

PERAMALAN PERTAMBAHAN PASIEN COVID-19 MENGGUNAKAN SUPPORT VECTOR REGRESSION

FORECASTING GROWTH OF COVID-19 PATIENTS USING SUPPORT VECTOR REGRESSION

Maryati Puji Lestari¹, Deden Jacob Witarsyah², Faqih Hamami³

^{1,2,3} Universitas Telkom, Bandung

¹maryaaa@student.telkomuniversity.ac.id, ²dedenw@telkomuniversity.co.id,

³faqihhamami@telkomuniversity.ac.id

Abstrak

Sejak awal Maret 2020 virus COVID-19 mulai masuk ke Indonesia. Melalui laman covid19.go.id ditunjukkan lebih dari 2 juta kasus di Indonesia sejak Maret lalu. Perkembangan kasus yang sangat cepat membuat kapasitas pelayanan Kesehatan semakin menipis. Menurut Yulinna (2021) pada laman CNN, Jawa Timur memiliki 145 rumah sakit rujukan COVID-19 dengan BOR (*Bed Occupancy Rate*) mencapai 70%[1]. BOR merupakan nilai yang digunakan untuk jumlah pemakaian tempat tidur yang dipakai. WHO telah menetapkan batas BOR ideal di angka 60%. Rumah sakit dengan nilai BOR yang tinggi dapat diartikan tidak mampu menyediakan tempat tidur untuk pasien bergejala COVID-19. Berdasarkan hal tersebut penelitian ini bertujuan untuk melakukan peramalan terhadap kasus COVID-19 di daerah Jawa Timur dimana hasil peramalan tersebut diharapkan bisa membantu pemerintahan Jawa Timur untuk mengoptimalkan layanan kesehatan untuk pasien COVID-19. Penelitian ini menggunakan *Support Vector Regression* dengan fungsi kernel Linear untuk pembuatan model dan *GridSearchCV* untuk mengoptimalkan parameter fungsi Kernel Linear. Kemudian dilakukan peramalan untuk tanggal selanjutnya menggunakan SVR-Linear dengan parameter optimal. Hasil dari penelitian ini adalah peramalan sebanyak 30 hari dari tanggal 1 Februari 2021 sampai 2 Maret 2021. Penggunaan SVR-linear menghasilkan nilai akurasi dengan MSE sebesar 89.6 dan MAPE sebesar 13.9%.

Kata kunci : COVID-19, *Support Vector Regression*, Peramalan, *Mean Absolute Percentage Error*, *Mean Square Error*.

Abstract

Since early March 2020 the COVID-19 virus has entered Indonesia. Since March, the website covid19.go.id has recorded over 2 million cases in Indonesia. The capacity of health services is becoming increasingly limited due to the rapid development of cases. According to Yulinna (2021) on the CNN page, East Java has 145 COVID-19 referral hospitals with a BOR (*Bed Occupancy Rate*) reaching 70%[1]. BOR is the value used for the number of beds used. The optimal BOR limit is 60 percent, according to the World Health Organization. Hospitals with high BOR values may be unable to accommodate patients with COVID-19 symptoms. Based on this, the goal of this study is to forecast COVID-19 cases in East Java, with the forecasting results expected to aid the East Java government in improving COVID-19 patient care. *Support Vector Regression* with *Linear Kernel Function* is used in this study for modeling, while *GridSearchCV* is used to optimize the linear Kernel function parameters. Then, using SVR-Linear with optimal parameters, forecasting for the following date. This research yielded a 30-day forecast from February 1, 2021 to March 2, 2021. The accuracy of SVR-linear is 89.6% MSE and 13.9 percent MAPE.

Keywords: COVID-19, *Support Vector Regression*, Forecasting, *Mean Absolute Percentage Error*, *Mean Square Error*.

1. Pendahuluan

Pada saat ini masyarakat sedang dihebohkan oleh kemunculan suatu virus yang menyerang pernafasan manusia. Kemunculan virus ini awalnya diketahui saat ada sejumlah pasien dirawat di rumah sakit dengan diagnosis awal *pneumonia* [2]. *World Health Organization* (WHO) memberi nama virus korona yang baru tersebut disebut dengan *Severe Acute Respiratory Syndrome Coronavirus-2* (SARSCoV-2) dan nama penyakitnya adalah *Coronavirus disease 2019* (COVID-19) [3]. Virus ini masuk ke Indonesia sejak awal Maret lalu hingga saat ini dan diperkirakan akan terus berlanjut hingga waktu yang belum bisa diprediksi. Melalui laman COVID19.go.id ditunjukkan bahwa ada total lebih dari 2 juta kasus di Indonesia sejak Maret 2020 lalu. Dengan jumlah pasien terinfeksi yang makin banyak dengan penyebaran yang sangat cepat maka penyakit tersebut bisa dikatakan

sebagai pandemi. Perkembangan kasus yang sangat cepat membuat kapasitas pelayanan Kesehatan semakin menipis. Menurut Yulinna (2021) pada laman CNN, Jawa Timur memiliki 145 rumah sakit rujukan untuk COVID-19 dengan BOR (Bed Occupancy Rate) mencapai 70 persen [1]. BOR merupakan nilai yang digunakan untuk jumlah pemakaian tempat tidur yang dipakai [4]. WHO telah menetapkan batas BOR ideal di angka 60%. Rumah sakit dengan nilai BOR yang tinggi dapat diindikasikan tidak mampu menyediakan tempat tidur untuk pasien bergejala COVID-19. Maka dengan meramalkan jumlah pasien COVID-19 di daerah Jawa Timur diharapkan bisa mengoptimalkan peran pemerintah untuk menambah layanan Kesehatan yang diperlukan untuk pasien COVID-19 [5]. Berdasarkan hal tersebut penelitian ini bertujuan untuk melakukan peramalan terhadap kasus COVID-19 di daerah Jawa Timur dimana hasil peramalan tersebut diharapkan bisa membantu pemerintahan Jawa Timur untuk mengoptimalkan layanan kesehatan untuk pasien COVID-19. Peneliti menggunakan *support vector regression* (SVR) untuk melakukan peramalan. Data yang ada akan diolah menggunakan SVR dengan fungsi kernel beserta parameternya dan menggunakan bahasa pemrograman *python*.

