

Penerapan Algoritma *Support Vector Regression* untuk Prediksi Jumlah Pasien Covid-19 di Provinsi Riau

Adyah Widiarni¹, Mustakim^{2,*}

^{1,2} Puzzle Research Data Technology (Predatech), Faculty of Science and Technology

^{1,2} Department of Information System, Faculty of Science and Technology

Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950324463@students.uin-suska.ac.id, ^{2,*}mustakim@uin-suska.ac.id

Submitted: 05/09/2021; Accepted: 25/09/2021; Published: 29/09/2021

Abstrak—Pada tahun 2019 tepatnya akhir Desember, terjadi wabah penyakit dengan penyebab yang tidak diketahui di Wuhan, Provinsi Hubei, China. World Health Organization memberi nama wabah penyakit tersebut adalah coronavirus yang disebabkan oleh Sindrom Pernapasan Akut Parah Coronavirus 2 (SARS-CoV-2) atau Covid-19. Covid-19 adalah suatu wabah penyakit yang sudah tersebar diberbagai daerah Indonesia seperti di Riau terdapat pada PT. Perkebunan Nusantara V (PTPN V). Maka dibutuhkan suatu cara agar dapat meningkatkan kesadaran dan kewaspadaan yakni dengan menyajikan suatu informasi menggunakan Data Mining dalam prediksi jumlah kasus dengan algoritma Support Vector Regression (SVR). Dilakukan proses prediksi menggunakan SVR dengan menentukan parameter SVR dan Kernel Linear. Algoritma SVR dapat memprediksi jumlah pasien Covid-19 pada 30 hari kedepan sehingga diperoleh tingkat Koefisien Korelasi (R) 85% dan nilai Mean Square Error (MSE) sebesar 0.196. Dari hasil percobaan tersebut terjadinya penurunan kasus pasien Covid-19 pada PT. Perkebunan Nusantara V di 30 hari kedepan, dengan perolehan nilai minimum sensitifitas yang terbaik adalah 0.09.

Kata Kunci: Covid-19; K-Fold Cross Validation; Prediksi; Support Vector Regression

Abstract—In 2019, at the end of December, there was an outbreak of a disease with an unknown cause in Wuhan, Hubei Province, China. The World Health Organization has named the outbreak of the disease as coronavirus caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) or Covid-19. Covid-19 is a disease outbreak that has spread in various regions of Indonesia, such as in Riau, at PT. Nusantara V Plantation (PTPN V). So we need a way to increase awareness and vigilance, namely by presenting information using Data Mining in predicting the number of cases with the algorithm Support Vector Regression (SVR). The prediction process is carried out using SVR by specifying the SVR and Kernel Linear parameters. The SVR algorithm can predict the number of Covid-19 patients in the next 30 days so that the Correlation Coefficient (R) level is 85% and the Mean Square Error (MSE) value is 0.196. From the results of the experiment, there was a decrease in cases of Covid-19 patients at PT. Perkebunan Nusantara V in the next 30 days, with the acquisition of the best minimum sensitivity value of 0.09.

Keywords: Covid-19; K-Fold Cross Validation; Prediction; Support Vector Regression

1. PENDAHULUAN

Tahun 2019 tepatnya akhir Desember, terjadi wabah penyakit dengan penyebab yang tidak diketahui di Wuhan, Provinsi Hubei, China. *World Health Organization* memberi nama coronavirus yang disebabkan oleh Sindrom Pernapasan Akut Parah Coronavirus 2 (SARS-CoV-2). Covid-19 dapat menginfeksi saluran pernapasan [1]. Pada Selasa, 11 Maret 2020 *World Health Organization* (WHO) mengumumkan nama resmi untuk wabah virus corona, yakni Covid-19 [2]. Penyebaran keberadaan Covid-19 telah mencapai sebagian besar negara di dunia, maka tanggal 16 Maret 2020 ditetapkan menjadi pandemik [3][4]. Covid-19 merupakan wabah penyakit yang sudah menyebar diberbagai daerah di Indonesia seperti di Provinsi Riau tepatnya di PT. Perkebunan Nusantara V (PTPN V).

Pada awalnya PTPN V merupakan Badan Usaha Milik Negara (BUMN) yang didirikan berdasarkan Peraturan Pemerintah Republik Indonesia (PPRI) No. 10 tahun 1996 tanggal 14 Februari 1996 tentang penyetoran modal Negara Republik Indonesia untuk pendirian Perusahaan. PTPN V merupakan Perusahaan Perkebunan Negara yang didirikan tanggal 11 Maret 1996 sebagai hasil konsolidasi kebun pengembangan PTP II, PTP IV, dan PTP V di Provinsi Riau. Perseroan mulai beroperasi secara efektif sejak tanggal 9 April 1996 dengan Kantor Pusat di Pekanbaru. Saat ini Kantor Pusat Perseroan berkedudukan di Jl. Rambutan No. 43 Pekanbaru, dengan berbagai unit usaha yang tersebar di berbagai Kabupaten di Provinsi Riau. Dilansir dari *website Kompas.com*, tanggal 22 September 2020 bahwasannya karyawan PTPN V terkonfirmasi positif Covid-19 [5]. Hingga tahun 2021 jumlah kasus Covid-19 PTPN V semakin meningkat sehingga Covid-19 harus dievaluasi dalam proses penanganannya. Berbagai upaya telah dilakukan oleh PTPN V untuk mencegah penyebaran Covid-19 dengan menerapkan Protokol Kesehatan.

