpha dapat memiliki nilai antara 0 sampai 1. Apabila nilai mendekati 1, maka model SES akan lebih memperhatikan observasi yang jaraknya dekat dengan waktu peramalan. Apabila nilainya mendekati 0, maka model akan lebih memperhatikan data histori yang jaraknya lebih jauh dengan waktu peramalan. Sehingga kesimpulannya adalah apabila nilai alpha mendekati 1, maka model mengindikasikan fast learning (data terbaru yang akan mempengaruhi hasil peramalan), jika nilai alpha mendekati 0, maka model mengindikasikan slow learning (observasi data terdahulu yang lebih jauh dari waktu peramalan akan lebih mempengaruhi hasil peramalan) (Shmueli, 2016:89).

2. Double Exponential Smoothing
   Double Exponential Smoothing merupakan turunan dari Exponential Smoothing dimana pada model ini ditambahkan fungsi untuk meramalkan data yang memiliki tren dalam rentang waktu yang bersifat univariat. Double Exponential Smoothing juga dikenal dengan nama Holt's Linear Trend Model, sesuai dengan nama penemunya yaitu Charles Holt.

Perbedaannya dengan SES adalah, pada Holt Model ditambahkan variabel baru yaitu beta dimana beta akan berfungsi sebagai *smoothing factor* yang mengatur perubahan nilai yang dipengaruhi oleh perubahan arah tren. Metode ini dapat memprediksi nilai dalam tren yang berubah baik secara *additive* dan *multiplicative*. Untuk tren *additive* dapat digunakan Double Exponential Smoothing dengan tren linier. Sedangkan untuk tren *multiplicative* dapat digunakan Double Exponential Smoothing dengan tren eksponensial.

Hasil peramalan menggunakan Linear Holt Model menunjukkan perubahan tren yang konstan (naik atau turun) secara tidak wajar di masa depan dan penambahan parameter dapat mengurangi tren di masa depan menjadi *sideways* atau datar (tidak memiliki tren) (Hyndman, 2013:183).

3. Triple Exponential Smoothing
   Triple Exponential Smoothing merupakan turunan dari Exponential Smoothing yang mendukung peramalan data univariat yang bersifat musiman. Metode ini juga dikenal dengan nama Holt-Winters Exponential Smoothing, sesuai dengan nama penemunya yaitu Charles Holt dan Peter Winters.

   Jika pada model sebelumnya hanya menggunakan variabel Alpha dan Beta, maka model ini menggunakan variabel tambahan lagi yang disebut dengan Gamma, dimana variabel Gamma berfungsi untuk mengatur pengaruh dari komponen data musiman. Maksud dari data musiman contohnya adalah data bulanan yang dibagi menjadi beberapa periode, misal 12 periode. Sehingga periode harus didefinisikan terlebih dahulu sebelum melakukan peramalan.

   Periode atau musim dapat dimodelkan sebagai proses *additive* atau *multiplicative* untuk perubahan secara linier maupun eksponensial. Untuk data *additive seasonal* dapat menggunakan Triple Exponential Smoothing dengan *linear seasonality*, sedangkan untuk data *multiplicative seasonal* dapat menggunakan Triple Exponential Smoothing dengan *exponential seasonality*.

   Triple Exponential Smoothing merupakan turunan paling mutakhir dari model Exponential Smoothing lainnya karena model ini dapat berkembang dan menghasilkan baik model double maupun single exponential smoothing. Sebagai model yang adaptive TES atau Holt-Winters model memungkinkan untuk adanya perubahan pola level, tren dan musim secara berkala (Shmueli, 2016:95).

2.1.9 LSTM

LSTM (Long Short Term Memory) atau yang dikenal dengan RNN (Recurrent Neural Network) merupakan salah satu algoritma *forecasting* yang terbilang baru dibandingkan dengan algoritma lainnya. LSTM memiliki kemampuan untuk mengingat nilai dari *record* sebelumnya pada *time series* dan

menggunakannya untuk keperluan meramalkan data yang akan datang (Namini, 2018).

Saat ini terdapat beberapa jenis Neural Network yang diketahui, yaitu :

1. Artificial Neural Network (ANN)
   Merupakan syaraf tiruan yang dibuat seperti cara kerja otak manusia yang memiliki tujuan untuk menyelesaikan suatu pekerjaan tertentu. Untuk implementasi pada jaringan ini biasa dilakukan memakai komponen elektronik dan disimulasikan menggunakan aplikasi pada komputer (Haykin, 2009).

   Sedangkan menurut Kurniawansyah (2018), ANN merupakan salah satu teknik *machine learning* yang terinspirasi dengan jaringan syaraf biologis otak manusia dimana Jaringan Syaraf Tiruan memiliki beberapa prosesor yang dibuat secara sederhana dan berhubungan satu sama lain yang disebut dengan neuron.

   Kemudian, penentuan jumlah lapisan dan neuron pada *hidden layer* dan hubungan antara syaraf tersebut adalah hal yang penting (Saritas & Yasar, 2019).

   Adapun kelebihan dari Jaringan Syaraf Tiruan menurut Haykin (2009) diantaranya adalah :

   a. Nonlinearity
      Sifat *nonlinearity* dibutuhkan apabila pembuatan sinyal input menggunakan mekanisme non-linear.

   b. Input-Output Mapping
      Metode pembelajaran supervised learning akan bergantung pada modifikasi dari bobot sinaptik Neural Network dengan menerapkan sekumpulan contoh data latih. Pelatihan dilakukan berkali-kali untuk semua data latih dan dilakukan hingga jaringan mencapai keadaan stabil (perbedaan antar bobot sinaptik nilainya kecil).

   c. Adaptivity
      Neural Network mampu untuk melakukan adaptasi bobot sinaptiknya untuk melakukan perubahan pada lingkungannya. Intinya, jaringan syaraf tiruan memiliki kemampuan untuk dilatih ulang agar dapat beroperasi di dalam lingkungan tertentu dan dapat menghadapi perubahan kecil di suatu lingkungan.

   d. Evidential Response
      Sebuah Neural Network dapat dirancang untuk mengenali informasi tentang pola apa yang harus diambil.

   e. Contextual Information

Knowledge yang digambarkan dengan susunan dan status dari Neural Network apakan dalam keadaan aktif atau tidak. Setiap komponen neuron di dalam jaringan memiliki kemungkinan untuk terpengaruh dengan aktivitas global neuron lain di dalam jaringan.

f. Fault Tolerance
   Penerapan Neural Network dalam bentuk perangkat keras memiliki kemungkinan untuk terjadinya toleransi kegagalan dimana performa akan menurun secara perlahan dalam kondisi operasi tertentu.

g. VLSI (*Very Large Scale Integrated*) Implementability
   Adanya sifat paralel dari Neural Network memungkinkan jaringan untuk menghitung tugas dengan cepat. Sehingga fitur ini membuat Neural Network baik untuk diimplementasikan menggunakan teknologi VLSI.

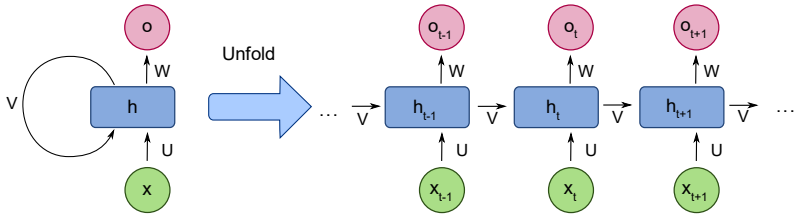
h. Uniformity of Analysis and Design
   Umumnya Neural Network memiliki sifat universal sebagai pengolah informasi. Hal ini karena Neural Network didukung oleh fitur-fitur seperti neuron, unversalitas untuk menerapkan teori dan algoritma pembelajaran dalam aplikasi yang berbeda, pembentukan jaringan modular melalui integrasi modul.

i. Neurobiological Analogy
   Analogi perancangan Neural Network dibuat berdasarkan cara kerja otak, sehingga bisa menjadi bukti hidup bahwa proses toleransi kegagalan tidak hanya terjadi secara fisik. Peneliti melihat Neural Network sebagai alat penelitian untuk menafsirkan fenomena neurobiologis.

Akan tetapi ANN masih memiliki kekurangan dimana ANN memerlukan waktu yang lama untuk melakukan training data dengan jumlah yang besar.

2. Recurrent Neural Network (RNN)
   Recurrent Neural Network merupakan jaringan syaraf tiruan yang mampu untuk mencari korelasi tersembunyi yang terdapat pada data dalam bidang *voice recognition*, *natural language processing*, dan *timeseries prediction* (Tian dkk, 2018). RNN merupakan salah satu model yang baik digunakan untuk melakukan prediksi data berbentuk deret waktu.

   RNN diperkenalkan pertama kali oleh Jeff Elman pada tahun 1990 dimana model ini merupakan variasi dari model terdahulunya yaitu ANN.



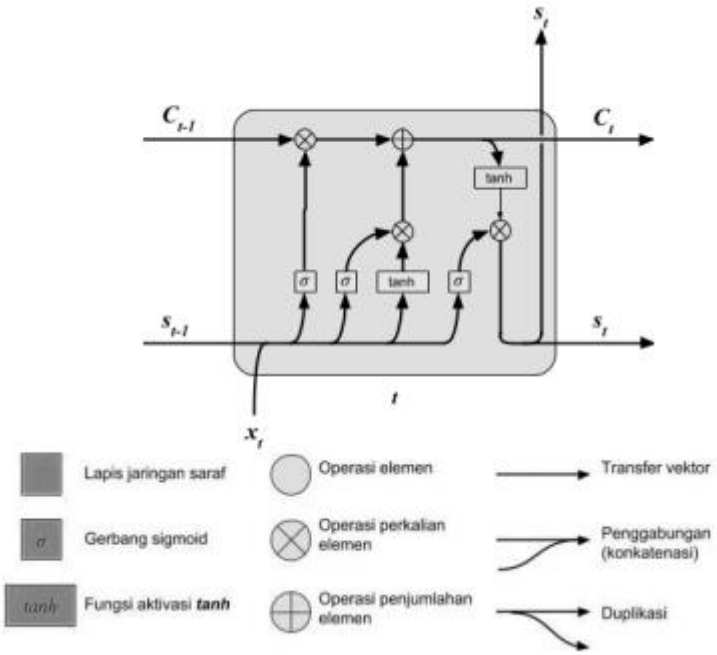
Gambar 2.1 Susunan Recurrent Neural Network

Gambar di atas adalah model dari Recurrent Neural Network yang digambarkan secara berurutan dimana x adalah Input Layer, h adalah Hidden Layer, dan o adalah Output Layer. Input Layer berfungsi untuk mengambil nilai awal dari sebuah *record*, kemudian selanjutnya terdapat Hidden Layer yang berfungsi sebagai bagian yang mengingat *record* sebelumnya. Output Layer merupakan bagian yang menjadi hasil prediksi pada node tersebut dan menjadi input untuk node selanjutnya. Susunan RNN digambarkan berurutan karena cara kerja RNN yang melakukan pengulangan.

Namun RNN masih memiliki kekurangan, yaitu tidak dapat mengingat deret untuk waktu yang lama.

3. Long-Short Term Memory (Special RNN)
   LSTM merupakan model terbaru yang dikembangkan dari RNN (bentuk spesial dari RNN) yang dapat melakukan *learning* pada *long-term dependencies*. Orang yang menemukan model ini untuk pertama kali adalah Hochreiter dan Schmidhuber pada tahun 1997.

   Menurut Ruales (2014), algoritma LSTM memiliki tingkat *error* yang lebih rendah dibandingkan RNN. Pada penelitian yang dilakukannya, ia menguji algoritma tersebut untuk mengklasifikasikan sentimen pada ulasan film. Hasilnya adalah algoritma LSTM memiliki tingkat *error* sebanyak 0.134 lebih kecil dibandingkan RNN yang saat itu menghasilkan tingkat *error* sebanyak 0.248.



Gambar 2.2 Susunan Long Short-Term Memory

Berdasarkan ilustrasi susunan dari LSTM pada gambar 2.2, setiap garis tersebut membawa vektor dari output satu node ke input node lainnya. Kemudian simbol lingkaran menggambarkan elemen operasi seperti penambahan atau perkalian elemen vektor. Selanjutnya terdapat simbol persegi yang menggambarkan *layer* dari jaringan syaraf dengan kemampuan *learning*. Selain itu terdapat simbol dua garis yang bergabung dan berpisah. Garis yang bergabung menggambarkan penyatuan dari dua vektor/matriks, sedangkan dua garis yang berpisah menandakan adanya duplikasi konten dimana salinan dari konten tersebut masuk ke dalam simpul yang berbeda. LSTM bekerja dengan menggunakan tiga struktur gerbang yang mencakup gerbang *input*, *hidden*, dan *output*.

Model ini memiliki kelebihan dibanding model sebelumnya, dimana LSTM dapat mengingat deret dalam waktu yang lebih lama dibandingkan dengan RNN.

2.2 Tinjauan Studi

Penelitian sebelumnya yang membahas permasalahan dalam domain yang sama telah dilakukan oleh beberapa peneliti menggunakan model yang berbeda-beda sehingga menghasilkan tingkat akurasi dan mean error yang berbeda-beda pula. Jika penelitian sebelumnya meneliti tentang prediksi nilai data dengan menerapkan model peramalan, maka pada penelitian ini penulis akan melakukan komparasi model yang digunakan para peneliti sebelumnya untuk melakukan peramalan sehingga didapatkan sebuah model yang memiliki performa terbaik dibandingkan model sebelumnya.

Oleh karena itu untuk mengatasi masalah tersebut, penulis melakukan perbandingan antara model-model yang digunakan pada penelitian terdahulu untuk mencari model yang menghasilkan akurasi terbaik dan mean error terkecil diantara model-model tersebut.

Sehingga penulis menggunakan hasil dari penelitian tersebut untuk direview dan dikomparasikan satu dengan lainnya untuk menemunkan performa terbaik diantara model-model tersebut.

Beberapa dari referensi yang dirangkum pada tabel di bawah merupakan penelitian-penelitian terdahulu yang pernah dilakukan yang akan berguna bagi penulis sebagai pedoman dari penelitian yang akan penulis lakukan sehingga memudahkan penulis untuk menemukan model yang memiliki performa terbaik untuk melakukan peramalan data jumlah penderita Covid-19 di masa depan. Tinjauan studi yang penulis lakukan adalah dari karya ilmiah berupa Jurnal yang digunakan sebagai acuan untuk menemukan model penelitian terbaik yang akan direkomendasikan pada hasil penelitian nanti.

Pertama, jurnal yang berjudul Forecasting Outbreak of COVID-19 in Turkey; Comparison of Box–Jenkins, Brown's Exponential Smoothing and Long Short-Term Memory Models, ditulis oleh Didem Guleryuz pada tahun 2021 memiliki tujuan untuk mencari model *forecasting* untuk data statistik kasus Covid-19 di Turki. Penelitian ini mengkomparasikan beberapa model untuk mencari model dengan akurasi tertinggi. Dataset yang digunakan adalah data statistic Covid-19 pada tanggal 12 Maret sampai 16 Mei 2020 di negara Turki yang diambil dari WHO. Sebelumnya, untuk negara Turki belum ada peramalan data untuk melihat gambaran data kasus Covid-19 untuk beberapa hari ke depan. Menurut penulis, model peramalan diperlukan agar pemerintah dan masyarakat dapat melakukan antisipasi dalam menghadapi lonjakan kasus yang tinggi. Kemudian penulis memilih tiga kandidat model yang akan digunakan untuk melakukan peramalan, yaitu ARIMA, Exponential Smoothing, dan Long-Short Term Memory. Dari ketiga pengujian terhadap model tersebut, model ARIMA memberikan hasil terbaik dengan AIC value terendah yaitu 12.0342, -2.51411, 12.0253, 3.67729, -4.24405, dan 3.66077. Dengan demikian, penulis memilih untuk mengusulkan model ARIMA untuk melakukan peramalan. Kemudian menurut penulis, pemerintah sebaiknya membuat keputusan yang akurat dan tindakan antisipasi dengan cara memperhatikan pengelolaan sumber daya untuk menangani wabah.

Kedua adalah jurnal yang berjudul Covid-19 Future Forecasting Using Supervised Machine Learning Models, ditulis oleh Furqan Rustam, Aijaz Ahmad Reshi, pada tahun 2020 memiliki tujuan untuk mencari model untuk memprediksi skenario pandemi Covid-19 selama 10 hari ke depan. Penelitian ini mengkomparasikan beberapa model untuk mencari model dengan akurasi tertinggi. Dataset yang digunakan adalah dataset dari repository Github yang disediakan oleh Center for System Science and Engineering, John Hopkins University. Pada penelitian ini, penulis mengusulkan beberapa model yang akan dipakai untuk melakukan peramalan data, yaitu Linear Regression, Long-Short Term Memory, Support Vector Machine, dan Exponential Smoothing. Dari keempat model tersebut, dihasilkan bahwa Exponential Smoothing memberikan performa terbaik dengan nilai R^2 0.99, nilai $R^2_{Adjusted}$ 0.99, nilai MSE 5970634.07, nilai MAE 1827.85, dan nilai RMSE 2443.48. Pada penelitian selanjutnya akan dilanjutkan

dengan eksplorasi metodologi peramalan menggunakan dataset yang terbaru dan realtime

Ketiga adalah jurnal yang berjudul Perkiraan Kasus COVID-19 di Indonesia dengan Algoritma Triple Exponential Smoothing, ditulis oleh Ruliyanta pada tahun 2020. Penelitian ini bertujuan untuk mencari model peramalan untuk meramalkan data statistik Covid-19 selama beberapa hari ke depan agar masyarakat dapat bersiap dan pemerintah dapat mengantisipasi hal tersebut dengan mempersiapkan anggaran untuk menangani lonjakan yang akan datang. Dataset yang diambil adalah data kasus Covid-19 yang berada di Indonesia dari Data Nasional Bersama dari Satgas Percepatan Penanganan Covid-19. Dalam penelitian ini, penulis melakukan peramalan data Covid-19 menggunakan model Triple Exponential Smoothing dan didapatkan hasil yaitu berdasarkan hasil peramalan, di Indonesia pada akhir 2020 COVID-19 akan terus bertambah dengan jumlah terkonfirmasi sebanyak 386.571 dan meninggal sebanyak 15.622 orang. Saran dari penulis adalah dengan tingginya hasil peramalan COVID-19 di Indonesia, dirasa perlu untuk melakukan langkah konkrit untuk mengurangi laju pertumbuhan COVID-19.