2. Dasar Teori

2.1 COVID-19

Pada 31 Desember 2019, 27 kasus pneumonia dengan etiologi yang tidak diketahui diidentifikasi di Kota Wuhan, provinsi Hubei di Cina. Pasien-pasien ini awalnya memiliki gejala klinis batuk kering, dispnea, demam, dan infiltrat paru bilateral pada penggambarannya. Melalui tes Swab pada tenggorokan yang dilakukan oleh Pusat Pengendalian dan Pencegahan Penyakit China (CCDC) pada 7 Januari 2020, dan kemudian dinamai *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2). Penyakit itu dinamai COVID-19 oleh *World Health Organization* (WHO) [6]. Pada 11 Maret 2020 *World Health Organization* (WHO) telah menyatakan wabah virus corona baru (COVID-19) sebagai pandemi global [7].

2.2 Data Mining

Data Mining adalah teknologi untuk menemukan struktur dan pola dalam kumpulan data yang besar [8]. Data Mining merupakan proses penting dimana metode digunakan untuk mengekstrak pola data yang valid. Langkah ini mencakup pilihan tugas *data mining* yang paling sesuai (seperti klasifikasi, regresi, pengelompokan, atau asosiasi) [9]. *Data mining* juga disebut sebagai *Knowledge Discovery in Databases* (KDDs) yang merupakan sebuah bidang studi yang dikenal dalam ilmu kehidupan dan perdagangan, namun penerapan *data mining* dalam konteks pendidikan masih terbatas [10]. Secara umum *data mining* memiliki beberapa Teknik penting yang dapat digunakan untuk melakukan *data mining*, diantaranya adalah *association rules*, *lassification*, *clustering*, dan *forecasting* [11].

2.3 Peramalan

Peramalan adalah analisa peramalan yang memakai *input* data berupa *data time series*. *Data time series* umumnya berisi data yang saling berdekatan dan saling berurutan [11]. Peramalan juga bisa disebut sebagai suatu kegiatan yang memperkirakan sebuah kejadian di masa yang akan datang menggunakan cara analisis terhadap data historis yang akan dibuat sebuah model pola dan menghasilkan data perkiraan [12]. Menurut Engelbrecht dan Van Greunen (2015) terdapat 5 tahapan penting yang harus dilakukan saat peramalan, yaitu penentuan sebuah masalah, mengumpulkan informasi yang diperlukan, dilakukan visualisasi data dan dianalisis untuk mendapatkan pola, tren, dan hubungan variabel, kemudian memilih model peramalan yang sesuai terhadap histori data, dan yang terakhir adalah menggunakan metode peramalan yang sudah dipilih dan dilakukan evaluasi terhadap metode yang telah digunakan [12].

2.4 Machine Learning

Machine learning merupakan sub-bidang ilmu komputer dan dapat diklasifikasikan sebagai metode kecerdasan buatan. Hal ini memiliki keuntungan yaitu model dapat memecahkan masalah yang tidak mungkin diwakilkan oleh algoritma secara langsung [13]. *Machine learning* disebut juga sebagai bidang studi yang memberi komputer kemampuan untuk belajar tanpa memprogram secara gambling. *Machine learning* telah banyak digunakan untuk berbagai tugas komputasi ketika tidak memungkinkan untuk menentukan dengan hitungan tangan [14]. Terdapat tiga kelas besar yang dimiliki oleh *machine learning* sehingga dapat menyelesaikan masalah secara jenius, yaitu *supervised learning*, *unsupervised learning*, *reinforcement learning*. Dengan hal tersebut juga memungkinkan kita untuk menyatakan terminologi dasar dan mengekspos beberapa alat dasar dari *machine learning*. Teknik ini telah digunakan selama beberapa tahun untuk membantu identifikasi partikel dan pemilihan peristiwa pada fisika partikel eksperimental [15].

2.5 Support Vector Regression

SVR (*Support Vector Regression*) adalah salah satu model dari SVM yang dapat diterapkan pada permasalahan regresi dan menghasilkan nilai berupa nilai kontinu atau riil [16]. SVR bertujuan untuk meminimalkan batas atas kesalahan generalisasi. Fungsi regresi dihasilkan dengan menerapkan sekumpulan fungsi linier berdimensi tinggi [17]. SVR mempunyai kelebihan untuk mengatasi *overfitting*, sehingga dapat menghasilkan sebuah model dengan tingkat kesalahan yang kecil juga memiliki skor akurasi yang bagus [18]. Salah satu keuntungan utama SVR adalah kompleksitas komputasinya tidak bergantung pada dimensi ruang *input*. Selain itu, SVR memiliki kemampuan generalisasi yang sangat baik, dengan akurasi prediksi yang tinggi [19]. SVR masih memiliki konsep yang sama dengan SVM, mencari garis pemisah terbaik di antara fungsi pemisah yang memiliki jumlah tak terhingga. Garis pemisah yang baik dapat memisahkan data juga memiliki batasan *margin* yang paling besar [20]. Ide dasar dari SVR adalah menentukan *dataset* yang dipecah menjadi data *training* dan data *testing*. Setelah itu dari data *training* tersebut ditetapkan suatu fungsi regresi dengan batasan penyimpangan tertentu sehingga bisa menghasilkan prediksi yang mendekati nilai aktual [21]. Contohnya jika ada n -data *training* dilambangkan dengan (x_i, y_i) dengan $i = 1, 2, 3, \dots, n$. Sedangkan $x_i = \{x_1, x_2, x_3, \dots, x_n\}^T \in \mathbb{R}^n$ adalah sebuah vektor dalam input space dan $y_i = \{y_1, y_2, y_3, \dots, y_n\} \in \mathbb{R}$ adalah sebuah nilai output dari x_i yang bersesuaian. Maka dari itu, fungsi regresi dari SVR dituliskan sebagai berikut

