

**IMPLEMENTASI METODE SVR (*SUPPORT VECTOR REGRESSION*)  
UNTUK PREDIKSI JUMLAH POSITIF COVID-19 DI JAWA TIMUR**

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

Disusun Oleh  
**KURNIA NOVITA SUROYO**  
**H92218046**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA**

**2022**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : KURNIA NOVITA SUROYO

NIM : H92218046

Program Studi : Matematika

Angkatan : 2018

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "IMPLEMENTASI METODE SVR (*SUPPORT VECTOR REGRESSION*) UNTUK PREDIKSI JUMLAH POSITIF COVID-19 DI JAWA TIMUR". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 18 Juli 2022

Yang menyatakan,



---

KURNIA NOVITA SUROYO  
NIM. H92218046

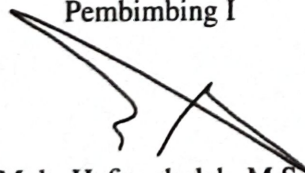
## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : KURNIA NOVITA SUROYO  
NIM : H92218046  
Judul Skripsi : IMPLEMENTASI METODE SVR (*SUPPORT VECTOR REGRESSION*) UNTUK PREDIKSI JUMLAH POSITIF COVID-19 DI JAWA TIMUR

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Dr. Moh. Hafiyusholeh, M.Si., M.PMat.  
NIP. 198002042014031001

Pembimbing II



Lutfi Hakim, M.Ag.  
NIP. 197312252006041001

Mengetahui,  
Ketua Program Studi Matematika  
UIN Sunan Ampel Surabaya



Yuniar Farida, M.T.  
NIP. 197905272014032002

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : KURNIA NOVITA SUROYO  
NIM : H92218046  
Judul Skripsi : IMPLEMENTASI METODE SVR (*SUPPORT VECTOR REGRESSION*) UNTUK PREDIKSI JUMLAH POSITIF COVID-19 DI JAWA TIMUR

Telah dipertahankan di depan Tim Penguji  
pada tanggal 18 Juli 2022

Mengesahkan,  
Tim Penguji

Penguji I



Ahmad Hanif Ashar, M.Si.  
NIP. 198601232014031001

Penguji II



Dian Candra Rini Novitasari, M.Kom.  
NIP. 198511242014032001

Penguji III



Dr. Moh. Hafiyusholeh, M.Si., M.PMat.  
NIP. 198002042014031001

Penguji IV



Lutfi Hakim, M.Ag.  
NIP. 197312252006041001

Mengetahui,

Dekan Fakultas Sains dan Teknologi  
UM Sunan Ampel Surabaya



  
Dr. A. Saepul Hamdani, M.Pd.  
NIP. 196507312000031002



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA  
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300  
E-Mail: perpustakaan@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : KURNIA NOVITA SUROYO  
NIM : H92218046  
Fakultas/Jurusan : SAINTEK / MATEMATIKA  
E-mail address : kurnianovitasuroyo96@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi  Tesis  Desertasi  Lain-lain (.....)  
yang berjudul :

IMPLEMENTASI METODE SVR (SUPPORT VECTOR REGRESSION) UNTUK PREDIKSI

JUMLAH POSITIF COVID-19 DI JAWA TIMUR

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 29 September 2022

Penulis

( KURNIA NOVITA S. )  
nama terang dan tanda tangan

## ABSTRAK

### IMPLEMENTASI METODE SVR (*SUPPORT VECTOR REGRESSION*) UNTUK PREDIKSI JUMLAH POSITIF COVID-19 DI JAWA TIMUR

Pada tahun 2020, virus korona melanda dan menggegerkan dunia karena penyebarannya yang cepat. Tingkat penyebaran virus covid-19 di daerah Jawa timur mengalami peningkatan 19,3% atau sebanyak 46.984 jiwa yang telah terkonfirmasi covid-19 pada bulan Oktober 2020. Sehingga menjadikan provinsi Jawa Timur menduduki urutan kedua setelah DKI Jakarta dalam peningkatan penyebaran kasus baru covid- 19. Melakukan prediksi atau peramalan merupakan elemen yang sangat penting dalam proses pengambilan keputusan. Sehingga dalam penelitian ini akan dilakukan proses prediksi pada kasus covid-19 yang bertujuan untuk memprediksi angka kasus covid-19 dimasa depan agar di waktu mendatang kestabilan akan tetap terjaga. Metode yang digunakan dalam penelitian ini yaitu metode SVR (*Support Vector Regression*), dimana digunakan algoritma *grid search optimization* untuk pengoptimalan nilai parameter kernel. Untuk menemukan kinerja *grid search* dalam mencari nilai parameter yang optimal akan digunakan *cross validation*. SVR memiliki kelebihan dalam mengatasi masalah *overfitting*, sehingga dapat digunakan untuk melakukan prediksi covid-19 di Jawa Timur. Kernel yang digunakan yaitu linier, RBF, polynomial, dan sigmoid. Tingkat akurasi terbaik yang dihasilkan dari ke-empat kernel tersebut yaitu kernel RBF dengan perbandingan data *training* dan *testing* yaitu sebanyak 60:40, dimana nilai akurasi MAPE dan RMSE yaitu sebesar 18,3326 dan 47.07387. Hasil tersebut jika diukur dalam MAPE yaitu memiliki nilai yang baik dalam memprediksi angka positif covid-19 di wilayah Jawa Timur.

**Kata kunci:** SVR (*Support Vector Regression*), *Grid Search Optimization*, Prediksi, Covid-19

## ABSTRACT

### IMPLEMENTATION OF THE SVR (*SUPPORT VECTOR REGRESSION*) METHOD FOR PREDICTION THE NUMBER OF POSITIVE COVID-19 IN EAST JAVA

Since 2019, the corona virus hit and stirred the world because fast spread. The rate of spread of the covid-19 virus in the East Java area has increased by 19.3% or as many as 46,984 people who have confirmed covid-19 in October 2020. So, made East Java province second after DKI Jakarta in increasing the spread of new cases of covid-19. Making prediction or forecasting is a very important element in the decision making process. So that in this study, a prediction process for Covid-19 cases will be carried out which aims to predict the number of covid-19 cases in the future so that in the future stability will be maintained. The method used in this study is the SVR (Support Vector Regression) method, where a grid search optimization algorithm is used to optimize the kernel parameter values. To find the performance of the grid search in finding the optimal parameter values will be used with cross Validation. SVR has the advantage of overcoming the overfitting problem, so it can be used to predict covid-19 in East Java. The kernels used are linear, RBF, polynomial, and sigmoid. The best level of accuracy produced from the four kernels is the RBF kernel with a comparison of training and testing data as much as 60:40, where the MAPE and RMSE accuracy values are 18.3326 and 47.07387. These results, when measured in MAPE, have a good value in predicting the positive number of covid-19 in the East Java region.

**Keywords:** SVR (*Support Vector Regression*), *Grid Search Optimization*, *Forecast*, *Covid-19*

## DAFTAR ISI

<b>HALAMAN JUDUL</b> . . . . .	<b>i</b>
<b>HALAMAN PERNYATAAN KEASLIAN</b> . . . . .	<b>ii</b>
<b>LEMBAR PERSETUJUAN PEMBIMBING</b> . . . . .	<b>iii</b>
<b>PENGESAHAN TIM PENGUJI SKRIPSI</b> . . . . .	<b>iv</b>
<b>LEMBAR PERSETUJUAN PUBLIKASI</b> . . . . .	<b>v</b>
<b>ABSTRAK</b> . . . . .	<b>vi</b>
<b>ABSTRACT</b> . . . . .	<b>vii</b>
<b>DAFTAR ISI</b> . . . . .	<b>1</b>
<b>DAFTAR TABEL</b> . . . . .	<b>3</b>
<b>DAFTAR GAMBAR</b> . . . . .	<b>4</b>
<b>I PENDAHULUAN</b> . . . . .	<b>5</b>
1.1. Latar Belakang Masalah . . . . .	5
1.2. Rumusan Masalah . . . . .	12
1.3. Tujuan Penelitian . . . . .	12
1.4. Manfaat Penelitian . . . . .	13
1.5. Batasan Masalah . . . . .	13
1.6. Sistematika Penulisan . . . . .	14
<b>II TINJAUAN PUSTAKA</b> . . . . .	<b>16</b>
2.1. Covid-19 di Jawa Timur . . . . .	16
2.2. <i>Support Vector Regression</i> (SVR) . . . . .	19
2.3. Penentuan Variabel Input . . . . .	30
2.4. Algoritma <i>Grid Search</i> . . . . .	32
2.5. <i>K-Fold Validation</i> . . . . .	33
2.6. Tingkat Akurasi . . . . .	35
2.7. Integrasi Keilmuan . . . . .	38
<b>III METODE PENELITIAN</b> . . . . .	<b>46</b>
3.1. Jenis Penelitian . . . . .	46



3.2. Jenis dan Sumber Data	46
3.3. Teknik Analisis Data	46
<b>IV HASIL DAN PEMBAHASAN</b>	<b>48</b>
4.1. Analisis Deskriptif	48
4.2. <i>Processing Data</i>	51
4.3. Menentukan Lag Berpengaruh Menggunakan Plot PACF	52
4.3.1. Penentuan Variabel	53
4.3.2. Pembagian Data <i>Training</i> dan Data <i>Testing</i>	59
4.4. Analisis Metode <i>Support Vector Regression</i>	60
4.4.1. Membangun Model SVR Sebelum dan Sesudah Proses <i>Grid Search Optimization</i>	61
4.4.2. Kernel Linier	65
4.4.3. Kernel Radial Basis Function (RBF)	66
4.4.4. Kernel Polynomial	68
4.4.5. Kernel <i>Tangent Hyperbolic</i> (Sigmoid)	69
4.5. Perbandingan Tingkat Akurasi	71
4.6. Peramalan Kasus Covid-19 di Wilayah Jawa Timur	73
4.7. Pembahasan	77
4.8. Integrasi Keilmuan	78
<b>V PENUTUP</b>	<b>80</b>
5.1. Kesimpulan	80
5.2. Saran	81
<b>DAFTAR PUSTAKA</b>	<b>81</b>

## DAFTAR TABEL

2.1 Akurasi Nilai MAPE	38
4.1 Aktual Positif Covid-19 di Jatim	49
4.2 Deskriptif Positif Covid-19 di Jatim	51
4.3 Penentuan Variabel	58
4.4 Pembagian Data <i>Training</i> dan <i>Testing</i>	60
4.5 Nilai Akurasi Tanpa Proses Pengoptimalan Parameter	62
4.6 Batas Untuk Nilai <i>Hyperparameter</i>	64
4.7 Nilai Akurasi Melalui Proses <i>Grid Search Optimization</i>	65
4.8 Akurasi Menggunakan Kernel Linier	66
4.9 Akurasi Menggunakan Kernel RBF	67
4.10 Akurasi Menggunakan Kernel Polynomial Derajat 1	69
4.11 Akurasi Menggunakan Kernel Polynomial Derajat 2	69
4.12 Akurasi Menggunakan Kernel Sigmoid	70
4.13 Perbandingan 60:40	71
4.14 Perbandingan 70:30	72
4.15 Perbandingan 80:20	72
4.16 Perbandingan 90:10	72
4.17 Perbandingan Data Aktual dan Prediksi	74
4.18 Peramalan Kasus Positif Covid-19 di Jawa Timur	76

## DAFTAR GAMBAR

2.1	Grafik Penyebaran Covid-19 di Jawa Timur	18
2.2	Konsep SVM	23
2.3	Grafik Positif Covid-19 di Jawa Timur	32
2.4	Konsep <i>K-fold Validation</i>	34
3.1	Plot Tahapan Penelitian	47
4.1	Grafik Positif Covid-19 di Jawa Timur	50
4.2	Grafik Positif Covid-19 di Jawa Timur	52
4.3	Tabel Inputan Pada Lag 11	54
4.4	Tabel Inputan Pada Lag 8	55
4.5	Tabel Penentuan Lag	57
4.6	Perbandingan Data Aktual dan Prediksi	75

UIN SUNAN AMPEL  
S U R A B A Y A

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Virus korona telah melanda dan menggegerkan dunia dikarenakan penyebarannya yang cepat. Pertama kali virus ini muncul diketahui dari pasar hewan dan *seafood* di kota Wuhan, yang terletak di Cina. Virus ini menyebar luas hampir keseluruhan negara di dunia salah satunya Indonesia (Nursofwa et al., 2020). Penularan virus covid-19 terjadi dengan sangat cepat dikarenakan penyebaran virus ini dapat menyebar melalui kontak fisik secara langsung maupun tidak langsung. Pada umumnya tanda seseorang yang telah terinfeksi covid-19 yaitu demam serta terjadi gangguan pernapasan seperti batuk, sesak napas, dan kesulitan bernapas. Gejala covid-19 dapat menyebabkan kondisi yang lebih parah seperti pneumonia, sindrom pernapasan akut, gagal napas hingga kematian (dkk. Arifin, 2021).

Pertama kali wabah virus covid-19 terdeteksi di Indonesia dengan jumlah positif yaitu sebanyak 2 orang pada tanggal 2 Maret 2020. Hanya dalam hitungan beberapa Minggu saja penyebaran wabah covid-19 di Indonesia melonjak tajam. Terhitung sampai hari Kamis 16 September 2021 kasus yang bertambah sebanyak 3.145 kasus. Sehingga total akumulasi positif covid-19 saat ini lebih dari 4,1 juta kasus atau setara dengan 4.181.309 kasus (<https://covid19.go.id/>). Dengan memperhitungkan kecepatan dan mudahnya virus ini yang masuk dan menyebar di Indonesia, maka dampak yang disebabkan juga sangat besar dan luas,

mencangkup masyarakat kecil hingga pemerintahan. Seperti halnya para tenaga kerja, kerentanan dalam segi hal ekonomi bukan satu-satunya kendala yang dihadapi, tetapi hal tersebut juga berpengaruh dalam segi kesehatan (Syahril, 2020).

Dikarenakan semakin cepatnya wabah covid-19 menyebar luas di Indonesia, pemerintah melakukan pengambilan kebijakan dalam pembatasan kegiatan masyarakat guna untuk menekan dan memutus rantai penyebaran virus covid-19 dimulai dari dilakukannya Pembatasan Sosial Berskala Besar (PSBB), PPKM, PPKM Mikro, PPKM Darurat, hingga PPKM level 4 (Soetjipto, 2020). Tindakan Pemerintah mulai menetapkan Pembatasan Sosial Berskala Besar (PSBB) sampai dengan PPKM level 4 ini juga sesuai dengan hadist nabi yang berbunyi:

إِذَا سَمِعْتُمْ بِهِ بِأَرْضٍ فَلَا تَقْدَمُوا عَلَيْهِ وَإِذَا وَقَعَ بِأَرْضٍ وَأَنْتُمْ بِهَا فَلَا تَخْرُجُوا فِرَارًا مِنْهُ

Artinya: "Jika kalian mendengar wabah tersebut menjangkiti suatu negeri, maka janganlah kalian menuju ke sana. Namun bila suatu negeri terkena wabah penyakit dan kalian sedang berada di dalamnya, maka janganlah kalian keluar dan lari darinya." (Hadist Shahih Al-Bukhari No.5289-Kitab Pengobatan)

Walaupun dalam konteks hadist tersebut membicarakan wabah penyakit lain, namun hal ini masih berkaitan dengan adanya pandemi covid-19 yang saat ini tengah menjadi wabah dan melanda di hampir seluruh dunia. Cara yang efektif dalam menangani penyakit menular yang penyebarannya sangat cepat yaitu dengan isolasi dan jaga jarak sehingga tidak memakan banyak korban. Masyarakat harusnya menjauhi zona merah penyakit ini dan tidak boleh keluar masuk di

dalamnya walaupun dalam hal beribadah (Suryadilaga, 2020).

Jumlah peningkatan kasus covid-19 juga terjadi di setiap wilayah di Indonesia, termasuk di Provinsi Jawa Timur. Tingkat penyebaran virus covid-19 di daerah Jawa timur mengalami peningkatan 19,3% atau sebanyak 46.984 jiwa yang telah terkonfirmasi covid-19 pada bulan Oktober 2020, sehingga menjadikan provinsi Jawa Timur menduduki urutan kedua setelah DKI Jakarta dalam penambahan tingkat penyebaran kasus baru covid-19 (Putri, Dinda Pratiwi Dwi, 2021). Sehingga hal ini menjadi salah satu penyebab Jawa Timur menjadi provinsi yang paling rawan dan jumlah pasien paling banyak terpapar virus covid-19.

Kenaikan angka positif covid-19 tersebut dikarenakan beberapa aspek salah satunya aspek ekonomi. Seperti contoh di Sidoarjo dapat terjadi kenaikan karena mayoritas penduduknya banyak yang menggantungkan kehidupannya dari sektor swasta dan sektor perdagangan. Sedangkan di daerah Malang, kenaikan kasus positif covid-19 dapat terjadi karena wilayah tersebut banyak tempat wisata yang mana sering dikunjungi wisatawan dari berbagai wilayah sekitar maupun wisatawan luar Jawa. Hal ini dapat menyebabkan daerah tersebut banyak mengalami kenaikan angka positif covid-19 karena rentanya antar individu yang saling berinteraksi secara langsung (Hapsari and Ulfa, 2018).

Selain itu faktor yang menjadi penyebab kenaikan angka positif covid-19 yang ada di Jawa Timur ialah jumlah penduduk yang tinggi sehingga menimbulkan tingginya interaksi sosial dalam bermasyarakat. Kepadatan penduduk juga mengakibatkan lamanya proses penanganan dalam pemberian vaksin, sehingga peluang pelonjakan kasus covid-19 juga semakin besar. Dari isu penularan dan penyebaran covid-19 di Jawa Timur, bisa dikatakan pemerintah pada dasarnya tidak mampu menghindari kematian akibat virus corona dan dampak ekonomi

akibat penyebaran virus ini. Namun, pemerintah tetap memiliki tanggung jawab untuk mengoptimalkan segala upaya guna melindungi masyarakat dari pandemi covid-19 melalui berbagai pedoman yang dikeluarkan. Seperti halnya setiap individu dianjurkan bahkan dipaksa untuk berubah dan beradaptasi dengan keadaan baru, inilah konsep pemerintahan yang akan diwujudkan mulai dari pelaksanaan karantina mandiri secara sukarela, penerapan jaga jarak atau *physical distancing*, hingga rakyat diharuskan memperbarui pola hidup yang lebih bersih dibandingkan sebelum adanya pandemi covid-19. Selain itu, banyak kegiatan kerja yang mengharuskan tatap muka, jadwal rapat dan sistem pembelajaran kini telah bergeser ke media *online*, artinya ini adalah sesuatu yang baru yang biasanya belum pernah dilakukan sebelumnya (Rahmawati et al., 2021).