Sesuai dengan Instruksi Presiden Nomor 6 Tahun 2020 yang berisi tentang sanksi untuk pelanggar protokol kesehatan. Sanksi berupa teguran secara tertulis, denda administratif serta penutupan usaha sementara bagi pelaku ekonomi [6]. Penerapan kebijakan tersebut bertujuan untuk mencegah penyebaran virus Covid-19, memastikan masyarakat agar tetap melaksanakan protokol kesehatan, dan mengharuskan masyarakat beradaptasi terhadap kebiasaan baru atau *new normal* [7]. Maka dibutuhkan suatu cara agar dapat meningkatkan kesadaran dan kewaspadaan yakni dengan menyajikan suatu informasi menggunakan Data Mining dalam prediksi jumlah kasus [8].

Data Mining merupakan teknik komputer yang digunakan untuk menganalisis serta mengekstraksi data sehingga menghasilkan pengetahuan [9]. Data Mining adalah suatu teknik memproses data dalam jumlah yang besar untuk menemukan pengetahuan yang menarik. Data mining terdiri dari 6 teknik diantaranya *Association Rule Mining* (ARM), Estimasi, Prediksi, Klasifikasi, Asosiasi dan Pengklusteran [10]. Salah satu algoritma yang dapat digunakan pada kasus prediksi ialah algoritma *Support Vector Regression* (SVR). Algoritma SVR digunakan untuk kasus regresi [11], diperkenalkan oleh Vapnik tahun 1995. Metode prediksi dapat mengatasi masalah pada data skala besar pada fase *training* dan mampu mengenali pola dari data *time series* disebut Algoritma SVR [12]. Algoritma SVR digunakan untuk memperoleh performansi yang baik dengan tingkat eror yang kecil untuk mengatasi overfitting [13]. Dalam beberapa kasus, algoritma SVR telah banyak diterapkan di kehidupan nyata dengan menghasilkan data yang akurat [14].

Algoritma SVR pernah dilakukan oleh Putri, Winahju dan Mashuri tahun 2020 dengan metode terbaik dalam memprediksi indeks batu bara PT XYZ adalah Algoritma SVR dengan *kernel-polynomial*. Algoritma SVR menggunakan kernel-polynomial menghasilkan nilai RMSE sebesar 0.619 dengan parameter sigma sebesar 0.100 dan nilai c sebesar 1. Tujuan penelitian tersebut untuk memprediksi variabel indeks batubara dengan variabel lain yang dapat mempengaruhinya dan mencari model terbaik yang terbentuk menggunakan nilai RMSE [15].

Penelitian terbaru yang dilakukan oleh Putra dan Kurniawati tahun 2021 dengan. Tujuan utama untuk memberikan informasi terbaru tentang wabah Covid-19 di Indonesia dengan memprediksi akumulasi kasus beberapa hari kedepan. Selain itu dapat meningkatkan kesadaran masyarakat tentang bahaya Covid-19. Berdasarkan penelitian tersebut algoritma SVR dapat memprediksi jumlah kasus Covid-19 selama 30 hari kedepan dengan tingkat akurasi 80% [16]. Dalam memprediksi algoritma SVR lebih akurat dibandingkan dengan Metode *Multilayer Perceptron* (MLP) dengan menggunakan data *time series* yang sedikit. Sedangkan saat menggunakan data *time series* yang cukup banyak menggunakan algoritma MLP lebih unggul dan akurat dibandingkan dengan algoritma SVR [17]. Algoritma SVR juga lebih unggul dibandingkan dengan algoritma ARIMA berdasarkan nilai *Root Mean Square Error* (RMSE). Dengan perolehan nilai RMSE pada algoritma SVR ialah 0.0533 sedangkan nilai RMSE algoritma ARIMA adalah 0.0672. Sehingga selisih kedua algoritma tersebut sebesar 0.0139 [13]. Berdasarkan uraian diatas serta didukung oleh penelitian sebelumnya, penelitian ini dilakukan prediksi jumlah kasus pasien Covid-19 di PTPN V dengan menggunakan penerapan SVR dengan tujuan utama adalah memprediksi jumlah kasus pasien Covid-19 untuk memberikan saran kepada PTPN V dalam mengurangi penyebaran Covid-19.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian terdiri dari tahapan Pengumpulan Data, tahap *Pre-Processing* Data, Pembagian Data *K-Fold Cross Validation*, Penerapan Algoritma SVR, serta tahapan Analisis dan Hasil Penelitian. Berikut Metodologi Penelitian yang dapat dilihat pada Gambar 1 [18].