$$f(x_i) = w \cdot x_i + b \quad (1)$$

dimana w adalah vektor bobot dan b adalah sebuah bias. Koefisien w dan b berfungsi untuk meminimalkan fungsi resiko. Pada persamaan (1) tersebut koefisiennya diperkirakan dengan cara meminimalkan fungsi resiko dengan persamaan

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

dengan syarat

$$\begin{aligned} y_i &\leq f(x_i) + \varepsilon + \xi_i \\ y_i &\geq f(x_i) - \varepsilon - \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \text{ dimana } i = 1, 2, \dots, n \end{aligned}$$

dimana $\|w\|$ merupakan regularisasi dari fungsi yang diperkecil agar membuat fungsi setipis mungkin. Dimana konstanta $C > 0$ merupakan nilai tawar dari ketipisan fungsi f dan batas atas penyimpangan lebih dari ε yang masih bisa ditoleransi. Di sisi lain ξ_i, ξ_i^* merupakan variabel *slack* positif untuk mengatasi masalah pembatas yang tidak memadai pada masalah optimasi. SVR memiliki masalah umum berupa meminimalkan fungsi kuadrat untuk diubah menjadi batasan. Batasan hasil penurunan fungsi kuadrat tersebut akan diselesaikan dengan persamaan *Lagrange Multiplier* hingga ditemukan rumus umum untuk SVR Linear setelah melalui penurunan rumus yang panjang dan rumit dimana x_i merupakan sebuah *support vector* dan x adalah *test vector* [22].

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b \quad (3)$$

Berikut merupakan fungsi kernel yang umum digunakan dalam perhitungan SVR [22].

$$1. \text{ Linear : } K(x_i, x) = x_i \cdot x \quad (4)$$

$$2. \text{ Radial Basis Function : } K(x_i, x) = \exp\left(\frac{\|x_i - x\|^2}{-2\sigma^2}\right) \gamma = \frac{1}{2\sigma^2} \quad (5)$$

$$3. \text{ Polynomial : } K(x_i, x) = (x_i^T x_j + t)^d \quad (6)$$

2.6 Python

Python adalah bahasa modern yang sepenuhnya didasarkan pada kelas yang terstruktur dengan baik (tidak seperti kebanyakan bahasa ilmiah lainnya seperti Matlab dan R), yang memudahkan pembuatan, pemeliharaan, dan penggunaan kembali kode berorientasi objek modular [23]. Python dianggap sebagai bahasa produktivitas tinggi karena sintaksnya yang sederhana, semantik yang fleksibel, dan jumlah perpustakaan yang besar. Di antara banyak pustaka yang berfokus pada kinerja di Python, NumPy adalah salah satu pustaka ilmiah penting dalam ekosistem Python karena menyediakan objek larik multi-dimensi (*array*) yang telah menjadi dasar penghitungan numerik yang efisien dalam Python [24]. Python merupakan bahasa pemrograman yang dapat

diakses dan digunakan dengan mudah baik oleh pemula maupun profesional, python juga merupakan Bahasa yang dapat digunakan baik dalam proyek besar maupun proyek kecil. Hal ini disebabkan sifat python yang mudah beradaptasi dan menjadi Bahasa yang berkembang dengan baik [25]. Dengan menggunakan python kita tidak perlu menginstal dan mengembangkan apapun bahkan tidak harus memiliki *runtime python* sendiri, hanya perlu web browser yang digunakan untuk menulis kode, mengakses materi pendukung dan menjalankannya [26]. Keunggulan dari python yaitu, dengan program yang gratis python menyediakan banyak pustaka dan kamus yang dapat membantu seorang *programmer* untuk menyelesaikan proyek mereka secara terorganisir dengan baik. Python juga dapat menyederhanakan *debugging* dengan fitur bawaannya. Selain itu, dalam python semua akan menjadi objek dan memiliki *namespace* sendiri. Sehingga python dapat digunakan untuk membuat aplikasi yang terlihat bagus dan berfungsi dengan baik bagi perorangan [25].

2.7 Mean Square Error

Mean Square Error (MSE) adalah sebuah perhitungan untuk menghitung rata-rata kesalahan berpangkat. MSE dapat dinyatakan dengan rumus

$$MSE = \frac{\sum(Aktual - Forecast)^2}{n-1} \quad (7)$$

Rumus (7) diartikan sebagai hasil pengurangan antara nilai aktual dan nilai *forecast* yang sudah dikuadratkan kemudian dibagi dengan n , n merupakan jumlah periode yang digunakan dalam perhitungan MSE [27].

2.8 Mean Absolute Error

Mean Absolute Error (MAPE) merupakan sebuah besaran untuk mengukur keakuratan dari hasil sebuah peramalan dengan cara menghitung nilai rata-rata eror dari hasil peramalan. MAPE memiliki satuan persentasi. Berikut Rumus dari MAPE.