Keempat adalah jurnal yang berjudul Optimization Method for Forecasting Confirmed Cases of COVID-19 in China, yang ditulis oleh Mohammed A. A. Al-qaness, dan Ahmed A. Ewees pada tahun 2020. Penelitian ini bertujuan untuk mencari akurasi yang lebih tinggi untuk model ANFIS. Dataset yang diambil adalah dataset resmi kasus Covid-19 dari WHO. Dalam penelitian ini, penulis menggunakan model gabungan, yaitu ANFIS, FPA, dan SSA dimana SSA digunakan untuk mengatasi kekurangan FPA seperti keterbatasan pada local optima. Penulis ingin meningkatkan performa dari ANFIS dengan cara menentukan parameter dari ANFIS melalui FPASSA. Hasil yang didapat adalah model FPASSA memiliki rata-rata nilai evaluasi terendah, yaitu RMSE sebanyak 5779, MAE sebanyak 4271, MAPE sebanyak 4.79, RMSRE sebanyak 0.07, R2 sebanyak 0.9645, dan waktu proses selama 23.30 detik. Namun model FPASSA masih memiliki nilai R2 yang lebih tinggi dari beberapa model lainnya sehingga diperlukan adanya penggabungan dengan model lain agar dapat menurunkan nilai R2.

Kelima adalah jurnal yang berjudul Modelling and Forecasting for the Number of Cases of the Covid-19 Pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods dan ditulis oleh Harun Yonar dan Aynur Yonar pada tahun 2020. Jurnal ini dibuat karena adanya keterlambatan yang dialami beberapa negara di dunia dalam menangani pandemi membuat banyak pasien Covid-19 tidak mendapatkan perawatan intensif. Salah satu penyebabnya adalah karena kurangnya antisipasi masyarakat dalam menghadapi tsunami Covid-19. Jurnal ini bertujuan untuk menyediakan model untuk melakukan peramalan data agar masyarakat memiliki gambaran ke depan sehingga dapat mempersiapkan diri menghadapi gelombang wabah Covid-19. Model yang diusulkan penulis adalah Curve Estimation, ARIMA, dan Exponential Smoothing. Hasil dari penelitian ini adalah Jepang (Holt Model), Jerman (ARIMA(1,4.0)) dan Prancis (ARIMA(0,1,3)) menyediakan data statistik yang signifikan namun tidak teruji secara klinis. Inggris (Holt Model), Canada (Holt Model), Italy (Holt Model) dan Turki (ARIMA(1,4,0)) menunjukkan hasil yang lebih reliabel. Di penelitian yang akan datang, dibutuhkan lebih banyak data dan evaluasi yang lebih sehat. Kemudian hasil dari penelitian ini

dapat digunakan sebagai gambaran agar masyarakat dapat melakukan antisipasi terhadap gelombang pandemik selanjutnya.

Dari beberapa penelitian terdahulu yang sudah dirangkum di atas, penulis juga telah merangkum penelitian-penelitian tersebut dalam bentuk tabel diantaranya sebagai berikut :

Tabel 2.2 Perbandingan Penelitian Terdahulu

No.	Wilayah	Atribut	Rentang Waktu	Metode Forecast	Parameter Evaluasi	Hasil	Tahun
1.	Negara (Turki)	1. Jumlah Kasus Keseluruhan. 2. Laju Pertumbuhan Kasus. 3. Jumlah Kasus Baru. 4. Jumlah Kematian Keseluruhan. 5. Laju Pertumbuhan Kematian. 6. Jumlah Kematian Baru.	12 Mar – 16 Mei 2020	ARIMA, LSTM, ES	RMSE, MAE, MAPE, AIC	1. Jumlah Kasus Keseluruhan : ARIMA (379.539). 2. Laju Pertumbuhan Kasus : LSTM (0.00218). 3. Jumlah Kasus Baru : ARIMA (377.386). 4. Jumlah Kematian Keseluruhan : ARIMA (5.97003). 5. Laju Pertumbuhan Kematian : LSTM (0.1046). 6.Jumlah Kematian Baru : ARIMA (6.03522).	2021
2.	Negara (Indonesia)	1. Total Kasus. 2. Kasus Baru. 3. Kematian Baru. 4. Total Kematian.	2 Maret – 1 Agustus 2020	Triple Exponential Smoothing	LCB, UCB	1. Total Kasus : - UCB : 511.157 - Mean : 386.571 - LCB : 261.986 2. Kasus Baru : - UCB : 5.777 - Mean : 4.159 - LCB : 2.541 3. Kematian baru : - UCB : 249 - Mean : 161 - LCB : 74 4. Total Kematian : - UCB : 32.356 - Mean : 15.662	2020

						- LCB : -1.033	
3.	Global	1. Kematian Baru. 2. Kasus Baru. 3. Kasus Sembuh.	22 Jan – 16 Feb 2020	LASSO, LR, SVM, ES	R^2, R^2 Adjusted, MSE, MAE, RMSE	1. Kematian Baru : ES (96%). 2. Kasus Baru : ES (98%), LASSO (98%). 3. Sembuh : ES (99%).	2020
4.	Negara (Ethiopia)	Total Kasus	14 Mar – 5 Jun 2020	Exponential Smoothing	RMSSE	1. Exponential Growth Model : 0.21839. 2. Simple Exponential Smoothing : 0.41142. 3. Double Exponential Smoothing : 0.03978.	2020
5.	Negara (Indonesia)	1. Kasus Baru. 2. Jumlah Sembuh. 3. Jumlah Meninggal.	2 Mar – 7 Apr 2020	Double ES	MAPE, MAD, MSD	1. MAPE : 8.998; MAD : 18.243; MSD : 686.766; 2. MAPE : 14.179; MAD : 2.7787; MSD : 19.351; 3. MAPE : 13.9526; MAD : 2.4309; MSD : 18.7697;	2020
6.	Global	Kasus Baru	22 Jan – 22 Mar 2020	ARIMA, ES	R2 Stationary, R2, RMSE, MAPE, MAE, MaxAPE, MaxAE	1. ARIMA : - R2 Stationary : 0.188 - R2 : 0.826 - RMSE : 6.171 - MAPE : 1.301 - MAE : 2.176 - MaxAPE : 653.246 - MaxAE : 28.823 2. ES : - R2 Stationary : 0.646 - R2 : 0.462 - RMSE : 4.474 - MAPE : 3.682	2020

						- MAE : 1.908 - MaxAPE : 308.867 - MaxAE : 17.851	
7.	Global	1. Kasus Baru. 2. Jumlah Sembuh. 3. Jumlah Meninggal.	22 Jan – 11 Mar 2020	Exponential Smoothing	Akurasi	- Persentase Error : 6,2% - Akurasi : 93,8%	2020
8.	Global	1. Kasus Baru. 2. Total Meninggal.	23 Mar – 30 Jul 2020	Nonlinear Autoregressive Artificial Neural Network	MAPE	1. MAPE : 8.65%; 2. MAPE : 12.33%;	2020
9.	Global	1. Kasus Baru. 2. Total Sembuh. 3. Total Meninggal.	Jan 2020 – Mar 2020	Single ES, Weighted MA	MAD, MSE, RMSE, MAPE	Kasus Baru : 1. WMA : - MAD : 4614.96 - MSE : 32413677.85 - RMSE : 5693.30 - MAPE : 11.16% 2. SES : - MAD : 3385.65 - MSE : 20050014.56 - RMSE : 4477.72 - MAPE : 9.68% Total Sembuh : 1. WMA : - MAD : 697.39 - MSE : 956531.43 - RMSE : 978.02 - MAPE : 19.30% 2. SES : - MAD : 517.54 - MSE : 523335.16 - RMSE : 723.42 - MAPE : 16.38%	2020

						Total Meninggal : 1. WMA : - MAD : 113.37 - MSE : 17056.99 - RMSE : 130.60 - MAPE : 12.28% 2. SES : - MAD : 82.44 - MSE : 9685.59 - RMSE : 98.42 - MAPE : 9.43%	
10.	Negara (Jepang)	Total Kasus	15 Jan – 29 Feb 2020	SEIR	R_0	$R_0 = 2.6$	2020
11.	Negara (India)	1. Kasus Baru. 2. Total Sembuh. 3. Total Meninggal.	30 Jan – 3 Apr 2020	Timeseries Forecasting	MAE, RMSE	1. Kasus Baru : - MAE : 485,45 - RMSE : 613,63 2. Total Sembuh : - MAE : 14,04 - RMSE : 19,31 3. Total Meninggal : - MAE : 11,27 - RMSE : 29,5	2020
12.	Negara (China)	Kasus Baru	31 Des 2019 – 19 Jan 2020	SIR	UCB, LCB	- UCB : 74.350 - Mean : 47.990 - LCB : 21.630	2020
13.	Negara (China)	Jumlah Kematian	12 Des 2019 – 28 Feb 2020	Logistic Model, Gompertz Model, Bertalanffy Model	R^2	1. Logistic Model : 0.9993 2. Gompertz Model : 0.9967 3. Bertalanffy Model : 0.9993	2020
14.	Negara (China)	Kasus Baru	20 Jan 2020 – 9 Feb 2020	SMSI	RMSE, MAE, MAPE	1. Subset Selection : - RMSE : 51.6671 - MAE : 34.0739 - MAPE : 0.0107 2. Forward Selection :	2020

						- RMSE : 70.0168 - MAE : 39.9790 - MAPE : 0.0113 3. Ridge Regression : - RMSE : 415.2922 - MAE : 279.6788 - MAPE : 0.0827 4. Lasso Regression : - RMSE : 519.7440 - MAE : 358.0979 - MAPE : 0.1032	
15.	Global	1. Kasus Baru. 2. Kasus Sembuh. 3. Kasus Meninggal.	Des 2019 – Feb 2020	ARIMA, ES, LR, SVM, PNN+cf, DT	RMSE	- ARIMA : 1016.2687 - ES : 962.9372 - LR : 520.1573 - SVM : 228.3925 - PNN+cf : 136.547 - DT : 1744.5256	2020
16.	Global	1. Kasus Baru. 2. Jumlah Meninggal. 3. Jumlah Sembuh. 4. Kasus Aktif.	Mar 2020 – Feb 2021	Proposed Model, SVM, LR, CNN	Akurasi	1. Proposed Model : 94,03% 2. SVM : 88,85% 3. LR : 86,93% 4. CNN : 93,17%	2021
17.	Negara (India)	1. Kasus Baru. 2. Jumlah Meninggal. 3. Jumlah Sembuh. 4. Kasus Aktif.	14 Mar 2020 – 03 Jul 2020	SEIAQRDT, SIRD, SEIR, LSTM	Akurasi	1. SEIAQRDT : 98,6343 2. SEIR : 92,8477 3. SIRD : 96,01 4. LSTM : 95,5818	2020
18.	Global	1. Kasus Baru. 2. Jumlah Meninggal. 3. Jumlah Sembuh.	22 Jan 2020 – 13 Apr 2020	ARIMA, ES, DES, MA, S-Curve	MAPE, MAD, MSD	1. ARIMA : MAPE (4.1), MAD (58.3), MSD (25319.5) 2. SES : MAPE (9.7), MAD (98.3), MSD (58982.5) 3. DES : MAPE (9.1), MAD (52.2), MSD (18909.5) 4. MA : MAPE (10), MAD (141), MSD (103715)	2020

						5. S-Curve : MAPE (66), MAD (1094), MSD (6798514)	
19.	Negara (India)	Kasus Baru	22 Jan 2020 – 25 Apr 2020	SVM	MAE, Akurasi	1. MAE : 17,57 2. Akurasi : 82,47%	2020
20.	Negara Bagian (Louisiana)	Kasus Baru	22 Jan 2020 – 19 Apr 2020	K-Means LSTM, SEIR	RMSE	1. K-Means LSTM : RMSE 601.20 2. SEIR : RMSE 3615.83	2020
21.	Negara (India)	Kasus Baru	Mar 2020 – Apr 2020	SIS[e] Model	R_0[b]	R_0[b] : 2,84	2020
22.	Negara (China)	Kasus Baru	23 Jan 2020 – 28 Feb 2020	ANN, KNN, SVR, ANFIS, PSO, GA, ABC, FPA, FPASSA	RMSE, MAE, MAPE, RMSRE, R2, Time	FPASSA : - RMSE : 5779 - MAE : 4271 - MAPE : 4.79 - RMSRE : 0.07 - R2 : 0.9645 - Time : 23.30	2020

Tabel di atas adalah rangkuman penelitian-penelitian terdahulu yang meneliti tentang prediksi data jumlah penderita Covid-19 dari berbagai rentang waktu dan wilayah. Penelitian-penelitian tersebut menggunakan model dan parameter evaluasi yang berbeda satu dengan yang lainnya namun menggunakan variabel yang sama untuk diprediksi sehingga menghasilkan kesimpulan penelitian yang berbeda-beda. Tujuan yang ingin dicapai penulis adalah untuk mencari model terbaik diantara model-model sebelumnya sehingga hasil dari prediksi menggunakan model terbaik akan lebih dapat diandalkan. Oleh karena itu penulis akan melakukan komparasi antara penelitian-penelitian terdahulu dari sisi atribut, model, dan parameter evaluasi.

Berikut adalah ringkasan hasil dari penelitian yang pernah dilakukan dari seluruh dunia.

2.3 Tinjauan Objek Penelitian

Pada subbab ini akan dijelaskan mengenai tempat penulis melakukan penelitian yang meliputi profil organisasi Kemenkes sebagai instansi utama yang menangani bidang kesehatan khususnya pandemi Covid-19 yang terjadi di Indonesia.

2.3.1 Profil Organisasi Kementerian Kesehatan

Kementerian Kesehatan merupakan sebuah instansi pemerintah yang menangani bidang Kesehatan. Kementerian Kesehatan pertama kali dibentuk pada tanggal 17 Agustus 1945 pada Kabinet Presidensial atau kabinet pertama yang dibentuk pada masa pemerintahan presiden Soekarno dengan Boentaran Martoadmojo sebagai Menteri pertamanya.

Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 25 Tahun 2020 tentang Organisasi dan Tata Kerja Kementerian Kesehatan Pasal 1, 2 dan 3, yang membahas Kedudukan, Tugas dan Fungsi Kementerian Kesehatan, Kementerian Kesehatan berada di bawah dan bertanggung jawab kepada Presiden dan Kementerian Kesehatan dipimpin oleh Menteri Kesehatan.

Kementerian Kesehatan memiliki tugas menyelenggarakan urusan pemerintahan bidang kesehatan untuk membantu Presiden dalam menyelenggarakan pemerintahan negara. Sedangkan untuk fungsi yang diselenggarakan oleh Kementerian Kesehatan adalah :

1. Perumusan, penetapan, dan pelaksanaan kebijakan di bidang kesehatan masyarakat, pencegahan dan pengendalian penyakit, pelayanan kesehatan, dan kefarmasian dan alat kesehatan;
2. Koordinasi pelaksanaan tugas, pembinaan, dan pemberian dukungan administrasi kepada seluruh unsur organisasi di lingkungan Kementerian Kesehatan;
3. Pengelolaan barang milik negara yang menjadi tanggung jawab Kementerian Kesehatan;
4. Pelaksanaan penelitian dan pengembangan di bidang kesehatan;
5. Pelaksanaan pengembangan dan pemberdayaan sumber daya manusia di bidang kesehatan serta pengelolaan tenaga kesehatan;
6. Pelaksanaan bimbingan teknis dan supervisi atas pelaksanaan urusan Kementerian Kesehatan di daerah;
7. Pengawasan atas pelaksanaan tugas di lingkungan Kementerian Kesehatan; dan
8. Pelaksanaan dukungan substantif kepada seluruh unsur organisasi di lingkungan Kementerian Kesehatan.

Adapun visi, misi, dan tujuan strategis Kementerian Kesehatan tahun 2020-2024 adalah :

1. Visi
   "Menciptakan manusia yang sehat, produktif, mandiri dan berkeadilan".

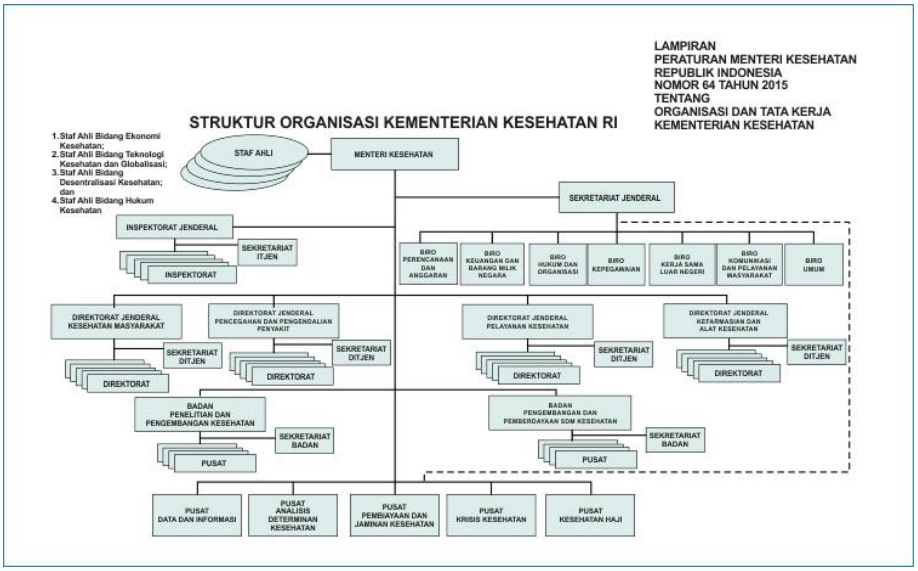
2. Misi
   a. Menurunkan angka kematian ibu dan bayi;
   b. Menurunkan angka stunting pada balita;
   c. Memperbaiki pengelolaan Jaminan Kesehatan Nasional; dan
   d. Meningkatkan kemandirian dan penggunaan produk farmasi dan alat kesehatan dalam negeri.

3. Tujuan Strategis
   a. Meningkatkan derajat kesehatan masyarakat melalui pendekatan siklus hidup;
   b. Penguatan pelayanan kesehatan dasar dan rujukan;
   c. Peningkatan pencegahan dan pengendalian penyakit dan pengelolaan kedaruratan kesehatan masyarakat;
   d. Peningkatan sumber daya kesehatan.

Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 25 Tahun 2020 tentang Organisasi dan Tata Kerja Kementerian Kesehatan Pasal 4 yang membahas Susunan Organisasi, Susunan Organisasi Kementerian Kesehatan terdiri atas :

1. Sekretariat Jenderal;
2. Direktorat Jenderal Kesehatan Masyarakat;
3. Direktorat Jenderal Pencegahan dan Pengendalian Penyakit;
4. Direktorat Jenderal Pelayanan Kesehatan;
5. Direktorat Jenderal Kefarmasian dan Alat Kesehatan;
6. Inspektorat Jenderal;
7. Badan Penelitian dan Pengembangan Kesehatan;
8. Badan Pengembangan dan Pemberdayaan Sumber Daya Manusia Kesehatan;
9. Staf Ahli Bidang Ekonomi Kesehatan;
10. Staf Ahli Bidang Teknologi Kesehatan dan Globalisasi;
11. Staf Ahli Bidang Desentralisasi Kesehatan;
12. Staf Ahli Bidang Hukum Kesehatan;
13. Pusat Data dan Informasi;
14. Pusat Analisis Determinan Kesehatan;
15. Pusat Pembiayaan dan Jaminan Kesehatan;
16. Pusat Krisis Kesehatan; dan
17. Pusat Kesehatan Haji.

Berikut adalah gambar Struktur Organisasi Kementerian Kesehatan menurut Peraturan Menteri Kesehatan Nomor 25 tahun 2020 tentang Organisasi dan Tata Kerja Kementerian Kesehatan :



Gambar 2.3 Struktur Organisasi Kementerian Kesehatan

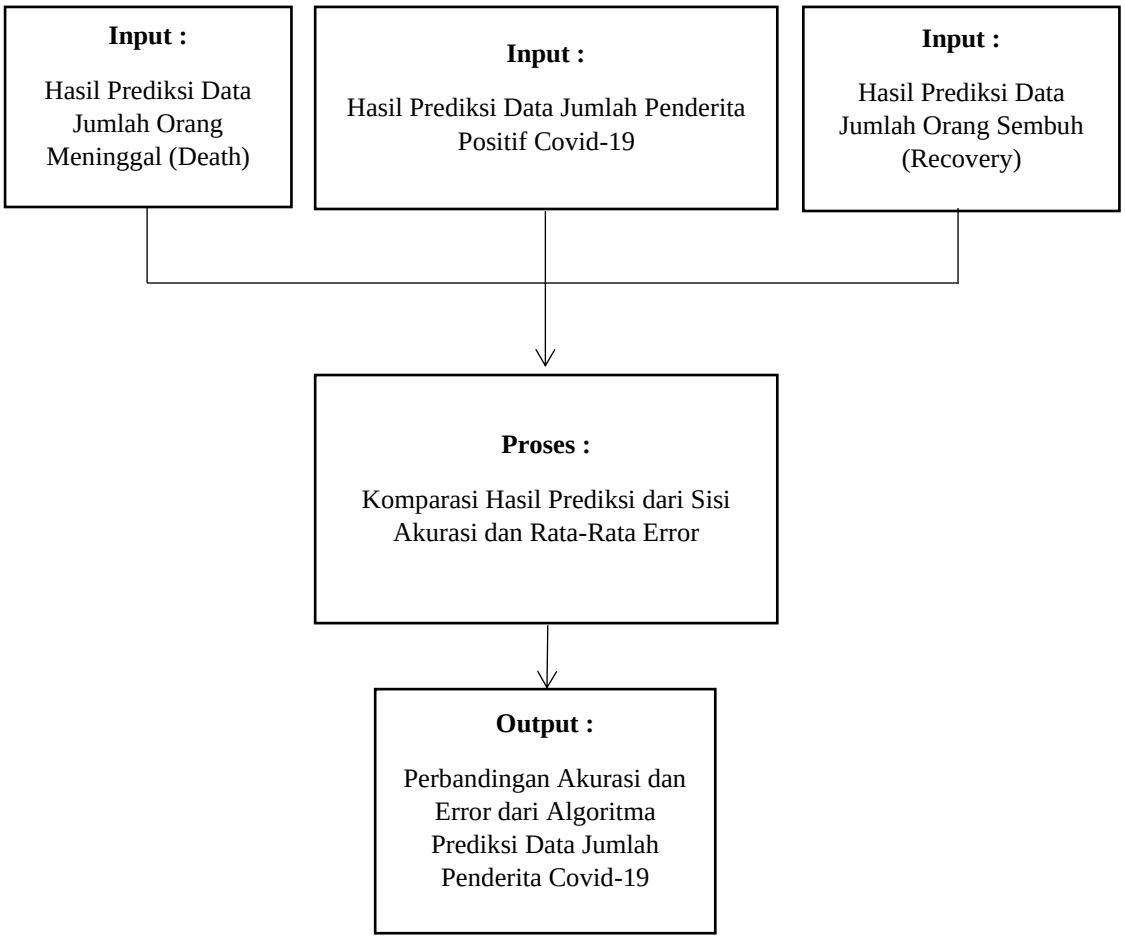
Kementerian Kesehatan memiliki 17 unit Eselon I yang terdiri dari 8 Direktorat Jenderal, 4 Staf Ahli, dan 5 Pusat. Selain itu juga terdapat Kantor Wilayah di seluruh provinsi di Indonesia yang bertanggung jawab kepada Sekretariat Jenderal dan Unit Pelaksana Teknis seperti Rumah Sakit Pemerintah dan Sekolah Kedinasan yang tersebar di seluruh wilayah Indonesia.

Struktur organisasi yang ada saat ini adalah struktur yang terbaru dan mungkin akan berubah di masa depan untuk menyesuaikan kebutuhan dari Kementerian Kesehatan.

2.4 Kerangka Konsep/Pola Pikir Pemecahan Masalah

Menurut Notoatmojo (2010), kerangka konsep merupakan justifikasi ilmiah terhadap penelitian yang dilakukan dan memberi landasan kuat terhadap judul yang dipilih sesuai dengan identifikasi masalahnya.

Kerangka konsep yang akan digunakan dalam penelitian ini adalah sebagai berikut :



Gambar 2.4 Kerangka Konsep Penelitian

Kerangka konsep penelitian ini terdiri dari 3 bagian (Input, Proses, Output), yaitu :

1. Input
   Bagian ini menunjukkan variabel yang menjadi input untuk penelitian ini, dimana penelitian ini menggunakan 3 variabel, yaitu :
   a. Hasil Prediksi Data Jumlah Orang Meninggal (Death)
      Variabel ini diambil dari hasil prediksi pada variabel jumlah meninggal menggunakan berbagai macam model prediksi, kemudian hasilnya akan dibandingkan dengan data aktual pada bagian Proses.
   b. Hasil Prediksi Data Jumlah Penderita Positif Covid-19
      Variabel ini diambil dari hasil prediksi pada variabel jumlah terkonfirmasi positif menggunakan berbagai macam model prediksi, kemudian hasilnya akan dibandingkan dengan data aktual pada bagian Proses.
   c. Hasil Prediksi Data Jumlah Orang Sembuh (Recovery)

Variabel ini diambil dari hasil prediksi pada variabel jumlah sembuh menggunakan berbagai macam model prediksi, kemudian hasilnya akan dibandingkan dengan data aktual pada bagian Proses.

2. Proses
   Bagian ini menunjukkan proses untuk melakukan komparasi hasil prediksi dari sisi akurasi dan mean error. Pertama akan dilakukan perbandingan antara data hasil prediksi dengan data aktual, kemudian akan dihitung akurasi dan mean error dari perbandingan tersebut. Selanjutnya akan dilakukan pengurutan berdasarkan hasil perhitungan rata-rata performa pada proses output.

3. Output
   Bagian ini menunjukkan hasil perbandingan antara model prediksi yang digunakan pada bagian sebelumnya dari sisi akurasi dan mean error. Hasil akhir dari penelitian ini adalah urutan performa model-model prediksi dan rekomendasi penggunaan model prediksi untuk melakukan peramalan data jumlah penderita Covid-19.

2.5 Hipotesis

Hipotesis adalah jawaban sementara yang kebenarannya harus diuji atau rangkuman kesimpulan secara teoritis yang diperoleh melalui tinjauan pustaka (Martono, 2010:57). Berdasarkan hasil analisa dari kerangka konsep/pola pikir pemecahan masalah pada subbab sebelumnya, dapat dibuat hipotesis sebagai berikut :

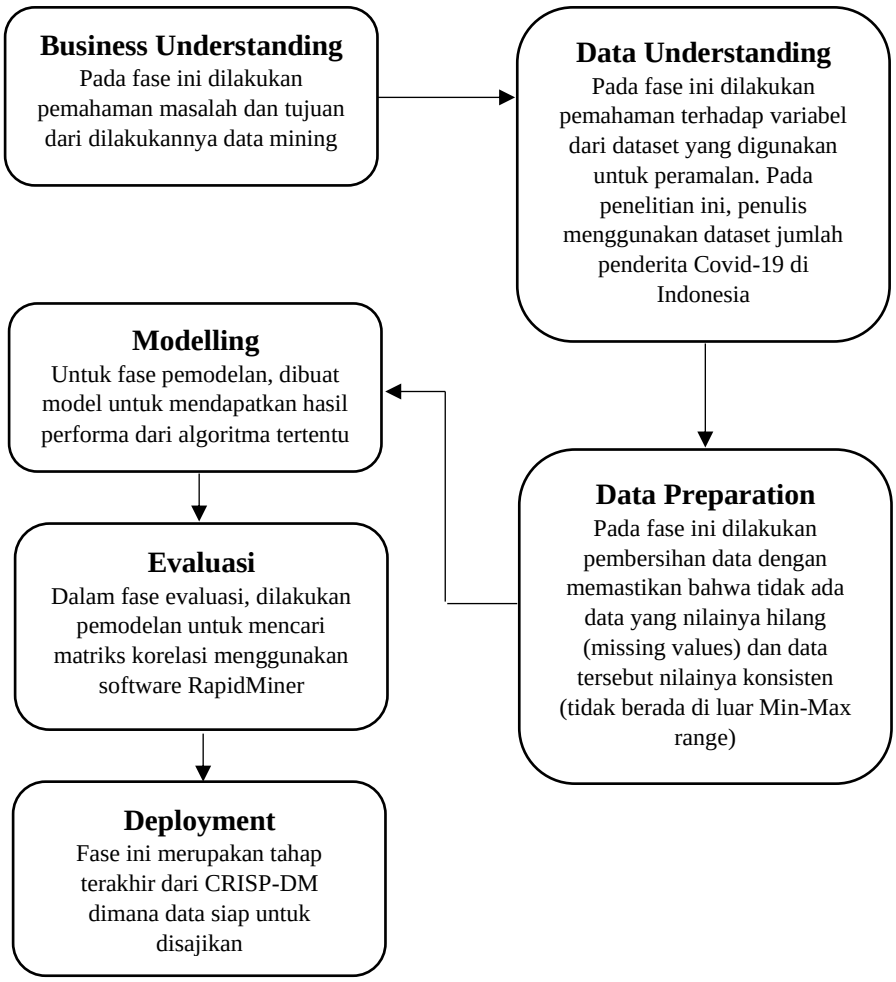
“Diperkirakan rekomendasi model peramalan data jumlah penderita Covid-19 yang dapat memberikan hasil prediksi dengan akurasi tertinggi dan rata-rata error terendah cenderung mengarah kepada model ARIMA karena berdasarkan hasil penelitian terdahulu, hasil prediksi menggunakan ARIMA memiliki nilai MAPE terkecil yaitu 1.301”.

BAB III
METODOLOGI DAN RANCANGAN/DESAIN PENELITIAN

3.1 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah Metode Kuantitatif dengan pendekatan deskriptif. Metode Kuantitatif merupakan suatu teknik yang berdasarkan pada filsafat positivisme dan digunakan untuk menganalisa populasi atau sampel tertentu, metode sampling yang digunakan biasanya dilakukan secara acak, pengumpulan data menggunakan instrumen penelitian analisa data bersifat kuantitatif atau statistik dengan maksud untuk melakukan pengujian terhadap hipotesis yang telah dikemukakan (Sugiyono, 2013:13). Sedangkan definisi deskriptif menurut Sugiyono (2012:29) adalah teknik yang berfungsi dari untuk mendeskripsikan atau menggambarkan objek penelitian melalui data/sampel yang terkumpul sesuai dengan keadaan data tersebut apa adanya tanpa melakukan analisis dan kesimpulan yang berlaku umum.

Secara umum, penelitian ini menggunakan metodologi CRISP-DM seperti sebagai berikut :



Gambar 3.1 Metodologi Penelitian

3.1.1 Metode Data Mining

Tahap awal penelitian dimulai dengan melakukan data mining dari dataset yang digunakan untuk penelitian ini, yaitu data jumlah penderita Covid-19 periode 2020-2021 di Indonesia. Metode yang akan digunakan untuk melakukan data mining adalah CRISP-DM. Pada CRISP-DM sendiri terdiri dari 6 fase, yaitu :

1. Fase Business Understanding
   Pada fase ini dilakukan pemahaman masalah dan tujuan dari dilakukannya data mining. Masalah yang terjadi saat ini adalah bahwa pemerintah dan masyarakat mengalami kesulitan dalam membendung lonjakan kasus Covid-19 yang terjadi begitu cepat, seperti terbatasnya jumlah fasilitas dan tenaga kesehatan yang tersedia saat ini. Sehingga dilakukanlah *forecasting* data jumlah penderita Covid-19 di masa depan sebagai salah satu solusi untuk mengatasi masalah tersebut. Untuk mendapatkan hasil peramalan data di masa depan, perlu dilakukan data mining untuk penderita Covid-19 di masa lalu. Tujuannya adalah agar data tersebut dapat dianalisa dan diproses menggunakan model forecasting sehingga didapatkan hasil prediksi data jumlah penderita Covid-19 selama beberapa hari ke depan.

2. Fase Data Understanding
   Pada fase ini dilakukan pemahaman terhadap variabel dari dataset yang digunakan untuk peramalan. Pada penelitian ini, penulis menggunakan dataset jumlah penderita Covid-19 di Indonesia, dimana dataset tersebut terdiri dari variabel sebagai berikut :

Tabel 3.1 Variabel pada Dataset yang Digunakan

Field	Tipe Data
Date	DATE
Country	STRING
Confirmed	INT
Recovered	INT
Death	INT

a. Variabel Date menunjukkan tanggal pada saat data tersebut diambil, menggunakan format “YY-m-d”.
b. Variabel Country menunujukkan negara dimana data tersebut diambil.
c. Variabel Confirmed menunjukkan jumlah orang yang terkonfirmasi positif Covid-19.
d. Variabel Recovered menunjukkan jumlah orang yang sudah terkonfirmasi positif Covid-19 dan sudah dinyatakan sembuh atau negatif.

e. Variabel Death menunjukkan jumlah orang yang sudah terkonfirmasi positif Covid-19 dan meninggal dengan status positif.

3. Fase Data Preparation

Pada fase ini dilakukan pembersihan data dengan memastikan bahwa tidak ada data yang nilainya hilang (missing values) dan data tersebut nilainya konsisten (tidak berada di luar Min-Max range). Berdasarkan analisa menggunakan software RapidMiner Studio, didapatkan hasil sebagai berikut :

Tabel 3.2 Hasil Statistik Dataset Menggunakan Rapid Miner Studio

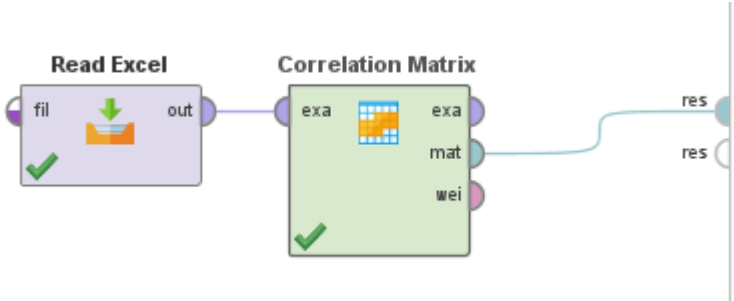
Nama	Tipe	Missing	Min/Least	Max/Most	Average
Date	Nominal	0	1	1	1
Country	Nominal	0	Indonesia	Indonesia	Indonesia
Confirmed	Integer	0	0	2527203	658290.136
Recovered	Integer	0	0	2084724	566971.207
Deaths	Integer	0	0	66464	18944.115

4. Fase Modelling

Untuk fase pemodelan, dibuat model untuk mendapatkan hasil performa dari setiap algoritma. Model dirancang untuk memproses data yang dibagi menjadi dua bagian, yaitu Data Training dan Data Testing. Data Training akan digunakan pada fase evaluasi, sedangkan Data Testing akan digunakan pada fase *deployment*. Pada bagian output juga dirancang parameter evaluasi untuk menilai tingkat error dari setiap algoritma agar dapat dilakukan perankingan. Pemodelan dilakukan menggunakan aplikasi Rapidminer untuk semua algoritma yang digunakan.

5. Fase Evaluation

Dalam fase evaluasi, dilakukan pemodelan untuk mencari matriks korelasi menggunakan software RapidMiner, penulis menggunakan model sebagai berikut :



Gambar 3.2 Model untuk Mencari Korelasi

Untuk mencari hubungan antar variabel, penulis menggunakan fitur Read Excel untuk mengimpor data dari format xlsx, kemudian menggunakan modul "Correlation Matrix" pada RapidMiner Studio. Kemudian menghubungkan output dari dataset ke input exa pada modul Correlation Matrix. Terakhir adalah menghubungkan output mat pada Correlation Matrix ke bagian res (result) untuk menampilkan hasilnya.

Dari proses tersebut, didapatkan hasil matriks korelasi sebagai berikut :

Attribut...	Date	Country	Confirm...	Recove...	Deaths
Date	1	?	?	?	?
Country	?	1	?	?	?
Confirmed	?	?	1	0.999	0.998
Recovered	?	?	0.999	1	0.997
Deaths	?	?	0.998	0.997	1

Gambar 3.3 Hasil Evaluasi Korelasi Antar Variabel

Variabel "Country" menunjukkan tidak memiliki korelasi dengan variabel lainnya karena variabel tersebut berisikan satu nilai saja, yaitu "Indonesia". Sedangkan variabel "Date" berisikan tanggal yang nilainya selalu unik karena setiap row menunjukkan tanggal yang berbeda. Namun karena variabel "Date" memiliki tipe Nominal, maka hasilnya adalah NaN. Tetapi variabel "Date" tetap dibutuhkan dalam pemrosesan dataset ini karena akan digunakan sebagai axis X untuk penyajian dalam bentuk grafik *timeseries* nanti.