Pandemi covid-19 termasuk kedalam bencana non alam yang mana untuk menghadapinya diperlukan upaya yang berguna untuk meminimalisir dampak yang ditimbulkan akibat pandemi covid-19 atau yang biasa disebut dengan mitigasi bencana. Oleh sebab itu, peneliti akan melakukan prediksi untuk era/masa *new normal* dikarenakan perlu adanya prediksi untuk masa depan yang bertujuan memperkirakan apakah ada kelonjakan atau penurunan dalam masa *new normal* ini, dengan seperti itu model prediksi yang didapat nantinya diharapkan bisa membantu pemerintah wilayah Jawa Timur dalam memprediksi banyaknya orang yang terpapar covid-19 sehingga pemerintah daerah dapat lebih siap dalam mengantisipasi lonjakan covid-19 diantaranya adalah dengan menyiapkan tenaga kesehatan, rumah sakit, rumah obat-obatan, maupun peralatan lain yang dapat menangani pasien yang terpapar virus covid-19. Serta pemerintah dapat menyiapkan bantuan terhadap korban atau keluarga yang terdampak covid-19.

Meskipun dalam Al-Quran ramalan tidak dijelaskan, akan tetapi Allah tetap

memerintah manusia untuk selalu senantiasa bekerja dan berikhtiar untuk mempersiapkan masa depan guna memperoleh kehidupan yang lebih baik dan aman di dunia maupun di akhirat kelak (Fairuzyana, 2016). Sebagaimana kisah dari nabi Yusuf yang terdapat pada ayat 46 surat Yusuf, yaitu :

يُوسُفُ أَيُّهَا الصِّدِّيقُ أَفْتِنَا فِي سَبْعِ بَقَرَاتٍ سِمَانٍ يَأْكُلُهُنَّ سَبْعُ عِجَافٍ وَسَبْعِ سُنبُلَاتٍ  
خُضْرٍ وَأُخَرَ يَبِيسٍ لَعَلِّي أَرْجِعُ إِلَى النَّاسِ لَعَلَّهُمْ يَعْلَمُونَ

Artinya: *Dia (pelayan) berkata saat berjumpa dengan Yusuf “Wahai Yusuf, orang yang sangat dipercaya, jelaskanlah kepada kami (takwil mimpiku) tentang tujuh ekor sapi gemuk yang dimakan oleh tujuh (ekor sapi) kurus dan tujuh tangkai (gandum) hijau yang (meliputi tujuh tangkai) lainnya yang kering agar aku kembali kepada orang-orang itu supaya mereka mengetahuinya”.* (QS: Yusuf: 46)

Kemudian Nabi Yusuf menyuruh mereka untuk bercocok tanam selama tujuh tahun berturut-turut, dikarenakan pada saat itu musimnya lagi subur dan baik untuk bercocok tanam. Kemudian setelah dipanen, Nabi Yusuf memerintahkan untuk membiarkan atau tidak mencabut tangkainya kecuali untuk dimakan. Setelah itu mereka harus berhemat dan menyimpan hasil panen mereka dengan baik. Hal ini supaya hasil panen mereka awet atau bertahan lama untuk menghadapi tujuh tahun selanjutnya, dimana tahun tersebut adalah musim yang tidak subur, sehingga mereka sulit untuk bercocok tanam.

Dari ayat yang telah dijelaskan diatas, bermakna bahwasanya manusia diajarkan untuk selalu senantiasa mengelola dan mengembangkan kekayaan yang ada, sehingga dapat merencanakan serta mempersiapkan kehidupan masa depan yang lebih baik untuk kepentingan banyak orang (Jalaluddin, 2021).



Melakukan prediksi atau peramalan merupakan elemen yang sangat penting dalam proses pengambilan keputusan. Biasanya peramalan dilakukan untuk mengurangi risiko dalam kondisi ketidakpastian tentang sesuatu yang akan terjadi di masa mendatang. Sehingga metode peramalan ini dapat meminimalkan sesuatu yang sifatnya tidak pasti. Penelitian ini difokuskan terhadap Provinsi Jawa Timur, sebagai salah satu daerah yang banyak terdampak covid-19. Dalam menentukan metode peramalan yang paling baik dapat dilakukan dengan melihat pola data historis yaitu pola horizontal, pola musiman, pola siklis, dan pola trend. Salah satu metode peramalan yang bersifat objektif yaitu *time series*. Data runtun waktu diperoleh dengan cara mengumpulkan, merekam atau mengamati secara berurutan dari waktu ke waktu dalam kurun waktu tertentu.

Banyak metode yang digunakan dalam melakukan prediksi atau peramalan menggunakan data *time series*, seperti metode naïve yaitu metode peramalan yang sangat sederhana karena data selanjutnya sama dengan data aktual sebelumnya (Kumila et al., 2019). Kemudian metode *average* yaitu metode yang paling sering digunakan dan standart karena mengambil rata-ratanya yang akan digunakan untuk meramalkan beberapa periode yang akan datang (Wardah and Iskandar, 2017). Selain itu metode *exponential* yang merupakan metode yang mengatur data masa lalu sesuai dengan musim data yang sedang terjadi (Lusiana and Yuliarty, 2020). Setelah itu metode regresi *time series* yaitu metode regresi yang paling sederhana karena hanya menggunakan variabel (t) sebagai prediktor (Sadewo, 2016). Tetapi ada juga metode *time series* yang mengarah ke teknik yang berbasis *machine learning* yang masih belum banyak digunakan oleh para peneliti untuk melakukan proses prediksi, baik digunakan dalam regresi maupun kasus klasifikasi (Lumbanraja et al., 2019).

Beberapa penelitian yang telah dilakukan sebagai acuan referensi terkait penelitian yang diambil, yaitu penelitian yang telah dilakukan oleh (Putri et al., 2020) mengenai metode *Ridge Regression* dan metode *Support Vector Regression*, dimana hasil prediksi SVR dengan kernel polynomial memiliki tingkat keakuratan yang lebih tinggi dibandingkan *Ridge Regression*. Selain itu, metode SVR juga memiliki hasil yang lebih baik dibandingkan RBF dan NN, dikarenakan SVR adalah salah satu metode yang baik dan sangat berguna untuk data yang distribusinya tidak diketahui (Karim, 2020). SVR juga lebih baik dalam pemodelan beban listrik di empat belas wilayah di Jawa Timur dibandingkan metode Arima (Widjaya et al., 2017).

Berdasarkan penelitian yang telah dilakukan sebelumnya, penulis tertarik untuk melakukan prediksi kasus positif covid-19 di wilayah Jawa Timur dengan menggunakan metode SVR (*Support Vector Regression*). SVR merupakan penerapan algoritma SVM dalam kasus regresi yang mana pada algoritma penerapannya menghasilkan *output* berupa bilangan riil. Sehingga Algoritma SVR memiliki keunggulan menunjukkan kinerja yang sangat baik untuk prediksi deret waktu (*time series*) yang bersifat linier maupun nonlinier. Dalam mengatasi permasalahan nonlinier dapat menggunakan bantuan kernel. Kernel yang biasanya digunakan yaitu, kernel Linier, kernel Polinomial, kernel Radial Basis Function (RBF) / Gaussian, dan kernel Tangent hyperbolic (sigmoid). Selain itu, algoritma SVR merupakan metode yang didasarkan pada teori pembelajaran statistik dan memberikan hasil yang menjanjikan yang akan lebih baik dari metode lainnya (Drajana, 2017). Konsep dari SVR sama dengan SVM yaitu memaksimalkan *hyperplane* sehingga galat yang dihasilkan lebih kecil.

Permasalahan yang biasanya terjadi ketika menggunakan metode SVR

adalah pada saat menentukan parameter model yang optimal. Dalam mengatasi permasalahan tersebut, salah satu cara yang dapat digunakan untuk menentukan parameter yang terbaik yaitu dengan menggunakan *Grid Search Optimization*. *Grid Search Optimization* ini akan membagi jangkauan parameter yang akan dioptimalkan kedalam grid dan melintasi semua titik sehingga didapatkan parameter yang optimal (Asyiva, 2019).

### 1.2. Rumusan Masalah

Berdasarkan hal di atas, penulis akan merumuskan masalah yang akan diselesaikan, yaitu:

1. Bagaimana penerapan metode peramalan dalam memprediksi kasus covid-19 menggunakan metode SVR dengan algoritma *Grid Search Optimization*?
2. Bagaimana hasil peramalan pertambahan pasien positif covid-19 dengan menggunakan metode SVR?
3. Bagaimana perbandingan tingkat akurasi yang dihasilkan oleh metode SVR menggunakan kernel linier, RBF, polynomial, dan sigmoid?

### 1.3. Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini yaitu:

1. Untuk mendapati bagaimana penerapan metode peramalan dalam memprediksi kasus covid-19 menggunakan metode SVR dengan algoritma *Grid Search Optimization*.
2. Untuk mendapati hasil peramalan pertambahan pasien positif covid-19 dengan menggunakan metode SVR.

3. Untuk mengetahui tingkat akurasi yang dihasilkan dari fungsi kernel terbaik guna untuk melakukan peramalan angka positif covid-19 di Jawa Timur dalam 1 bulan kedepan.

#### 1.4. Manfaat Penelitian

Hasil penelitian yang telah dilakukan diharapkan dapat bermanfaat untuk :

1. Bagi penulis, bisa melatih kemampuan penulis dalam menganalisis perkara dari fakta dan data yang diadaptasi dengan pengetahuan yang didapatkan selama kuliah, serta dapat menambah wawasan kepada para pembaca.
2. Bagi Pembaca, dapat menambah wawasan serta dapat dijadikan referensi bahan dalam melakukan penelitian selanjutnya
3. Bagi Pemerintah, bisa membagikan manfaat bagi para pengambil keputusan yang dapat berguna sebagai dasar pengambilan kebijakan dalam penanganan covid-19 di wilayah Jawa Timur.

#### 1.5. Batasan Masalah

Dalam penelitian ini batasan masalah yang perlu diperhatikan yaitu sebagai berikut :

1. Data yang digunakan pada penelitian ini yaitu data jumlah kasus harian covid-19 yang ada di wilayah Jawa Timur mulai dari 10 Februari 2022 sampai dengan 31 Juli 2022.
2. Metode yang digunakan dalam penelitian ini yaitu *support vector regression* dengan menggunakan algoritma *grid search optimization*.

3. Fungsi kernel yang digunakan pada penelitian ini yaitu kernel linier, RBF, polynomial, dan Sigmoid.
4. Data penelitian ini didapatkan dari web pemerintah yang menangani kasus harian covid-19 yang ada di wilayah Jawa Timur.

### **1.6. Sistematika Penulisan**

Untuk memberikan gambaran jelas tentang permasalahan yang akan dikaji dalam penulisan ini, maka disajikan sistematika penyusunan sebagai berikut :

#### **1. BAB I PENDAHULUAN**

Pada bab I akan diuraikan tentang Latar Belakang terkait masalah yang akan diteliti, Rumusan Masalah, Tujuan Penelitian, Manfaat Penelitian, Batasan Masalah, dan Sistematika Penulisan.

#### **2. BAB II KAJIAN PUSTAKA**

Pada bab II berisi tentang landasan teori yang akan dipakai sebagai acuan dalam menganalisa dan memprediksi positif covid-19 yang ada di Jawa Timur.

#### **3. BAB III METODE PENELITIAN**

Pada bab III berisi tentang jenis penelitian, jenis dan sumber data, variabel-variabel yang digunakan dalam penelitian, serta pembahasan mengenai sistematika dalam teknik analisis data.

#### **4. BAB IV HASIL DAN PEMBAHASAN**

Berisi tentang hasil dan pembahasan mengenai rumusan masalah pada penelitian ini.

## 5. BAB V PENUTUP

Berisi tentang kesimpulan penelitian serta saran dari peneliti kepada peneliti-peneliti selanjutnya.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Covid-19 di Jawa Timur

Virus *corona* termasuk sekelompok virus yang dapat menyerang hewan dan manusia, dikarenakan virus ini berasal dari *subfamili Orthocoronavirinae* dalam *famili Coronaviridae* dan *Ordo Nidovirales*. Pada manusia, gejala yang ditimbulkan akibat virus ini serupa dengan infeksi pada penyakit SARS dan MERS, tetapi pada covid-19 perkembangannya masih sangat bersifat masif. Virus ini pertama kali ada di negara China di daerah kota Wuhan, tepatnya pada Provinsi Hubei pada akhir tahun 2019. Dalam kurun waktu 30 hari saja, virus ini dapat menyebar dengan pesat disegala penjuru kota di Negara China. Hampir diseluruh negara saat ini sedang dilanda penyebaran penularan virus covid-19. Indonesia termasuk salah satu negara yang mendapat dampak negatif dari mewabahnya virus ini. Virus ini dapat mengakibatkan hal yang fatal, terutama bagi mereka yang sebelumnya pernah menderita gangguan pernapasan akan mengalami sindrom gangguan pernapasan akut meskipun sudah dinyatakan sembuh dari virus ini. Hal itu karena efek jangka panjang dari infeksi covid-19 setelah melewati serangkaian pemulihan yang mana fungsi paru-paru penderita akan mengalami penurunan sebanyak 20% sampai 30%. Selain itu, gejala awal yang ditimbulkan dari virus ini yaitu penderita merasa seperti pusing serta terjadinya gangguan pada indera penciuman dan perasa (Wahidah et al., 2020).

Terjadinya penularan virus ini biasanya terjadi lewat droplet dan kontak

secara langsung dengan orang yang terinfeksi virus covid-19 yang kemudian virus tersebut masuk ke dalam mukosa yang sedang terbuka (Virania et al., 2021). Sebuah analisis pernah dilakukan dengan cara mengukur tingkat penularan berdasarkan dari masa inkubasi, gejala, dan durasi antara gejala dengan pasien yang sedang diisolasi. Pada analisa tersebut didapatkan bahwa penularan yang terjadi pada 1 pasien mampu menginfeksi 3 orang disekitarnya, namun kemungkinan penularan pada masa inkubasi menyebabkan masa kontak pasien ke orang sekitar menjadi lebih lama sehingga resiko tertular dari 1 pasien mungkin lebih besar. Virus *corona* merupakan zoonosis, sehingga mengakibatkan penularan virus dapat terjadi dari hewan ke manusia (Davies, 2020).

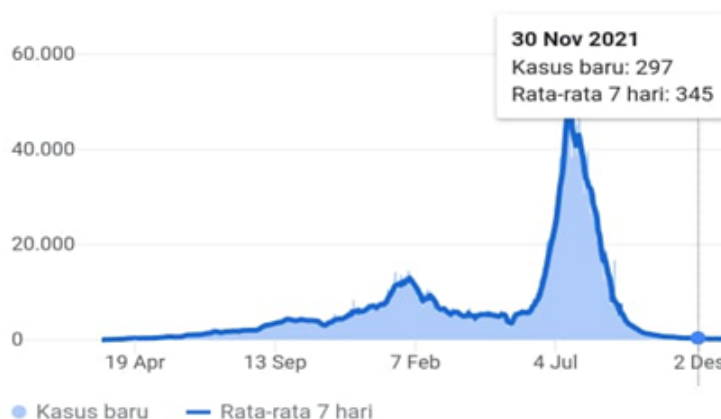
Salah satu provinsi di Indonesia yang memiliki jumlah pasien positif *corona* (covid-19) terbesar ke-3 yaitu provinsi Jawa Timur. Dimana pada tanggal 19 Mei 2020 kasus positif berjumlah 1673 (73,35%) , kasus sembuh sebanyak 375 (16,44%) dan meninggal sebanyak 224 (9,82%)(Uiya, 2020). Selain itu, Jawa Timur benar-benar gawat dengan menjadi provinsi dengan total kasus 12.321 dan jumlah kasus aktif mencapai 7.196 jiwa yang mana hal ini mengakibatkan Jawa Timur menjadi peringkat 1 provinsi dengan jumlah kasus positif covid-19 di Indonesia pada tanggal 1 Juli 2020. Kemudian pada tanggal 26 November 2020 total kasus covid-19 di Jawa Timur mencapai 60.190 jiwa. Dari total kasus tersebut terdapat kasus positif (aktif) sebanyak 2.784 jiwa dan tingkat kesembuhan mengalami kenaikan yaitu sebanyak 53.131 jiwa (Rohman et al., 2021).

Kenaikan angka tersebut disumbang dari beberapa wilayah yang ada di Jawa Timur seperti halnya di kota Surabaya yang menjadi kota terbanyak yang menyumbang kasus covid-19. Dilihat dari laman [news.detik.com](http://news.detik.com) pada per November 2021 jumlah kasus aktif di Jawa Timur yaitu sebanyak 298 kasus,



dimana kota Surabaya menyumbang sebanyak 23 kasus. Setelah Surabaya, daerah yang menyumbang kasus covid-19 yaitu Malang dan Ngawi dimana memiliki 20 dan 17 kasus. Kenaikan angka ini dikarenakan banyaknya masyarakat yang melakukan perjalanan luar kota dikarenakan Pembelajaran Tatap Muka (PTM) yang telah dilakukan.

Berdasarkan data dari situs resmi covid-19 provinsi Jawa Timur diperoleh data terbaru per November 2021, menunjukkan grafik kasus positif covid-19 di Jawa Timur sebagai berikut:



**Gambar 2.1 Grafik Penyebaran Covid-19 di Jawa Timur**

Sumber: JHU CSSE COVID-19 Data

Berdasarkan grafik diatas membuktikan bahwa kasus covid-19 di Jawa Timur masih ada dan bahkan sewaktu-waktu bisa melonjak tajam. Tidak ada pengobatan khusus untuk menyembuhkan infeksi akibat virus *corona*. Pada umumnya penderita akan sembuh dengan sendirinya dengan pengobatan yang bisa dilakukan di dalam rumah. Namun, terdapat beberapa upaya yang mampu dilakukan untuk meredakan gejala akibat infeksi virus *corona*. Meskipun tantangan dalam mengendalikan covid-19 termasuk pada kecepatan

penyebarannya dan tingkat penularannya yang tinggi, padahal pengobatan untuk penyakit ini masih sedikit yang diketahui serta vaksinnnya masih terus dikembangkan. Sehingga dibutuhkan upaya yang lebih kuat dari pemerintah serta masyarakat untuk mematuhi protokol kesehatan sehingga dapat menekan laju sebaran covid-19. Selain itu, dalam menghadapi pandemi covid-19 yang berkembang pesat, diperlukan juga tahapan analisis yang mampu memberikan serta menyajikan informasi yang berkualitas dan tepat waktu bagi para pengambil keputusan di Indonesia.

## **2.2. Support Vector Regression (SVR)**

*Support Vector Machine* (SVM) pertama kali dikembangkan oleh Boser, Guyon, Vapnik kemudian dipresentasikan pertama kali pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep dasar SVM sebenarnya adalah kombinasi serasi berdasarkan teori-teori komputasi yang sudah ada pada puluhan tahun sebelumnya (Assaffat, 2016a). Selama dekade terakhir, SVM menjadi salah satu algoritma Machine learning yang paling populer serta metode yang kuat saat digunakan untuk melakukan prediksi baik dalam pola klasifikasi dan regresi atau yang biasa disebut dengan SVR, dimana metode ini memiliki tingkat keberhasilan yang tinggi saat diterapkan di berbagai bidang. Sehingga banyak dari kalangan komunitas *machine learning* berminat untuk mempelajari dan mengembangkan SVM karena kinerjanya yang sangat baik dalam berbagai masalah pembelajaran (Assaffat, 2016b).

*Support Vector Regression* (SVR) adalah sistem pembelajaran yang menerapkan fungsi-fungsi linear pada sebuah fitur ruang hipotesis dengan dimensi yang tinggi. Konsep algoritma SVR dapat menghasilkan nilai peramalan yang bagus karena SVR mempunyai kemampuan menyelesaikan masalah *overfitting*.