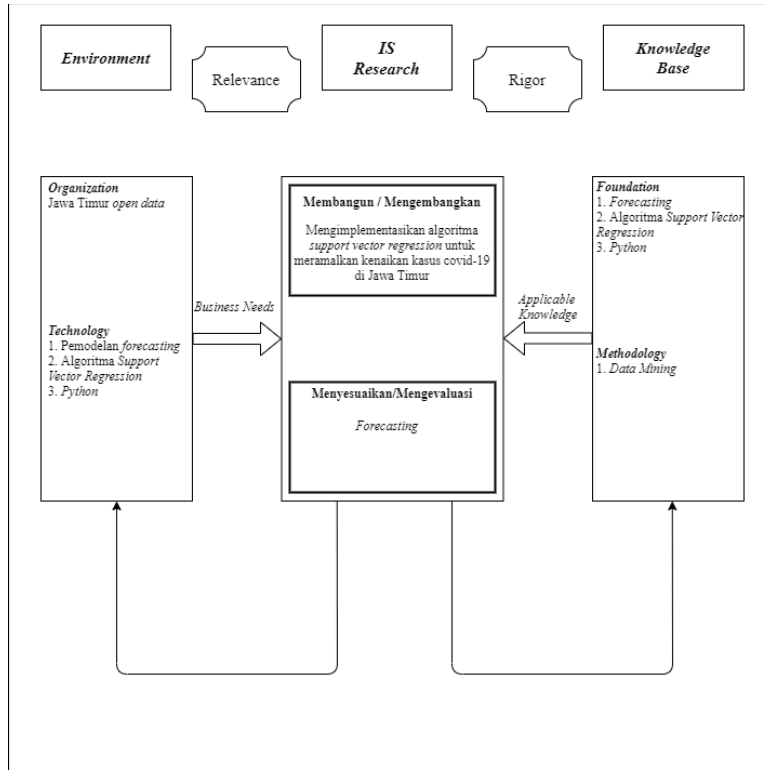
$$MAPE = \frac{\sum\left(\frac{|Aktual - Forecast|}{Aktual}\right)}{n} * 100 \quad (9)$$

Rumus (8) diartikan sebagai hasil pengurangan antara nilai aktual dan *forecast* yang sudah dimutlak-kan lalu dibagi dengan nilai aktualnya sendiri. Sedangkan n merupakan jumlah periode yang digunakan untuk menghitung MAPE [28].

3. Metodologi Penelitian

3.1 Model Konseptual

Model konseptual adalah sebuah model rancangan sebuah konsep pemikiran yang menggambarkan hubungan antara faktor yang terkait dalam penelitian dan menunjukkan hubungan yang relevan dengan tujuan penelitian. Dalam penelitian ini menggunakan 3 siklus utama *Relevance Cycle*, *Design Cycle*, dan *Rigor Cycle* (Mary dkk., 2018). Berikut model yang digunakan dalam penelitian ini dapat diilustrasikan dengan Gambar 1.

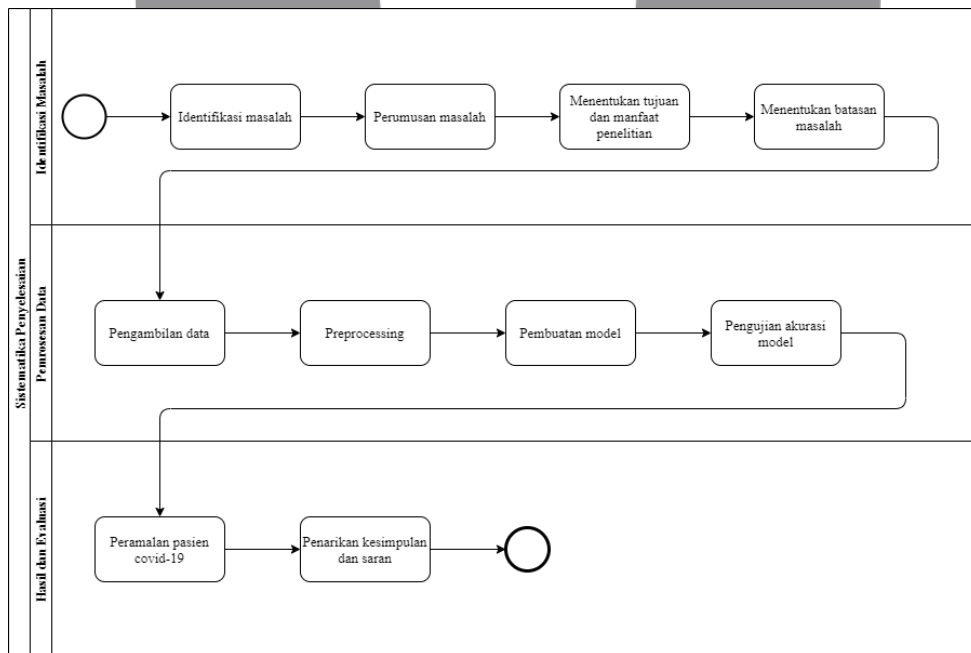


Gambar 1 Model Konseptual

Pada Gambar 1 dijelaskan tentang data yang dibutuhkan dalam penelitian ini agar permasalahan yang ada dapat dianalisis dan dipahami dengan mudah. Berdasarkan Gambar 1, penelitian ini mengambil data *open data*. Metode yang akan digunakan untuk penelitian ini adalah *Support Vector Regression (SVR)*.

3.2 Sistematika Penyelesaian Masalah

Metode yang akan digunakan pada penelitian ini dapat dilihat pada alur dalam diagram. Dari gambar diagram yang ada terdapat tiga tahapan yang dilakukan identifikasi masalah, pemrosesan data, dan hasil dan evaluasi.



Gambar 2 Sistematika Penelitian

Dengan adanya Gambar III.1 dapat menunjukkan bagaimana penelitian ini akan dijalankan untuk melakukan peramalan jumlah pasien COVID-19 daerah Jawa Timur. Selain itu dapat dilihat pada gambar tersebut bahwa pada sistem akan dilakukan *data training* dan *data testing*. Dengan demikian diagram tersebut telah mendeskripsikan bagaimana langkah-langkah Tugas Akhir yang sedang dikerjakan menggunakan *time series* dengan algoritma *Support Vector Regression*.