Pada fase ini juga dilakukan pengujian terhadap beberapa algoritma *machine learning* untuk melakukan prediksi kepada variabel Terkonfirmasi Positif, Sembuh, dan Meninggal. Model yang didesain untuk masing-masing algoritma akan berbeda tergantung dari jenis algoritma yang diuji, namun hasil output yang dihasilkan akan menggunakan parameter evaluasi yang sama yaitu RMSE dan Absolute Error.

Data yang akan digunakan pada fase ini adalah data pada porsi Training. Setelah mendapatkan hasil evaluasi dari Training yang diuji, akan dilakukan pengurutan peringkat berdasarkan parameter evaluasi yang digunakan untuk masing-masing algoritma. Sehingga didapatkan algoritma terbaik yang akan digunakan pada fase berikutnya.

6. Fase Deployment

   Fase ini merupakan tahap terakhir dari CRISP-DM dimana data siap untuk disajikan. Pada tahap ini, variabel "Country" dihilangkan dari dataset karena tidak memiliki pengaruh terhadap variabel lain.

Pada fase ini juga akan dilakukan pengujian pembali pada Data Testing dari algoritma terbaik yang sudah diuji pada tahap sebelumnya. Pengujian kali ini tidak menggunakan aplikasi RapidMiner namun akan menggunakan prototipe aplikasi yang dibangun secara mandiri menggunakan bahasa pemrograman Python 3.9.5 menggunakan framework Flask.

3.2 Sampling/Metode Pemilihan Sampel

Teknik pemilihan sampel yang digunakan pada penelitian ini adalah Sensus atau Sampling Jenuh. Teknik Sensus atau Sampling Jenuh merupakan teknik pengambilan sampel dimana kita menggunakan semua anggota dari populasi menjadi sampel (Sugiyono, 2008:78). Alasan penggunaan teknik ini adalah karena data yang digunakan untuk dianalisa adalah karena data yang didapat menunjukkan data jumlah penderita Covid-19 dari awal kemunculan wabah pada Januari 2020 sampai Oktober 2021 sesuai dengan batasan penelitian ini.

3.3 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari :

1. Data Sekunder

   Data sekunder untuk penelitian ini adalah berupa data kasus positif, pasien meninggal dan pasien sembuh dari penyakit Covid-19 yang berasal dari seluruh dunia dan diambil dari situs www.kaggle.com.

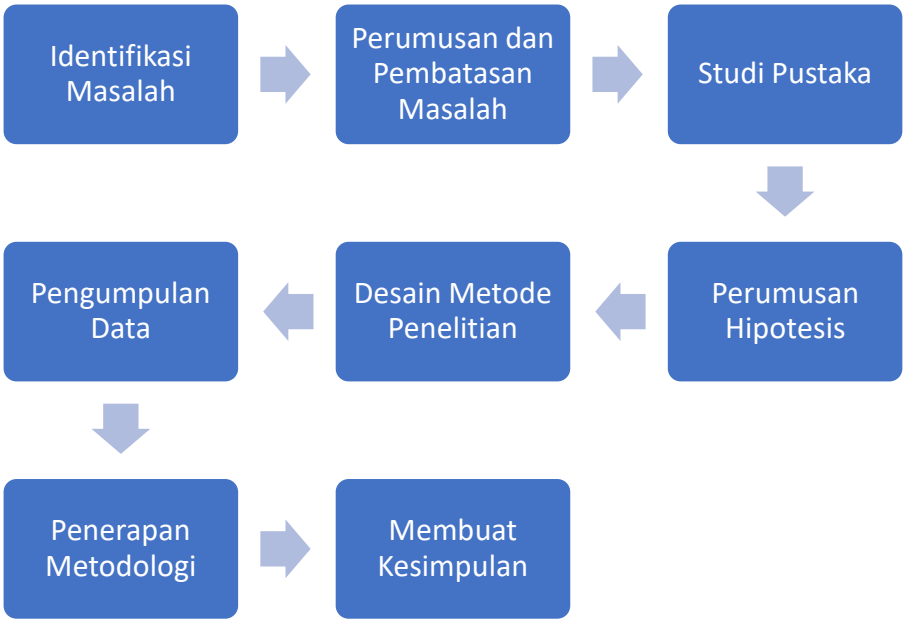
Metode pengumpulan yang dilakukan dalam penelitian ini adalah :

1. Studi Pustaka

   Penulis melakukan pengumpulan informasi dan data yang berasal dari karya tulis ilmiah seperti tesis, jurnal, dan buku.

3.4 Langkah-langkah Penelitian

Penelitian ini terdiri dari 8 tahap, yang dimulai dari Identifikasi Masalah hingga Pembuatan Kesimpulan. Secara detail, tahapan-tahapan tersebut dapat dilihat pada gambar dan penjelasan berikut :



Gambar 3.4 Tahapan Penelitian

3.4.1 Identifikasi Masalah

Pada tahap ini dilakukan penentuan topik penelitian dan identifikasi masalah yang terjadi dengan melakukan observasi pada lingkungan tempat penulis bekerja, dengan mengidentifikasi isu, fenomena atau permasalahan yang ada.

3.4.2 Merumuskan dan Membatasi Masalah

Berangkat dari masalah yang sudah diidentifikasi sebelumnya, penulis melakukan perumusan masalah agar penelitian yang dilakukan memiliki tujuan yang jelas. Selain itu masalah yang ada perlu diberikan pembatasan agar penyelesaian yang kita bahas tidak keluar dari topik pembahasan.

3.4.3 Melakukan Studi Kepustakaan

Meninjau literatur dilakukan untuk mengetahui apakah penelitian yang sejenis dengan penelitian yang akan kita lakukan sudah pernah dibuat sebelumnya, selain itu kita juga dapat mencari referensi dari penelitian-penelitian sebelumnya. Kita juga dapat memahami aspek-aspek tertentu dari isu, fenomena atau masalah yang pernah diselesaikan pada penelitian sebelumnya. Biasanya proses ini dilakukan melalui buku, jurnal, tesis dan publikasi-publikasi lainnya.

3.4.4 Merumuskan Hipotesis atau Pertanyaan Penelitian

Di tahap ini dilakukan perumusan hipotesis berdasarkan metodologi yang akan kita gunakan yang sudah kita kemukakan pada saat menjelaskan latar belakang penelitian. Hipotesis yang diberikan hendaknya harus disampaikan berdasarkan model penelitian yang digunakan.

3.4.5 Menentukan Desain dan Metode Penelitian

Metodologi dan model penelitian dapat kita buat sesuai dengan kebutuhan penelitian yang akan kita lakukan. Penulis mencari referensi mengenai desain model dan metodologi yang akan dilakukan dari beberapa penelitian sebelumnya sejenis dengan penelitian ini.

3.4.6 Menyusun Instrumen dan Mengumpulkan Data

Selanjutnya adalah membuat instrumen untuk mengumpulkan data yang dibutuhkan untuk penelitian. Data yang dikumpulkan nantinya akan diproses menggunakan metodologi yang akan dirancang pada tahap selanjutnya.

3.4.7 Menerapkan Metodologi untuk Memproses Data

Tahap ini merupakan tahap dimana proses analisis data dilakukan. Proses analisis data dilakukan sesuai dengan tahapan pada metodologi yang sudah dirancang dan menggunakan alat analisis yang sudah ditentukan pada tahap sebelumnya sehingga dihasilkan sebuah output yang dapat menjawab pertanyaan pada rumusan masalah.

3.4.8 Memberikan Kesimpulan

Tahap terakhir adalah membuat konklusi atau kesimpulan dari hasil penelitian yang kita lakukan. Selain itu dimuat juga saran untuk menunjukkan kekurangan dari penelitian kita sehingga dapat diperbaiki atau dikembangkan pada penelitian selanjutnya

3.5 Jadwal Penelitian

Berikut adalah tabel timeline untuk pelaksanaan penelitian ini :

Tabel 3.3 Timeline Pelaksanaan Penelitian

No.	Uraian	Agustus				September				Oktober				November				Desember			
		Minggu ke																			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Identifikasi Masalah	■																			
2.	Merumuskan Masalah		■																		
3.	Studi Pustaka			■	■																
4.	Merumuskan Hipotesis				■																
5.	Menentukan Model Penelitian					■	■	■													
6.	Mengumpulkan Data								■	■	■	■									
7.	Memproses Data												■	■	■	■	■	■	■	■	
8.	Membuat Kesimpulan																				■

Waktu penelitian akan dilakukan selama 5 bulan yang terhitung dari bulan Agustus sampai bulan Desember. Adapun rincian waktu yang digunakan adalah sebagai berikut :

1. Minggu ke 1 Bulan Agustus : Identifikasi Masalah dilakukan bersamaan dengan tahap penyusunan Proposal Penelitian.
2. Minggu ke 2 Bulan Agustus : Perumsuan Masalah dilakukan bersamaan dengan tahap penyusunan Proposal Penelitian.
3. Minggu ke 3 sampai 4 Bulan Agustus : Studi Pustaka dilakukan bersamaan dengan tahap penyusunan Proposal Penelitian, adapun setelah Proposal selesai disusun, Studi Pustaka akan tetap dilakukan sesuai dengan kebutuhan penelitian.
4. Minggu ke 4 Bulan Agustus : Dilakukan perumusan Hipotesis penelitian berdasarkan tinjauan teori pada tahap sebelumnya.
5. Minggu ke 1 sampai 3 Bulan September : Penentuan Model Penelitian dilakukan berdasarakan hasil dari *Preliminary Research* untuk mencari model terbaik yang akan digunakan sebagai alat untuk memproses data.
6. Minggu ke 4 Bulan September sampai Minggu ke 3 Bulan Oktober : Melakukan Pengumpulan Data.
7. Minggu ke 4 Bulan Oktober sampai Minggu ke 3 Bulan Desember : Melakukan Pemrosesan Data menggunakan Model yang sudah ditentukan pada tahap sebelumnya.
8. Minggu ke 4 Bulan Desember : Menyimpulkan hasil dari penelitian yang sudah dilakukan pada tahap sebelumnya.

BAB IV
PEMBAHASAN

4.1 Business Understanding

Berdasarkan data dari John Hopkins University Coronavirus Research Center, perkembangan jumlah kasus baru penduduk terkonfirmasi positif virus corona khususnya di Indonesia terlihat menurun sejak tanggal 18 Juli 2021. Namun hal ini berbanding terbalik dengan beberapa negara di Eropa yang justru mengalami peningkatan yang cukup signifikan di akhir tahun 2021 ini dengan negara Slovenia sebagai negara dengan peningkatan kasus tertinggi pada bulan November 2021.

Salah satu cara yang dilakukan untuk mengantisipasi lonjakan kasus di masa depan adalah dengan melakukan prediksi data jumlah kasus positif menggunakan data di masa lalu. Beberapa peneliti dari seluruh dunia telah melakukan penelitian untuk melakukan prediksi jumlah kasus positif, kesembuhan dan kematian dari para penderita Covid-19 dalam berbagai cakupan wilayah dari seluruh dunia.

Berdasarkan hasil penelitian yang dilakukan, para peneliti menggunakan berbagai macam rentang waktu dari data yang dikumpulkan, dan menggunakan model serta parameter evaluasi yang berbeda untuk melakukan prediksi terhadap variabel yang sama. Data yang digunakan umumnya memiliki rentang waktu 2 minggu sampai 2 bulan. Variabel yang umumnya diprediksi adalah Kasus Positif Harian (Daily Case), Jumlah Sembuh (Recovery), dan Jumlah Kematian (Death). Variabel tersebut kemudian diproses menggunakan berbagai model seperti ARIMA, LSTM, Exponential Smoothing, Linear Regression, SVM dan model-model lainnya. Selanjutnya hasilnya dievaluasi menggunakan berbagai macam parameter, dimana yang umum digunakan adalah MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MSE (Mean Square Error), RMSE (Root Mean Square Error), dan parameter evaluasi lainnya. Perbedaan model dan teknik evaluasi inilah yang membuat hasil dari prediksi setiap penelitian berbeda-beda, karena tingkat akurasi yang dihasilkan oleh setiap model pasti berbeda.

Oleh karena itu, diperlukan sebuah penelitian untuk melakukan komparasi model dan parameter evaluasi dari penelitian-penelitian sebelumnya untuk mendapatkan model dengan hasil akurasi terbaik.

4.2 Data Understanding

Pada penelitian ini, penulis menggunakan dataset yang diambil dari repository pada John Hopkins University Coronavirus Research Center. Dari dataset tesebut akan diambil sampel data dari 2 bulan terakhir. Data yang akan diambil adalah data jumlah penderita Covid-19 dari seluruh dunia. Variabel yang digunakan adalah Jumlah Penderita Positif Harian, Jumlah Penderita Sembuh, dan Jumlah Penderita Meninggal. Data tersebut kemudian akan diolah menggunakan *software* Rapid Miner dengan pengujian menggunakan beberapa algoritma yaitu

ARIMA (1,0,1), Exponential Smoothing, Linear Regression, Support Vector Regression, dan LSTM.

4.3 Data Preparation

Untuk dataset yang akan digunakan dalam penelitian ini menggunakan data yang didownload dari Kaggle menggunakan *repository* dari John Hopkins University. Dataset tersebut berbentuk csv dan diberi nama worldwide-aggregate.csv. Dataset memiliki *record* kasus penderita Covid-19 sejak tanggal 22 Januari 2020 sampai 4 November 2021, dengan field Confirmed (Kasus Baru), Recovered (Sembuh), Deaths (Meninggal), dan Increase Rate (Laju Pertumbuhan Kasus). Data yang akan digunakan berjumlah 652 row, dengan pembagian antara Training dan Testing sebanyak 80:20 dimana untuk training menjadi 522 row dan Testing menjadi 130 row.

Berikut adalah gambaran dataset yang akan digunakan dalam penelitian ini :

Row No.	Date	Confirmed	Recovered	Deaths	Increase rate
1	Jan 22, 2020	557	30	17	?
2	Jan 23, 2020	655	32	18	17.594
3	Jan 24, 2020	941	39	26	43.664
4	Jan 25, 2020	1434	42	42	52.391
5	Jan 26, 2020	2118	56	56	47.699
6	Jan 27, 2020	2927	65	82	38.196
7	Jan 28, 2020	5578	108	131	90.571
8	Jan 29, 2020	6167	127	133	10.559
9	Jan 30, 2020	8235	145	171	33.533
10	Jan 31, 2020	9927	225	213	20.546
11	Feb 1, 2020	12038	287	259	21.265
12	Feb 2, 2020	16787	476	362	39.450
13	Feb 3, 2020	19887	627	426	18.467
14	Feb 4, 2020	23898	857	492	20.169
15	Feb 5, 2020	27643	1130	564	15.671

Gambar 4.1 Dataset Kasus Penderita Covid-19 dari Seluruh Dunia

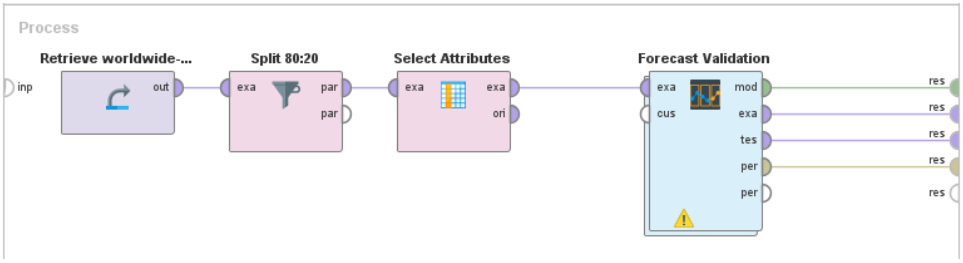
4.4 Modelling

Pada tahap ini dilakukan desain model yang akan digunakan untuk melakukan pengujian terhadap data training dan testing. Pengujian terhadap data

testing dan training akan dilakukan menggunakan 5 algoritma yang memiliki performa terbaik dari keseluruhan algoritma yang digunakan pada penelitian terdahulu berdasarkan pembahasan pada bab sebelumnya. Algoritma yang digunakan adalah ARIMA, Exponential Smoothing, Linear Regression, SVR, dan LSTM. Rasio data training dan testing yang digunakan pada pengujian ini adalah 80:20. Berikut adalah pembahasan dari setiap algoritma :

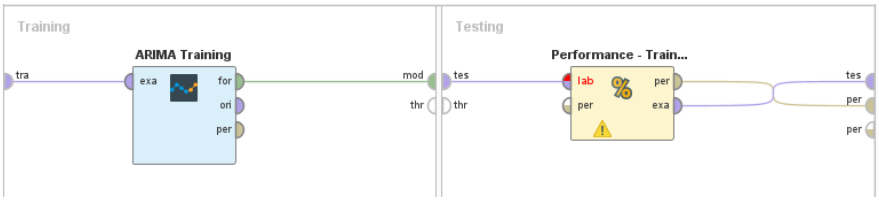
4.4.1 ARIMA

Percobaan pada algoritma pertama dilakukan pada algoritma ARIMA. Percobaan ini dilakukan sebanyak tiga kali, dimana masing-masing dari percobaan tersebut menggunakan nilai p,d, dan q dengan nilai (1,0,1), (1,0,0), dan (0,0,1). Percobaan dilakukan menggunakan aplikasi RapidMiner versi 9, menggunakan skema desain model seperti berikut ini :



Gambar 4.2 Desain Model ARIMA

1. Pada tahap pertama dilakukan penarikan data dari repository dengan nama "*worldwide-aggregate*".
2. Tahap selanjutnya adalah melakukan splitting dataset menjadi 2 bagian, yaitu Training dan Testing yang diporsi sebanyak 80:20.
3. Selanjutnya adalah memproses data yang sudah dibagi tersebut menggunakan algoritma ARIMA.
4. Lalu dilakukan validation terhadap data training dan diuji menggunakan algoritma ARIMA sebelum dilakukan pengukuran performa.