2.2 Covid-19

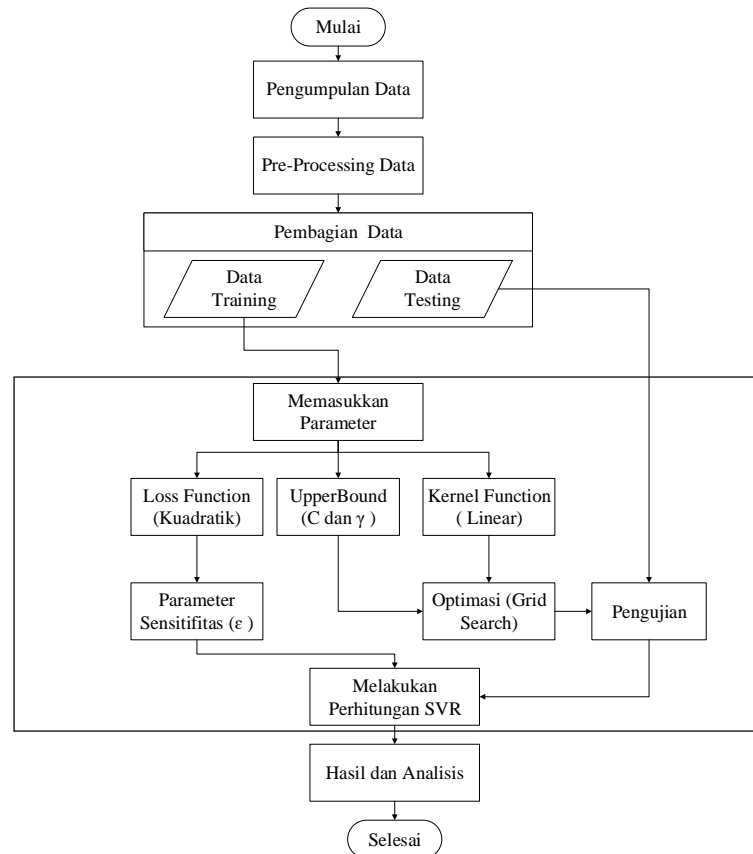
Covid-19 merupakan jenis penyakit yang disebabkan oleh Sindrom Pernapasan Akut Parah Coronavirus 2 (SARS-CoV-2). Disebut SARS-CoV-2 karena terdapat hubungan filogenetik dengan SARS-CoV [19]. Virus ini dapat menginfeksi saluran pernapasan seperti flu, *Middle East Respiratory Syndromen* (MERS), dan *Severe Acute Respiratory Syndrome* (SARS). Covid-19 ditemukan pertama kali di wilayah Wuhan, Hubei, China pada tahun 2019 [20]. Gejala umum yang ditimbulkan dari Covid-19 adalah demam dan kesulitan bernapas. Selain itu gejala terburuk dapat menyebabkan pneumonia, gagal ginjal, gangguan pernapasan akut hingga dapat menyebabkan kematian [21].

2.3 Data Mining

Data Mining merupakan teknik komputer yang digunakan untuk menganalisis serta mengekstraksi data sehingga menghasilkan pengetahuan [9]. Data Mining adalah suatu teknik memproses data dalam jumlah yang besar untuk menemukan pengetahuan yang menarik. *Knowledge Discovery in Database* (KDD) merupakan sebutan Data Mining dalam jurnal ilmiah [22]. Data mining terdiri dari 6 teknik diantaranya ARM, Estimasi, Prediksi, Klasifikasi, Asosiasi dan Pengklusteran [10].

2.4 Prediksi

Sebuah seni dan ilmu [23] yang digunakan untuk peramalan di masa depan pada jangka waktu pendek disebut Prediksi. Terdapat dua jenis prediksi yaitu prediksi kualitatif dan prediksi kuantitatif. Prediksi kualitatif bersifat subjektif karena berdasarkan pada pengalaman empiris, emosi manusia, dan lain-lain. Sedangkan prediksi kuantitatif bersifat objektif karena berasal dari data aktual serta data tersebut diolah dengan metode tertentu [12].



Gambar 1. Tahapan Penelitian

2.5 Support Vector Regression (SVR)

Algoritma SVR merupakan salah satu metode regresi dalam memprediksi ukuran, jumlah dan panjang penyebaran kasus covid-19, memiliki performansi tinggi sehingga dapat memberikan suatu prediksi dengan tingkat kesalahan yang kecil [16]. Algoritma SVR dapat diterapkan dalam memecahkan masalah non-linier seperti permasalahan dalam memprediksi *time series*. Algoritma SVR dapat digunakan dalam data mining diantaranya ialah klasifikasi, identifikasi, regresi dan analisis *time-series* [24].