*Overfitting* adalah perilaku data saat fase pelatihan atau *training* yang menghasilkan akurasi prediksi hampir sempurna (Maulana and Kumalasari, 2019). SVR dilatih dengan menerapkan algoritma pembelajaran untuk mendukung teori optimasi. SVR memiliki landasan teori yang ketat dan merupakan sejenis metode pembelajaran mesin yang didasarkan pada prinsip minimalisasi risiko struktural melalui perbaikan dan pengembangan berkelanjutan (Fahrusyiana, 2019). Tujuan pada algoritma SVR adalah menemukan garis pemisah atau bisa disebut dengan *hyperplane* terbaik. *Hyperplane* ialah garis pemisah terbaik antara dua buah kelas karena terletak tepat diantara kedua kelas. Untuk mencari *hyperplane* terbaik dapat ditemukan dengan cara mengukur *margin* dari *hyperplane* dan mencari titik maksimumnya. Pada *hyperplane*, *margin* yang paling dekat antara *hyperplane* dengan data pada dua buah kelas yang berbeda disebut dengan *support vector* (Marhamah, 2020).

Seperti pada konsep SVM, ide dasar dari SVR adalah mengubah ruang input menjadi ruang berdimensi tinggi melalui transformasi nonlinier, kemudian menemukan hubungan linear antara variabel input dan output dalam ruang dimensi tinggi ini. Jika pada SVM konsepnya ialah membagi data menjadi dua buah kelas, sedangkan untuk konsep dari SVR ialah bagaimana caranya agar semua data masuk kedalam satu zona dengan tetap meminimalisasi nilai epsilon  $\varepsilon$ . Diasumsikan terdapat data *training*  $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$  menggunakan data *points*  $x_i (i = 1, 2, \dots, n)$ , dimana  $x_i \in R^d$  merupakan vektor *input* data ke- $i$ ,  $d$  merupakan dimensi,  $y_i$  adalah nilai target, dan  $n$  merupakan jumlah data yang digunakan. Metode SVR digunakan untuk mencari fungsi  $f(x)$  yang mempunyai deviasi maksimal sebesar  $\varepsilon$  untuk mendapatkan target  $y_i$  dari semua data *training*. Fungsi regresi yang sempurna jika fungsi regresi tersebut memiliki batas deviasi

( $\varepsilon$ ) sama dengan 0. Sedangkan model SVR pada fungsi regresi linier atau persamaan umum yang digunakan dalam metode SVR adalah sebagai berikut

(Cahyono et al., 2019):

$$f(x) = w\phi(x) + b \quad (2.1)$$

Dimana:

$w$  = adalah nilai bobot yang berbentuk vektor kolom berukuran (px1)

$b$  = merupakan nilai bias yang berupa konstanta

$\phi(x)$  = titik didalam *feature space* F, hasil dari pemetaan  $x$  di dalam *input space*

$x$  = vektor *input*

$f(x)$  = fungsi regresi

Dalam mendapatkan generalisasi yang baik untuk regresi  $f(x)$  maka dapat dilakukan dengan cara meminimalkan fungsi resiko pada koefisien  $w$ . Salah satu cara untuk mencari nilai tersebut adalah dengan meminimalkan bentuk *Euclidean*  $\|w\|^2$ . Hal tersebut dapat ditulis secara matematis dengan permasalahan *convex optimization* yang dijelaskan seperti rumus berikut ini:

$$\frac{1}{2} \|w\|^2 \quad (2.2)$$

dengan syarat

$$y_i - \langle w \cdot x \rangle - b \leq \varepsilon$$

$$\langle w \cdot x \rangle + b - y_i \leq \varepsilon$$

Pada persamaan (2.2) semua titik diasumsikan berada pada dan dalam rentang  $f(x) \pm \varepsilon$ . Terdapat kondisi dimana nilai kesalahan diperbolehkan melebihi dari batas nilai  $\varepsilon$  yang dikarenakan titik-titik yang mungkin keluar dari rentang  $f(x) \pm \varepsilon$ . Dalam kondisi seperti ini maka dibutuhkan *soft margin* atau biasa disebut dengan variabel *slack* ( $\xi_i + \xi_i^*$ ) sehingga persamaan (2.2) berubah menjadi persamaan dibawah ini:

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2.3)$$

dengan syarat

$$y_i - \langle w \cdot x \rangle - b \leq \varepsilon + \xi_i$$

$$\langle w \cdot x \rangle + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

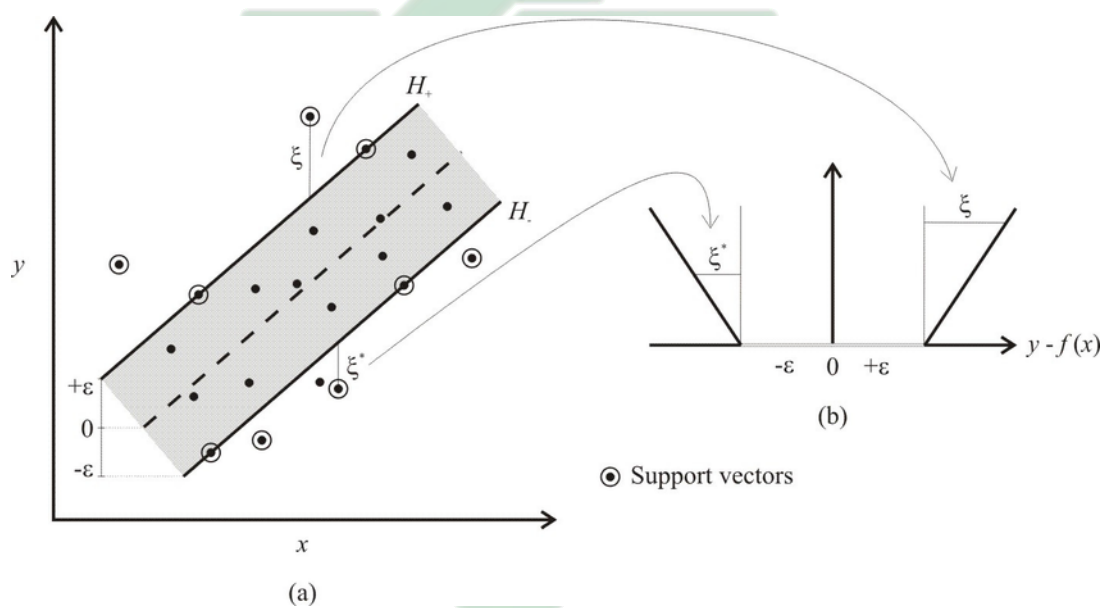
Konstanta  $C > 0$  menentukan nilai pinalti antara *flatness* dari fungsi  $f(x)$  dan seberapa besar tingkat deviasi kesalahan dari batas  $\varepsilon$  yang masih bisa ditoleransi. Semua nilai yang lebih besar dari  $\varepsilon$  akan dikenakan pinalti sebesar  $C$ . Sedangkan  $\xi_i, \xi_i^* \geq 0$  adalah variabel *slack* positif yang bertujuan untuk mengatasi masalah pembatas yang tidak layak dalam masalah optimasi (Hendayanti et al., 2019).

*Loss function* adalah fungsi yang menunjukkan hubungan antara *error* dengan bagaimana *error* ini dikenai pinalti, perbedaan *loss function* akan menghasilkan formula SVR yang berbeda. *Loss function* yang paling sederhana adalah  $\varepsilon$  - *intensitive loss function*  $|\xi_i|_\varepsilon$  sebagai sebuah pendekatan *Huber's loss function* yang memungkinkan serangkaian *support vector* akan diperoleh (Sriyana

et al., 2019). Berikut ini adalah formulasi  $\varepsilon$ -intensitive loss function  $|\xi_i|_\varepsilon$ :

$$|\xi_i|_\varepsilon \begin{cases} 0, & |\xi_i| \leq \varepsilon \\ |\xi_i| - \varepsilon, & \text{lainya} \end{cases} \quad (2.4)$$

Berikut disajikan gambar mengenai konsep metode SVR.



Gambar 2.2 Konsep SVM

Sumber: (Nugroho and Purqon, 2015)

Pada gambar (2.2) merupakan ilustrasi dari konsep *Support Vector Regression* dan  $\varepsilon$ -intensitive loss function. Dimana semua titik akan dikenai penalti jika berada di luar margin. Nilai  $\varepsilon$  dapat menentukan performansi dari model SVR. Jika nilai  $\varepsilon$  yang kecil berkaitan dengan nilai yang tinggi pada variabel *slack* dan akurasi pada aproksimasi yang tinggi. Sebaliknya, nilai yang tinggi untuk  $\varepsilon$  berkaitan dengan nilai yang kecil pada variabel *slack* dan membuat akurasi menjadi rendah. Nilai *slack* yang tinggi akan mengakibatkan kesalahan empiris

pada perhitungan (Yudhawan, 2020).

Koefesien nilai  $w$  dan  $b$  berfungsi untuk meminimalkan fungsi resiko yang menjadi masalah pemograman kuadratik (*quadratic programming*). Dimana dalam menentukan nilai parameter  $w$  dan  $b$  akan lebih mudah diselesaikan menggunakan *lagrange multiplier*, karena dapat meminimumkan suatu fungsi kuadrat dengan syarat suatu pertidaksamaan linier. Solusi optimal untuk persamaan (2.3) dengan pembatas pada persamaan (2.4) dapat diuraikan dengan persamaan *lagrange multiplier* sebagai berikut:

$$\begin{aligned}
 L_p(w, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \eta_i, \eta_i^*) = & \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) - \\
 & \left( \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + \langle w \cdot x \rangle + b) \right) - \\
 & \left( \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle w \cdot x \rangle - b) \right) - \\
 & \left( \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \right)
 \end{aligned} \tag{2.5}$$

Dengan  $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$  adalah koefesien *lagrange*. Dimana untuk mendapatkan solusi yang optimal, maka dapat dilakukan dengan cara turunan parsial dari  $L_p$  terhadap  $w, b, \xi_i, \xi_i^*$ , sehingga didapatkan sebagai berikut:

$$\begin{aligned}
 \frac{\partial L_p}{\partial w} = 0 \longrightarrow w &= \sum_{i=1}^n (\alpha_i + \alpha_i^*) x_i \\
 \frac{\partial L_p}{\partial w} = w - \sum_{i=1}^n (\alpha_i + \alpha_i^*) x_i &= 0
 \end{aligned} \tag{2.6}$$

$$\frac{\partial L_p}{\partial b} = 0 \longrightarrow w = \sum_{i=1}^n (\alpha_i + \alpha_i^*) \tag{2.7}$$

$$\begin{aligned}\frac{\partial L_p}{\partial \xi_i} &= 0 \longrightarrow \alpha_i + \eta_i = C \\ \frac{\partial L_p}{\partial \xi_i} &= C - \alpha_i - \eta_i = 0\end{aligned}\quad (2.8)$$

$$\begin{aligned}\frac{\partial L_p}{\partial \xi_i^*} &= 0 \longrightarrow \alpha_i^* + \eta_i^* = C \\ \frac{\partial L_p}{\partial \xi_i^*} &= C - \alpha_i^* - \eta_i^* = 0\end{aligned}\quad (2.9)$$

Setelah itu substitusikan persamaan diatas kedalam persamaan lagrange, sehingga diperoleh persamaan dual untuk problem optimasi dari SVR, sebagai berikut:

$$\begin{aligned}L_d(\alpha_i, \alpha_i^*) &= -\frac{1}{2} \sum_{i=1}^n (\alpha_i - \alpha_i^*)(\alpha_j, \alpha_j^*) \langle x_i \cdot x_j \rangle - \\ &\quad \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*)\end{aligned}\quad (2.10)$$

dengan kendala

$$\begin{aligned}\sum_{i=1}^n (\alpha_i - \alpha_i^*) &= 0 \\ 0 &\leq \alpha_i, \alpha_i^* \leq C \\ \forall_i &= 1, 2, \dots, N\end{aligned}$$

Berdasarkan penurunan rumus pada persamaan diatas, maka diperoleh persamaan untuk solusi optimal dari kendala estimasi parameter  $w$  dalam bentuk koefisien lagrange yaitu  $\alpha_i, \alpha_i^*$ . Solusi permasalahan ini diturunkan dari vektor  $w$



kemudian disubstitusikan ke dalam fungsi  $f(x)$ , maka dapat ditulis kembali:

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \quad (2.11)$$

Selanjutnya nilai  $b$  dapat dimaksimalkan dari kondisi:

$$\alpha_i (\varepsilon + \xi_i + y_i - \langle w \cdot x \rangle - b) = 0 \quad (2.12)$$

$$\alpha_i^* (\varepsilon + \xi_i - y_i + \langle w \cdot x \rangle + b) = 0 \quad (2.13)$$

$$(C - \alpha_i) \xi_i = 0$$

$$(C - \alpha_i^*) \xi_i^* = 0$$

Sehingga nilai estimasi dari  $b$  yang diperoleh adalah sebagai berikut:

$$b = y_i - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot \langle x_i \cdot x_j \rangle + \varepsilon \quad (2.14)$$

dan

$$b^* = y_i - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot \langle x_i \cdot x_j \rangle - \varepsilon$$

Untuk  $0 \leq \alpha_i, \alpha_i^* \leq C$

Dari persamaan (2.14) didapatkan bahwa  $C = \alpha_i$  sehingga nilai  $w$  dihitung bergantung dengan nilai *support vector*-nya. Setelah mendapatkan nilai  $w$  dan  $b$ , selanjutnya mensubstitusikan persamaan (2.1) dan persamaan (2.11). Sehingga

didapatkan fungsi regresi SVR dalam kasus linier sebagai berikut:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot \langle x_i \cdot x_j \rangle + b \quad (2.15)$$

SVR dapat digunakan untuk kasus nonlinier dikarenakan kebanyakan data dalam dunia nyata pada umumnya bersifat *linear separable*. Permasalahan ini dapat diatasi dengan mentransformasikan terlebih dahulu nilai  $x$  ke dalam *feature space*. Hal ini dapat dilakukan dengan pendekatan alternatif yaitu melalui pemetaan vektor  $x$  dari *input space* ke *feature space* dengan dimensi yang lebih tinggi melalui suatu fungsi  $\phi$ , sehingga  $\phi : x \rightarrow \phi(x)$ . Sehingga fungsi regresi SVR untuk kasus nonlinier dapat dituliskan sebagai berikut:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i) \phi(x_j) + b \quad (2.16)$$

Setelah input vektor  $x$  sudah ditransformasikan oleh fungsi  $\phi$  dan berada didalam *feature space*, maka fungsi transformasi dapat diatasi dengan *kernel trick* (Purnama and Setianingsih, 2020). Selain itu, performansi model atau besar dan kecilnya nilai akurasi yang akan dihasilkan pada model yang dibangun pada proses pelatihan dengan metode SVR sangat bergantung pada fungsi kernel dan parameter yang dipakai. Sehingga fungsi kernel harus diatur dengan benar karena dapat mempengaruhi akurasi pada regresi. *Kernel trick* merupakan metode yang dilakukan dengan cara perkalian antara input fitur data  $x$  yang kemudian diganti dengan fungsi kernel dengan persamaan sebagai berikut (Cahyono et al., 2019):

$$(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = K(x_i, x_j) \quad (2.17)$$

Dengan fungsi kernel tersebut mampu mendefinisikan secara implisit transformasi  $\phi$ . Sehingga didapatkan fungsi akhir dari persamaan 2.17 menjadi:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (2.18)$$

Fungsi  $K(x_i, x_j)$  merupakan *kernel trick* yang sering digunakan dalam metode SVM maupun SVR. Dalam metode SVR, fungsi kernel yang biasanya digunakan yaitu kernel linier, kernel polinomial, kernel radial basis function (RBF) / gaussian, dan kernel tangen hyperbolic (sigmoid). Fungsi kernel yang dipakai untuk substitusi *product* dalam *feature space* akan sangat bergantung pada data (Darsyah, 2014).

Berikut adalah beberapa penjelasan mengenai fungsi kernel secara matematis, yaitu:

#### 1. Kernel Linier

Kernel linier merupakan fungsi kernel yang paling sederhana dikarenakan cara hyperplane atau pembatas antar kelas hanya menggunakan garis lurus atau saat data yang akan dianalisis telah terpisah secara linier.

$$K = (x, y) = x \cdot y \quad (2.19)$$

#### 2. Kernel Polynomial

Kernel Polinomial ini memiliki dua parameter berbeda dari kernel lain yaitu parameter  $c$  sebagai parameter bebas dan parameter  $d$  sebagai parameter derajat atau kuadrat. Parameter  $c$  disebut homogen apabila nilai  $c$  sama dengan nol dan parameter kuadrat pada umumnya diisi dengan 2 (Nugraha

and Purnamasari, 2019).

$$K(x, y) = (x \cdot y + c)^d \quad (2.20)$$

### 3. Kernel Radial Basis Function (RBF)

Pada umumnya kernel RBF dapat dimanfaatkan pada semua jenis data. Fungsi grafik pada kernel ini berbentuk lonceng. Kernel RBF memiliki dua parameter yaitu *gamma* dan *cost* dalam melakukan analisis optimasinya. Parameter *gamma* dapat digunakan untuk mengontrol kecepatan *learning* dan parameter *cost* dapat dipakai untuk membatasi nilai alfa di setiap sampel ketika proses *training*.

$$K(x, y) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (2.21)$$

### 4. Kernel Tangent Hyperbolic (Sigmoid)

Kernel Tangent hyperbolic yaitu fungsi trigonometri hiperbolik yang memiliki bentuk huruf "S". kernel ini adalah *output* dari gabungan fungsi-fungsi eksponen (Sriyana et al., 2019).

$$K(x, y) = \tan(\sigma(x \cdot y) + c) \quad (2.22)$$

$x, y$  adalah pasangan dua data berdasarkan seluruh bagian dari data *training*, dimana  $x$  adalah data *input* sedangkan  $y$  adalah data *input* yang ingin memprediksi *output*. Parameter  $\sigma, c, d, > 0$  ialah konstanta  $\|x - x_i\|^2$  yang merupakan kuadrat jarak antara vektor  $x$  dan  $y$ . Fungsi kernel dapat dipakai untuk substitusi *dot product* berdasarkan fitur dimensi lama ke dimensi baru yang mana hal itu sangat bergantung

pada kondisi data. Biasanya *k-fold validation* digunakan dalam pemilihan fungsi kernel. Pemilihan fungsi kernel yang tepat merupakan hal yang sangat penting lantaran akan menentukan *feature space* (Audina et al., 2021).