4. Analisis dan Perancangan

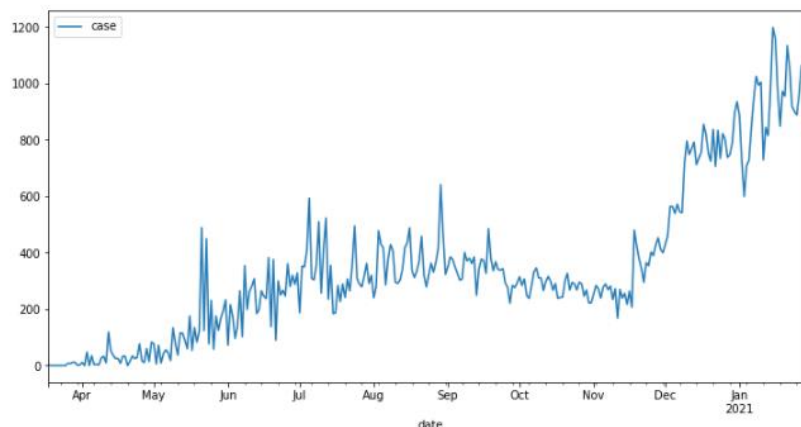
4.1 Pengambilan Data

Dilakukan pengambilan data melalui laman yang disediakan pemerintah untuk melihat data terbaru harian COVID-19 di covid19.go.id. Melalui laman tersebut dapat dilihat jumlah kasus COVID-19 seluruh Indonesia dan dapat menyaring tampilan agar dapat ditampilkan untuk kasus harian di Jawa Timur saja. Setelah laman sudah menampilkan kasus harian untuk daerah Jawa Timur saja, dilakukan pengambilan data dari awal munculnya kasus yaitu tanggal 18 Maret 2020 sampai tanggal 31 Januari 2021. Data yang digunakan dalam penelitian ini memiliki 2 atribut yaitu tanggal dan kasus. Berikut contoh data kasus harian di Jawa Timur :

Tabel 1 Data Kasus Harian COVID-19 di Jawa Timur

date	case
18/03/2020	2
19/03/2020	0
20/03/2020	0
21/03/2020	0
22/03/2020	0
23/03/2020	0
24/03/2020	0
25/03/2020	0
26/03/2020	8

Pada Tabel IV.1 ditampilkan 10 data pertama dari kasus harian COVID-19 di Jawa Timur. Kemudian pada Gambar IV.2 merupakan grafik kenaikan kasus Maret 2020 sampai Januari 2021.



Gambar 3 Grafik Kasus Harian COVID-19 Jawa Timur

4.2 Preprocessing

Proses preprocessing yang pertama dilakukan adalah menyesuaikan format tanggal yang digunakan agar bisa terbaca oleh sistem. Format awal tanggal yang digunakan pada data setnya yaitu menggunakan format

“dd/mm/yyyy” diubah menjadi format “yyyy-mm-dd”. Berikut pada Gambar IV.2 ditampilkan bagaimana data sebelum dan sesudah dilakukan transformasi tanggal.

	date	case		date	case
0	18/03/2020	2	→	2020-03-18	2
1	19/03/2020	0		2020-03-19	0
2	20/03/2020	0		2020-03-20	0
3	21/03/2020	0		2020-03-21	0
4	22/03/2020	0		2020-03-22	0

Gambar 4 Transformasi Tanggal pada Dataset

Proses kedua untuk *preprocessing* adalah penambahan data *moving average*. *Moving average* adalah sebuah metode yang digunakan untuk mengaktifkan nilai ramalan di masa depan dengan menggunakan jumlah rata-rata dari data aktual. Sifat khusus yang dimiliki oleh *moving average* yaitu, membuat peramalan dengan menggunakan data sebelumnya dengan jangka waktu tertentu, jika semakin panjang *moving average* maka akan menghasilkan nilai yang semakin halus. Rumus umum dari *moving average* adalah [28] :

$$F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} \dots}{n} \quad (10)$$

Dalam rumus (10), F_t yang merupakan masa yang akan dicari nilai ramalannya. Kemudian $A_{t-1} + A_{t-2} + A_{t-3} \dots$ adalah data aktual dari masa sebelum masa yang akan dicari nilai ramalnya, lalu n adalah jumlah masa yang akan dihitung peramalan dengan periode tertentu.

Tabel 2 Data Kasus Setelah *Featurization*

	date	case	ma_7days	ma_14days	ma_21days	ma_28days
0	2020-03-18	2	0.000000	0.0	0.0	0.0
1	2020-03-19	0	0.000000	0.0	0.0	0.0
2	2020-03-20	0	0.000000	0.0	0.0	0.0
3	2020-03-21	0	0.000000	0.0	0.0	0.0
4	2020-03-22	0	0.000000	0.0	0.0	0.0
5	2020-03-23	0	0.000000	0.0	0.0	0.0
6	2020-03-24	0	0.000000	0.0	0.0	0.0
7	2020-03-25	0	0.285714	0.0	0.0	0.0
8	2020-03-26	8	0.000000	0.0	0.0	0.0
9	2020-03-27	7	1.142.857	0.0	0.0	0.0

Pada Tabel IV.2 ditunjukkan penambahan *feature* untuk *moving average* hari ke-7, ke-14, ke-21, dan hari ke-28. Fungsi *moving average* pada penelitian ini adalah sebagai sampel data latih atau data independen sebagai X .