Tujuan utama menerapkan algoritma SVR adalah untuk menemukan sebuah fungsi regresi $y = f(x)$ dengan nilai deviasi yang besar pada rentang ϵ sehingga dapat memprediksi nilai aktual untuk semua data latih. Sebuah parameter yang mewakili radius disekitar fungsi regresi disebut Variabel ϵ . SVR tidak memperlakukan adanya sebuah kesalahan jika kesalahan tersebut kurang dari ϵ . Berikut bentuk persamaan umum pada fungsi regresi adalah

$$f(x, w) = w^T \varphi(x) + b \tag{1}$$

Keterangan:

- $\varphi(x)$: titik didalam feature space F , hasil dari pemetaan x di dalam input space
- w : vector bobot, yang mempunyai dimensi l
- b : bias
- x : vektor input
- $f(x)$: fungsi regresi

Cara untuk meminimalkan fungsi resiko pada estimasi koefisien w dan b dapat didefinisikan pada persamaan (2), (3), dan (4).

$$R = \min \frac{1}{2} \|w\|^2 + C \frac{1}{\gamma} ((\sum_{i=1}^l L\epsilon(y_i, f(x_i)))) \tag{2}$$

dengan batasan

$$y_i - w\varphi(x_i) - b \leq \epsilon \quad w\varphi(x_i) - y_i + b \leq \epsilon \quad i=1,2,3,\dots,l \tag{3}$$

Dimana,

$$L\epsilon(y_i, f(x_i)) = 0$$



$$\{ 0, \text{ untuk } |y_i - f(x_i)| < \varepsilon \mid |y_i - f(x_i)| - \varepsilon, \text{ untuk } |y_i - f(x_i)| \geq \varepsilon \} \quad (4)$$

- $L\varepsilon$: Loss function yang bertipe ε -insentive *loss function*
- R : Fungsi Resiko (*Risk Function*)
- $\|w\|$: merupakan normalisasi w
- ε : deviasi atau derajat toleransi terhadap error disebut Epsilon
- C : Nilai penalti jika deviasi lebih besar dari batas error

Terdapat persamaan kernel linear yang umum digunakan dalam SVR. Berikut adalah bentuk persamaan umum kernel linear adalah [25].

$$K(x_i, x) = x_i^T x \quad (5) [25]$$

2.6 K- Fold Cross Validation

K-Fold Cross Validation adalah metode yang baik digunakan dan praktis untuk memprediksi atau perkiraan dari tingkat kesalahan berdasarkan algoritma pembelajaran. Metode ini dapat digunakan saat membagi data menjadi dua yaitu *training set* dan *test set*. Pada tahap validasi, metode ini dapat menggunakan nilai k sesuai keinginan dengan membagi data secara acak [25].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dalam penelitian ini data diperoleh dari PTPN V pada tanggal 19 Juli 2021 hingga 31 Agustus 2021 dengan jumlah kasus pasien Covid-19. Berdasarkan data yang diperoleh maka jumlah kasus pasien terkonfirmasi Covid-19 sebanyak 6876 orang. Pada tahap *preprocessing* data, *cleaning*, *transformasi* dan normalisasi terhadap data-data yang digunakan. Pada penelitian ini diperoleh data hasil *cleaning* seperti pada Tabel 1.

Tabel 1. Data Hasil *Cleaning*

| No | Unit Kerja Distrik | Bulan & Tanggal | | | | | | | | |
|-------|-----------------------|-----------------|-------|-------|-------|-------|-------|-------|---------|-------|
| | | Juli | | | | | | | Agustus | |
| | | 19 | 20 | 21 | 22 | 23 | 24 | | 30 | 31 |
| 1 | Kantor Direksi | 22 | 25 | 25 | 33 | 33 | 29 | | 13 | 13 |
| 2 | Kantor Direksi Barat | 4 | 4 | 4 | 4 | 5 | 5 | | 3 | 3 |
| 3 | Kantor Direksi Timur | 4 | 4 | 10 | 10 | 10 | 10 | | 6 | 5 |
| 4 | Kebun Air Molek 2 | 4 | 4 | 4 | 4 | 5 | 5 | | 3 | 3 |
| 5 | Kebun Lubuk Dalam | 4 | 4 | 4 | 4 | 2 | 5 | | 8 | 8 |
| 6 | Kebun Sei Batu Langka | 4 | 4 | 4 | 4 | 5 | 5 | | 3 | 3 |
| | | | | | | | | | | |
| 33 | Plasma TPU/TME | 4 | 4 | 4 | 4 | 5 | 5 | | 3 | 3 |

3.2 Normalisasi Data

Tahap Normalisasi yang diterapkan pada penelitian ini adalah menggunakan *min-max normalization*. Tujuan melakukan normalisasi yaitu untuk menghasilkan data maksimum (nilai yang tertinggi) dan minimum (nilai yang rendah) untuk mencapai nilai keseimbangan. Hasil dari normalisasi sata tersebut dapat dilihat pada Tabel 2.