### 2.3. Penentuan Variabel Input

Sebelum melakukan prediksi covid-19 menggunakan metode SVR, dilakukan tahap *processing* data yaitu penentuan inputan atau penyusunan data. Format penentuan inputan atau variabel terbagi menjadi dua yaitu target value dan atribut. Nilai value ini adalah hasil dari nilai prediksi sedangkan nilai atribut adalah nilai-nilai yang berpengaruh terhadap nilai prediksi. Atribut ini dapat dikatakan juga sebagai variabel X atau variabel bebas, sedangkan target value merupakan variabel Y atau biasa disebut variabel terikat. Pada penelitian ini, data yang digunakan yaitu data runtun waktu (*time series*) sehingga analisis runtun waktu yaitu pada waktu ( $X_t$ ) yang bergantung pada beberapa waktu sebelumnya ( $X_{t-k}$ ). Penentuan inputan pada metode SVR yaitu menggunakan grafik pada plot PACF (*Partial Autocorrelation Function*) karena PACF dapat digunakan untuk mengukur tingkat korelasi antara X dan Y. Fungsi PACF dinotasikan dengan  $\phi_{kk}$ . Cara menggunakan inputan pada metode SVR menggunakan plot PACF yaitu fungsi autokorelasi parsial dapat diturunkan berdasarkan model regresi dengan variabel  $X_{t+k}$  sebagai variabel terikat dari sebuah proses stasioner dengan rata-rata nol yang diregresikan dengan variabel  $X_{t+k-1}, X_{t+k-2}, \dots, X_t$  pada lag ke k, sehingga fungsi autokorelasi parsial dapat dihitung berdasarkan rumus sebagai berikut:

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j} \quad (2.23)$$

$$\hat{\phi}_{kk} = \frac{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{bmatrix}} \quad (2.24)$$

Dengan garis batas pada PACF =  $2\sqrt{\text{Var}(\hat{\phi}_{kk})}$

Dimana:

$j = 1, 2, \dots, k$  dengan nilai  $\hat{\phi}_{11} = \hat{\rho}_1$

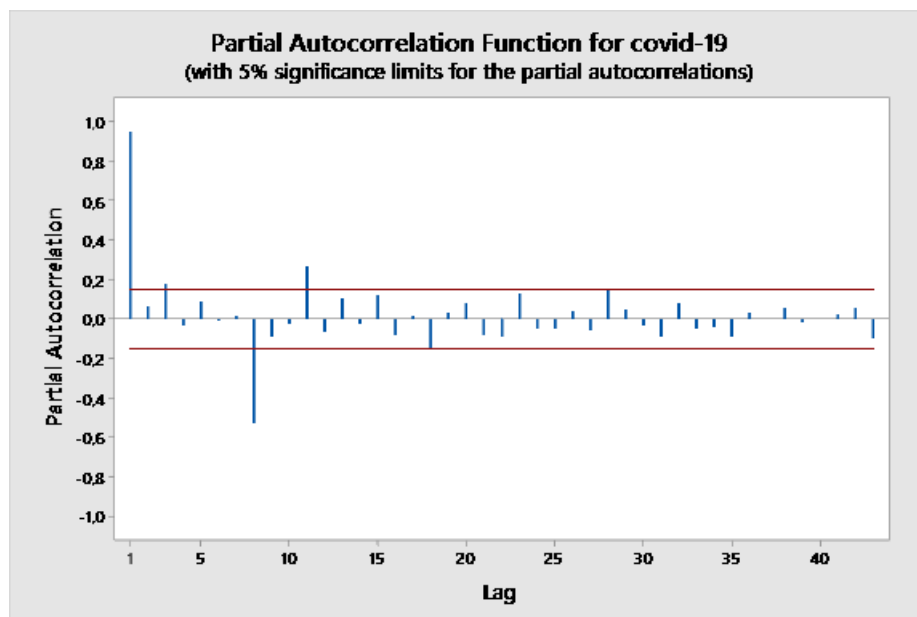
$\hat{\phi}_{k-1,j}$  = merupakan fungsi autokorelasi parsial pada lag ke  $k + 1$  dengan  $j$

$\hat{\phi}_{k+1,k+1}$  = merupakan fungsi autokorelasi parsial pada lag ke  $k + 1$  dengan  $k + 1$

$\hat{\rho}_k$  = merupakan fungsi autokorelasi pada lag ke  $k + 1$

UIN SUNAN AMPEL  
S U R A B A Y A

Untuk mempermudah dalam memahami, dapat dilihat contoh plot PACF pada gambar dibawah ini berdasarkan plot yang ada pada penelitian ini:



Gambar 2.3 Grafik Positif Covid-19 di Jawa Timur

Penentuan variabel input pada data yaitu dengan melihat lag-lag yang keluar dari data atau selang kepercayaan (garis yang berwarna merah). Lag-lag yang keluar tersebut dapat dikatakan lag yang signifikan pada plot PACF dari data covid-19 di Jawa Timur. Sehingga untuk penentuan inputan dapat dilihat lag-lag mana saja yang keluar dari selang kepercayaan, kemudian lag-lag tersebut akan diproses berdasarkan rumus yang sudah ditentukan (Ispriyanti et al., 2020).

#### 2.4. Algoritma *Grid Search*

Algoritma merupakan suatu metode yang efektif guna menyelesaikan masalah dimana masalah yang akan diselesaikan dituntut secara sistematis, terstruktur dan logis. Proses dalam pencarian *hyperparameter* terbaik biasa disebut dengan *grid search*. *Grid search* merupakan metode optimasi parameter yang dalam penggunaannya dikombinasikan menggunakan fungsi *k-fold validation* sehingga bisa memprediksi data yang tidak diketahui secara akurat (Sulistiana, 2020). Optimasi sendiri merupakan salah satu disiplin ilmu dalam matematika

yang fokus untuk mendapatkan nilai minimum atau maksimum secara sistematis dari suatu fungsi, peluang, maupun pencarian nilai lainya dalam berbagai kasus (Ardyan et al., 2017).

Algoritma *grid search* akan mengambil model atau objek yang ingin dilatih. Kemudian setiap kemungkinan nilai parameter yang tidak sinkron akan dicoba dan selanjutnya dari setiap model nilai parameter akan dihitung nilai galat atau erornya, sehingga peneliti dapat menentukan nilai parameter yang terbaik dengan melihat nilai galat terendah yang akhirnya dapat digunakan untuk model SVR. Dalam proses pengoptimalan pada algoritma *grid search*, akan membagi rentang parameter ke dalam grid dan disemua titik agar mendapatkan parameter yang optimal. Sehingga dalam perhitunganya, metode ini membutuhkan waktu yang bergantung pada parameter awal pada proses iterasi (Arimuko et al., 2019). Penerapan dalam algoritma *grid search* harus didampingi oleh sejumlah metrik kineja, yang biasanya dapat diukur dengan menggunakan *cross validation* pada data pelatihan (Fahrusyiana, 2019).

## 2.5. K-Fold Validation



*K-fold cross validation* merupakan teknik yang bisa digunakan jika mempunyai jumlah data yang terbatas (jumlah kelas tidak banyak). *K-fold cross validation* adalah suatu metode yang dipakai untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan iterasi atau perulangan pada atribut masukan yang diacak sehingga didapatkan sistem yang teruji untuk beberapa atribut input yang acak (Pitria, 2019). Uji *Cross-Validation* (CV) adalah suatu metode statistika yang dapat digunakan untuk mengevaluasi kinerja pada algoritma dimana data akan dipisahkan sebagai dua subset yaitu data proses *training* dan data *testing* (Widjaya et al., 2017). Data yang sudah melalui proses



SVR akan menghasilkan *support vector* pada setiap data, kemudian vektor terbaik dari setiap data akan diberikan kelas atau label pada proses SVR, setelah itu data yang telah diberikan label atau kelas akan diuji menggunakan *k-fold validation* (Lubis et al., 2020).

*K-fold cross validation* merupakan salah satu jenis CV yang berdasarkan pada ukuran dataset. Salah satu contoh *k-fold cross-validation* yang banyak digunakan ialah 10 fold *cross-validation* dikarenakan ada sebuah penelitian yang menunjukkan bahwa 10 fold *cross-validation* dapat memberikan perkiraan akurasi yang kurang bias dibandingkan menggunakan model lainnya. Berikut merupakan contoh pembagian dataset pada proses 10-fold *cross validation* sebagai berikut (Hulu, 2020):

	Dataset dibagi menjadi 10 bagian secara <i>random</i> (acak)										Akurasi	
	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%		
Percobaan 1	10%											a1
Percobaan 2		10%										a2
Percobaan 3			10%									a3
Percobaan 4				10%								a4
Percobaan 5					10%							a5
Percobaan 6						10%						a6
Percobaan 7							10%					a7
Percobaan 8								10%				a8
Percobaan 9									10%			a9
Percobaan 10										10%		a10

Keterangan gambar:  
 = data testing  
 = data training

Gambar 2.4 Konsep *K-fold Validation*

Sumber: (Rachman, 2018)

Kinerja dari *K-fold validation* yaitu sebagai berikut:

1. Total data dibagi menjadi N bagian
2. Fold ke-1 pada bagian ke-1 merupakan data uji atau *testing* dan sisanya

merupakan data latih. Kemudian, berdasarkan bagian data tersebut yaitu menghitung nilai akurasi pada kesamaan atau perkiraan hasil pengukuran dengan data sebenarnya berdasarkan banyaknya data. Berikut adalah rumus yang digunakan untuk menghitung nilai akurasi:

$$\text{Akurasi} = \frac{\sum \text{Data Uji Benar Klasifikasi}}{\sum \text{Total Data Uji}} \times 100\% \quad (2.25)$$

3. Fold ke-2 pada bagian ke-2 merupakan data uji atau *testing* dan sisanya merupakan data latih. Kemudian menghitung nilai akurasi berdasarkan data sebenarnya.
4. Demikian seterusnya proses tersebut sampai mencapai fold ke-k. Setelah itu, menghitung rata-rata akurasi dari k buah akurasi di atas. Kemudian rata-rata akurasi tersebut menjadi akurasi terakhir.

## 2.6. Tingkat Akurasi

Pada dasarnya prakiraan merupakan suatu prediksi atau dugaan akan terjadinya suatu peristiwa dimasa yang akan mendatang. Prakiraan dapat disebut dengan peramalan secara ilmiah dengan menggunakan beberapa metode. Dengan istilah lain, peramalan bisa diartikan sebagai proses untuk memperkirakan kebutuhan pada masa yang akan tiba dengan menggunakan data pada masa lampau yang bisa memberikan *output* peramalan yang dipercaya ketetapannya dengan menggunakan metode-metode peramalan yang bertujuan untuk meminimalisir resiko pada kesalahan (Sihite, 2020).

Peramalan umumnya dibagi sebagai tiga kelompok bagian yaitu jangka pendek, jangka menengah, dan jangka panjang. Peramalan jangka pendek biasanya

di manfaatkan untuk memprediksi peristiwa dengan memakai periode hari, minggu, sampai bulan ke depan. Peramalan jangka menengah merupakan pendekatan peramalan dengan memanfaatkan data waktu dari satu tahun hingga dua tahun ke depan. Sedangkan peramalan jangka panjang dapat dilakukan untuk mengetahui peristiwa lebih dari dua tahun kedepan. Pada peramalan, data yang digunakan berupa data *time series* atau metode deret waktu yang berdasarkan data historis pada masa lampau yang menjadikan keluaran berupa prediksi tentang peristiwa di masa yang akan datang (Ahmad, 2020).

Setelah melakukan *forecasting* atau peramalan, untuk langkah selanjutnya yaitu dengan cara mengukur tingkat kesalahan peramalan (*forecast error*). Hasil prediksi yang akurat ialah prediksi yang mampu meminimalkan kesalahan pada peramalan. Pada umumnya, kesalahan dari peramalan (*forecast error*) dapat dihitung menggunakan data sebenarnya yang dikurangi dengan besarnya ramalan.

1. *Root Mean square error* ialah kesalahan rata-rata akurasi dari nilai MSE yang dikuadratkan (Maulana et al., 2019). *Root Mean square error* (RMSE) adalah cara lain untuk mengevaluasi metode peramalan dimana setiap kesalahan atau sisa dikuadratkan dan kemudian dijumlahkan dengan jumlah pada observasi (Wahyuni, 2019). Nilai RMSE dapat dihitung dengan persamaan sebagai berikut:

$$\text{RMSE} = \sqrt{\sum_{t=1}^n \frac{(X_t - Y_t)^2}{n}} \quad (2.26)$$

Keterangan:

$n$  = Banyaknya data

$t$  = Periode peramalan

$X_t$  = Data aktual pada periode  $t$

$Y_t$  = Nilai peramalan pada periode  $t$

2. *Mean Absolute Percentage Error* (MAPE) merupakan rata-rata persentase kesalahan mutlak yang dihitung dengan cara mencari nilai kesalahan mutlak disetiap periode, kemudian dibagi dengan nilai pengamatan aktual dan presentase kesalahan mutlak. MAPE memiliki dua keuntungan, yakni kesalahan positif dan negatif dari perhitungan akan dijaga nilai absolutnya, serta dikarenakan kesalahan relatif tidak bergantung pada skala variabel dependen, sehingga pada pengukuran ini memungkinkan peneliti membandingkan akurasi perkiraan antara data *time series* berbeda skala (Caraka et al., 2017). Nilai MAPE bisa dihitung menggunakan persamaan menjadi berikut:

$$\text{MAPE} = \left( \frac{100\%}{n} \right) \sum_{t=1}^n \frac{|X_t - Y_t|}{X_t} \quad (2.27)$$

Keterangan:

$n$  = Banyaknya data

$t$  = Periode peramalan

$X_t$  = Data aktual pada periode  $t$

$Y_t$  = Nilai peramalan pada periode  $t$

Pada peramalan menggunakan MAPE, jika mempunyai nilai kurang dari 10% maka bisa dikategorikan kemampuan pada peramalannya sangat baik. Tetapi jika diketahui nilai MAPE lebih dari 50% maka dikategorikan kemampuan pada peramalan yang tidak baik (Ningtyas, 2018).

**Tabel 2.1 Akurasi Nilai MAPE**

Nilai MAPE	Akurasi Prediksi
$MAPE \leq 10\%$	Tinggi
$10\% < MAPE \leq 20\%$	Baik
$20\% < MAPE \leq 50\%$	Masih baik digunakan
$MAPE > 50\%$	Rendah

## 2.7. Integrasi Keilmuan

Hadirnya virus *corona* menjadikan terganggunya tatanan kehidupan manusia di dunia pendidikan, ekonomi, sosial budaya, dan keberagaman. Jika dilihat dari sisi keberagaman, pandemi telah mengganggu berbagai kegiatan ibadah seperti sholat berjamaah atau ritual mingguan umat kristen dikarenakan banyak masjid dan gereja atau tempat ibadah lainnya yang tutup serta beberapa tempat ibadah yang membatasi dan menerapkan standar protokol kesehatan covid-19. Selain itu dalam segi ekonomi dan sosial budaya, adanya pandemi ini menjadikan guncangan akibat kebiasaan masyarakat Indonesia yang memiliki tradisi dan kebiasaan yang melibatkan banyak orang dikarenakan manusia tercipta sebagai makhluk sosial, dimana mereka saling membutuhkan antara satu sama lain, seperti saling berinteraksi untuk memenuhi kebutuhan sosial tengah menjadi problematika untuk saat ini.

Maka untuk mencegah penularan virus covid-19 ini diperlukan beberapa

tindakan sehingga angka penyebaran virus ini dapat diturunkan. Maka dari itu diperlukan kesadaran yang tinggi dari masyarakat untuk ikut serta membantu pemerintah dalam menekan angka positif covid-19. Hal yang dapat dilakukan yaitu sebagai berikut:

#### 1. Kebersihan Diri dan Lingkungan

Secara definisi, menjaga kebersihan diri merupakan perawatan diri yang dilakukan untuk menjaga Kesehatan fisik dan rohani. Sehingga dapat dilihat bahwa kebersihan diri tidak hanya bisa berimbas dalam hal-hal fisik tetapi juga hal-hal yg bersifat mental. Dalam masa pandemi seperti saat ini, kesehatan mental juga perlu dijaga supaya bisa melewati masa pandemi dengan baik (Jiwandono et al., 2020).

Kebersihan fisik ialah keadaan yang terbebas dari kotoran, termasuk diantaranya yaitu sampah, debu, dan bau. Pada zaman modern, proses penularan penyakit yang diakibatkan dari mikroba dapat ditemukan oleh Louis Pasteur. Dengan menjaga kebersihan juga dapat terhindar dan terbebas dari virus atau bakteri dan bahan kimia yang berbahaya. Kebersihan merupakan suatu tanda bahwa keadaan saat itu dalam keadaan yang baik atau *higiene*. Manusia sangat dianjurkan menjaga kebersihan diri dan kebersihan lingkungan supaya terhindar dari berbagai virus atau kuman yang dapat membahayakan diri sendiri maupun orang lain. (Iskandar, 2018).

Kebersihan juga terdapat dalam sebuah hadist, yaitu:

إِنَّ اللَّهَ طَيِّبٌ يُحِبُّ الطَّيِّبَ نَظِيفٌ يُحِبُّ النَّظَافَةَ كَرِيمٌ يُحِبُّ الْكَرَمَ جَوَادٌ يُحِبُّ الْجُودَ  
فَنَظِّفُوا أَرَاهُ قَالَ أَفْنَيْتَكُمْ وَلَا تَشَبَّهُوا بِالْيَهُودِ

Artinya: ”*Sesungguhnya Allah Ta’ala itu baik (dan) menyukai kebaikan, bersih (dan) menyukai kebersihan, mulia (dan) menyukai kemuliaan, bagus (dan) menyukai kebagusan. Oleh sebab itu, bersihkanlah lingkunganmu.*” (Hadits Jami’ At-Tirmidzi No.2723-Kitab Adab)

Kebersihan lingkungan merupakan kebersihan yang ada di lingkungan masyarakat seperti kebersihan pada tempat tinggal, tempat bekerja, dan berbagai tempat umum. Kebersihan di tempat tinggal dapat dilakukan dengan cara membersihkan setiap ventilasi dan segala perabotan yang ada di tempat tinggal, seperti dengan menyapu lantai kemudian mengepelnya menggunakan pembersih lantai yang mengandung disinfektan serta mencuci alat-alat yang masak yang telah digunakan. Selain itu, sering membersihkan kamar mandi dan toilet sekurangnya setiap satu Minggu sekali. Kebersihan lingkungan dapat dimulai dengan menjaga kebersihan halaman dan selokan, membuang sampah pada tempatnya, serta membersihkan jalanan dari sampah (Nur, 2020).

## 2. Penggunaan Masker

Salah satu cara agar melindungi diri dari penyebaran dan penularan covid-19 ialah dengan menggunakan Alat Pelindung Diri (APD). Beberapa jenis APD yang diwajibkan dan disarankan guna mencegah penularan covid-19 yaitu dengan menggunakan masker (*mask*) dan pelindung wajah (*face shield*). Masker adalah suatu alat yang dapat berfungsi untuk melindungi pengguna dari partikel sekitar yang berbahaya seperti droplet dan airborne yang mana dapat masuk melalui mulut dan hidung. Droplet atau percikan air ludah biasanya terjadi saat batuk dan bersin. Selain terhirup secara langsung droplet juga dapat menempel di permukaan suatu benda. Sehingga ketika

seseorang menyentuh benda tersebut peluang untuk terkontaminasi virus yang ada di droplet besar apabila orang tersebut menyentuh mulut atau area hidungnya.