Gambar 4.3 Validation Data Training pada algoritma ARIMA

5. Kemudian dilakukan pengukuran performa dari data, menggunakan parameter evaluasi RMSE dan Absolute Error.

Hasil dari percobaan menggunakan skema tersebut adalah sebagai berikut :

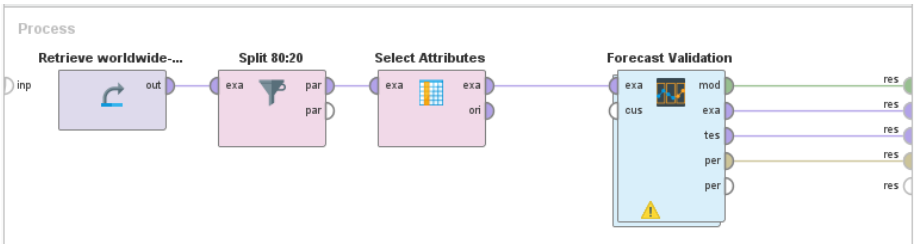
Tabel 4.1 Hasil Percobaan Menggunakan ARIMA

RMSE	1.719.164,787
Absolute Error	1.719.164,787 +/- 12.254.096,763

4.4.2 Exponential Smoothing

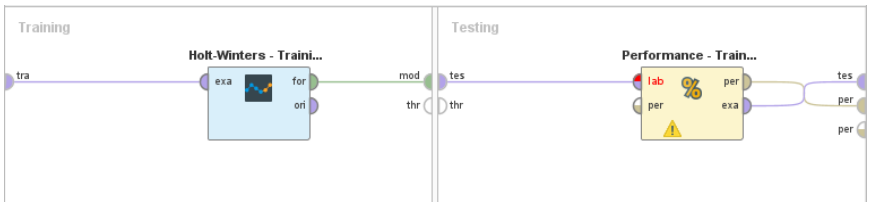
Percobaan pada algoritma kedua dilakukan pada algoritma Triple Exponential Smoothing. Koefisien alpha, beta dan gamma yang digunakan secara berurutan adalah (0,2; 0,1; 0,5).

Percobaan dilakukan menggunakan aplikasi RapidMiner versi 9, menggunakan skema desain model seperti berikut ini :



Gambar 4.4 Desain Model Exponential Smoothing

1. Pada tahap pertama dilakukan penarikan data dari repository dengan nama "*worldwide-aggregate"*.
2. Tahap selanjutnya adalah melakukan splitting dataset menjadi 2 bagian, yaitu Training dan Testing yang diporsi sebanyak 80:20.
3. Selanjutnya adalah memproses data yang sudah dibagi tersebut menggunakan algoritma Exponential Smoothing.
4. Lalu dilakukan validation terhadap data training dan diuji menggunakan algoritma Exponential Smoothing sebelum dilakukan pengukuran performa.



Gambar 4.5 Validation Data Training pada algoritma Exponential Smoothing

5. Kemudian dilakukan pengukuran performa dari kedua data, menggunakan parameter evaluasi RMSE dan Absolute Error.

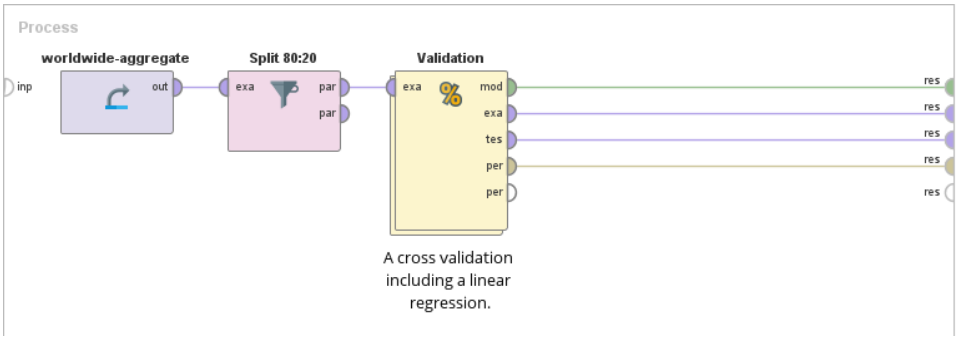
Hasil dari percobaan menggunakan skema tersebut adalah sebagai berikut :

Tabel 4.2 Hasil Percobaan Menggunakan Exponential Smoothing

Training RMSE	6.667.288,427
Training Absolute Error	6.667.288,427 +/- 18.308.027,257

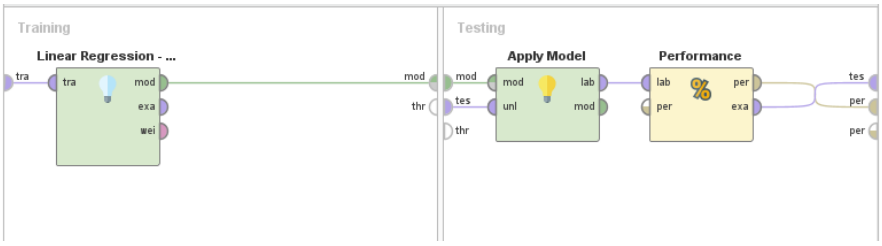
4.4.3 Linear Regression

Percobaan pada algoritma ketiga dilakukan pada algoritma Linear Regression. Percobaan dilakukan menggunakan aplikasi RapidMiner versi 9, menggunakan skema desain model seperti berikut ini :



Gambar 4.6 Desain Model Linear Regression

1. Pada tahap pertama dilakukan penarikan data dari repository dengan nama "*worldwide-aggregate"*.
2. Tahap selanjutnya adalah melakukan splitting dataset menjadi 2 bagian, yaitu Training dan Testing yang diporsi sebanyak 80:20.
3. Selanjutnya adalah memproses data yang sudah dibagi tersebut menggunakan algoritma Linear Regression.
4. Lalu dilakukan validation terhadap data training dan diuji menggunakan algoritma Linear Regression sebelum dilakukan pengukuran performa.



Gambar 4.7 Validation Data Training pada algoritma Linear Regression

5. Kemudian dilakukan pengukuran performa dari kedua data, menggunakan parameter evaluasi RMSE dan Absolute Error.

Hasil dari percobaan menggunakan skema tersebut adalah sebagai berikut :

Tabel 4.3 Hasil Percobaan Menggunakan Linear Regression

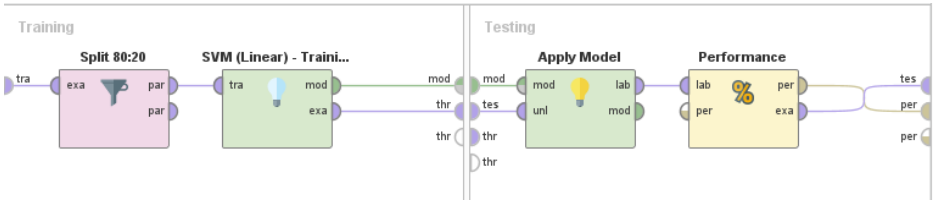
Training RMSE	5.251.387,456
Training Absolute Error	4.250.956,550 +/- 3.083.251,307

4.4.4 Support Vector Regression

Percobaan pada algoritma keempat dilakukan pada algoritma Support Vector Regression. Percobaan dilakukan menggunakan aplikasi RapidMiner versi 9, menggunakan skema desain model seperti berikut ini :

Gambar 4.8 Desain Model Support Vector Regression

1. Pada tahap pertama dilakukan penarikan data dari repository dengan nama "*worldwide-aggregate*".
2. Tahap selanjutnya adalah melakukan splitting dataset menjadi 2 bagian, yaitu Training dan Testing yang diporsi sebanyak 80:20.
3. Selanjutnya adalah memproses data yang sudah dibagi tersebut menggunakan algoritma Support Vector Regression.
4. Lalu dilakukan validation terhadap data training dan diuji menggunakan algoritma Support Vector Regression sebelum dilakukan pengukuran performa.



Gambar 4.9 Validation Data Training pada algoritma Support Vector Regression

5. Kemudian dilakukan pengukuran performa dari kedua data, menggunakan parameter evaluasi RMSE dan Absolute Error.

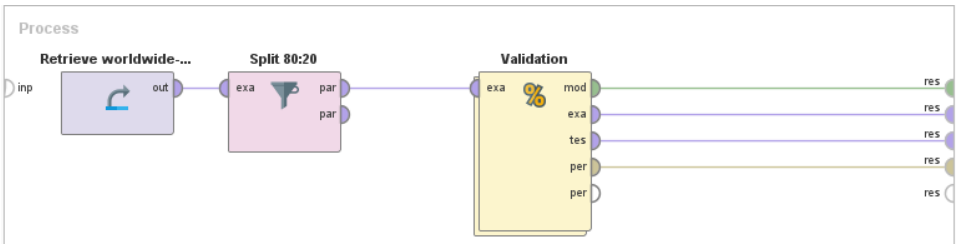
Hasil dari percobaan menggunakan skema tersebut adalah sebagai berikut :

Tabel 4.4 Hasil Percobaan Menggunakan Support Vector Regression

RMSE	80.917.464,754
Absolute Error	80.335.182,149 +/- 42.855.538,242

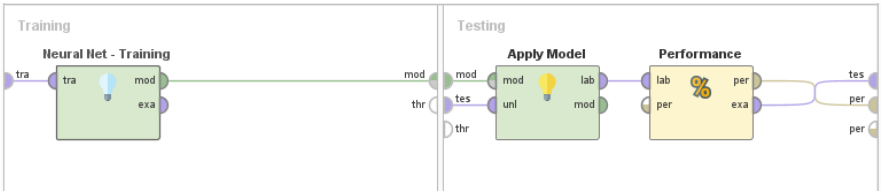
4.4.5 LSTM

Percobaan pada algoritma kelima dilakukan pada algoritma LSTM (Neural Net). Percobaan dilakukan menggunakan aplikasi RapidMiner versi 9, menggunakan skema desain model seperti berikut ini :



Gambar 4.10 Desain Model LSTM (Neural Net)

1. Pada tahap pertama dilakukan penarikan data dari repository dengan nama "*worldwide-aggregate*".
2. Tahap selanjutnya adalah melakukan splitting dataset menjadi 2 bagian, yaitu Training dan Testing yang diporsi sebanyak 80:20.
3. Selanjutnya adalah memproses data yang sudah dibagi tersebut menggunakan algoritma LSTM Neural Net.
4. Lalu dilakukan validation terhadap data training dan diuji menggunakan algoritma LSTM sebelum dilakukan pengukuran performa.



Gambar 4.11 Validation Data Training pada algoritma LSTM

5. Kemudian dilakukan pengukuran performa dari kedua data, menggunakan parameter evaluasi RMSE dan Absolute Error.

Hasil dari percobaan menggunakan skema tersebut adalah sebagai berikut :

Tabel 4.5 Hasil Percobaan Menggunakan LSTM

RMSE	1.573.495,937
Absolute Error	1.569.215,161 +/- 1.252.593,676

4.5 Evaluation

Pada tahap ini dilakukan evaluasi dari hasil pengolahan data testing dan training pada langkah sebelumnya. Evaluasi dibagi menjadi dua bagian yaitu Data Testing dan Data Training. Kemudian dari kedua bagian itu akan dilakukan evaluasi menggunakan parameter evaluasi yang diterapkan pada percobaan pada tahap sebelumnya. Berdasarkan hasil percobaan pada lima algoritma yang sudah ditentukan sebelumnya, didapatkan hasil sebagai berikut :

4.5.1 Data Training

Pada bagian ini merupakan evaluasi Data Training yang berjumlah 522 row. Evaluasi dilakukan menggunakan parameter RMSE (Root Mean Square Error) untuk mengetahui akar dari tingkat kesalahan yang diukur dari selisih antara data prediksi dengan data aktual di masa lalu. Berikut adalah perngkat dari hasil percobaan yang sudah dilakukan menggunakan Rapidminer menggunakan kelima algoritma :

Tabel 4.6 Peringkat Hasil Percobaan dari Semua Algoritma pada Training menggunakan RMSE

No.	Algoritma	Nilai
1	LSTM	1.573.495,937
2	ARIMA (1,0,1)	1.719.164,787
3	Linear Regression	5.251.387,456
4	Exponential Smoothing	6.667.288,427
5	SVR	80.917.464,754

Selanjutnya pada bagian ini merupakan evaluasi Data Training yang berjumlah 522 row. Evaluasi dilakukan menggunakan parameter Absolute Error untuk mengetahui tingkat kesalahan yang diukur dari selisih antara data prediksi dengan data aktual di masa lalu. Berikut adalah perngkat dari hasil percobaan yang sudah dilakukan menggunakan Rapidminer menggunakan kelima algoritma :

Tabel 4.7 Peringkat Hasil Percobaan dari Semua Algoritma pada Training menggunakan Absolute Error

No.	Algoritma	Nilai
1	LSTM	1.569.215,161 +/- 1.252.593,676
2	ARIMA (1,0,1)	1.719.164,787 +/- 12.254.096,763
3	Linear Regression	4.250.956,550 +/- 3.083.251,307
4	Exponential Smoothing	6.667.288,427 +/- 18.308.027,257
5	SVR	80.335.182,149 +/- 42.855.538,242

4.6 Deployment

Berdasarkan hasil dari evaluasi pada tahap sebelumnya, didapatkan bahwa algoritma LSTM (Long Short-Term Memory) memiliki performa yang terbaik dibandingkan dengan algoritma lainnya. Hal ini ditunjukkan dengan adanya hasil dari percobaan dimana algoritma LSTM memiliki tingkat kesalahan terkecil dibandingkan dengan algoritma lainnya.

Pada subbab ini akan dilakukan pengujian untuk Data Testing menggunakan algoritma dengan akurasi terbaik yang sudah diuji pada tahap sebelumnya menggunakan Data Training dengan validasi. Sehingga pada tahap terakhir ini akan dilakukan pengujian Data Testing menggunakan algoritma LSTM.

Pada pengujian kali ini tidak dilakukan menggunakan aplikasi RapidMiner, akan tetapi dilakukan menggunakan aplikasi yang dikembangkan secara mandiri. Pengujian dilakukan menggunakan aplikasi yang dikembangkan menggunakan Bahasa Python versi 3.9.5 menggunakan *code editor* Visual Studio Code. Environment yang akan digunakan untuk menjalankan aplikasi akan mengunakan virtual environment dari python menggunakan perintah pipenv shell. Kemudian akan menggunakan beberapa paket seperti :

1. Flask
   Merupakan framework yang digunakan untuk menjalankan aplikasi Python dalam bentuk *web based* yang bekerja secara terintegrasi dengan Gunicorn yang akan berfungsi sebagai server HTTP.
2. Gunicorn

Gunicorn merupakan paket yang berfungsi seperti Apache dan WAMP pada PHP, yaitu sebagai server HTTP untuk aplikasi Python sehingga dapat ditampilkan melalui *web browser*.

3. Pandas
   Pandas digunakan untuk membaca data, dalam hal ini file .csv dari *storage*. Pada aplikasi ini pengambilan data dilakukan dengan membaca file, bukan dari database.
4. SKLearn
   SKLearn (*Sci Kit Learn*) merupakan paket untuk menjalankan fungsi prediksi/*forecast* menggunakan Python. Paket ini akan bekerja dan dikombinasikan dengan paket lain seperti Keras dan Tensorflow untuk *forecasting* menggunakan Long-Short Term Memory yang akan ditampilkan dalam bentuk grafik garis.
5. Keras
   Keras berfungsi sebagai antarmuka untuk *library* paket Tensorflow. *Library* yang disediakan oleh Keras adalah fungsi-fungsi yang umumnya berhubungan dengan *Neural Network*.
6. Tensorflow
   Tensorflow merupakan paket yang didukung oleh Keras. Tensorflow berisi fungsi-fungsi untuk melakukan pembelajaran mesin. Namun Tensorflow lebih berfokus pada *training* dan jaringan syaraf.
7. Matplotlib
   Matplotlib merupakan paket yang berisi fungsi-fungsi yang berhubungan dengan matematika dan fungsi untuk memproyeksikan plot ke dalam grafik.

4.6.1 Data Splitting

Pada tahap sebelumnya telah dilakukan pengujian menggunakan dataset Training dengan porsi 80% dari keseluruhan data atau sebanyak 522 row. Sehingga pada tahap ini dilakukan pengujian pada dataset Testing yang berjumlah 20% dari keseluruhan data. Jumlah row yang digunakan pada pengujian ini adalah sebanyak 130 row.

Mula-mula aplikasi akan meminta input berupa tanggal dan persentase data training dan pada percobaan di tahap ini akan digunakan keseluruhan dataset dengan menginput tanggal dari awal (22 Januari 2020) sampai akhir dataset (4 November 2021). Adapun untuk nilai persentase data training yang diinput adalah sebanyak 80 persen.

Selanjutnya, hasil akhir yang ditampilkan aplikasi akan dievaluasi menggunakan parameter evaluasi. Parameter evaluasi yang digunakan pada percobaan ini adalah RMSE (Root Mean Square Error).

4.6.2 Model Config

Pada tahap ini dilakukan konfigurasi dari model yang digunakan. Sebelum menentukan konfigurasi, telah dilakukan *trial and error* pada masing-masing parameter untuk menemukan tingkat kesalahan terendah. Setelah didapatkan tingkat kesalahan terendah, maka ditetapkan konfigurasi yang digunakan sebagai berikut :

1. Jumlah Neuron
   LSTM merupakan salah satu variasi dari RNN (*Recurrent Neural Network*) dimana LSTM menggunakan bantuan neuron untuk melakukan prediksi terhadap suatu nilai tertentu. Untuk konfigurasi jumlah neuron pada percobaan ini akan diatur menjadi 50 neuron.
2. *Batch Size*
   Batch Size merupakan jumlah dari sampel yang diproses sebelum dilakukan pembaharuan model. Nilai dari *Batch Size* harus sama dengan atau lebih kecil dari jumlah sampel yang dimiliki dalam dataset. Untuk konfigurasi *batch size* yang digunakan adalah 15.
3. *Epoch*
   Epoch merupakan banyaknya pass yang dilakukan terhadap keseluruhan dataset dalam neural network. Untuk konfigurasi *epoch* yang digunakan adalah 1.

4.6.3 Hasil Pemrosesan dengan Aplikasi

Tahap selanjutnya adalah memproses dataset menggunakan aplikasi prototipe. Sebelum memasukkan konfigurasi input, telah dilakukan *trial and error* terlebih dahulu menggunakan berbagai macam nilai pada parameter input.