Tabel 2 Hasil Normalisasi Data

| No | Unit Kerja Distrik | Bulan & Tanggal | | | | | | | | |
|----|----------------------|-----------------|--------|--------|--------|--------|--------|-------|---------|--------|
| | | Juli | | | | | | | Agustus | |
| | | 19 | 20 | 21 | 22 | 23 | 24 | | 30 | 31 |
| 1 | Kantor Direksi | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | | 1.0000 | 1.0000 |
| 2 | Kantor Direksi Barat | 0.1429 | 0.1250 | 0.1250 | 0.0938 | 0.1250 | 0.1429 | | 0.1667 | 0.1667 |
| 3 | Kantor Direksi Timur | 0.1429 | 0.1250 | 0.3750 | 0.2813 | 0.2813 | 0.3214 | | 0.4167 | 0.3333 |
| 4 | Kebun Air Molek 2 | 0.1429 | 0.1250 | 0.1250 | 0.0938 | 0.1250 | 0.1429 | | 0.1667 | 0.1667 |



| No | Unit Kerja Distrik | Bulan & Tanggal | | | | | | | | | |
|-------|-----------------------|-----------------|--------|--------|--------|--------|--------|-------|---------|--------|--|
| | | Juli | | | | | | | Agustus | | |
| | | 19 | 20 | 21 | 22 | 23 | 24 | | 30 | 31 | |
| 5 | Kebun Lubuk Dalam | 0.1429 | 0.1250 | 0.1250 | 0.0938 | 0.0313 | 0.1429 | | 0.5833 | 0.5833 | |
| 6 | Kebun Sei Batu Langka | 0.1429 | 0.1250 | 0.1250 | 0.0938 | 0.1250 | 0.1429 | | 0.1667 | 0.1667 | |
| | | | | | | | | | | | |
| 33 | Plasma TPU/TME | 0.1429 | 0.1250 | 0.1250 | 0.0983 | 0.1250 | 0.1429 | | 0.1667 | 0.1667 | |

3.3 Pembagian Data K-Fold Cross Validation

Pada tahap pembagian data ini, data dibagi menjadi *training set* dan *test set*. Sehingga diperoleh pembagian data berupa data training dan data testing. Jumlah data *time series* sebanyak 44 kolom data yang dibagi menggunakan teknik *K-Fold Cross Validation* menjadi 4 Fold, dan masing-masing pembagian sebanyak 11 kolom data. Dari jumlah data *time series* tersebut terdiri dari data training, output training, data testing dan output testing.

3.4 Uji Coba

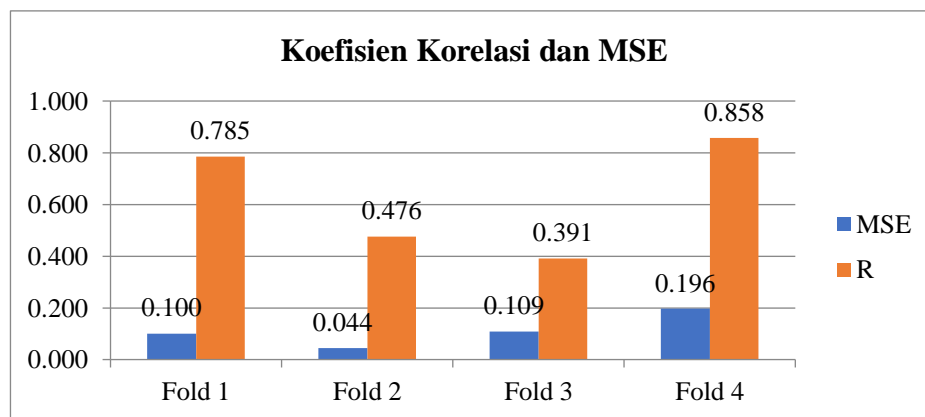
Uji Coba dilakukan setelah melakukan pembagian data menggunakan teknik *K-Fold Cross Validation* dengan menghitung berdasarkan data training dan data testing. Pada Percobaan ini menggunakan Kernel Linear dengan parameter C dan Gamma. Nilai parameter C yang digunakan ialah 0.125, 0.25, 0.5, 0.75, 1, 2, 4, 8, 16 dan nilai parameter Gamma yang digunakan adalah 0.25, 0.5, 0.75, 1, 2, 4, 8, 16, 32.

Berdasarkan nilai parameter C dan Gamma pada kernel Linear, maka menghasilkan nilai MSE dan R dan hasil prediksi disetiap Fold. Nilai MSE dan nilai R pada setiap fold dapat dilihat pada Tabel 3.