Meskipun begitu di mata agama, bersin merupakan sebuah nikmat dari Allah SWT untuk manusia supaya agar terhindar dari bibit penyakit. Menahan bersin juga dapat berbahaya bagi tubuh dikarenakan fungsi bersin yaitu membuang atau mengeluarkan benda asing yang dapat menyebabkan gangguan pada hidung dan tenggorokan. Benda asing tersebut bisa berupa virus, debu, bakteri, dan gas kimia. Dikarenakan hasil dari bersin adalah sekumpulan bibit penyakit, maka banyak cara dan tuntunan adab ketika bersin. Salah satunya yaitu dengan cara merendahkan suara dan menutup mulut dengan pakaian atau tangan ketika saat sedang bersin. Hal ini dikuatkan dengan hadist seperti berikut:

أَنَّ النَّبِيَّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ كَانَ إِذَا عَطَسَ غَطَّى وَجْهَهُ بِيَدِهِ أَوْ بِثَوْبِهِ وَعَضَّ بِهَا  
صَوْتَهُ قَالَ أَبُو عِيسَى هَذَا حَدِيثٌ حَسَنٌ صَحِيحٌ

UIN SUNAN AMPEL  
S U R A B A Y A

Artinya : *"Bahwasanya apabila Nabi Muhammad SAW sedang bersin, beliau menutup wajah dengan tangan atau dengan kain-Nya sambil merendahkan suaranya."* (Hadits Jami' At-Tirmidzi No. 2669-Kitab Adab)

Dari hadist diatas dijelaskan bahwasanya ketika bersin, seharusnya kita senantiasa menutup mulut dan hidung, seperti yang telah dilakukan oleh Rasulullah. Hal ini dapat menjadikan lingkungan sekitar kita bersih dari



virus atau bakteri yang telah kita keluarkan, sehingga kesehatan orang lain juga ikut terjaga. Seperti memakai masker adalah salah satu contoh penerapan yang efektif untuk menjaga kesehatan kita maupun kesehatan orang lain yang sedang berada di sekitar kita (Ahsan, 2009).

Dalam bidang kesehatan, masker mempunyai fungsi secara generik untuk mencegah kontaminasi virus maupun penyakit lain serta bertujuan agar dapat terhindar dari paparan virus yang akan masuk ke tubuh ataupun virus dari tubuh yang dapat menularkannya ke orang lain. Pada pemakaian sehari-hari, masker biasanya dipakai untuk melindungi wajah dari paparan debu atau polusi udara waktu berada pada luar ruangan (Alam, 2020).

### 3. Menjaga Jarak (*social distancing*)

Aktivitas di luar rumah adalah aktivitas yang biasanya dilakukan manusia entah dalam hal memenuhi kebutuhan hidup seperti bekerja, sekolah, atau bahkan hanya sekadar jalan-jalan. Usia yang sangat rentan terpapar virus covid-19 adalah pada usia dewasa, dikarenakan pada usia ini, rata-rata banyak yang melakukan aktivitas di luar rumah seperti halnya bekerja, bersekolah, ataupun aktivitas lainnya serta kurangnya kesadaran dalam mematuhi protokol kesehatan, salah satunya menjaga jarak.

Hal ini perlu menjadi perhatian, karena semakin meningkat frekuensi kegiatan yang dilakukan di luar rumah (tempat keramaian) apalagi dengan tidak menjaga jarak maka resiko untuk terpapar covid-19 akan semakin tinggi (Nurhayatun et al., 2021). Selain itu menjaga jarak antar individu satu dengan individu lainnya dapat meminimalkan resiko dari penularan droplet secara langsung. Sehingga menjaga jarak dalam situasi covid-19 merupakan perbuatan yang baik, dikarenakan dapat menjaga diri dan orang lain dari

virus yang dapat menyebabkan timbulnya berbagai penyakit. Jarak antar individu yang disarankan yaitu sekitar 2 meter atau lebih. Hal ini juga terdapat pada surah al Baqarah ayat 195 yang berbunyi sebagai berikut:

وَأَنْفِقُوا فِي سَبِيلِ اللَّهِ وَلَا تُلْقُوا بِأَيْدِيكُمْ إِلَى التَّهْلُكَةِ وَأَحْسِنُوا إِنَّ اللَّهَ يُحِبُّ الْمُحْسِنِينَ

Artinya : ”Berinfaklah di jalan Allah, janganlah jerumuskan dirimu ke dalam kebinasaan, dan berbuatbaiklah. Sesungguhnya Allah menyukai orang-orang yang berbuat baik.” (QS: Al-Baqarah: 195)

Pesan moral yang terkandung dalam ayat diatas yaitu bahwa hendaknya kita melakukan suatu tindakan seperti tidak mendekati sesuatu yang dapat mengakibatkan kebinasaan sehingga dapat terjamin keselamatan dari hal membahayakan yang dapat menyebabkan kematian. Dalam situasi sulit seperti pandemi covid-19 ini, ayat diatas bisa dijadikan legitimasi atas larangan untuk mengadakan pertemuan dan program yang dihadiri oleh orang banyak, misal konferensi, seminar, atau rapat, mengunjungi orang yang sedang sakit, melakukan kontak fisik secara langsung seperti berjabat tangan dengan orang yang terpapar covid-19, atau melakukan sesuatu yang dapat mengakibatkan terjangkit virus yang mengerikan ini. Disamping larangan yang ditegakkan, terdapat pula anjuran yang dapat dilakukan untuk meminimalisir penyebaran dari virus covid-19 ini, seperti anjuran melakukan *social distancing*, karantina, dan isolasi mandiri.

#### 4. Selalu Berfikiran Positif dan Tidak Cemas

Kondisi pandemi dapat memunculkan kekawatiran pada masyarakat, dikarenakan jumlah pasien yang terjangkit covid-19 dan angka kematian

yang terus bertambah. Hal ini dapat mempengaruhi kesehatan mental masyarakat sehingga akan menimbulkan rasa panik dan rasa takut. Banyak masyarakat yang mengalami kecemasan jika akan beraktifitas di luar rumah serta bertemu atau berinteraksi dengan orang lain. Rasa kecemasan ini muncul karena reaksi normal dari stres akibat situasi covid-19 yang mengancam. Berfikiran tidak positif atau selalu berfikir negatif dapat menjadikan hubungan antar individu jadi kurang baik dan harmonis.

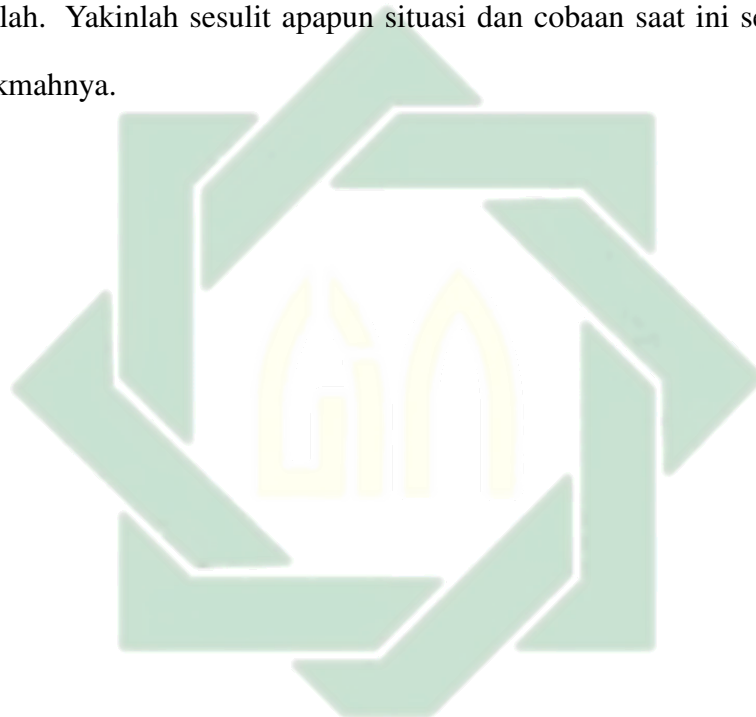
Dalam situasi seperti ini kecemasan dan berpikiran negatif seharusnya mampu dikelola dengan baik agar tidak menimbulkan kepanikan yang berlebihan atau bahkan sampai mengancam kesehatan kejiwaan seseorang. Selain dapat mengancam gangguan kejiwaan, kecemasan juga dapat mengakibatkan penurunan pada sistem imun di dalam tubuh, sehingga mudah bagi tubuh untuk lelah dan terserang berbagai penyakit. (Vibriyanti, 2020). Jadi dalam hal ini, berfikir negatif terhadap sesuatu yang dirasa mengancam dan merugikan sehingga dapat menimbulkan rasa kecemasan, perlu diubah dengan berfikiran positif seperti sabar dan bersyukur atas situasi dampak covid-19 dan selalu mengambil sisi positif dari setiap kejadian dan rencana Allah. Seperti halnya pakar ilmu kedokteran era emas kejayaan islam yaitu Ibnu Sina mengatakan:

أَلْوَهُمْ نِصْفُ الدَّاءِ، وَالْإِطْمِئْنَانُ نِصْفُ الدَّوَاءِ، وَالصَّبْرُ بَدَايَةُ الشِّفَاءِ

Artinya: "Kepanikan adalah separuh penyakit, ketenangan adalah separuh obat, dan kesabaran adalah permulaan kesembuhan."

Cobaan terbesar saat ini adalah bagaimana cara kita agar dapat mengontrol

tindakan serta selalu berfikir positif dan tenang di tengah pandemi covid-19 yang sulit ini. Sumber ketenangan adalah kedekatan kita kepada Allah. Hanya dengan mengingat Allah hati menjadi tenang. Cara mengingat tentu dalam arti yang luas, mematuhi perintah-Nya dan menjauhi larangan-Nya. Seberat apapun masalah, lebih baik kita kembali kepada Allah. Yakinlah sesulit apapun situasi dan cobaan saat ini semua akan ada hikmahnya.



**UIN SUNAN AMPEL**  
**S U R A B A Y A**

## BAB III

### METODE PENELITIAN

#### 3.1. Jenis Penelitian

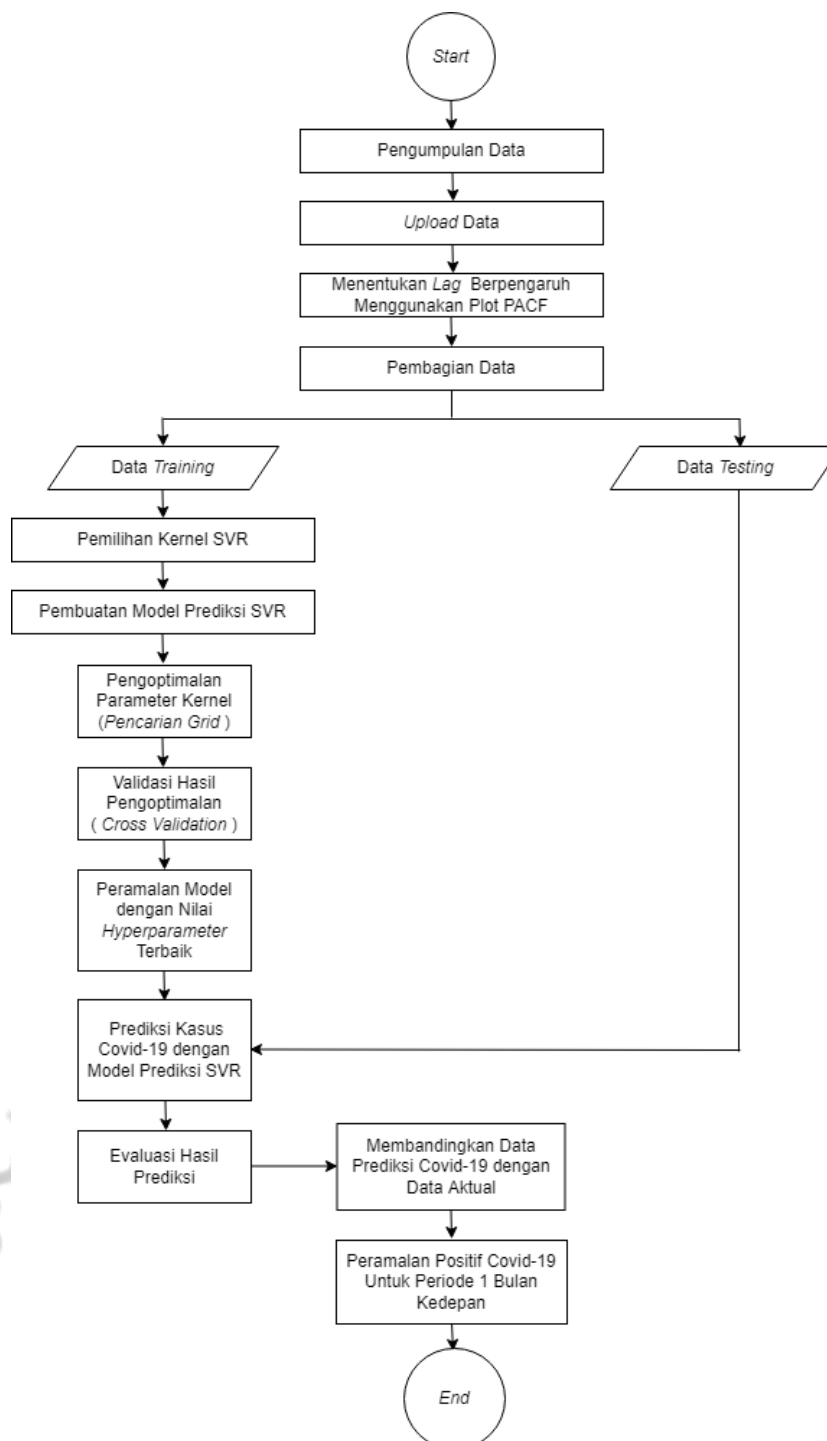
Dalam penelitian ini penulis memakai metode pendekatan penelitian secara kuantitatif, dimana jenis penelitian yang digunakan termasuk kedalam penelitian terapan (*applied research*) yang bertujuan agar permasalahan yang telah diidentifikasi dapat mendapatkan solusi dengan menggunakan pengetahuan ilmiah dan proses dalam menyelidikannya yang sistematis.

#### 3.2. Jenis dan Sumber Data

Sumber data kuantitatif pada penelitian ini menggunakan data sekunder tentang penyebaran positif covid-19 di Jawa Timur dari 10 Februari 2022 sampai dengan 31 Juli 2022. Data ini diperoleh melalui akses web dari akun resmi pemerintahan provinsi Jawa Timur yang bisa diakses secara umum di alamat situs ([Pemerintah Provinsi Jawa Timur, 2022](#)) Data jumlah positif covid-19 tersebut berupa data historis yang berdasarkan rumus *time series*.

#### 3.3. Teknik Analisis Data

Pada tahap *processing*, data terlebih dahulu dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Presentase pembagian pada data *training* dan *testing* yaitu 60 : 40, 70 : 30, 80 : 20, dan 90 : 10.



**Gambar 3.1** Plot Tahapan Penelitian

Gambar (3.1) merupakan alur dalam proses analisis data yang akan dilakukan pada penelitian ini.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1. Analisis Deskriptif

Analisis deskriptif data ini digunakan untuk mengetahui karakteristik atau gambaran umum dari data positif covid-19 yang ada di wilayah Jawa Timur. Data yang digunakan dalam penelitian ini berupa data *time series* mengenai kasus positif covid-19 di wilayah Jawa Timur dari periode 10 Februari 2022 hingga pada periode 31 Juli 2022. Data *time series* merupakan jenis data yang dikumpulkan menurut urutan waktu dalam suatu rentang waktu tertentu. Berikut merupakan data *time series* dari kasus positif covid-19 di Jawa Timur (4.1):

UIN SUNAN AMPEL  
S U R A B A Y A

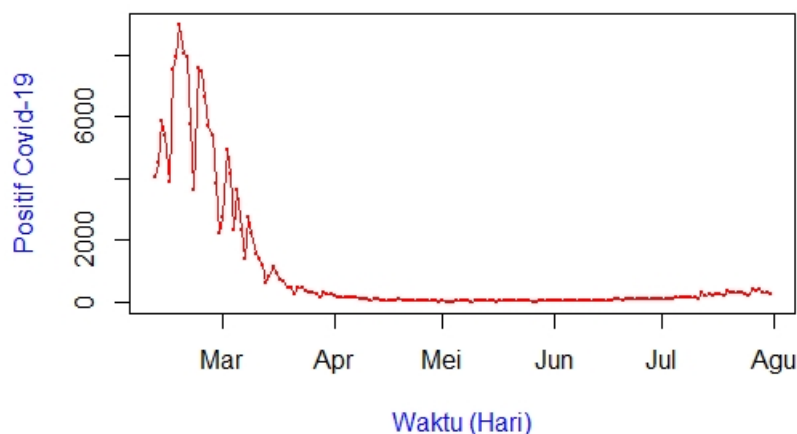
**Tabel 4.1 Aktual Positif Covid-19 di Jatim**

<b>Tanggal</b>	<b>Positif covid-19</b>
10-Februari-2022	4054
11-Februari-2022	4506
12-Februari-2022	5880
13-Februari-2022	5376
14-Februari-2022	3885
15-Februari-2022	7530
16-Februari-2022	7919
17-Februari-2022	8977
18-Februari-2022	8037
19-Februari-2022	7946
:	:
:	:
31-Juli-2022	258

Data pada penelitian ini (4.1) diambil dalam kurun hari mulai tanggal 10 Februari 2022 hingga tanggal 31 Juli 2022, sehingga banyak data pada penelitian ini yaitu 172 data. Untuk melihat pergerakan pada data covid-19 di Jawa Timur, maka ditampilkan plot data sebagai berikut:



**Grafik Kasus Positif Covid-19 di Jawa Timur**



**Gambar 4.1 Grafik Positif Covid-19 di Jawa Timur**

Berdasarkan informasi dari grafik (4.1), plot data covid-19 dari permulaan gelombang 3 pandemi covid-19 mewabah sampai dengan tanggal 12 Maret 2022, data cenderung mengalami kelonjoran yang mana menghasilkan rentang yang lumayan jauh. Kenaikan pada data dapat terjadi lantaran periode sebelumnya sampai dengan 12 Maret 2022, tidak sedikit masyarakat yang merasa bosan dengan adanya disiplin protokol kesehatan sehingga banyak masyarakat yang menjalani aktivitas dengan melanggar protokol kesehatan, salah satunya berkerumun saat mengadakan acara atau perayaan tanpa menggunakan masker dan menjaga jarak. Selain itu, kenaikan dapat terjadi karena adanya covid varian baru yang muncul mulai dari varian Alpha, Beta, Delta, dan omicron. Berbeda pada saat memasuki tanggal 19 Maret 2022 data tersebut mulai mengalami kestabilan, dimana rentang data saat kenaikan tidak terlalu jauh.