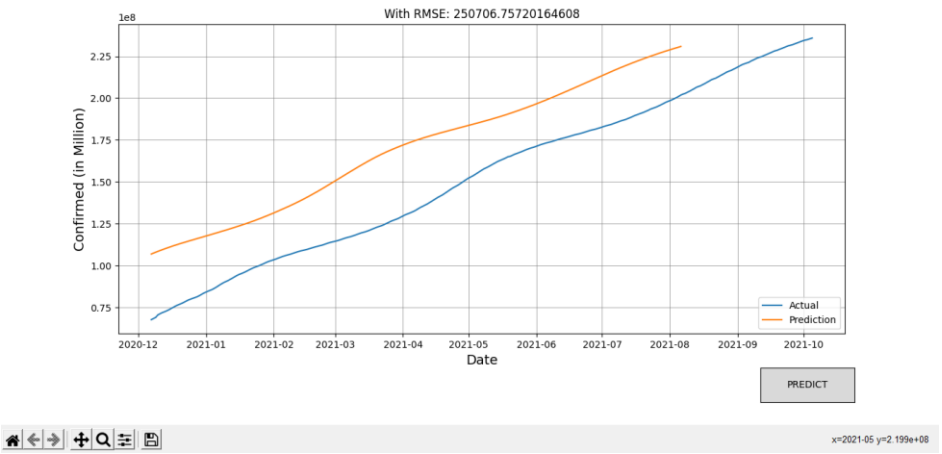
Setelah didapatkan tingkat kesalahan terendah, maka ditetapkanlah nilai parameter yang dijadikan input. Adapun konfigurasi input yang dimasukkan pada saat eksekusi adalah sebagai berikut :

1. **Rentang Waktu** : 05 Februari 2020 – 05 Oktober 2021
2. **Persentase Training** : 80%

Dari konfigurasi berikut, setelah dilakukan eksekusi, maka didapatkan hasil sebagai berikut :

a. Kasus Terkonfirmasi Positif

Pada variabel Terkonfirmasi Positif akan dilakukan *forecasting* sampai 30 hari ke depan. Proses *forecasting* data dibagi menjadi beberapa tahap, setiap tahapnya akan menghasilkan *forecasting* selama 5 hari ke depan. Setiap kali hasil *forecasting* baru muncul, maka akan menghasilkan nilai RMSE yang berbeda dan umumnya nilai RMSE akan turun untuk setiap tahap *forecast* yang dilakukan.



Gambar 4.12 Hasil Pengolahan dengan Jumlah Training 80% (Variabel Konfirmasi Positif)

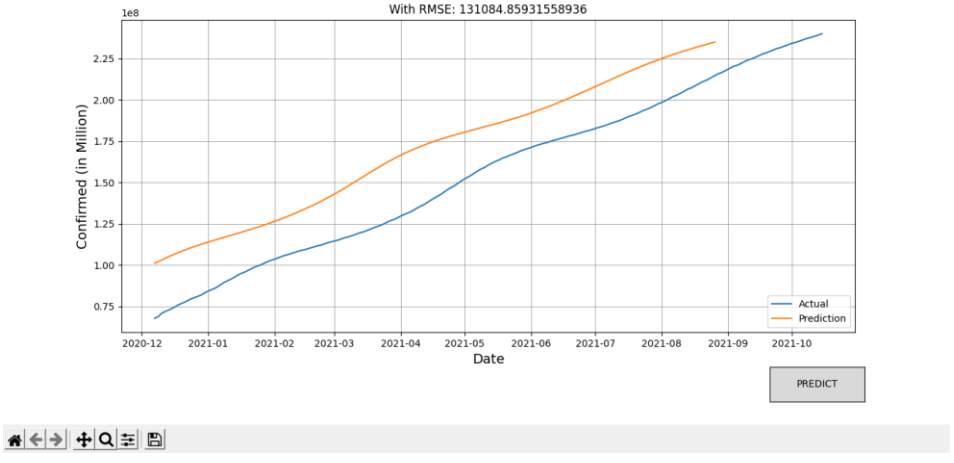
Hasil yang didapatkan dari pengolahan adalah prediksi untuk kasus terkonfirmasi positif Covid-19 dari seluruh dunia memiliki akar tingkat kesalahan sebanyak 250.706,7572.

Hasil tersebut adalah berdasarkan prediksi dari kasus yang sudah ada menggunakan aplikasi. Dan pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan untuk setiap tahap prediksi.

Gambar 4.13 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Pertama)

Gambar 4.13 adalah hasil pengolahan prediksi pertama, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap pertama. Nilai RMSE yang dihasilkan pada proses kedua/prediksi pertama ini adalah sebanyak 190.526,2055.

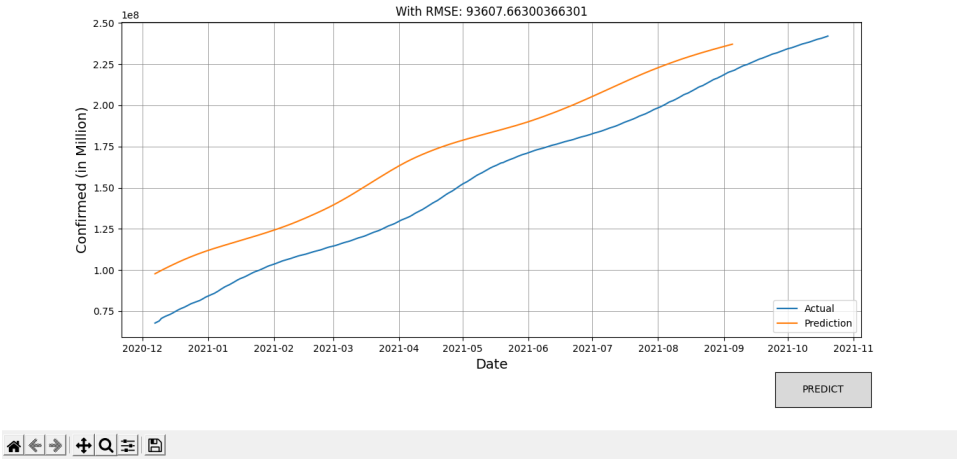
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi pertama atau 10 hari ke depan dari proses pengolahan pertama.



Gambar 4.14 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Kedua)

Gambar 4.14 adalah hasil pengolahan prediksi kedua, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap pertama. Nilai RMSE yang dihasilkan pada proses ketiga/prediksi kedua ini adalah sebanyak 131.064,8593.

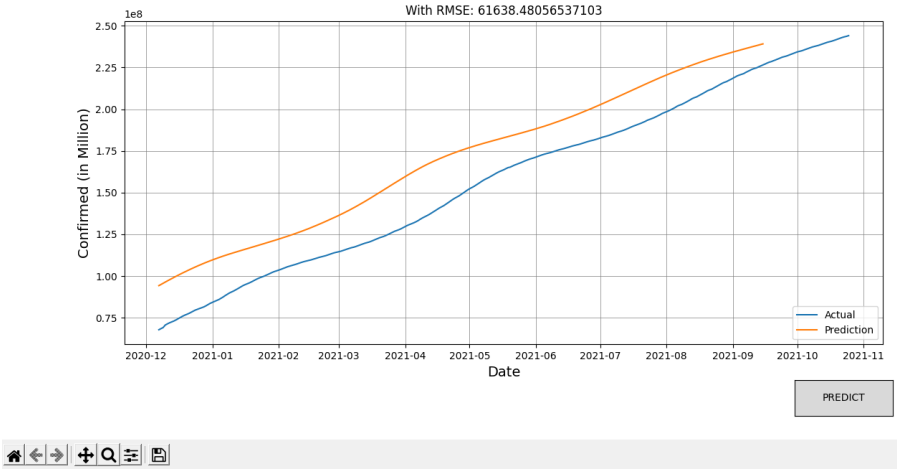
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi kedua atau 15 hari ke depan dari proses pengolahan pertama.



Gambar 4.15 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Ketiga)

Gambar 4.15 adalah hasil pengolahan prediksi ketiga, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap kedua. Nilai RMSE yang dihasilkan pada proses keempat/prediksi ketiga ini adalah sebanyak 93.607,6630.

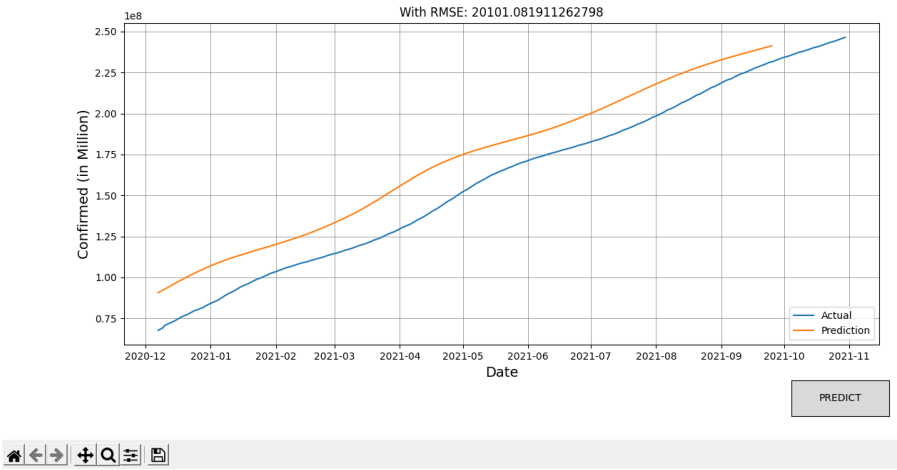
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi ketiga atau 20 hari ke depan dari proses pengolahan pertama.



Gambar 4.16 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Keempat)

Gambar 4.16 adalah hasil pengolahan prediksi keempat, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap ketiga. Nilai RMSE yang dihasilkan pada proses kelima/prediksi keempat ini adalah sebanyak 61.638,4805.

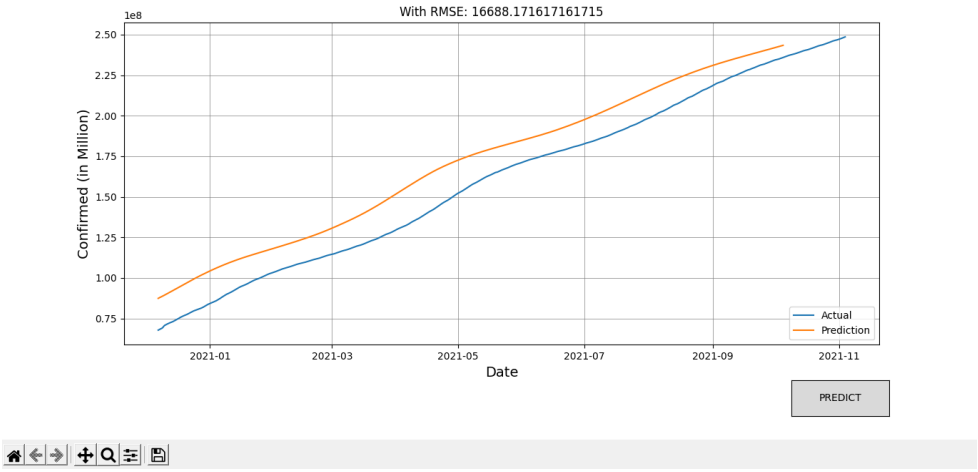
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi keempat atau 25 hari ke depan dari proses pengolahan pertama.



Gambar 4.17 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Kelima)

Gambar 4.17 adalah hasil pengolahan prediksi kelima, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap keempat. Nilai RMSE yang dihasilkan pada proses keenam/prediksi kelima ini adalah sebanyak 20.101,0819.

Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi kelima atau 30 hari ke depan dari proses pengolahan pertama.



Gambar 4.18 Hasil Pengolahan dengan Jumlah Training 80% (Prediksi Keenam)

Gambar 4.18 adalah hasil pengolahan prediksi keenam, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap kelima. Nilai RMSE yang dihasilkan pada proses ketujuh/prediksi keenam ini adalah sebanyak 16.688,1716.

Tahap ini merupakan tahap terakhir dimana aplikasi hanya dapat memprediksi sampai pada 30 hari ke depan untuk variabel Konfirmasi Positif. Namun banyaknya jumlah hari yang dapat diprediksi ke depan dipengaruhi oleh jumlah row yang menjadi data aktual pada proses *forecasting* ini, yaitu data penderita yang direkam pada dataset dari bulan Januari 2020 sampai bulan November 2021.

b. Kasus Sembuh

Pada variabel Sembuh akan dilakukan *forecasting* sampai 30 hari ke depan. Proses *forecasting* data dibagi menjadi beberapa tahap, setiap tahapnya akan menghasilkan *forecasting* selama 5 hari ke depan. Setiap kali hasil *forecasting* baru muncul, maka akan menghasilkan nilai RMSE

yang berbeda dan umumnya nilai RMSE akan turun untuk setiap tahap *forecast* yang dilakukan.

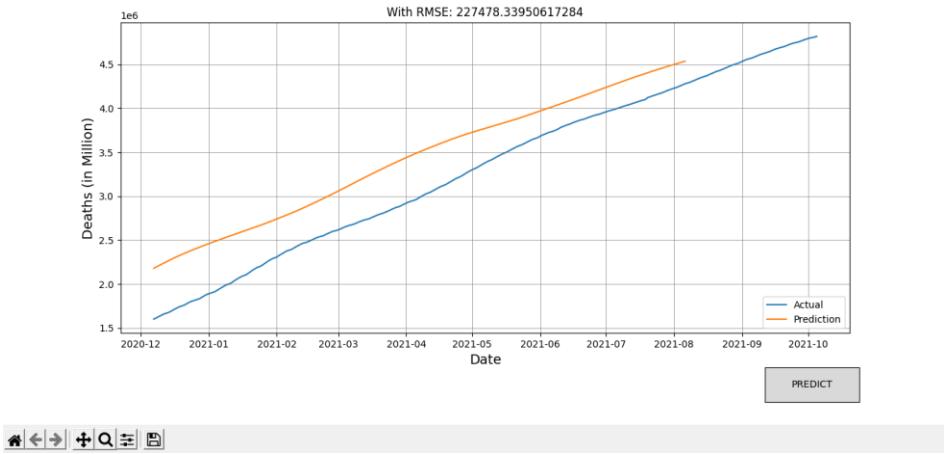
Gambar 4.19 Hasil Pengolahan dengan Jumlah Training 80% (Variabel Sembuh)

Gambar 4.19 adalah hasil pengolahan dari variabel Sembuh, dimana menghasilkan tingkat kesalahan dengan nilai RMSE sebesar 3.722,9814.

Tahap ini tidak dapat dilakukan prediksi karena berdasarkan dataset yang didapat, nilai kasus sembuh sejak bulan Agustus ke depan adalah 0. Sehingga setiap prediksi yang dilakukan selama beberapa hari ke depan akan selalu menghasilkan nilai 0.

c. Kasus Meninggal

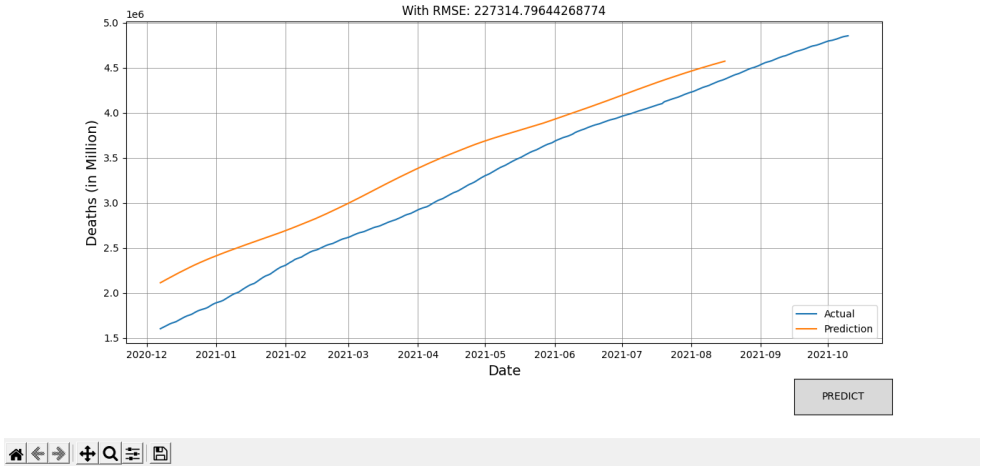
Pada variabel Meninggal akan dilakukan *forecasting* sampai 30 hari ke depan. Proses *forecasting* data dibagi menjadi beberapa tahap, setiap tahapnya akan menghasilkan *forecasting* selama 5 hari ke depan. Setiap kali hasil *forecasting* baru muncul, maka akan menghasilkan nilai RMSE yang berbeda dan umumnya nilai RMSE akan turun untuk setiap tahap *forecast* yang dilakukan.



Gambar 4.20 Hasil Pengolahan dengan Jumlah Training 80% (Variabel Kematian)

Hasil yang didapatkan dari pengolahan adalah prediksi untuk kasus kematian akibat positif Covid-19 dari seluruh dunia memiliki tingkat kesalahan yang direpresentasikan oleh RMSE sebesar 227.478,3395.

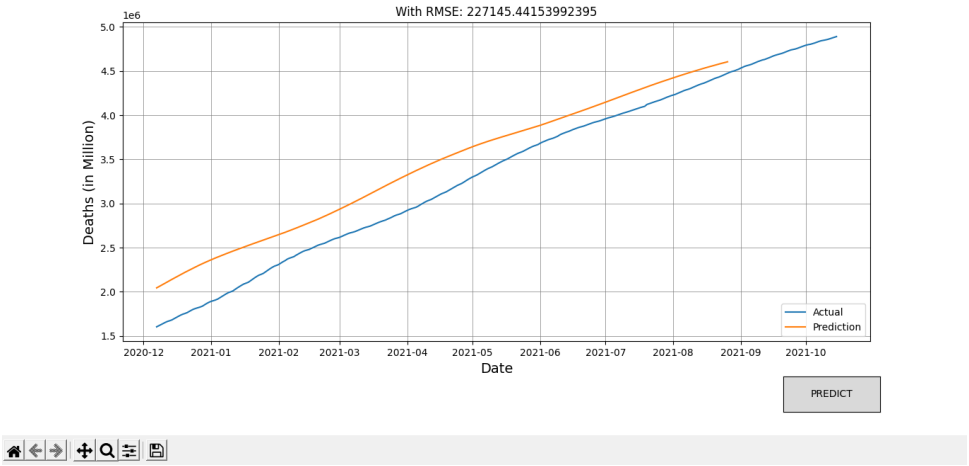
Hasil tersebut adalah berdasarkan prediksi dari kasus yang sudah ada menggunakan aplikasi. Dan pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan untuk setiap tahap prediksi.



Gambar 4.21 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Pertama)

Gambar 4.21 adalah hasil pengolahan prediksi pertama, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada tahap pertama. Nilai RMSE yang dihasilkan pada proses prediksi pertama ini adalah sebanyak 227.314,7964.