Tabel 3. Hasil Percobaan Kernel Linear pada Fold-1

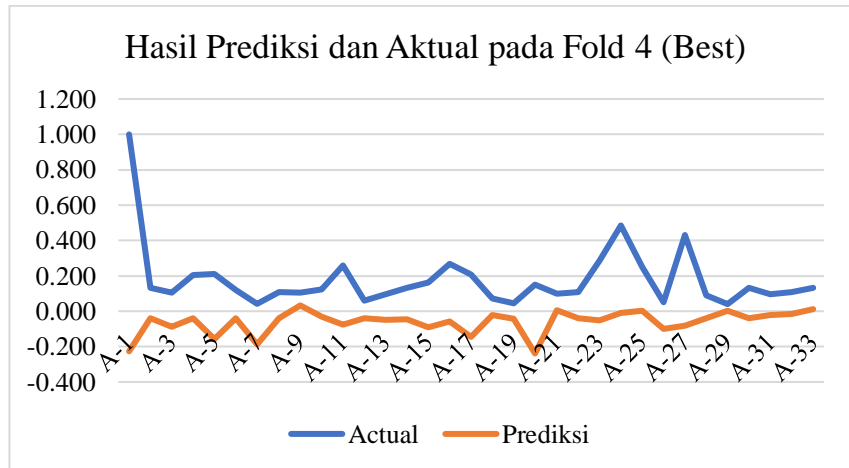
| Fold | C | 0.125 | 0.25 | 0.5 | 1 | 2 | 4 | 8 | 16 |
|--------|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| | | Gamma | 0.25 | 0.5 | 1 | 2 | 4 | 8 | 16 |
| Fold 1 | MSE | 0.101 | 0.101 | 0.100 | 0.100 | 0.101 | 0.141 | 0.174 | 0.316 |
| | R | 0.776 | 0.785 | 0.785 | 0.708 | 0.561 | 0.624 | 0.633 | 0.622 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Fold 4 | MSE | 0.142 | 0.146 | 0.150 | 0.175 | 0.196 | 0.229 | 0.072 | 0.072 |
| | R | 0.824 | 0.824 | 0.824 | 0.834 | 0.858 | 0.855 | 0.399 | 0.399 |

Dari hasil uji coba menggunakan nilai parameter C dan parameter Gamma, maka diperoleh nilai MSE dan R terbaik disetiap Fold. Pada fold-1 diperoleh nilai MSE (0.100) dan nilai R (0.785), fold-2 diperoleh nilai MSE (0.044) dan nilai R (0.476). Untuk fold-3 diperoleh nilai MSE (0.109), nilai R (0.391), dan fold-4 dnegan nilai MSE (0.196) dan nilai R (0.858). Hasil uji coba tersebut dapat dilihat pada Gambar 2.



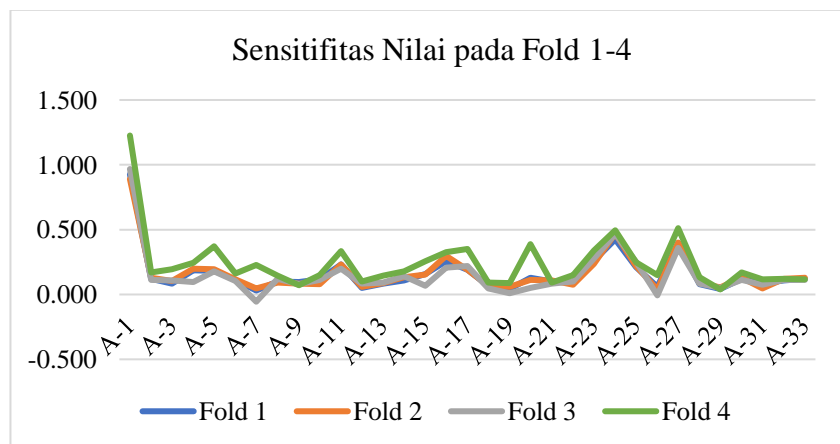
Gambar 2. Hasil nilai MSE dan R terbaik pada masing-masing Fold

Dari hasil nilai MSE dan R terbaik pada masing-masing fold maka diperoleh nilai terbaik berada pada fold-4. Dengan perolehan nilai MSE 0.196 dengan nilai akurasi sebesar 85%. Berikut adalah hasil prediksi dan nilai aktual pada fold-4 yang dapat dilihat pada Gambar 3