Setelah itu dilakukan deskripsi data untuk mengetahui karakteristik dari data positif covid-19 di Jawa Timur. Berikut merupakan tabel deskriptif dari data positif covid-19 di wilayah Jawa Timur (4.2):

Tabel 4.2 Deskriptif Positif Covid-19 di Jatim

Variabel	Positif covid-19
Jumlah	172
Mean	955
Median	113
Std. Dev	1366,399
Min	8
Max	8977

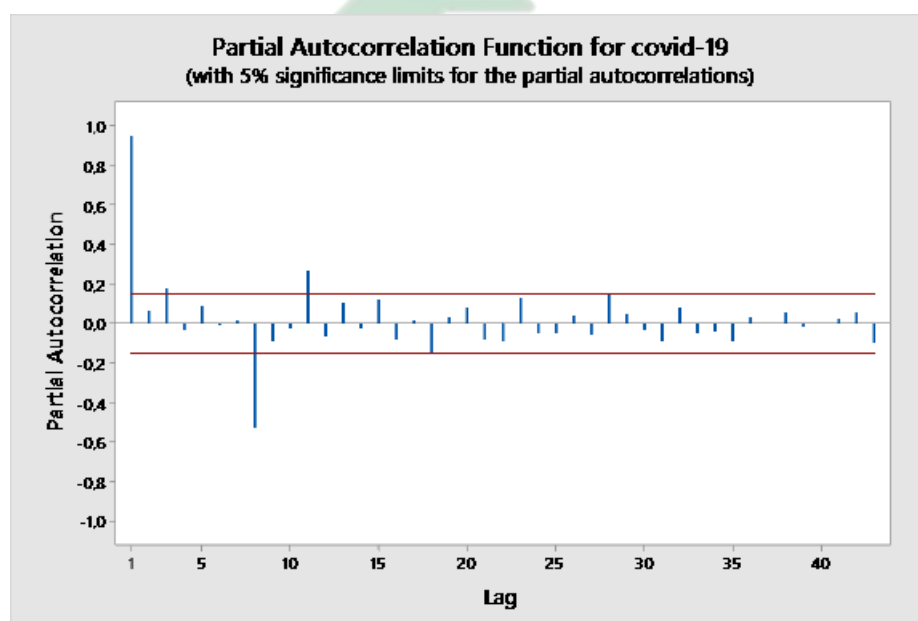
Dapat dilihat pada tabel (4.2), bahwa dengan jumlah data covid-19 sebanyak 172 data didapatkan bahwa rata-rata warga yang dinyatakan positif covid-19 di Jawa Timur sebanyak 955. Median dari data tersebut diperoleh sebesar 113, kemudian didapatkan nilai standart deviasi atau jarak antara data terhadap nilai rata-ratanya sebesar 1366,399. Setelah itu untuk nilai minimum dan maksimum pada data covid-19 diperoleh sebesar 8 dan 8977.

#### 4.2. Processing Data

Pada processing data ini dapat dilakukan dengan tujuan agar model yang dihasilkan dari perhitungan SVR dapat menghasilkan tingkat akurasi yang baik. Tahapan pengolahan data pada penelitian ini menggunakan beberapa proses seperti menentukan lag yang berpengaruh menggunakan plot PACF yang akan digunakan untuk variabel bebas pada data, kemudian dilakukan pembagian data kasus positif covid menjadi data *training* dan data *testing*, dan perencanaan data aktual sebagai peramalan masa mendatang.

### 4.3. Menentukan Lag Berpengaruh Menggunakan Plot PACF

Sebelum dilakukan proses analisis menggunakan metode SVR, langkah pertama yang dilakukan yaitu dengan mengkonversi data covid-19 ke dalam bentuk *time lag* yang nantinya akan dijadikan sebagai data *input* dalam model SVR. Penentuan banyaknya *time lag* ini berdasarkan pada plot PACF. Berikut merupakan hasil plot PACF pada data covid-19 di wilayah Jawa Timur (4.4):



Gambar 4.2 Grafik Positif Covid-19 di Jawa Timur

Pada gambar (4.4) plot PACF akan digunakan untuk menentukan model *Autoregressive* (AR) yang mengansumsikan bahwa data saat ini dipengaruhi oleh data sebelumnya, sesuai dengan data *time series*. Jumlah *node input* yang akan digunakan yaitu berdasarkan lag-lag yang signifikan pada plot PACF. Dapat dilihat pada gambar diatas, bahwa lag yang signifikan yaitu berupa garis yang keluar dari selang kepercayaan (garis lurus berwarna merah). Pada data positif covid-19 di Jawa Timur lag yang signifikan berdasarkan plot PACF yaitu pada lag 1,3,8, dan 11, sehingga pada penelitian ini variabel prediktor (X) yang digunakan sebagai

inputan adalah  $X_{t-1}$ ,  $X_{t-3}$ ,  $X_{t-8}$ , dan  $X_{t-11}$ .

#### 4.3.1. Penentuan Variabel

Pada penelitian ini, data kasus positif covid-19 di Jawa Timur akan dibagi menjadi 5 variabel, yaitu variabel X sebanyak 4 dan variabel Y sebanyak 1. Hal ini ditentukan karena data lag dari atribut yang berpengaruh adalah  $X_{t-1}$ ,  $X_{t-3}$ ,  $X_{t-8}$ , dan  $X_{t-11}$ . Dimisalkan  $X_t$  merupakan nilai *target value*, sehingga hasil pemodelan dapat ditunjukkan pada persamaan (2.24), sehingga:

$$X = \begin{bmatrix} X_{t-11} & X_{t-8} & X_{t-3} & X_{t-1} \\ X_{t-10} & X_{t-9} & X_{t-2} & X_t \\ \vdots & \vdots & \vdots & \vdots \\ X_{t-11+n} & X_{t-8+n} & X_{t-3+n} & X_{t-1+n} \end{bmatrix}, Y = \begin{bmatrix} X_t \\ X_{t+1} \\ \vdots \\ X_n \end{bmatrix} \quad (4.1)$$

Dari persamaan tersebut, dijelaskan bahwasanya cara penentuan lag 11 didapatkan dengan cara penurunan dari data pertama ke data ke data 11. Seperti contoh:

UIN SUNAN AMPEL  
S U R A B A Y A

Covid-19	Variabel
4054	X1
4506	X2
5880	X3
5376	X4
3885	X5
7530	X6
7919	X7
8977	X8
8037	X9
7946	10
5766	X11
3621	X12
7569	X13
7486	X14
6643	X15
5729	X16
5413	X17
3861	X18
2233	X19
2761	X20
:	:
:	:
258	X172

Gambar 4.3 Tabel Inputan Pada Lag 11

Dikarenakan banyaknya data pada penelitian ini yaitu sebanyak 172 data, sehingga pada lag 11 data inputanya yaitu  $172 - 11 = 161$  data. Penentuan inputan pada lag 11 dimulai dari data ke-12 sampai dengan data ke-172. Sehingga inputan yang dihasilkan sebanyak 161 data. Setelah itu untuk inputan pada lag 8 yaitu

dilakukanya penurunan data sampai data ke-8. Seperti contoh sebagai berikut:

Covid-19	Variabel
4054	X1
4506	X2
5880	X3
5376	X4
3885	X5
7530	X6
7919	X7
8977	X8
8037	X9
7946	10
5766	X11
3621	X12
7569	X13
7486	X14
6643	X15
5729	X16
5413	X17
3861	X18
2233	X19
2761	X20
:	:
:	:
258	X172

Gambar 4.4 Tabel Inputan Pada Lag 8

Setelah itu untuk inputan pada lag 8 yaitu  $172 - 8 = 164$  data. Sehingga inputan pada lag 8 bisa dimulai dari data ke-9 sampai data ke-172. Sehingga banyak inputan pada lag 8 yaitu sebanyak 164. Pada inputan lag 3 dan lag 1 cara

menentukan banyak inputan data yaitu sama seperti penentuan pada lag 11 dan lag 8. Sehingga banyak data pada inputan lag 3 yaitu  $172 - 3 = 169$  data, yaitu dimulai dari data ke-4 sampai data ke-172. Setelah itu untuk inputan lag 1 yaitu  $172 - 1 = 171$  data, sehingga pada lag 1 dimulai dari data ke-2 sampai ke-172. Sehingga didapatkan data pada lag yang keluar yaitu sebagai berikut (4.5):



Lag 1	Lag 3	Lag 8	Lag 11	Covid-19	Variabel
				4054	X1
4054				4506	X2
4506				5880	X3
5880	4054			5376	X4
5376	4506			3885	X5
3885	5880			7530	X6
7530	5376			7919	X7
7919	3885			8977	X8
8977	7530	4054		8037	X9
8037	7919	4506		7946	X10
7946	8977	5880		5766	X11
5766	8037	5376	4054	3621	X12
3621	7946	3885	4506	7569	X13
7569	5766	7530	5880	7486	X14
7486	3621	7919	5376	6643	X15
:	:	:	:	:	:
:	:	:	:	:	:
389	412	310	181	324	X170
324	364	286	344	316	X171
316	389	285	327	258	X172
258	324	272	310		X173
	316	220	286		X174
	258	412	285		X175
		364	272		X176
		389	220		X177
		324	412		X178
		316	364		X179
		258	389		X180
			324		X181
			316		X182
			258		X183

Gambar 4.5 Tabel Penentuan Lag

Setelah data pada lag yang keluar berdasarkan plot PACF sudah didapatkan, maka langkah selanjutnya yaitu mengambil data, dimana data tersebut mempunyai nilai dengan matriks yang sama. Dapat dilihat pada tabel berwarna



kuning pada tabel (4.5) merupakan data pada semua lag yang mempunyai matriks yang sama. Sehingga semua inputan dimulai dari data ke-12 sampai dengan data ke-172. Banyak data yang sudah dihitung berdasarkan nilai lag yang keluar pada plot PACF yaitu sebanyak 161 data. Dikarenakan lag yang keluar paling tinggi adalah lag ke 11, sehingga *output* atau  $X_t$  dimulai pada data ke-12 sampai data ke-172. Dari rumus (4.1) didapatkan variabel pada penelitian ini yaitu:

**Tabel 4.3 Penentuan Variabel**

$X_{t-11}$	$X_{t-8}$	$X_{t-3}$	$X_{t-1}$	$X_t$
4056	5376	8037	5766	3621
4506	3885	7946	3621	7569
5880	7530	5766	7569	7486
5376	7919	3621	7486	6643
3885	8977	7569	6643	5729
7530	8037	7486	5729	5413
7919	7946	6643	5413	3861
8977	5766	5729	3861	2233
8037	3621	5413	2233	2761
7946	7569	3861	2761	4911
5766	7486	2233	4911	4135
3621	6643	2761	4135	2331
7569	5729	4911	2331	3651
:	:	:	:	:
:	:	:	:	:
327	285	389	316	258

Berdasarkan tabel (4.3), penentuan inputan pada penelitian ini yaitu berdasarkan plot pacf, dimana pada plot tersebut, lag yang signifikan yaitu pada lag 1,3,8, dan 11. Sehingga  $X_{t-11}$  sebagai variabel  $X_1$  yang artinya lag yang keluar pada data ke-11,  $X_{t-8}$  sebagai variabel  $X_2$  yaitu lag yang keluar pada data ke-8,  $X_{t-3}$  sebagai variabel  $X_3$  yaitu lag yang keluar pada data ke-3,  $X_{t-1}$  sebagai variabel  $X_4$  yaitu lag yang keluar pada data ke-4, dan  $X_t$  merupakan target atau bisa disebut variabel  $Y$ . Sehingga jumlah data setelah dilakukannya proses inputan (pembentukan variabel) berdasarkan lag yang keluar pada plot PACF yaitu menjadi 161 data.

#### 4.3.2. Pembagian Data *Training* dan Data *Testing*

Sebelum melakukan analisis dengan menggunakan *support vector regression*, maka data terlebih dahulu dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Presentase pembagian pada data *training* dan *testing* dalam penelitian ini yaitu sebanyak 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Pembuatan data *training* bertujuan untuk meningkatkan kinerja dari SVR terhadap data *testing* dalam upaya menentukan parameter terbaik dalam pembangunan model yang akan digunakan sebagai peramalan. Berikut merupakan pembagian data yang disajikan dalam tabel seperti dibawah ini:

Tabel 4.4 Pembagian Data *Training* dan *Testing*

Presentase		Tanggal		Data		Total
<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	
60%	40%	10 Februari 2022	17 Mei 2022	97	64	161
70%	30%	10 Februari 2022	2 Juni 2022	113	48	161
80%	20%	10 Februari 2022	18 Juni 2022	129	32	161
90%	10%	10 Februari 2022	4 Juli 2022	145	16	161

Berdasarkan tabel (4.4), data *training* memiliki jumlah presentase yang lebih besar dibandingkan data *testing* dikarenakan agar mesin pembelajaran lebih terlatih untuk mempelajari model. Hal tersebut dilakukan supaya mesin pembelajaran dapat membentuk model, setelah model didapatkan maka selanjutnya akan dilatih menggunakan data *testing* untuk memberikan prediksi pada data *testing* yang lebih optimal. Pembagian data *training* atau data latih maupun data *testing* dilakukan secara runtun. Selanjutnya data *training* akan dilatih dengan metode *support vector regression* sehingga terbentuk suatu model dengan kombinasi parameter yang terbaik. Setelah itu, data *testing* digunakan untuk menguji hasil dari model yang terbentuk dari pelatihan data *training*.

#### 4.4. Analisis Metode *Support Vector Regression*

Metode SVR merupakan penerapan dari SVM untuk kasus regresi, dimana SVR menggunakan ruang antara titik data sebagai margin kesalahan dan prediksi titik berikutnya yang paling mungkin dalam dataset. Pada data *time series*, metode ini dapat menjanjikan karena menggunakan fungsi risiko yang terdiri dari kesalahan empiris dan istilah rutinitas yang berasal dari struktur minimalisasi risiko prinsip jenis kernel. Selain itu, memiliki kemampuan yang sangat baik

dalam menggeneralisasi sampel kecil sehingga baik untuk memprediksi. Proses penentuan parameter merupakan bagian tersulit dalam tahapan perhitungan metode SVR karena terdapat beberapa parameter yang dikombinasikan. Pada kasus regresi parameter yang digunakan yaitu parameter *epsilon* dan parameter *cost*, dimana kedua parameter tersebut berpotensi mempengaruhi model akhir pada SVR yang terbentuk. Kemudian untuk menghindari *overfitting* pada data *training* ini akan digunakan *cross validation*.

#### 4.4.1. Membangun Model SVR Sebelum dan Sesudah Proses *Grid Search Optimization*

Pembangunan model peramalan SVR diawali dengan melatih model peramalan pada data training. Sebelum membangun model SVR di *software R*, *package* yang dibutuhkan yaitu *e1071*. Setelah *package e1071* sudah di *install*, maka langkah selanjutnya yaitu membangun model berdasarkan data *training* dengan nilai *cost*, *gamma*, dan *epsilon* yaitu 1, 0.25, dan 0.1. Dengan menggunakan model *default* yang telah disediakan sistem tanpa mengubah kernel yang ada pada *software R*, didapatkan bahwa model SVR tanpa melalui proses tuning atau proses pengoptimalan parameter dengan *grid search optimization* dan pemilihan parameter, menghasilkan *number support vector* sebesar 34. Dimana secara otomatis kernel yang digunakan dalam R yaitu kernel *radial* dengan parameter *cost* sebesar 1.00, *gamma* sebesar 0.25, dan *epsilon* sebesar 0.10. Untuk membentuk model persamaan data, terlebih dahulu mencari nilai  $w$  dengan menggunakan rumus pada persamaan (2.11), serta bantuan dari *software computer* sehingga didapatkan nilai  $w$  sebesar 5,849695 untuk  $X_1$ , 11,75892 untuk  $X_2$ , 10,97858 untuk  $X_3$ , dan 14,16956 untuk  $X_4$ . Selanjutnya menghitung nilai  $b$  dari persamaan (2.14), didapatkan nilai  $b$  sebesar -1,830045. Setelah mendapatkan

nilai  $w$  dan  $b$ , untuk membangun fungsi model yaitu dengan menggunakan persamaan (2.18). Sehingga didapatkan hasil akurasi sebagai berikut:

$$\begin{aligned}
 f(x) &= \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \\
 f(x) &= wK(x_i, x_j) + b \\
 f(x) &= \left[ 5,85 \quad 11,76 \quad 10,98 \quad 14,17 \right] K(x_i, x_j) - 1,830045 \quad (4.2)
 \end{aligned}$$

Setelah model didapatkan, maka langkah selanjutnya yaitu menerapkan model dari data *training* untuk diramalkan dengan menggunakan data *testing* serta menghitung nilai akurasi dari model, maka didapatkan hasil akurasi sebagai berikut:

**Tabel 4.5 Nilai Akurasi Tanpa Proses Pengoptimalan Parameter**

Pembagian Data	MAPE	RMSE
60:40	148,5732	89,89576
70:30	58,05624	75,76935
80:20	27,5813	70,5205
90:10	18,99759	83,19228

Pemilihan model terbaik dalam penelitian ini menggunakan kriteria MAPE dan RMSE. Dari hasil (4.5) model yang didapatkan tanpa melalui proses pengoptimalan parameter atau *tuning hyperparameter* tersebut menunjukkan bahwa prediksi yang dilakukan oleh SVR belum memberikan hasil yang optimal dikarenakan hasil tersebut tidak dapat menjangkau secara keseluruhan data yang ada serta tidak dapat menunjukkan hasil yang optimal. *Tuning* merupakan proses untuk menentukan parameter dalam mendapatkan model terbaik yang dapat dipilih

sebagai model untuk prediksi. Proses parameter *tuning* berfokus pada pemilihan nilai parameter kernel seperti *gamma*, *cost*, *degree*, dan *epsilon*. Dalam menentukan nilai parameter yang tepat perlu dilakukan secara berulang.

Berdasarkan permasalahan tersebut, maka dalam pencarian model terbaik untuk melakukan prediksi data positif covid-19, perlu dilakukan pengoptimalan model menggunakan metode *grid search optimization* (GSO) yang telah tersedia pada *software R*, dimana *grid search* mempunyai kelebihan untuk meningkatkan performa model melalui pengoptimalan *hyperparameter*. Setelah mendapatkan *hyperparameter* paling optimal, yaitu dengan akurasi model terbaik, maka selanjutnya model kembali diuji untuk dievaluasi menggunakan data uji atau *testing*.

Parameter pada kernel SVR disebut sebagai *hyperparameter* karena tidak secara langsung dipelajari dalam estimator. *Hyperparameter* pada kernel inilah yang nilainya akan dioptimalkan. Proses pengoptimalan nilai *hyperparameter* diawali dengan memberikan nilai batas bawah dan batas atas pada setiap *hyperparameter*. Nilai-nilai batas ini didapatkan dengan melihat hasil penelitian sebelumnya yang menyatakan bahwa performa *grid search* dalam mengoptimalkan nilai *hyperparameter* akan semakin cepat dan maksimal apabila nilai batas yang diberikan untuk masing-masing *hyperparameter* memiliki kelipatan secara eksponensial. Sehingga didapatkan variasi untuk nilai *hyperparameter* pada penelitian ini yaitu (4.6):

Tabel 4.6 Batas Untuk Nilai *Hyperparameter*

Hyperparameter	Nilai Percobaan
Cost (C)	$2^{-5}, 2^{-4}, \dots, 2^5$
Gamma ( $\gamma$ )	$2^{-5}, 2^{-4}, \dots, 2^5$
Epsilon( $\epsilon$ )	0; 0, 1; 0, 2; ...; 1

*Tuning* nilai *hyperparameter* dilakukan dengan menggunakan data *training*. Untuk mengetahui model peramalan dengan kombinasi nilai batas *hyperparameter* manakah yang dapat menghasilkan tingkat akurasi tertinggi digunakan *cross validation*. *Cross validation* akan mengevaluasi setiap kombinasi nilai *hyperparameter* yang telah dicoba oleh *grid search* dan dapat menentukan manakah nilai *hyperparameter* terbaik yang dapat menghasilkan model peramalan dengan tingkat akurasi tertinggi.