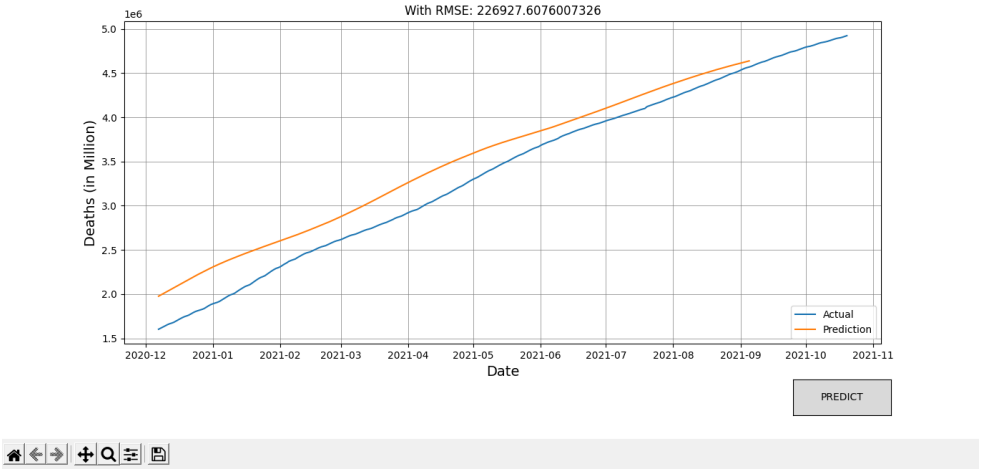
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi pertama atau 10 hari ke depan dari proses pengolahan pertama.



Gambar 4.22 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Kedua)

Gambar 4.22 adalah hasil pengolahan prediksi kedua, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada prediksi pertama. Nilai RMSE yang dihasilkan pada proses kedua/prediksi pertama ini adalah sebanyak 227.145,4415.

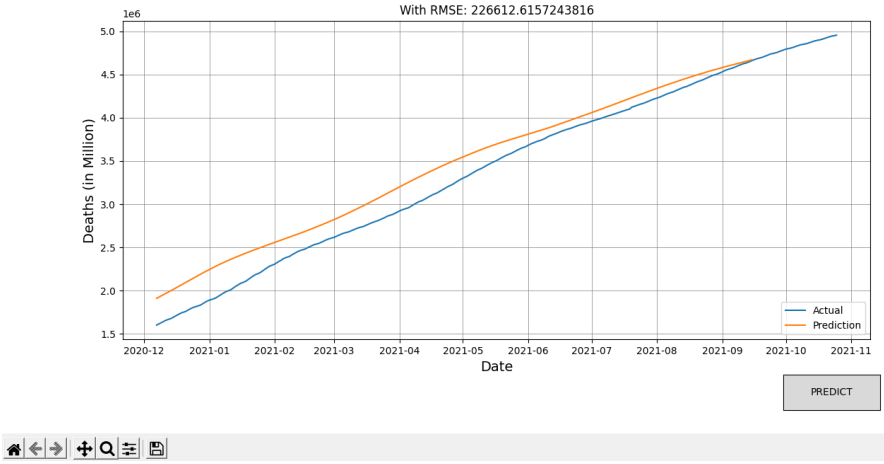
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi kedua atau 15 hari ke depan dari proses pengolahan pertama.



Gambar 4.23 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Ketiga)

Gambar 4.23 adalah hasil pengolahan prediksi ketiga, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada prediksi kedua. Nilai RMSE yang dihasilkan pada proses ketiga/prediksi kedua ini adalah sebanyak 226.927,6076.

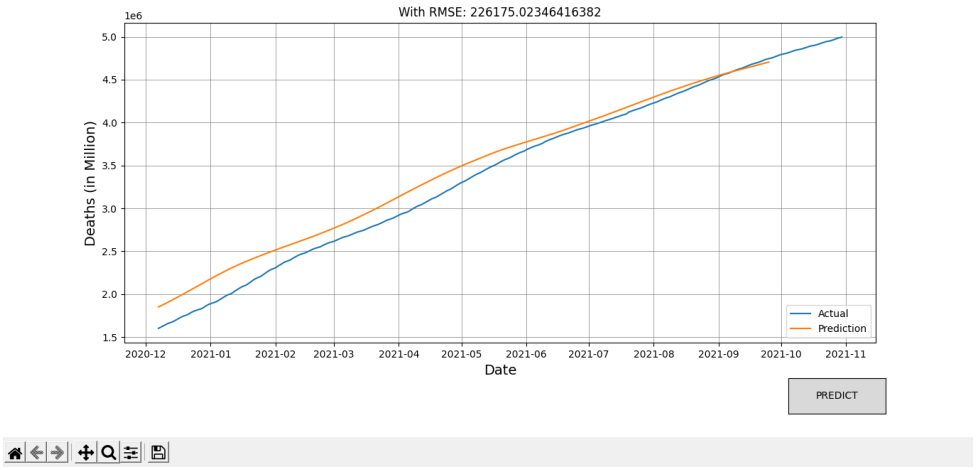
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi ketiga atau 20 hari ke depan dari proses pengolahan pertama.



Gambar 4.24 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Keempat)

Gambar 4.24 adalah hasil pengolahan prediksi keempat, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada prediksi ketiga. Nilai RMSE yang dihasilkan pada proses keempat/prediksi ketiga ini adalah sebanyak 226.612,6157.

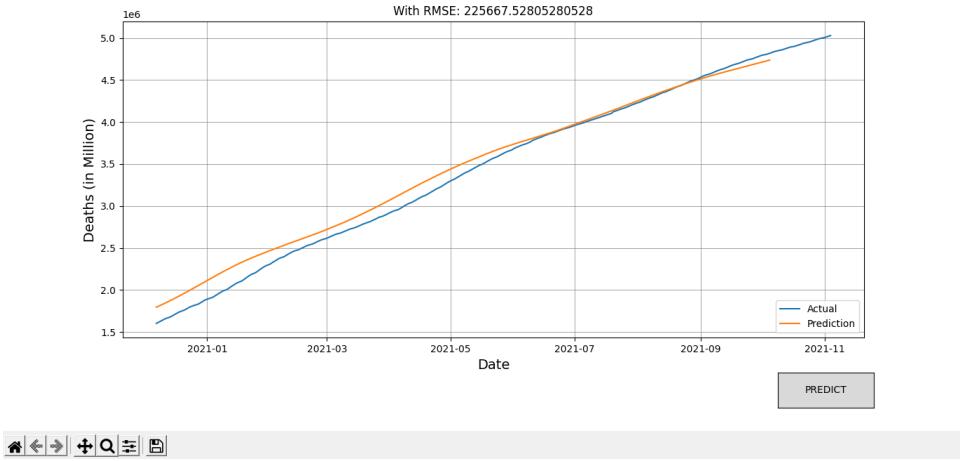
Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi keempat atau 25 hari ke depan dari proses pengolahan pertama.



Gambar 4.25 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Kelima)

Gambar 4.25 adalah hasil pengolahan prediksi kelima, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada prediksi keempat. Nilai RMSE yang dihasilkan pada proses kelima/prediksi keempat ini adalah sebanyak 226.175,0234.

Pada tahap berikutnya, aplikasi akan memprediksi kasus yang akan datang selama 5 hari ke depan dari prediksi kelima atau 30 hari ke depan dari proses pengolahan pertama.



Gambar 4.26 Hasil Pengolahan untuk Variabel Kematian dengan Jumlah Training 80% (Prediksi Keenam)

Gambar 4.26 adalah hasil pengolahan prediksi keenam, dimana menghasilkan tingkat kesalahan yang lebih kecil dibandingkan proses pengolahan pada prediksi kelima. Nilai RMSE yang dihasilkan pada proses keenam/prediksi kelima ini adalah sebanyak 225.667,5280.

Tahap ini merupakan tahap terakhir dimana aplikasi hanya dapat memprediksi sampai pada 35 hari ke depan untuk variabel Kematian. Namun banyaknya jumlah hari yang dapat diprediksi ke depan dipengaruhi oleh jumlah row yang menjadi data aktual pada proses *forecasting* ini, yaitu data penderita yang direkam pada dataset dari bulan Januari 2020 sampai bulan November 2021.

Dari seluruh pengolahan yang dilakukan didapatkan hasil sebagai berikut :

Tabel 4.8 Hasil Pengolahan Data Menggunakan Aplikasi Prototipe

Variabel Terkonfirmasi Positif		
1.	Pengolahan Pertama	250.706,7572
2.	Prediksi Pertama (+5 hari)	190.526,2055
3.	Prediksi Kedua (+10 hari)	131.064,8593
4.	Prediksi Ketiga (+15 hari)	93.607,6630
5.	Prediksi Keempat (+20 hari)	61.638,4805
6.	Prediksi Kelima (+25 hari)	20.101,0819

7.	Prediksi Keenam (+30 hari)	16.688,1716
Variabel Sembuh		
1.	Pengolahan Pertama	3.722,9814
Variabel Meninggal		
1.	Pengolahan Pertama	227.478,3395
2.	Prediksi Pertama (+5 hari)	227.314,7964
3.	Prediksi Kedua (+10 hari)	227.145,4415
4.	Prediksi Ketiga (+15 hari)	226.927,6076
5.	Prediksi Keempat (+20 hari)	226.612,6157
6.	Prediksi Kelima (+25 hari)	226.175,0234
7.	Prediksi Keenam (+30 hari)	225.667,5280

Berdasarkan hasil pemrosesan yang sudah dilakukan, hasil prediksi menunjukkan bahwa untuk setiap variabel yang diprediksi akan terus mengalami kenaikan. Hal ini dikarenakan tren yang berlaku pada setiap variabel adalah naik, dimana jumlah nilai pada setiap rownya dalam dataset mayoritas mengalami kenaikan nilai. Sehingga pada pemrosesan menggunakan algoritma LSTM, hasil prediksi yang dihasilkan adalah untuk beberapa minggu ke depan, setiap nilai dari masing-masing variabel masih akan terus mengalami kenaikan.

Hasil prediksi pertama untuk variabel terkonfirmasi positif menunjukkan kenaikan menjadi 230 juta kasus dengan nilai RMSE 190.526,2055, pada prediksi kedua naik lagi menjadi 235 juta kasus dengan nilai RMSE 131.084,8593, kemudian pada prediksi ketiga naik menjadi 238 juta kasus dengan nilai RMSE 93.607,6630, lalu keempat naik menjadi 240 juta kasus dengan RMSE 61.638,4805, kelima naik menjadi 243 juta kasus dengan RMSE 20.101,0819, keenam naik menjadi 248 juta kasus dengan RMSE 16.688,1716.

Untuk variabel sembuh, tidak dapat dilakukan prediksi karena data sejak bulan Agustus 2021 ke depan yang didapat dari repositori menunjukkan nilai 0 sehingga nilai prediksi yang dihasilkan akan selalu menunjukkan nilai 0.

Terakhir adalah variabel meninggal, hasil prediksi pertama menunjukkan kenaikan menjadi 4,8 juta kasus dengan nilai RMSE 227.314,7964, pada prediksi kedua menunjukkan kenaikan menjadi 4,85 juta kasus dengan nilai RMSE 227.145,4415, lalu pada prediksi ketiga menunjukkan

kenaikan menjadi 4,91 juta kasus dengan nilai RMSE 226.927,6076, pada prediksi keempat menunjukkan kenaikan menjadi 4,97 juta kasus dengan nilai RMSE 226.612,6157, lalu pada prediksi kelima menunjukkan kenaikan menjadi 5,02 juta kasus dengan RMSE 226.175,0234, dan pada prediksi keenam menunjukkan kenaikan menjadi 5,08 juta kasus dengan RMSE 225.667,5280.

Hasil prediksi juga menunjukkan hasil RMSE yang dihasilkan pada setiap tahapan prediksi mengalami penurunan yang berarti tingkat kesalahan pada setiap tahap prediksi menjadi semakin kecil. Penurunan nilai RMSE ini berlaku pada semua variabel yang diprediksi.

Akan tetapi, hasil prediksi yang dilakukan tersebut adalah pengolahan yang dilakukan hanya dengan mengandalkan nilai variabel masa lalu saja sebagai nilai input. Hasil prediksi yang akan didapat apabila terdapat variabel lain yang mungkin akan mempengaruhi nilai prediksi bisa saja berubah dan tidak menutup kemungkinan hasil prediksi akan menunjukkan hasil menurun apabila seandainya ada variabel lain yang digunakan sebagai variabel input untuk melakukan prediksi, sebagai contoh misalnya adalah dengan adanya vaksin, program baru khusus dari pemerintah untuk mencegah penyebaran virus Covid-19 dan hal-hal lainnya yang dapat menghambat penyebaran virus Covid-19.

Sehingga hasil akhir yang didapatkan dari proses komparasi dan *forecasting* yang sudah dilakukan adalah bahwa algoritma LSTM merupakan algoritma terbaik untuk melakukan *forecasting* data penderita Covid-19 di seluruh dunia saat ini. Karena berdasarkan hasil perbandingan pada tahap evaluasi menunjukkan bahwa algoritma LSTM menghasilkan prediksi dengan tingkat kesalahan terendah dari algoritma lainnya.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan uraian-uraian yang telah dijabarkan pada bab sebelumnya, maka kesimpulan dari penelitian penulis yang berjudul "Analisis Perbandingan Metode Machine Learning untuk Memprediksi Kasus Covid-19 di Dunia" adalah :

1. Bahwa penelitian sebelumnya yang bertujuan untuk memprediksi data jumlah penderita Covid-19 di berbagai negara menggunakan model yang berbeda-beda menunjukkan hasil dan akurasi yang berbeda antara satu dengan lainnya walaupun menggunakan variabel yang sama. Dengan adanya perbedaan hasil tersebut, perlu dilakukan perbandingan untuk mencari model dengan hasil akurasi tertinggi dan rata-rata error terendah.
2. Hasil perbandingan antara algoritma ARIMA, Linear Regression, Exponential Smoothing, SVR, dan LSTM menunjukkan bahwa algoritma LSTM memiliki tingkat kesalahan terendah dari algoritma lainnya.
3. Hasil prediksi kasus penderita Covid-19 menggunakan data penderita di seluruh menggunakan algoritma LSTM menunjukkan bahwa jumlah penderita akan terus mengalami kenaikan dikarenakan tren yang terjadi di masa lalu menunjukkan kenaikan jumlah penderita.

5.2 Saran

Dari uraian pada bab-bab sebelumnya dan hasil penelitian yang sudah didapat, penulis memberikan saran-saran sebagai berikut :

1. Bahwa dengan adanya rekomendasi model prediksi dengan akurasi tertinggi akan memberikan hasil peramalan yang lebih dapat dipercaya.
2. Hasil prediksi yang dihasilkan dari pengolahan menggunakan aplikasi prototipe hanya dihitung berdasarkan nilai dari variabel Terkonfirmasi Positif, Meninggal, dan Sembuh di masa lalu. Sehingga pada penelitian selanjutnya diharapkan dapat membuat prediksi dengan variabel input yang dapat memberikan pengaruh terhadap hasil akhir prediksi seperti "Vaksin", "Program Pencegahan Penyebaran Virus Covid dari Pemerintah", "Tingkat Kesadaran Masyarakat Akan Pentingnya Mencegah Penyebaran Virus Covid", dan lain sebagainya.
3. Seiring berkembangnya waktu di masa depan, diharapkan ada model-model baru yang dapat dibandingkan dengan model terbaik saat ini dan diharapkan model yang baru dapat memberikan akurasi lebih baik dibandingkan model terbaik saat ini.

DAFTAR PUSTAKA

Al-Qaness, M. A. A., Ewees, A. A., Fan, H., & Aziz, M. A. El. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. *Applied Sciences*, *9*(3). https://doi.org/10.3390/JCM9030674

Arhami, M. (2005). *Konsep Dasar Sistem Pakar*. Andi.

Azevedo, A., & Santos, M. F. (2008). KDD , SEMMA AND CRISP-DM : A PARALLEL OVERVIEW Ana Azevedo and M . F . Santos. *IADIS European Conference Data Mining*, 182–185. http://recipp.ipp.pt/handle/10400.22/136%0Ahttp://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf

Balaji, S. (2012). Waterfall vs v-model vs agile : A comparative study on SDLC. *WATEERFALL Vs V-MODEL Vs AGILE : A COMPARATIVE STUDY ON SDLC*, *2*(1), 26–30.

Brady, M., & Loonam, J. (2010). Exploring the use of entity□relationship diagramming as a technique to support grounded theory inquiry. *Qualitative Research in Organizations and Management: An International Journal*, *5*(3), 224–237. https://doi.org/10.1108/17465641011089854

Budiharto, W. (2014). *Robotika Modern*. Andi.

Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. *IEEE Access*, *8*, 71326–71338. https://doi.org/10.1109/ACCESS.2020.2985763

Chandra, C., & Budi, S. (2020). Analisis Komparatif ARIMA dan Prophet dengan Studi Kasus Dataset Pendaftaran Mahasiswa Baru. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(2), 278–287. https://doi.org/10.28932/jutisi.v6i2.2676

Data mining. (2001). *Pc Ai*, *15*(6), 11. https://doi.org/10.1002/9781118900772.etrds0071

Dorraki, M., Fouladzadeh, A., Salamon, S. J., Allison, A., Coventry, B. J., & Abbott, D. (2019). Can C-Reactive Protein (CRP) Time Series Forecasting be Achieved via Deep Learning? *IEEE Access*, *7*, 59311–59320. https://doi.org/10.1109/ACCESS.2019.2914473

Guleryuz, D. (2021). Forecasting outbreak of COVID-19 in Turkey; Comparison of Box–Jenkins, Brown's exponential smoothing and long short-term memory models. *Process Safety and Environmental Protection*, *149*(December 2019), 927–935. https://doi.org/10.1016/j.psep.2021.03.032

Harini, S. (2020). Identification COVID-19 Cases in Indonesia with The Double Exponential Smoothing Method. *Jurnal Matematika "MANTIK,"* *6*(1), 66–75. https://doi.org/10.15642/mantik.2020.6.1.66-75

He, F., Deng, Y., & Li, W. (2020). Coronavirus disease 2019: What we know? *Journal of Medical Virology*, *92*(7), 719–725. https://doi.org/10.1002/jmv.25766

Hoffer, J. (2012). *Modern Systems Analysis and Design, 6/e* (7th ed.). Pearson Prentice Hall.

Iii, B. a B. (2008). *Bab iii metodologi penelitian*. *i*, 16–28.