4.3 Pembuatan Model

Sebelum masuk ke pembuatan model dilakukan pembagian data dilakukan pemisahan data yang terdiri dari dua data yaitu X sebagai variabel independen dan y sebagai variabel dependen. Variabel independen merupakan sebuah variabel yang mempengaruhi variabel lain sedangkan variabel dependen adalah variabel yang dipengaruhi oleh variabel independen [29]. Data X terdiri dari hasil *moving average* yang sudah dihasilkan pada tahap *preprocessing*, sedangkan data y adalah data kasus harian COVID-19. Kemudian dilakukan pembagian data menjadi 2 data yaitu data uji dan data latih. Peneliti menggunakan beberapa rasio pembagian data yang dilakukan pengujian terhadap model. Rasio pembagian data yang digunakan adalah rasio 60:40, 70:30, 80:20, dan 90:10. Rasio tersebut akan di uji satu persatu terhadap model kemudian dibandingkan dengan satu sama lain lalu hasil evaluasi paling baik akan digunakan untuk pengujian model selanjutnya dan akan

digunakan sebagai peramalan.

Pemilihan SVR sebagai model yang dipakai dalam penelitian ini karena memiliki akurasi yang cukup baik dan cocok dengan data *time series* dengan data kontinu. Pada proses pembuatan model dilakukan pemilihan fungsi kernel dan parameternya. Terdapat tiga fungsi kernel pada penelitian ini yaitu, RBF, Polynomial, dan Linear. Dalam proses ini menghasilkan bahwa fungsi kernel Linear paling cocok dengan model data yang digunakan. Kemudian dilakukan GridSearchCV untuk mencari parameter terbaik dari sebuah model fungsi kernel. Parameter yang dihasilkan oleh GridSearchCV kemudian akan digunakan untuk membuat model baru yang lebih baik daripada model sebelum dilakukan pengoptimalan parameter.

4.4 Pengujian Akurasi Model

Setelah dilakukan pemodelan maka langkah selanjutnya adalah pengujian akurasi model menggunakan dua nilai eror yaitu nilai *mean square error* (MSE) dan *mean absolute percent error* (MAPE). Kedua nilai tersebut dihitung dari data uji dan data latih setelah dilakukan pemodelan. Suatu metode dikatakan sebuah metode yang baik untuk peramalan jika nilai MSE dan MAPE semakin kecil. Pada proses ini juga dilakukan perhitungan nilai *R-square* atau koefisien determinasi yang digunakan untuk mengetahui hubungan antara data dependen dan data independen. Jika nilai *R-square* semakin mendekati angka 1 maka dianggap metode yang baik [30].

4.5 Peramalan Pasien COVID-19

Tahap ini merupakan tahap terakhir dan merupakan munculnya hasil dari peramalan pertumbuhan pasien COVID-19. Ditahap ini peneliti memasukkan 30 data baru yang berisi data berupa tanggal dari 1 Februari 2021 sampai 2 Maret 2021. Setelah memasukkan data tanggal baru, peneliti akan memprediksi pertumbuhan pasien COVID-19 menggunakan SVR dengan fungsi kernel sesuai dengan hasil pemilihan yang diperoleh saat pembuatan model dan menggunakan parameter yang sesuai dengan hasil dari GridSearchCV.

5. Implementasi dan Pengujian

5.1 Pemodelan dengan Parameter Default

Pembangunan model diawali dengan membagi data set menjadi data uji dan data latih. Pada penelitian ini, penulis membagi data uji dan data latih menjadi beberapa rasio kemudian dicari rasio mana yang terbaik untuk model. Kemudian menentukan dan menyesuaikan model SVR sesuai dengan data latih yang ada. Setelah itu dilakukan pengujian pada data uji dengan menggunakan parameter *default* dan perhitungan nilai MAPE dan MSE. Berikut pada Tabel akan ditampilkan hasil nilai MAPE dan MSE dari data uji :

Tabel 3 Perbandingan Pemodelan Data Berdasarkan Pembagian Data

Pembagian Data	MSE	MAPE
60:40	89,57	16,2
70:30	93	14,3
80:20	104,1	12,2
90:10	137,8	14,2

Berdasarkan Tabel 3 dibandingkan nilai MSE dan nilai MAPE yang dihasilkan dari tiap data uji. Dimana jika semakin kecil nilai MSE maka dianggap sebagai nilai terbaik. Sehingga setelah dibandingkan dari hasil tersebut, maka peneliti menyimpulkan fungsi linear pada rasio 70:30 memiliki nilai MSE dan MAPE paling tepat berdasarkan tabel MAPE yang ada sehingga dapat disimpulkan jika nilai MAPE 14,3% memiliki kemampuan model peramalan baik.

5.2 Pemodelan dengan Tuned Parameter

Pada penelitian ini digunakan GridSearchCV untuk algoritma pengoptimalan parameter. GridSearchCV memiliki kelebihan salah satunya yaitu hampir tidak pernah (jatuh) dalam optimasi lokal. GridSearchCV juga sangat kompatibel dengan setiap algoritma pembelajaran, sehingga penulis menggunakan GridSearchCV untuk menemukan parameter terbaik. Setelah didapatkan fungsi kernel yang akan dipakai, kemudian dilakukan

proses pencarian parameter terbaik melalui GridSearch. Pertama dilakukan proses inialisasi parameter estimator untuk Kernel Linear, peneliti menggunakan skala $C = [0.1, 1, 10, 100]$ dan skala untuk $\gamma = [1, 0.1, 0.01, 0.001]$ yang akan dieksplorasi dan dievaluasi oleh GridSearchCV. Kemudian skala tersebut dicocokkan dengan model dan menghasilkan :

Tabel 4 Hasil Pengoptimalan Parameter Menggunakan GridSearch

	C	Gamma
SVR-Linear	0.1	1

Hasil dari Tabel V. 6 akan digunakan untuk membuat model SVR-Linear baru dan dicocokkan dengan data uji dan data latih. Kemudian dicari hasil evaluasi melalui nilai MAPE dan MSE-nya.