Pada penelitian ini algoritma *grid search* menggunakan *10-fold validation*. Penerapan *10-fold validation* sudah merupakan standar dari metode validasi yang canggih dan praktis. Selain itu lebih efisien serta mampu meningkatkan sedikit nilai kinerjanya. Pada 10 subset data tersebut, masing-masing dari CV akan menggunakan 9 *fold* untuk pelatihan dan 1 *fold* untuk pengujian.

Parameter terbaik setelah dilakukan proses *tuning* yaitu menghasilkan nilai cost sebesar 4, gamma sebesar 0.5, dan epsilon sebesar 0 dengan *number support vector* dan *best performance* bernilai 161 dan 267741,2.

Setelah parameter terbaik didapatkan, selanjutnya yaitu membangun model dengan mencari nilai  $w$  dan  $b$  menggunakan *software* R. Setelah nilai  $w$  dan  $b$  didapatkan, maka model yang terbentuk berdasarkan persamaan (2.18) setelah

dilakukan *tuning* parameter yaitu:

$$f(x) = \left[ 10,01 \quad 18,03 \quad 23,12 \quad 24,70 \right] K(x_i, x_j) - 2,158143 \quad (4.3)$$

Setelah model dari data *training* sudah didapatkan, selanjutnya akan diramalkan dengan menggunakan data *testing* serta menghitung akurasi dari model, maka didapatkan hasil akurasi sebagai berikut:

**Tabel 4.7 Nilai Akurasi Melalui Proses *Grid Search Optimization***

<b>Pembagian Data</b>	<b>MAPE</b>	<b>RMSE</b>
60:40	18,3326	47,07387
70:30	18,61822	54,2284
80:20	16,49343	64,70499
90:10	14,99808	73,04026

Terdapat beberapa kernel yang biasanya digunakan dalam pembangunan model pada metode SVR. Fungsi kernel yaitu mengubah data dari ruang nonlinier ke ruang linier. Trik kernel memungkinkan SVR untuk menemukan kecocokan dan kemudian data dipetakan ke ruang aslinya.

#### **4.4.2. Kernel Linier**

Kernel linier merupakan sebuah kernel yang paling sederhana, dimana kernel ini digunakan ketika data yang dianalisis sudah terpisah secara linier. Dalam melakukan analisis menggunakan kernel linear, parameter yang digunakan untuk optimasi biasanya yaitu parameter C atau *cost*. Pengoptimalan parameter C dapat dilakukan dengan cara *trial and error*.

Berdasarkan hasil pencarian parameter *hyperplane* dengan menggunakan



algoritma *grid search*, diperoleh parameter *hyperplane* yang terbaik yaitu untuk  $cost = 0,125$ ,  $gamma = 0,03125$ , dan nilai  $epsilon = 0$ , sehingga didapatkan nilai *number support vectornya* sebesar 161 dan *best performance* sebesar 263247. Setelah parameter dari setiap pembagian data didapatkan, maka selanjutnya akan dilakukan peramalan untuk kasus positif covid-19 di Jawa Timur dengan menggunakan data *testing*. Selanjutnya menghitung akurasi dari model yang telah didapatkan. Berikut merupakan tabel akurasi untuk metode SVR dengan kernel linier:

**Tabel 4.8 Akurasi Menggunakan Kernel Linier**

Kernel	Pembagian Data	Tingkat Akurasi	
		RMSE	MAPE
Linier	60:40	51,75471	19,9218
	70:30	59,65747	20,91292
	80:20	71,29911	19,10158
	90:10	83,12458	18,0664

Tabel diatas merupakan perbandingan tingkat akurasi dari nilai yang dihasilkan pada pembagian data untuk data *training* dan *testing*. Kemudian melakukan prediksi kasus positif covid-19 di Jawa Timur menggunakan data *testing* untuk menguji model. Dalam pengujian tersebut didapatkan nilai akurasi MAPE dan RMSE terbaik yaitu 19,9218 dan 51,75471. Hasil akurasi terbaik tersebut didapatkan dari pembagian data sebesar 60:40.

#### 4.4.3. Kernel Radial Basis Function (RBF)

Kernel RBF merupakan salah satu dari beberapa kernel untuk kasus non linier, dimana data tidak terpisah secara linear. Kernel ini memetakan suatu data

ke dimensi yang lebih tinggi dan membentuk kurva yang fleksibel sehingga dapat mengikuti pola data yang digunakan. Pada kernel RBF, parameter yang digunakan yaitu *gamma* dan *cost*.

Berdasarkan hasil pencarian parameter hyperplane dengan menggunakan algoritma *grid search*, diperoleh parameter hyperplane yang terbaik yaitu untuk  $cost = 4$ ,  $gamma = 0,5$ , dan nilai  $epsilon = 0$ , sehingga didapatkan nilai *number support vectornya* sebesar 161 dan *best performance* sebesar 272642,7.

Setelah parameter dari setiap pembagian data didapatkan, maka selanjutnya akan dilakukan peramalan untuk kasus positif covid-19 di Jawa Timur dengan menggunakan data *testing*. Maka didapatkan hasil akurasi menggunakan kernel RBF, yaitu:

**Tabel 4.9 Akurasi Menggunakan Kernel RBF**

Kernel	Pembagian Data	Tingkat Akurasi	
		RMSE	MAPE
RBF	60:40	47,07387	18,3326
	70:30	54,2284	18,61822
	80:20	64,70499	16,49343
	90:10	73,04026	14,99808

Tabel diatas merupakan perbandingan tingkat akurasi dari nilai yang dihasilkan pada pembagian data *training* dan *testing*. Kemudian melakukan prediksi kasus positif covid-19 di Jawa Timur menggunakan data *testing* untuk menguji model. Dalam pengujian tersebut didapatkan nilai akurasi MAPE dan RMSE terbaik yaitu 18,3326 dan 47,07387. Hasil akurasi terbaik tersebut didapatkan dari pembagian data sebesar 60:40.

#### 4.4.4. Kernel Polynomial

Kernel polynomial termasuk kedalam fungsi kernel non linier yang memiliki fungsi khusus untuk memetakan ke *feature space* yang biasanya berbentuk kurva bola. Parameter dari fungsi kernel ini yaitu *cost* dan *degree* ( $d$ ), dimana derajat memiliki fungsi untuk mencari nilai optimal pada setiap dataset. Pada umumnya nilai ( $d = 2$ ). Tetapi dalam penelitian ini nilai  $d$  yang akan digunakan yaitu 1 dan 2. Semakin besar nilai ( $d$ ) maka akurasi sistem yang dihasilkan akan fluktuatif dan kurang stabil. Hal ini dapat terjadi karena semakin tinggi nilai parameter ( $d$ ) maka semakin melengkung garis *hyperplane* yang dihasilkan. Sama seperti fungsi kernel sebelumnya, dalam pengoptimalan parameter dapat dilakukan dengan cara *trial and error*.

Berdasarkan hasil pencarian parameter *hyperplane* dengan menggunakan algoritma *grid search*, diperoleh parameter *hyperplane* yang terbaik pada degree 1 yaitu untuk  $cost = 8$ ,  $gamma = 0,03125$ , dan nilai  $epsilon = 0,1$ , sehingga didapatkan nilai *number support vectornya* sebesar 26 dan *best performance* sebesar 270164,8. Sedangkan untuk parameter *hyperplane* yang terbaik pada degree 2 yaitu untuk  $cost = 16$ ,  $gamma = 0,03125$ , dan nilai  $epsilon = 0,1$ , sehingga didapatkan nilai *number support vectornya* sebesar 78 dan *best performance* sebesar 384362,1.

Setelah parameter dari setiap pembagian data didapatkan, maka selanjutnya akan dilakukan peramalan untuk kasus positif covid-19 di Jawa Timur dengan menggunakan data *testing*. Maka didapatkan hasil akurasi menggunakan kernel polynomial, yaitu:

Tabel 4.10 Akurasi Menggunakan Kernel Polynomial Derajat 1

Kernel	Pembagian Data	Tingkat Akurasi	
		RMSE	MAPE
Polynomial	60:40	48,92445	56,79729
	70:30	51,16121	28,38106
	80:20	58,14117	21,86117
	90:10	64,93975	16,16689

Tabel 4.11 Akurasi Menggunakan Kernel Polynomial Derajat 2

Kernel	Pembagian Data	Tingkat Akurasi	
		RMSE	MAPE
Polynomial	60:40	124,6222	168,6037
	70:30	120,1889	68,54427
	80:20	134,0642	42,75838
	90:10	180,6578	54,60399

Tabel diatas merupakan perbandingan tingkat akurasi dari nilai yang dihasilkan pada pembagian data untuk data *training* dan *testing*. Kemudian melakukan prediksi kasus positif covid-19 di Jawa Timur menggunakan data *testing* untuk menguji model. Dalam pengujian tersebut didapatkan nilai akurasi terbaik pada kernel polynomial yaitu menggunakan derajat 1, dimana hasil akurasi terbaik tersebut didapatkan dari pembagian data sebesar 90:10.

#### 4.4.5. Kernel *Tangent Hyperbolic* (Sigmoid)

Kernel sigmoid merupakan sebuah kernel untuk kasus non linier, dimana data tidak terpisah secara linear. Pada kernel sigmoid, parameter yang digunakan

yaitu *gamma*, *epsilon*, dan *cost*.

Berdasarkan hasil pencarian parameter *hyperplane* dengan menggunakan algoritma *grid search*, diperoleh parameter *hyperplane* yang terbaik untuk *cost* = 1, *gamma* = 0,03125, dan nilai *epsilon* = 0,4, sehingga didapatkan nilai *number support vectornya* sebesar 16 dan *best performance* sebesar 494953,8.

Setelah parameter dari setiap pembagian data didapatkan, maka selanjutnya akan dilakukan peramalan untuk kasus positif covid-19 di Jawa Timur dengan menggunakan data *testing*. Maka didapatkan hasil akurasi menggunakan kernel sigmoid, yaitu:

**Tabel 4.12 Akurasi Menggunakan Kernel Sigmoid**

Kernel	Pembagian Data	Tingkat Akurasi	
		RMSE	MAPE
Sigmoid	60:40	288,8261	401,479
	70:30	293,7956	228,3575
	80:20	297,6051	150,9145
	90:10	298,6417	98,98588

Tabel diatas merupakan perbandingan tingkat akurasi dari nilai yang dihasilkan pada pembagian data untuk data *training* dan *testing*. Kemudian melakukan prediksi kasus positif covid-19 di Jawa Timur menggunakan data *testing* untuk menguji model. Dalam pengujian tersebut didapatkan nilai akurasi MAPE dan RMSE terbaik yaitu 98,98588 dan 298,6417. Hasil akurasi terbaik tersebut didapatkan dari pembagian data sebesar 90:10.

#### 4.5. Perbandingan Tingkat Akurasi

Setelah didapatkan model peramalan dengan nilai *hyperparameter* yang terbaik, aktivitas selanjutnya adalah melakukan prediksi menggunakan model tersebut untuk data *testing*, hal ini dilakukan untuk menguji model prediksi dengan nilai *hyperparameter* terbaik. Evaluasi model prediksi dilakukan untuk melihat seberapa akurat model peramalan terbaik yang telah dibangun sebelumnya. *Input* pada proses evaluasi hasil peramalan ini adalah data aktual pada data *testing* beserta hasil prediksinya. Evaluasi ini penting untuk dilakukan agar dapat menilai model prediksi SVR yang telah dibangun dan dilatih menggunakan data *training*, apakah model yang dibangun cocok dan baik digunakan untuk meramalkan kasus positif covid-19 di wilayah Jawa Timur. Berikut merupakan hasil akurasi dari beberapa kernel.

**Tabel 4.13 Perbandingan 60:40**

Kernel	Tingkat Akurasi	
	RMSE	MAPE
Linier	51,75471	19,9218
RBF	47,07387	18,3326
Polynomial	48,92445	56,79729
Sigmoid	288,8261	401,479

Tabel 4.14 Perbandingan 70:30

Kernel	Tingkat Akurasi	
	RMSE	MAPE
Linier	59,65747	20,91292
RBF	54,2284	18,61822
Polynomial	51,16121	28,38106
Sigmoid	293,7956	228,3575

Tabel 4.15 Perbandingan 80:20

Kernel	Tingkat Akurasi	
	RMSE	MAPE
Linier	71,29911	19,10158
RBF	64,70499	16,49343
Polynomial	58,14117	21,86117
Sigmoid	297,6051	150,9145

Tabel 4.16 Perbandingan 90:10

Kernel	Tingkat Akurasi	
	RMSE	MAPE
Linier	83,12458	18,0664
RBF	73,04026	14,99808
Polynomial	64,93975	16,16689
Sigmoid	298,6417	98,98588

Dari perbandingan tabel akurasi diatas, didapatkan nilai akurasi terkecil yaitu metode SVR dengan menggunakan kernel RBF (*Radial Basis Function*)

perbandingan 60:40. Maka langkah selanjutnya yaitu melakukan peramalan untuk beberapa periode menggunakan model SVR kernel RBF dengan parameter  $cost = 4$ ,  $gamma = 0,5$ , dan nilai epsilon-nya = 0 untuk 1 bulan berikutnya.

#### 4.6. Peramalan Kasus Covid-19 di Wilayah Jawa Timur

Dari hasil percobaan beberapa parameter untuk membentuk model *support vector regression* yang telah dilakukan, maka tahap selanjutnya yaitu melakukan peramalan angka kasus positif covid-19 di wilayah Jawa Timur dengan menggunakan model terbaik yang telah dibangun sebelumnya. Berikut merupakan tabel perbandingan data prediksi dan data aktual covid-19 pada data *testing* dengan menggunakan salah satu model terbaik yang telah ditentukan.



UIN SUNAN AMPEL  
S U R A B A Y A

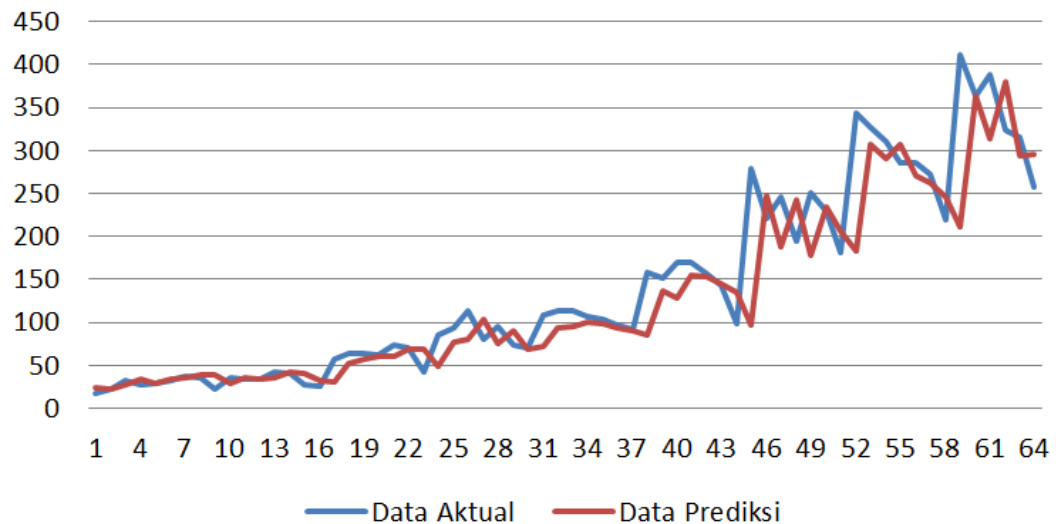


**Tabel 4.17 Perbandingan Data Aktual dan Prediksi**

No	Data Aktual	Prediksi
1	18	23,70019
2	22	22,10854
3	33	26,73955
4	27	33,58966
5	29	29,49591
6	33	34,11352
7	38	35,78033
8	36	39,09487
9	23	38,88623
10	35	29,05077
11	34	36,42220
12	34	34,78896
13	43	36,03948
:	:	:
:	:	:
64	258	295,05912

Berikut akan disajikan plot perbandingan data aktual dan prediksi:

## Plot Data Aktual dan Prediksi



**Gambar 4.6 Perbandingan Data Aktual dan Prediksi**

Berdasarkan plot diatas, didapatkan hasil prediksi hampir mendekati data aktual, sehingga model dengan parameter yang digunakan tergolong baik untuk meramalkan kasus covid-19 di wilayah Jawa Timur.

Setelah itu akan dilakukan peramalan positif covid-19 untuk 1 bulan kedepan dengan menggunakan model terbaik yang telah didapatkan, yaitu sebagai berikut:

Tabel 4.18 Peramalan Kasus Positif Covid-19 di Jawa Timur

Tanggal	Ramalan	Tanggal	Ramalan
1 Agustus 2022	100,21504	17 Agustus 2022	235,00207
2 Agustus 2022	98,22298	18 Agustus 2022	206,74454
3 Agustus 2022	93,00285	19 Agustus 2022	183,03476
4 Agustus 2022	89,60364	20 Agustus 2022	306,88829
5 Agustus 2022	85,67325	21 Agustus 2022	290,28527
6 Agustus 2022	136,17945	22 Agustus 2022	307,88872
7 Agustus 2022	128,82259	23 Agustus 2022	270,03885
8 Agustus 2022	155,32448	24 Agustus 2022	262,27253
9 Agustus 2022	152,72755	25 Agustus 2022	246,25882
10 Agustus 2022	145,2368	26 Agustus 2022	210,64432
11 Agustus 2022	135,14577	27 Agustus 2022	363,76685
12 Agustus 2022	97,09679	28 Agustus 2022	314,60177
13 Agustus 2022	248,19028	29 Agustus 2022	380,48391
14 Agustus 2022	188,30708	30 Agustus 2022	294,59127
15 Agustus 2022	241,78501	31 Agustus 2022	295,05912
16 Agustus 2022	177,65973		

Tabel diatas merupakan tabel peramalan positif covid-19 yang ada di wilayah Jawa Timur untuk 1 bulan kedepan yakni dimulai pada periode 1 Agustus 2022 sampai dengan 31 Agustus 2022.

#### 4.7. Pembahasan

Pada penelitian menggunakan metode *support vector regression* dengan pengoptimalan parameter menggunakan *grid search optimization* untuk prediksi kasus covid-19 di wilayah Jawa Timur dimana data aktual tersebut didapatkan dengan cara mengunduh dari *website* resmi pemerintah mengenai covid-19 yang bersifat terbuka.

Berdasarkan pengujian yang telah dilakukan menggunakan data dari periode 10 Februari 2022 hingga pada periode 31 Juli 2022, hasil proses menunjukkan bahwa menambahkan algoritma *grid search* untuk pengoptimalan parameter, sangat berpengaruh yang mana hal tersebut menunjukkan tingkat akurasi yang jauh lebih baik dibandingkan tidak menggunakan proses pengoptimalan parameter, dikarenakan hal tersebut sangat penting untuk mengetahui nilai parameter yang paling optimal, sehingga menghasilkan tingkat *error* paling rendah untuk prediksi. Selain itu pemilihan kernel untuk *hyperparameter* terbaik juga sangat berpengaruh pada hasil akurasi prediksi. Dimana dalam penelitian ini terdapat 4 kernel yang digunakan yaitu kernel linier, RBF (*Radial Basis Function*), polinomial, dan kernel sigmoid.