Jagait, R. K., Fekri, M. N., Grolinger, K., & Mir, S. (2021). Load Forecasting Under Concept Drift: Online Ensemble Learning with Recurrent Neural Network and ARIMA. *IEEE Access*, *9*, 1–1. https://doi.org/10.1109/access.2021.3095420

Kadek Tutik A., G. A., Delima, R., & Proboyekti, U. (2011). Penerapan Forward Chaining Pada Program Diagnosa Anak Penderita Autisme. *Jurnal Informatika*, *5*(2), 46–60. https://doi.org/10.21460/inf.2009.52.73

Kristianto, R. P., Utami, E., & Lutfi, E. T. (2017). *Penerapan Algoritma Forecasting Untuk Prediksi Penderita Demam Berdarah Dengue Di Kabupaten Sragen*. *October*, 55–60.

Moiseev, G. (2021). Forecasting oil tanker shipping market in crisis periods: Exponential smoothing model application. *Asian Journal of Shipping and Logistics*, *xxxx*. https://doi.org/10.1016/j.ajsl.2021.06.002

Murad, D. F., Kusniawati, N., & Asyanto, A. (2013). Aplikasi Intelligence Website Untuk Penunjang Laporan Paud Pada Himpaudi Kota Tangerang. *CCIT Journal*, *7*(1), 44–58. https://doi.org/10.33050/ccit.v7i1.168

Octavia, Tanti and Yulia, Yulia dan Lydia, L. (2015). Peramalan stok barang untuk membantu pengambilan keputusan pembelian barang pada toko bangunan xyz dengan metode arima. *Seminar Nasional Informatika (SEMNASIF)*, *1*(semnasIF), 2–7.

Perc, M., Gorišek Miksić, N., Slavinec, M., & Stožer, A. (2020). Forecasting COVID-19. *Frontiers in Physics*, *8*(April), 1–5. https://doi.org/10.3389/fphy.2020.00127

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS ONE*, *15*(3), 1–8. https://doi.org/10.1371/journal.pone.0231236

Ruliyanta, R., & Nugroho, E. R. (2020). Forecast of COVID-19 Cases in Indonesia with the Triple Exponential Smoothing Algorithm. *Jurnal Ilmiah Giga*, *23*(2), 61. https://doi.org/10.47313/jig.v23i2.933

Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B. W., Aslam, W., & Choi, G. S. (2020). COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access*, *8*, 101489–101499. https://doi.org/10.1109/ACCESS.2020.2997311

Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, *12*(1), 217–222. http://www.ijisr.issr-

journals.org/

Teshome. (2020). *Forecasting the Number of Coronavirus (COVID-19) Cases in Ethiopia Using Exponential Smoothing Times Series Model Author : TeshomeHailemeskelAbebe Ambo University , Department of Economics , Ambo , Ethiopia Email : teshome251990@gmail.com*. *December 2019*.

Wandah, W. (2017). *Desain dan Program Multimedia Pembelajaran Interaktif* (1st ed.). Penerbit Cerdas Ulet Kreatif.

Yonar, H. (2020). Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. *Eurasian Journal of Medicine and Oncology*, *April*. https://doi.org/10.14744/ejmo.2020.28273

Yu, H., Ming, L. J., Sumei, R., & Shuping, Z. (2020). A Hybrid Model for Financial Time Series Forecasting-Integration of EWT, ARIMA with the Improved ABC Optimized ELM. *IEEE Access*, *8*, 84501–84518. https://doi.org/10.1109/ACCESS.2020.2987547

Zhao, H., Merchant, N. N., McNulty, A., Radcliff, T. A., Cote, M. J., Fischer, R. S. B., Sang, H., & Ory, M. G. (2021). COVID-19: Short term prediction model using daily incidence data. *PLoS ONE*, *16*(4 April), 1–14. https://doi.org/10.1371/journal.pone.0250110

LAMPIRAN

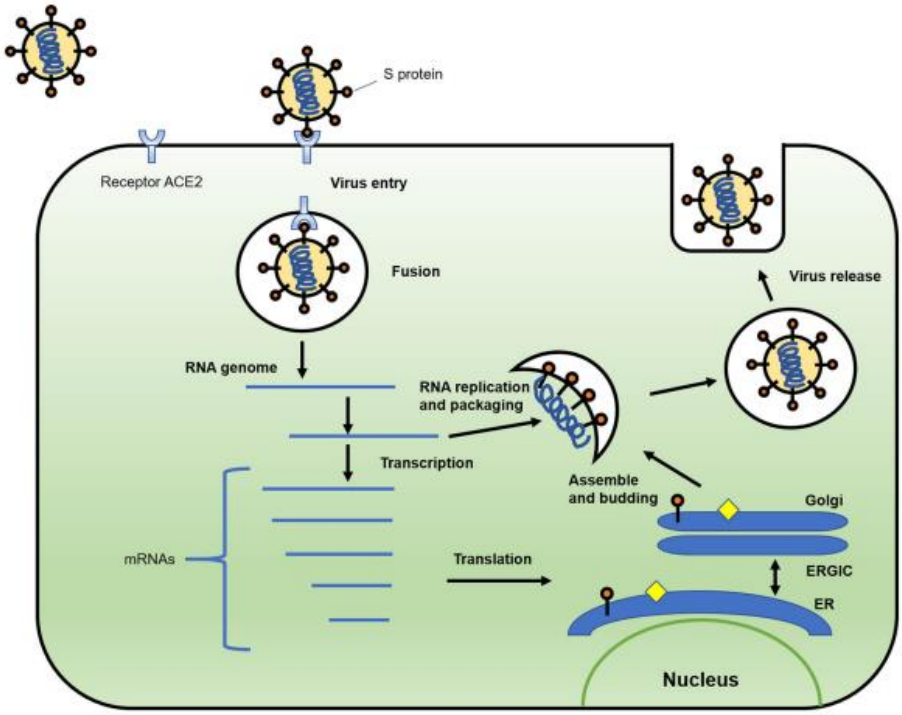


FIGURE 1 Schematic model of SARS-CoV-2 life cycle. S protein binds to the cellular receptor ACE2 to facilitate the entry of the virus. After the fusion of viral and plasma membranes, virus RNA undergoes replication and transcription. The proteins are synthesized. Viral proteins and new RNA genome are subsequently assembled in the ER and Golgi, followed by budding into the lumen of the ERGIC. New virions are released through vesicles. ACE2, angiotensin-converting enzyme 2; ER, endoplasmic reticulum; ERGIC, endoplasmic reticulum-Golgi intermediate compartment

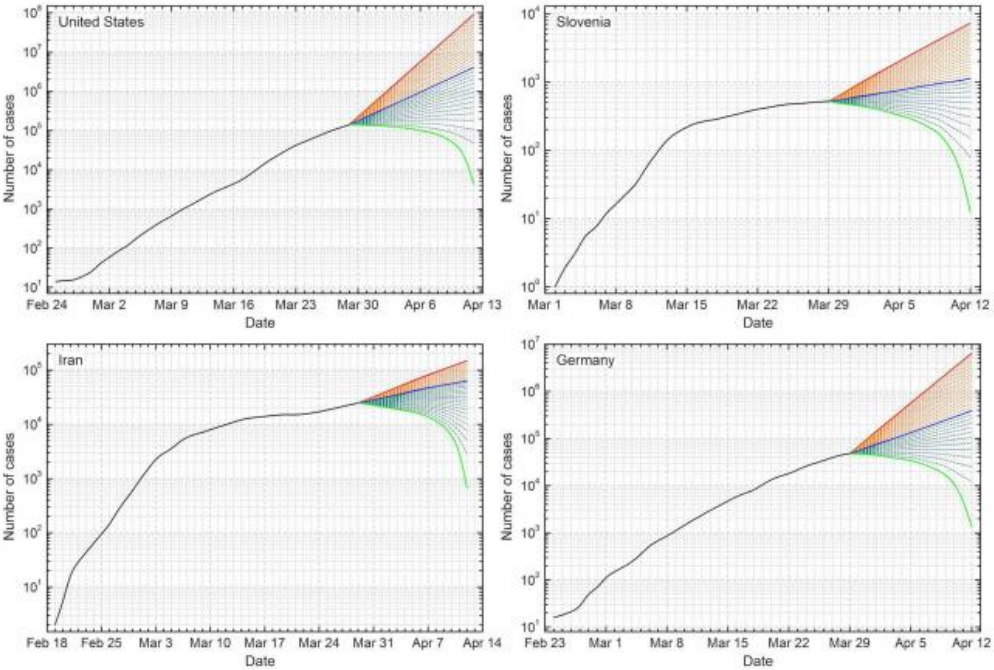


FIGURE 1 | Forecasts of COVID-19 cases for the United States, Slovenia, Iran, and Germany. Black solid line denotes the actual data, which were for this analysis last updated March 29th. From this date onward we thus have the predicted values, where the solid blue line denotes the continuation of the trend of the past 14 days, i.e., if nothing would change. The uppermost solid red line denotes the prediction obtained if the maximal daily growth rate recorded during the past 14 days G_7 would increase by 20%, while the lowermost green line denotes the prediction if the daily growth rate would drop to zero from March 29th onward. Orange and olive dashed lines denote predictions for equally spaced decreasing daily growth rates from top to bottom. Plateaus in the next 14 days would be reached if: United States target daily growth rate ≈ 5.9% (4th line from the bottom), Slovenia target daily growth rate ≈ 3.7% (7th line from the bottom), Iran target daily growth rate ≈ 3.6% (10th line from the bottom), and Germany target daily growth rate ≈ 5.5% (5th line from the bottom).

Table 2. The goodness of fit criteria of the Box-Jenkins and exponential smoothing models

		Model Fit Statistics								Ljung-Box Q (18)		
	Model Type	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	DF	p
Turkey	ARIMA(1,4,0)	0.205	0.995	6.171	2.578	2.176	1692.189	28.823	3.711	12.477	17	0.045
Germany	ARIMA(1,1,0)	0.188	0.826	394.416	6.389	147.556	653.246	1501.333	12.023	10.541	17	0.049
Italy	Holt	0.843	0.892	589.553	7.878	254.685	2719.729	3334.104	12.894	24.690	16	0.075
Japan	Holt	0.821	0.462	16.403	3.682	10.864	1032.755	60.545	5.730	19.645	16	0.037
Canada	Holt	0.841	0.777	29.212	8.001	11.152	692.214	133.395	6.884	60.943	16	0.000
Russia	Brown	0.646	0.908	4.474	6.595	1.908	308.867	17.851	3.064	20.297	17	0.032
United Kingdom	Holt	0.825	0.650	149.566	6.979	60.291	20263.194	805.081	10.150	16.715	16	0.040
France	ARIMA (0,1,3)	0.667	0.837	232.490	1.301	90.070	1148.194	1126.799	11.034	24.276	16	0.038

Gambar di atas adalah hasil dari evaluasi mean error dari model ARIMA dan Exponential Smoothing dengan parameter evaluasi Stationary R-squared, R-squared, RMSE, MAPE, MAE, MaxAPE, MaxAE, dan Normalized BIC.

Variables	Models	RMSE	MAE	MAPE	AIC
The Number of Total Case	ARIMA(1,2,4)	379.539	282.306	26.0422	12.03
	ARIMA(2,1,4)	380.117	276.566	122.887	12.07
	ARIMA(0,2,0)	418.859	299.371	4.63986	12.08
	ARIMA(4,2,0)	394.781	285.83	4.51657	12.08
	ARIMA(1,1,0)	418.21	294.491	4.61463	12.1
	EXPOS λ = 0.9912	415.848	289.797	7.29059	12.09
	LSTM	535.96	489.30	0.35	14.66
The Growth Rate of Total Cases	ARIMA(0,1,2)	0.275737	0.141242	10.0591	−2.51
	ARIMA(0,1,3)	0.344714	0.162714	9.97369	−2.04
	ARIMA(0,1,4)	0.362201	0.157848	9.65822	−1.91
	ARIMA(1,2,4)	0.357786	0.197014	14.8608	−1.9
	ARIMA(0,2,3)	0.420685	0.291096	23.3476	−1.64
	EXPOS λ = 0.1601	0.997909	0.397147	19.5974	0.027
	LSTM	0.00218	0.00161	0.15858	0.486
The Number of New Case	ARIMA(1,1,4)	377.866	278.118	163.11	12.03
	ARIMA(2,0,4)	377.386	272.416	337.864	12.05
	ARIMA(0,1,0)	415.521	294.619	17.479	12.06
	ARIMA(4,1,0)	391.426	281.324	18.6604	12.06
	EXPOS λ = 0.3585	435.701	305.3	240.838	12.19
	LSTM	1047.47	998.91	60.07	147.43
The Number of Total Death	ARIMA(0,2,2)	6.09454	4.70935		3.677
	ARIMA(2,2,0)	6.10109	4.6473		3.679
	ARIMA(1,2,1)	6.13436	4.6938		3.69
	ARIMA(2,0,2)	5.97003	4.33173		3.699
	ARIMA(2,1,1)	6.0851	4.44364		3.705
	EXPOS λ = 0.9999	6.4762	4.90641		3.768
	LSTM	24.179	22.40	0.59	8.132
The Growth Rate of Total Death	ARIMA(0,2,3)	0.114303	0.0506061	4.06235	−4.24
	ARIMA(0,2,4)	0.116274	0.0454201	3.64399	−4.18
	ARIMA(2,2,2)	0.11858	0.0539604	4.29279	−4.14
	ARIMA(4,1,4)	0.116602	0.0530026	4.14388	−4.05
	ARIMA(3,2,3)	0.129159	0.0552289	4.36873	−3.91
	EXPOS λ = 0.2	0.387387	0.160508	10.4269	−1.87
	LSTM	0.1046	0.1023	10.01119	−1.17
The Number of New Death	ARIMA(0,1,2)	6.04441	4.63418		3.661
	ARIMA(2,1,0)	6.05087	4.5735		3.663
	ARIMA(0,2,1)	6.17175	4.82092		3.671
	ARIMA(2,0,1)	6.03522	4.37296		3.689
	EXPOS λ = 0.6074	6.32014	4.65942		3.719
	LSTM	6.723	5.9351	11.3374	6.859

Gambar di atas adalah hasil dari evaluasi mean error dari model ARIMA dan Long Short-Term Memory. Terdapat 6 variabel yang diukur dan dievaluasi menggunakan parameter RMSE, MAE, MAPE dan AIC.

Table 1
Comparison of Naïve, ACM-B, ARIMA, GARCH, SVR, MLP on TCE datasets.

	Naïve					ACM-B				
	MAE	RMSE	MAPE	sMAPE	MASE	MAE	RMSE	MAPE	sMAPE	MASE
2015	14 965	20 521	0,4137	0,2299	14,515	9 487	16 075	0,2565	0,1570	9,064
2016	9 338	11 876	0,6540	0,2505	12,666	5 075	7 833	0,3534	0,1461	6,574
2017	5 522	7 330	1,2102	0,3517	9,803	2 973	4 323	0,8057	0,2373	5,353
2018	6 782	10 330	0,9513	0,3619	10,027	4 306	8 253	0,4885	0,2591	6,365
AVG.	**9 152**	**12 514**	0,8073	0,2985	**11,753**	**5 460**	**9 121**	0,4760	0,1999	6,839
	ARIMA					GARCH				
	MAE	RMSE	MAPE	sMAPE	MASE	MAE	RMSE	MAPE	sMAPE	MASE
2015	13 532	19 684	0,3183	0,1939	14,659	14 602	19 961	0,4414	0,1985	13,642
2016	7 938	11 625	0,6356	0,2533	10,545	8 687	12 055	0,7286	0,2628	13,292
2017	6 164	7 942	1,6747	0,4352	11,864	6 204	7 338	2,4645	0,4129	10,892
2018	8 182	12 372	1,0921	0,4192	11,247	8 110	11 643	1,2487	0,3951	14,238
AVG.	**8 953**	**12 905**	0,9301	0,3253	**12,078**	**9 400**	**12 749**	1,2207	0,3173	**13,015**
	SVR					MLP				
	MAE	RMSE	MAPE	sMAPE	MASE	MAE	RMSE	MAPE	sMAPE	MASE
2015	9 760	13 494	0,2468	0,1324	9,217	6 588	8 521	0,1942	0,0923	6,445
2016	7 024	9 183	0,6417	0,1996	9,753	5 879	7 949	0,4913	0,1707	7,807
2017	4 365	5 364	1,2885	0,2915	7,819	4 811	5 812	1,8749	0,2933	8,268
2018	6 679	9 706	1,0876	0,3296	10,495	5 800	7 630	1,2409	0,3075	8,368
AVG.	**6 956**	**9 436**	0,8161	0,2382	9,321	**5 769**	**7 477**	0,9503	0,2159	7,7219

Gambar di atas adalah hasil dari evaluasi mean error dari model Naïve, ACM-B, ARIMA, GARCH, SVR dan MLP. Hasil dievaluasi menggunakan parameter RMSE, MAE, MAPE, sMAPE dan MASE.

Cases	Significance Level	Prediction Interval	Range	True value
Recovery	80%	14403.95	-14375.95 to 14431.95	-23474.35
	90%	18511.33	-18483.33 to 18539.33	-23474.35
	95%	22056.05	-22028.05 to 22084.05	-23474.35
New Confirmed	80%	94320.84	-93765.84 to 94875.84	-64476.148
	90%	121217.02	-120662.02 to 121772.02	-64476.14
	95%	144428.79	-143873.79 to 144983.79	-64476.14
Death	80%	4719.35	-4702.35 to 4736.35	-3488.37
	90%	6065.10	-6048.10 to 6082.10	-3488.37
	95%	7226.50	-7209.50 to 7243.50	-3488.37

Table 4. Computational results for COVID-19.

Method	RMSE	MAE	MAPE	RMSRE	R2	Time
ANN	8750	5413	13.09	0.204	0.8991	-
KNN	12,100	7671	8.32	0.130	0.7710	-
SVR	7822	5354	8.40	0.080	0.8910	-
ANFIS	7375	5523	5.32	0.09	0.9032	-
PSO	6842	4559	5.12	0.08	0.9492	24.18
GA	7194	4963	5.26	0.08	0.9575	27.02
ABC	8327	6066	6.86	0.10	0.7906	46.80
FPA	6059	4379	5.04	0.07	0.9439	23.41
FPASSA	**5779**	**4271**	**4.79**	**0.07**	**0.9645**	**23.30**

Gambar di atas adalah hasil evaluasi dari model ANN, KNN, SVR, ANFIS, PSO, GA, ABC, FPA dan FPASSA menggunakan parameter evaluasi RMSE, MAE, MAPE, RMSRE, R2 dan waktu proses.