5.3 Evaluasi Nilai Akurasi Model

Setelah penggunaan parameter sesuai dengan hasil GridSearchCV, peneliti melakukan pengevaluasian atas hasil yang ada, diperoleh nilai MSE dan MAPE sebagai berikut:

Tabel 5 Perbandingan Nilai MSE dan MAPE dari Pemodelan Sebelum dan Sesudah *Tuning* Parameter

	MSE	MAPE(%)
Sebelum <i>Tuning</i> Parameter	93	14.3
Sesudah <i>Tuning</i> Parameter	89.6	13.9

Berdasarkan Tabel 5 ditunjukkan perbandingan nilai MSE dan MAPE sebelum dan sesudah *tuning* parameter dengan GridSearch. Sebelum *tuning* parameter diperoleh nilai MSE sebesar 93 dan nilai MAPE 14.3%. Sedangkan setelah *tuning* Parameter diperoleh hasil MSE sebesar 89.6 dan hasil nilai MAPE sebesar 13.9%. Dapat diartikan model memiliki kemampuan peramalan yang baik. Jika dibandingkan kedua hasil MSE dan MAPE sebelum dan sesudah *tuning* parameter maka, nilai MSE dan MAPE setelah dilakukan *tuning* parameter memiliki nilai yang lebih kecil daripada sebelum dilakukan *tuning* parameter. Sehingga dapat diartikan bahwa nilai MSE dan MAPE sesudah dilakukan *tuning* parameter memiliki nilai yang lebih baik dibandingkan sebelum dilakukan *tuning* parameter.

5.4 Hasil Peramalan

Berdasarkan uji yang dilakukan saat pembangunan model diatas, didapatkan data untuk pelatihan dan pengujian sebesar 70:30 dengan nilai MSE dan MAPE sebesar 93 dan 14.3% dapat dikatakan prediksi cukup akurat. Berikut hasil prediksi untuk tanggal 1 Februari 2021 sampai dengan 2 Maret 2021 yang didapat dari model SVR dengan menggunakan kernel Linear dengan parameter yang sudah disetel setelah proses GridSearchCV menjadi SVR-Linear dengan $C = 0.1$ dan $\Gamma = 1$ dan pembagian data sebesar 70:30, dapat dilihat pada Tabel 6 untuk hasil peramalan untuk 10 tanggal pertama pada data kasus COVID-19 di daerah Jawa Timur.

Tabel 6 Perbandingan Hasil Prediksi dan Data Aktual

Date	Aktual	Prediksi
01/02/2020	669	889.25
02/02/2020	809	892.76
03/02/2020	719	886.61
04/02/2020	595	878.30
05/02/2020	753	863.35
06/02/2020	693	843.55
07/02/2020	569	834.30
08/02/2020	666	830.43
09/02/2020	447	824.26

6. Kesimpulan

Adapun kesimpulan yang dapat diperoleh dari penelitian kali ini adalah :

- Peramalan pertumbuhan pasien COVID-19 di daerah Jawa Timur menggunakan metode *Support Vector Regression*. Menggunakan data yang diperoleh dari *Open Data* pada laman covid19.go.id dengan data dari tanggal 18 Maret 2020 sampai dengan 31 Januari 2021 berjumlah 319 data. Data berbentuk *csv* dengan 2 kolom berisikan tanggal dan jumlah pasien perhari. Pengolahan data menggunakan *Google Colaboratory* dengan bahasa pemrograman *python*. Pada tahap awal dilakukan tahap *preprocessing* untuk mengubah format tanggal dan menambahkan *featurization* yang digunakan sebagai data prediktor. Kemudian langkah selanjutnya adalah dilakukan pemodelan menggunakan SVR dan fungsi kernel. Pada tahap pemodelan dilakukan pencarian fungsi kernel terbaik yaitu fungsi kernel Linear. Selanjutnya dilakukan proses pencarian parameter terbaik menggunakan GridSearchCV. Setelah dilakukan pengujian model peramalan terhadap data uji dan data latih. Hasil ramalan kemudian divalidasi menggunakan metode MAPE dan MSE.
- Evaluasi hasil nilai akurasi diukur menggunakan MSE dan MAPE. Pemodelan SVR menggunakan fungsi kernel Linear dengan parameter optimal $C = 0.1$ dan $\gamma = 1$ menghasilkan nilai MSE sebesar 89.6 dan nilai MAPE sebesar 13.9%. Dengan hasil MSE dibawah angka nilai eror 100 dapat diartikan bahwa model layak untuk digunakan peramalan. Sedangkan untuk nilai MAPE sebesar 13.9% membuat model memiliki kemampuan peramalan yang baik.

Referensi:

- [1] S. Yulinnas, "RS Covid-19 di Jatim Hampir Penuh, Kapasitas Bakal Ditambah," *cnnindonesia.com*, 2021. [Online]. Available: <https://www.cnnindonesia.com/nasional/20210105131617-20-589632/rs-covid-19-di-jatim-hampir-penuh-kapasitas-bakal-ditambah>.
- [2] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *J. Autoimmun.*, vol. 109, no. February, p. 102433, 2020, doi: 10.1016/j.jaut.2020.102433.
- [3] Yuliana, "Corona virus disease (COVID-19)," *WELLNESS Heal. Mag.*, vol. 2, no. February, pp. 187–192, 2020, doi: 10.2307/j.ctvzxxb18.12.
- [4] H. Supriyanto, "the Implementation of System Dynamics in Hospital Services for Improving the Inpatient Rooms Utilization," *Kursor*, vol. 10, no. 1, 2019, doi: 10.28961/kursor.v10i1.206.
- [5] L. T. Tambun, "Tingkat Keterisian Tempat Tidur Pasien Covid-19 MASih Mencukupi," *Berita Satu*, 2020. [Online]. Available: <https://www.beritasatu.com/kesehatan/681319/tingkat-keterisian-tempat-tidur-pasien-covid19-masih%02mencukupi>.
- [6] C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, no. February, pp. 71–76, 2020, doi: 10.1016/j.ijssu.2020.02.034.
- [7] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta Biomed.*, vol. 91, no. 1, pp. 157–160, 2020, doi: 10.23750/abm.v91i1.9397.
- [8] P. VIKRAMA, P and Radha Krishna, "Data Mining Data mining," *Min. Massive Datasets*, vol. 2, no. January 2013, pp. 5–20, 2005, doi: 10.1002/9781118445112.stat06466.pub2.
- [9] S. Garcia, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining. Intelligent Systems Reference Library. 2015*, vol. 10. 2015.
- [10] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, vol. 5, no. c, pp. 15991–16005, 2017, doi: 10.1109/ACCESS.2017.2654247.
- [11] Y. Mahena, M. Rusli, and E. Winarso, "Prediksi Harga Emas Dunia Sebagai Pendukung Keputusan Investasi Saham Emas Menggunakan Teknik Data Mining," *Kalbiscientia J. Sains dan Teknol.*, vol. 2, no. 1, pp. 36–51, 2015.
- [12] H. A. Engelbrecht and M. Van Greunen, "Forecasting methods for cloud hosted resources, a comparison," *Proc. 11th Int. Conf. Netw. Serv. Manag. CNSM 2015*, pp. 29–35, 2015, doi: 10.1109/CNSM.2015.7367335.
- [13] C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, 2017, doi: 10.1016/j.renene.2016.12.095.
- [14] Q. Cao, R. Banerjee, S. Gupta, J. Li, W. Zhou, and B. Jeyachandra, "Data driven production forecasting using machine learning," *Soc. Pet. Eng. - SPE Argentina Explor. Prod. Unconv. Resour. Symp.*, 2016, doi: 10.2118/180984-ms.
- [15] G. Carleo *et al.*, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, no. 4, p. 45002, 2019, doi: 10.1103/RevModPhys.91.045002.
- [16] O. Bonita, L. Muflikhah, and R. K. Dewi, "Prediksi Harga Batu Bara Menggunakan Support Vector Regression (SVR)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6603–6609, 2018.
- [17] P. F. Pai, K. P. Lin, C. S. Lin, and P. T. Chang, "Time series forecasting by a seasonal support vector regression model," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4261–4265, 2010, doi:

- 10.1016/j.eswa.2009.11.076.
- [18] R. P. Furi, M. Si, and D. Saepudin, "Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus : IHSG dan JII," *ISSN 2355-9365 e-Proceeding Eng.*, vol. 2, no. 2, pp. 1–10, 2015.
- [19] M. Awad and R. Khanna, *Efficient Learning Machines*. 2015.
- [20] X. Huang, A. Maier, J. Hornegger, and J. A. K. Suykens, "Indefinite kernels in least squares support vector machines and principal component analysis," *Appl. Comput. Harmon. Anal.*, vol. 43, no. 1, pp. 162–172, 2017, doi: 10.1016/j.acha.2016.09.001.
- [21] D. I. Purnama and O. P. Hendarsin, "Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR)," *Jambura J. Math.*, vol. 2, no. 2, pp. 49–59, 2020, doi: 10.34312/jjom.v2i2.4458.
- [22] B. H. Prakoso, "Implementasi Support Vector Regression pada Prediksi Inflasi Indeks Harga Konsumen," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 155–162, 2019, doi: 10.30812/matrik.v19i1.511.
- [23] F. Milano, "A python-based software tool for power system analysis," *IEEE Power Energy Soc. Gen. Meet.*, pp. 1–5, 2013, doi: 10.1109/PESMG.2013.6672387.
- [24] S. K. Lam, A. Pitrou, and S. Seibert, "Numba: a LLVM-based Python JIT compiler," *Proc. Second Work. LLVM Compil. Infrastruct. HPC - LLVM '15*, pp. 1–6, 2015.
- [25] C. Ozgur, T. Colliau, G. Rogers, Z. Hughes, and E. B. Myer-Tyson, "MatLab vs. Python vs. R," *J. Data Sci.*, vol. 15, no. 3, pp. 355–372, 2017.
- [26] S. H. Edwards, D. S. Tilden, and A. Allevalo, "Pythy," pp. 641–646, 2014, doi: 10.1145/2538862.2538977.
- [27] N. K. Sukerti, "Peramalan Deret Waktu Menggunakan S-Curve dan Quadratic Trend Model," *Konf. Nas. Sist. Inform.*, pp. 592–597, 2015.
- [28] M. A. Maricar, "Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ," *J. Sist. dan Inform.*, vol. 13, no. 2, pp. 36–45, 2019.
- [29] L. Liana, "Penggunaan MRA dengan Spss untuk Menguji Pengaruh Variabel Moderating terhadap Hubungan antara Variabel Independen dan Variabel Dependen," *J. Teknol. Inf. Din.*, vol. XIV, no. 2, pp. 90–97, 2009.
- [30] S. A. Arismutia, "Pengaruh Penghargaan Finansial Dan Pertimbangan Pasar Kerja Terhadap Minat Mahasiswa Akuntansi Untuk Berkarir Menjadi Akuntan Publik (Studi Pada Mahasiswa Program Studi Akuntansi STIE INABA Bandung)," *J. Indones. Membangun*, vol. 16, no. 2, pp. 46–68, 2017.