Dari 4 kernel tersebut didapatkan *hyperparameter* terbaik untuk memprediksi kasus positif menggunakan kernel RBF dengan pembagian untuk data *training* dan *testing* sebesar 60:40. Nilai MAPE dan RMSE yang didapatkan yaitu sebesar 18,3326 dan 47,07387. Hasil akurasi tersebut jika diukur menggunakan tabel MAPE yaitu mempunyai nilai yang baik jika digunakan untuk peramalan kasus covid-19 di Jawa Timur.

#### 4.8. Integrasi Keilmuan

Di dalam dunia islam, masa depan juga sangat diperhatikan. Hal ini sesuai dalam Alquran yaitu Surat Al Hasyr (59) ayat 18 yang berbunyi,

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

Artinya : *"Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan."* (QS: Al-Hasyr: 18)

Dari ayat diatas, bahwasanya merencanakan hal untuk masa depan sangat penting, karena Allah menyuruh orang-orang beriman untuk selalu mempersiapkan hari esok dengan sebaik mungkin. Dimana hal tersebut dapat dilakukan dengan cara meninjau ulang hal-hal yang sudah terjadi sebelumnya. Masa depan itu tidak hanya di dunia, tetapi tidak kalah pentingnya adalah di akhirat. Masa depan itu harus direncanakan dengan baik. Apalagi perencanaan itu sangat bermanfaat untuk orang lain.

Peramalan dalam islam juga diperbolehkan, sebagaimana yang pernah dilakukan oleh Nabi Yusuf mengenai mimpi sang raja. Tanpa berfikir panjang, nabi Yusuf segera menafsirkan mimpi tersebut dengan detailnya, serta memberikan solusi tentang bagaimana cara menghadapi mimpi yang telah ditafsirkan tersebut. Nabi Yusuf berpesan agar selama tujuh tahun dimana masa tanah masih subur untuk bercocok tanam dalam rangka menghadapi krisis pangan selama tujuh tahun mendatang. Sehingga raja dan rakyat mesir dapat melalui masa sulit tersebut

dengan selamat karena berkat solusi yang telah diberikan oleh nabi yusuf. Di sini ada pelajaran penting bahwa diperbolehkan melakukan peramalan atau memprediksi suatu hal dimasa depan untuk menjaga stabilitas di waktu yang akan mendatang.



## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Dari penelitian yang sudah dilakukan, berdasarkan dari hasil pengolahan yang telah didapatkan dengan metode *support vector regression* menggunakan algoritma *grid search optimization* dengan jumlah data *training* yang digunakan untuk membangun model prediksi, yaitu sebagai berikut:

1. Hasil analisis yang didapatkan yaitu bahwa melakukan pengoptimalan parameter jauh lebih baik dibanding tidak melakukan pengoptimalan parameter, dikarenakan ketika parameter tidak di optimal terlebih dahulu, maka hasil kurva yang dibangun tidak akan fit dengan data aktual sehingga tingkat akurasi atau *error* yang dihasilkan lebih besar dan hal tersebut tidak baik untuk memprediksi.
2. Setelah itu berdasarkan hasil uji yang telah dilakukan dengan empat kernel yang berbeda, maka didapatkan hasil prediksi kasus covid-19 di wilayah Jawa Timur yang mempunyai tingkat akurasi terbaik adalah dengan menggunakan kernel RBF, dengan nilai parameter  $cost = 4$ ,  $gamma = 0,5$ ; dan  $epsilon = 0$ . Pada pembagian data *training* dan *testing* sebesar 60:40, dimana nilai akurasi MAPE dan RMSE yaitu sebesar 18,3326 dan 47,07387.
3. Hasil peramalan covid-19 di wilayah Jawa Timur mengalami kenaikan dan penurunan, walaupun begitu masyarakat dianjurkan untuk senantiasa

menjaga dan mentaati protokol kesehatan agar tidak terjadi kelonjakan jumlah kasus positif covid-19 di wilayah Jawa Timur untuk periode selanjutnya.

## 5.2. Saran

Berdasarkan hasil analisa yang telah dilakukan, maka peneliti tentunya memiliki beberapa saran yang bisa digunakan peneliti selanjutnya untuk mengembangkan penelitian ini sehingga menjadi lebih baik lagi. Adapun saran tersebut yaitu:

1. Disarankan mengombinasikan metode lain yang relevan dengan SVR untuk menghasilkan hasil yang lebih baik.
2. Untuk penelitian selanjutnya diharapkan melakukan optimasi parameter untuk mendapatkan hasil yang lebih baik, seperti metode GA atau PSO.
3. Menambah variasi parameter yang digunakan untuk melakukan optimasi parameter agar mendapatkan parameter yang optimal dan model terbaik untuk peramalan kasus positif covid-19 di Jawa Timur.

UIN SUNAN AMPEL  
S U R A B A Y A



## DAFTAR PUSTAKA

- Ahmad, F. (2020). Penentuan Metode Peramalan Pada Produksi Part New Granada Bowl ST di PT.X. *Jurnal Integrasi Sistem Industri*, 7(1):31–39.
- Ahsan, U. N. (2009). Hadis-hadis Tentang Anjuran Membaca Tahmid Setelah Bersin dan Mendoakan Orang Yang Bersin (Telaah Ma'anil Hadis). Technical report, Universitas Islam Negeri Sunan Kalijaga, Yogyakarta.
- Alam, A. S. L. S. (2020). Faktor Yang Mempengaruhi Kepatuhan Masyarakat Terhadap Penggunaan Maskerdalam Pencegahan Covid-19 Di Kecamatan Biringkanayakota Makassartahun 2020. Technical Report Kepatuhan, Universitas Hasanuddin.
- Arduyan, S., Mulyono, and Suyitno, A. (2017). Implementasi Algoritma Dijkstra Dalam Pencarian Rute Terpendek Tempat Wisata Di Kabupaten Gunungkidul Dengan Program Visual Basic. *UNNES Journal of Mathematics*, 6(2):108–116.
- Arimuko, A., Wibawa, A. S. W., and Firmansyah, A. (2019). Analisis Perbandingan Penentuan Hiposentrum Menggunakan Metode Grid Search, Geiger, dan Random Search: Studi Kasus pada Letusan Gunung Sinabung 2017. *Diffraction*, 1(2):22–28.
- Assaffat, L. (2016a). Pengaruh Jumlah Data Latih SVM Pada Peramalan Beban Listrik Bulanan Di Sektor Industri. *Techno*, 17(2):88–93.
- Assaffat, L. (2016b). Prediksi Beban Listrik Harian Pada Sektor Industri Berbasis

- SVM Dengan Kernel Polinomial. *Universty Research Coloquium*, pages 108–115.
- Asyiva, A. (2019). Prediksi Laju Inflasi Indonesia Menggunakan Metode Support Vector Regression Dengan Kernel Radial Basis Function. Technical report, Universitas Islam Negeri Syarif Hidayatullah, Jakarta.
- Audina, B., Fatekurohman, M., and Riski, A. (2021). Peramalan Arus Kas dengan Pendekatan Time Series Menggunakan Support Vector Machine. *Indonesian Journal of Applied Statistics*, 4(1):34.
- Cahyono, R. E., Sugiono, J. P., and Tjandra, S. (2019). Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen. *JTIM : Jurnal Teknologi Informasi dan Multimedia*, 1(2):106–116.
- Caraka, R. E., Yasin, H., and Basyiruddin, A. W. (2017). Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis. *Jurnal Matematika*, 7(1):43.
- Darsyah, M. Y. (2014). Klasifikasi Tuberkulosis Dengan Pendekatan Metode Supports Vector Machine ( SVM ). *Statistika*, 2(2):37–41.
- Davies, P. D. (2020). Multi-drug resistant tuberculosis. *Jurnal Respirologi Indonesia*, 40(2):9–12.
- dkk. Arifin, Z. N. P. A. (2021). Upaya Pencegahan Penularan dan Penyebaran Virus Corona (Covid-19) di Lingkungan Suradadi Timur Karang Baru Kota Mataram. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (PKM)*, 4:1252–1258.
- Drajana, I. C. R. (2017). Metode Support Vector Machine Dan Forward Selection

- Prediksi Pembayaran Pembelian Bahan Baku Kopra. *Ilkom Jurnal Ilmiah*, 9(2):116–123.
- Fahrusyiana, G. E. (2019). Peramalan Angka Penderita Penyakit Tuberkulosis di Provinsi Jawa Timur Menggunakan Metode Support Vector Machine. Technical report, Institut Teknologi Sepuluh Nopember, Surabaya.
- Fairuzyana, S. (2016). Ramalan Bintang Menurut Perspektif Al-Qur'an (Kajian Tafsir Tematik). Technical report.
- Hapsari, A. D. and Ulfa, B. (2018). Tipologi Wilayah Peri Urban Kabupaten Sidoarjo Berdasarkan Aspek Fisik, Sosial, dan Ekonomi. *Jurnal Teknik ITS*, 7(2).
- Hendayanti, N. P. N., Suniantara, I. K. P., and Nurhidayati, M. (2019). Penerapan Support Vector Regression (SVR) Dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik Ke Bali. *Jurnal Varian*, 3(1):43–50.
- Hulu, S. (2020). *Analisis Kinerja Metode Cross Validation Dan K-Nearest Neighbor Dalam Klasifikasi Data*. PhD thesis.
- Iskandar, A. A. (2018). Pentingnya Memelihara Kebersihan dan Keamanan Lingkungan Secara Partisipatif Demi Meningkatkan Gotong Royong dan Kualitas Hidup Warga. *Jurnal Ilmiah Pena*, 1(1):79 – 84.
- Ispriyanti, D., Prahutama, A., Tarno, T., Warsito, B., Yasin, H., and Anggara, P. (2020). Perbandingan Model Jaringan Syaraf Tiruan Dengan Algoritma Levenberg-Marquadt Dan Powell-Beale Conjugate Gradient pada Kecepatan Angin Rata-Rata Di Kota Semarang. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 8(2):127.

- Jalaluddin, M. (2021). Konsep Menanggulangi Krisis Pangan dalam Al-Qur'an. Technical report, Universitas Islam Negeri Sunan Ampel Surabaya, Surabaya.
- Jiwandono, I. S., Nurhasanah, H., Rosyidah, A. N. K., Anar, A. P., and Maulyda, M. A. (2020). Mengatasi Problematika Covid-19 Di Kalangan Mahasiswa: Webinar Peningkatan Pengetahuan Mahasiswa Terkait Kebersihan Diri. *Jurnal Pendidikan dan Pengabdian Masyarakat*, 3(3):176–181.
- Kumila, A., Sholihah, B., Evizia, E., Safitri, N., and Fitri, S. (2019). Perbandingan Metode Moving Average dan Metode Naïve Dalam Peramalan Data Kemiskinan. *JTAM — Jurnal Teori dan Aplikasi Matematika*, 3(1):65.
- Lubis, A. A., Wong, N. P., and Sinaga, F. M. (2020). Prediksi Akurasi Perusahaan Saham Menggunakan SVM dan K-Fold Cross Validation. *Jurnal SIFO Mikroskil*, 21(1):11–19.
- Lumbanraja, F. R., Sani, R. S., Kurniawan, D., and Irawati, A. R. (2019). Implementasi Metode Support Vector Machine Dalam Prediksi Persebaran Demam Berdarah Di Kota Bandar Lampung. *Jurnal Komputasi*, 7(2):63 – 73.
- Lusiana, A. and Yuliarty, P. (2020). Penerapan Metode Peramalan (Forecasting) Pada Permintaan Atap di PT X. *Industri Inovatif : Jurnal Teknik Industri*, 10(1):11–20.
- Marhamah (2020). Implementasi Algoritma Support Vector Machine (SVM) Dalam Memprediksi Kelulusan Jalur Masuk Perguruan Tinggi Banda Aceh (Studi Kasus Mahasiswa Baru Tahun Ajaran 2019). Technical Report May, Universitas Islam Negeri Ar-Raniry Darussalam Banda Aceh, Banda Aceh.
- Maulana, N. D., Setiawan, B. D., and Dewi, C. (2019). Implementasi Metode

- Support Vector Regression ( SVR ) Dalam Peramalan Penjualan Roti ( Studi Kasus : Harum Bakery ). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3):2986–2995.
- Maulana, R. and Kumalasari, D. (2019). Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham Ggrm. *Jurnal Informatika Kaputama (JIK)*, 3(1):22–28.
- Ningtiyas, S. R. (2018). Aplikasi Metode Double Exponential Smoothing Holt dan Arima Untuk Meramalkan Voluntary Counselling and Testing (VCT) Odha di Provinsi Jawa Timur. *The Indonesian Journal of Public Health*, 13(January):156–168.
- Nugraha, A. S. and Purnamasari, K. K. (2019). Penerapan Metode Support Vector Machine Pada Part of Speech Tag Bahasa Indonesia. (112).
- Nugroho, N. A. and Purqon, A. (2015). Analisis 9 Saham Sektor Industri di Indonesia Menggunakan Metode SVR. *Seminar Kontribusi Fisika, Bandung*, pages 295–300.
- Nur, N. I. (2020). Implementasi Nilai-Nilai Islam Terhadap Kebersihan Lingkungan di Desa Bululoe Kecamatan Turatea Kabupaten Jeneponto. Technical Report May, Pendidikan Agama Islam, Agama Islam, Universitas Muhammadiyah Makassar, Makassar.
- Nurhayatun, S., Fitriyanti, N., and Sonia, D. (2021). Analisis Average Length of Stay (AVLOS) Kasus Covid-19 di Rumah Sakit X Bandung. *Jurnal Ilmiah Manusia dan Kesehatan*, 4(3):1–8.
- Nursofwa, R. F., Sukur, M. H., Kurniadi, B. K., and ., H. (2020). Penanganan

- Pelayanan Kesehatan Di Masa Pandemi Covid-19 Dalam Perspektif Hukum Kesehatan. *Inicio Legis*, 1(1):1–17.
- Pemerintah Provinsi Jawa Timur (2022). Data Covid-19.
- Pitria, P. (2019). Analisis Sentimen Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naïve Bayes. *Teknik Informatika Universitas Komputer Indonesia*, pages 1–8.
- Purnama, D. I. and Setianingsih, S. (2020). Support Vector Regression (SVR) Model for Forecasting Number of Passengers on Domestic Flights at Sultan Hasanudin Airport Makassar. *Jurnal Matematika, Statistika dan Komputasi*, 16(3):391–403.
- Putri, R. A., Winahju, W. S., and Mashuri, M. (2020). Penerapan Metode Ridge Regression dan Support Vector Regression (SVR) untuk Prediksi Indeks Batubara di PT XYZ. *Jurnal Sains dan Seni ITS*, 9(1):64–71.
- Putri, Dinda Pratiwi Dwi., d. B. S. (2021). Analisis Korelasi Variasi Gas Nitrogen Dioksida Terhadap Kenaikan Kasus Positif Terkonformasi Covid – 19 Studi Kasus Provinsi Jawa Timur. *Jurnal Ilmiah Teknik Geomatika IMAGI*, 1(2):1–13.
- Rachman, D. (2018). Klasifikasi penduduk miskin provinsi nusa tenggara timur menggunakan algoritma naïve bayes dan cross validation skripsi. Technical report, Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa Bekasi, Bekasi.
- Rahmawati, Y., Anugrah, F. F., Hati, E. M., and Roziqin, A. (2021). Kampung Tangguh: Wujud Kolaborasi antar-Stakeholder dalam Merespons Pandemi Covid-19. *Journal of Social Development Studies*, 2(1):39–51.

- Rohman, Z. A., Rohmawati, A. A., and Indwiarti (2021). Prediksi Penyebaran Covid-19 Harian di Jawa Timur Menggunakan Model Vector Autoregressive Moving Average ( VARMA ). *e-Proceeding of Engineering*, 8(5):11150–11163.
- Sadewo, E. (2016). Perbandingan Beberapa Metode Time Series Pada Peramalan Jumlah Kunjungan Wisatawan Mancanegara : Studi Kasus Di Kabupaten Karimun, Provinsi Kepulauan Riau. *ResearchGate*, (Desember).
- Sihite, G. A. (2020). Perbandingan Metode Single Moving Average dan Double Exponential Smoothing Dalam Peramalan Pengambilan Uang di Mesin ATM. Technical Report May, Universitas Sanata Dharma, Yogyakarta.
- Soetjipto, N. (2020). *Ketahanan UMKM Jawa Timur Melintasi Pandemi Covid-19*. Yogyakarta.
- Sulistiana (2020). Optimasi Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna Meningkatkan Akurasi Review Pada Situs E-Commerce. Technical report.
- Suryadilaga, M. A. (2020). Pemahaman Hadis Tentang Covid-19 dalam Perspektif Integrasi-Interkoneksi Amin Abdullah. *Khazanah: Jurnal Studi Islam dan Humaniora*, 18(2):173–196.
- Syahrial (2020). Dampak Covid-19 Terhadap Tenaga Kerja. *Jurnal Ners*, 4(2):21–29.
- Ulya, H. N. (2020). Alternatif Strategi Penanganan Dampak Ekonomi Covid-19 Pemerintah Daerah Jawa Timur Pada Kawasan Agropolitan. *El-Barka: Journal of Islamic Economics and Business*, 3(1):80–109.

- Vibriyanti, D. (2020). Kesehatan Mental Masyarakat: Mengelola Kecemasan Di Tengah Pandemi Covid-19. *Jurnal Kependudukan Indonesia*, 2902:69.
- Virania, T. A., Choiruddin, A., and Ratnasari, V. (2021). Analisis Risiko Penyebaran Kasus Covid-19 di Surabaya Raya Menggunakan Model Thomas Cluster Process. *Inferensi*, 4(1):57.
- Wahidah, I., Athallah, R., Hartono, N. F. S., Rafqie, M. C. A., and Septiadi, M. A. (2020). Pandemi Covid-19: Analisis Perencanaan Pemerintah dan Masyarakat dalam Berbagai Upaya Pencegahan. *Jurnal Manajemen dan Organisasi*, 11(3):179–188.
- Wahyuni, R. (2019). *Peramalan Tingkat Produksi Tanaman Pangan dan Tanaman Perkebunan Rakyat Kabupaten Bulukumba Menggunakan Metode Exponential Smoothing (Pemulusan Eksponensial)*. PhD thesis, UIN Alauddin Makassar.
- Wardah, S. and Iskandar, I. (2017). Analisis Peramalan Penjualan Produk Keripik Pisang Kemasan Bungkus : Home Industry Arwana Food Tembilahan). *J@ti Undip : Jurnal Teknik Industri*, 11(3):135.
- Widjaya, A., Hiryanto, L., and Handhayani, T. (2017). Prediksi Masa Studi Mahasiswa Dengan Voting Feature Interval 5 Pada Aplikasi Konsultasi Akademik Online. *Computatio : Journal of Computer Science and Information Systems*, 1(1):25.
- Yudhawan, D. H. (2020). Implementasi Support Vector Regression Untuk Peramalan Harga Saham Perusahaan Pertambangan Di Indonesia. Technical report, Universitas Islam Indonesia, Yogyakarta.