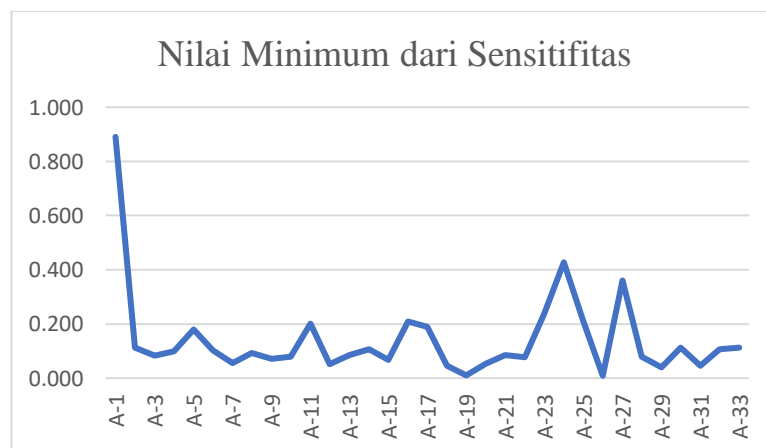
Gambar 3. Hasil Prediksi dan Aktual pada Fold-4

Berdasarkan percobaan yang telah dilakukan, terjadinya penurunan kasus yang terdapat pada masing-masing kebun di PTPN V dengan perolehan nilai menggunakan nilai aktual. Nilai aktual merupakan nilai yang diperoleh dari hasil rata-rata pada data disetiap periodenya. Nilai aktual tersebut digunakan untuk mendapatkan nilai sensitifitas pada masing-masing fold. Hingga diperoleh nilai sensitifitas pada masing-masing fold yang terdapat pada Gambar 4.



Gambar 4. Sensitifitas Nilai pada Fold 1-4

Pada Gambar 4 terdapat A-1 hingga A-33 yang menjelaskan nama kebun yang terdapat di PTPN V. Hingga diperoleh nilai sensitifitas terbaik pada masing-masing fold secara berturut-turut fold-1 (0.031), fold-2 (0.038), fold-3 (-0.056), dan fold-4 (0.039). Berdasarkan nilai sensitifitas yang telah didapatkan, maka menghasilkan nilai minimum yang digunakan untuk mencari nilai terbaik pada percobaan yang telah dilakukan.. Sehingga perolehan tersebut dapat dilihat pada Gambar 5.



Gambar 5. Hasil Perolehan Nilai Minimum dari Sensitifitas

4. KESIMPULAN

Berdasarkan hasil dari pengolahan data serta analisis yang dilakukan bahwa hasil prediksi jumlah pasien Covid-19 di PTPN V menggunakan algoritma SVR mengalami penurunan dari jumlah pasien Covid-19 sebelumnya. Walaupun mengalami penurunan, PTPN V harus tetap menerapkan protokol kesehatan secara ketat agar tidak terjadinya peningkatan jumlah kasus pasien Covid-19. Pada penelitian ini metode terbaik dalam memprediksi jumlah kasus pasien Covid-19 menggunakan Kernel Linear dan pembagian data menggunakan teknik *K-Fold Cross Validation*. Proses pembagian data tersebut menjadi 4 Fold. Sehingga menghasilkan tingkat akurasi (R) 85% dengan tingkat error (MSE) 0.196 yang diperoleh dari fold-4. Dari hasil percobaan tersebut terjadinya penurunan kasus pasien Covid-19 pada PT.Perkebunan Nusantara V di 30 hari kedepan, dengan perolehan nilai minimum yang terbaik adalah 0.09.

UCAPAN TERIMAKASIH

Ucapan terimakasih kepada PT. Perkebunan Nusantara V (PTPN V) yang memberikan kesempatan untuk melakukan penelitian, Keluarga besar *Puzzle Research Data Technology* dan teman-teman seperjuangan yang selalu memberikan masukan, dorongan dan semangat dalam penyelesaian penelitian ini.

REFERENCES

- [1] F. He, Y. Deng, and W. Li, "Coronavirus disease 2019: What we know?," *J. Med. Virol.*, vol. 92, no. 7, pp. 719–725, 2020, doi: 10.1002/jmv.25766.
- [2] R. R. Pratiwi, D. A. Artha, and H. Nurlaili, "Analisa Yuridis Penetapan Covid 19 Sebagai Kedaruratan Kesehatan Masyarakat Ditinjau Dari Peraturan Perundang-Undangan di Indonesia," *Inicio Legis*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.21107/il.v1i1.8827.
- [3] S. Ningsih, "Persepsi Mahasiswa Terhadap Pembelajaran Daring Pada Masa Pandemi Covid-19," *JINOTEP (Jurnal Inov. dan Teknol. Pembelajaran) Kaji. dan Ris. Dalam Teknol. Pembelajaran*, vol. 7, no. 2, pp. 124–132, 2020, doi: 10.17977/um031v7i22020p124.
- [4] R. R. Al Hakim, "Pencegahan Penularan Covid-19 Berbasis Aplikasi Android Sebagai Implementasi Kegiatan KKN Tematik Covid-19 di Sokanegara Purwokerto Banyumas," *Community Engagem. Emerg. J.*, vol. 2, no. 1, pp. 7–13, 2021, doi: 10.37385/ceej.v2i1.125.
- [5] I. T. Kompas. com, Kontributor Pekanbaru, "21 Karyawan PTPN V Riau Positif Covid-19, Ini Upaya Memutus Penyebarannya." [Online]. Available: <https://regional.kompas.com/read/2020/09/22/17272981/21-karyawan-ptpn-v-riau-positif-covid-19-ini-upaya-memutus-penyebarannya>.
- [6] J. Ilmiah and K. Sandi, "Pendahuluan Metode," vol. 10, pp. 356–363, 2021, doi: 10.35816/jiskh.
- [7] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sistemasi*, vol. 10, no. 1, p. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [8] D. Purwitasari *et al.*, "Time Series Analysis for Understanding Local Policy Impact of COVID-19 Cases in East Java," *CENIM 2020 - Proceeding Int. Conf. Comput. Eng. Network, Intell. Multimed. 2020*, no. Cenim 2020, pp. 52–57, 2020, doi: 10.1109/CENIM51130.2020.9297933.
- [9] M. C. M. Prasad, L. F. Florence, and A. Arya3, "A Study on Software Metrics based Software Defect Prediction using Data Mining and Machine Learning Techniques," *Int. J. Database Theory Appl.*, vol. 8, no. 3, pp. 179–190, 2015, doi: 10.14257/ijda.2015.8.3.15.
- [10] L. Kusriani, "Algoritma Data Mining." 2009.
- [11] H. ; sugito Septiningrum, Lutfia; Yasin, "PREDIKSI INDEKS HARGA SAHAM GABUNGAN MENGGUNAKAN SUPPORT VECTOR REGRESSION (SVR) DENGAN ALGORITMA GRID SEARCH," vol. 4, pp. 315–321, 2015.
- [12] L. Surtiningsih, M. T. Furqon, and S. Adinugroho, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2578–2586, 2018.
- [13] A. Kurniawati and A. Arima, "Analisis Prediksi Harga Saham PT. Astra International Tbk Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR)," vol. 20, no. September, pp. 417–423, 2021.
- [14] K. O. Akande, T. O. Owolabi, S. O. Olatunji, and A. Abdurraheem, "A Novel Homogenous Hybridization Scheme for Performance Improvement of Support Vector Machines Regression in Reservoir Characterization," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, pp. 1–10, 2016, doi: 10.1155/2016/2580169.
- [15] R. A. Putri, W. S. Winahju, and M. Mashuri, "Penerapan Metode Ridge Regression dan Support Vector Regression (SVR) untuk Prediksi Indeks Batubara di PT XYZ," vol. 9, no. 1, pp. 64–71, 2020.
- [16] A. B. Raharjo, Z. Z. Dinanto, D. Sunaryono, and D. Purwitasari, "Prediksi Akumulasi Kasus Terkonfirmasi Covid-19 Di Indonesia Menggunakan Support Vector Regression," *Techno.Com*, vol. 20, no. 3, pp. 372–381, 2021, doi: 10.33633/te.v20i3.5062.
- [17] R. Sistem, M. Perceptron, and L. Sederhana, "JURNAL RESTI Analisis Pola Prediksi Data Time Series menggunakan Support Vector," vol. 1, no. 10, pp. 282–287, 2019.
- [18] Mustakim, A. Buono, and I. Hermadi, "SUPPORT VECTOR REGRESSION UNTUK PREDIKSI," vol. 12, no. 2, pp. 179–188, 2015.
- [19] M. Ciotti, M. Ciccozzi, A. Terrinoni, W. C. Jiang, C. Bin Wang, and S. Bernardini, "The COVID-19 pandemic," *Crit. Rev. Clin. Lab. Sci.*, vol. 57, no. 6, pp. 365–388, 2020, doi: 10.1080/10408363.2020.1783198.



- [20] H. Lu, C. W. Stratton, and Y. W. Tang, “Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle,” *J. Med. Virol.*, vol. 92, no. 4, pp. 401–402, 2020, doi: 10.1002/jmv.25678.
- [21] M. L. Holshue *et al.*, “First Case of 2019 Novel Coronavirus in the United States,” *N. Engl. J. Med.*, vol. 382, no. 10, pp. 929–936, 2020, doi: 10.1056/nejmoa2001191.
- [22] G. Benoît, “Data mining,” *Annual Review of Information Science and Technology*, vol. 36. pp. 265–310, 2002, doi: 10.1002/aris.1440360107.
- [23] N. D. Maulana, B. D. Setiawan, and C. Dewi, “Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2986–2995, 2019.
- [24] S. Saadah, “Support Vector Regression (SVR) Dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR / USD,” vol. 5, no. 1, pp. 85–92, 2021.
- [25] M. Mustakim, C. Bella, and Y. R. Pratama, “Prediksi Jumlah Tunggakan Pajak Kendaraan Jatuh Tempo Menggunakan Algoritma Support Vector Regression,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind. 11*, no. November 2017, pp. 1–11, 2019.