

Prediksi Jumlah Perceraian Menggunakan Metode Support Vector Regression (SVR)

Eka Suryani Indra Septiawati*, Elvia Budianita, Fitri Insani, Lola Oktavia

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11950121684@students.uin-suska.ac.id, ²elvia.budianita@uin-suska.ac.id, ³fitri.insani@uin-suska.ac.id, ⁴lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: 11950121684@students.uin-suska.ac.id

Submitted: 23/11/2023; Accepted: 30/11/2023; Published: 30/11/2023

Abstrak—Peningkatan jumlah perceraian menjadi tantangan sosial yang semakin signifikan di Indonesia, termasuk di Kota Pekanbaru. Dampak dari perceraian ini pada kelompok remaja dapat memberikan pengaruh negatif terhadap kesejahteraan emosional dan psikologis mereka, serta kemampuan mereka dalam berinteraksi sosial dan dalam proses belajar. Penelitian ini menggunakan data perceraian bulanan dari tahun 2015 hingga April 2023 untuk melakukan analisis time series dan menerapkan metode Support Vector Regression (SVR) guna memprediksi jumlah perceraian di Kota Pekanbaru. Tiga jenis kernel SVR, yaitu linear, polynomial, dan radial basis function (RBF), dievaluasi dan dibandingkan untuk menemukan kernel dengan hasil Mean Squared Error (MSE) terbaik. Melalui analisis grid search, nilai optimal parameter untuk setiap kernel ditentukan. Hasil pengujian menunjukkan bahwa model SVR dengan kernel polynomial memberikan prediksi yang lebih akurat dengan MSE sebesar 0.010228, dibandingkan dengan kernel linear (MSE = 0.012767) dan kernel RBF (MSE = 0.010812).

Kata Kunci: Perceraian; Prediksi; Support Vector Regression (SVR); Kernel

Abstract—The increasing number of divorces poses an increasingly significant social challenge in Indonesia, including in the city of Pekanbaru. The impact of these divorces on the adolescent population can have negative effects on their emotional and psychological well-being, as well as their ability to interact socially and engage in the learning process. This study utilizes monthly divorce data from 2015 to April 2023 to conduct time series analysis and applies the Support Vector Regression (SVR) method to predict the number of divorces in the city of Pekanbaru. Three types of SVR kernels, namely linear, polynomial, and radial basis function (RBF), are evaluated and compared to find the kernel with the best Mean Squared Error (MSE) results. Through grid search analysis, optimal parameter values for each kernel are determined. The test results indicate that the SVR model with a polynomial kernel provides more accurate predictions with an MSE of 0.010228, compared to the linear kernel (MSE = 0.012767) and the RBF kernel (MSE = 0.010812).

Keywords: Divorce; Prediction; Support Vector Regression (SVR); Kernel

1. PENDAHULUAN

Perceraian adalah proses hukum yang mengakhiri ikatan pernikahan yang sah dan resmi antara suami dan istri, dan dampaknya terhadap keluarga dan anak-anak yang terlibat dapat sangat merugikan. Perceraian bisa berpengaruh negatif terhadap prestasi anak remaja, menimbulkan gangguan emosional dan psikologis yang mengganggu kemampuan mereka untuk fokus dan termotivasi dalam belajar. Selain itu, mereka mungkin mengalami kesulitan dalam berinteraksi dengan teman sebaya atau bahkan dalam membangun hubungan dengan orang lain, yang berdampak pada kemampuan mereka untuk bersosialisasi dan belajar dari pengalaman social [1]. Fenomena ini sering terjadi di negara kita yaitu Indonesia.

Kasus perceraian di Indonesia semakin meningkat secara signifikan. Data dari Badan Pusat Statistik Indonesia pada tahun 2021 mencatat adanya 447.743 kasus perceraian, menunjukkan peningkatan sebesar 53,50% dibandingkan tahun sebelumnya, di mana hanya terdapat 291.677 kasus perceraian. Lebih lanjut, dari jumlah tersebut, 337.343 kasus merupakan perceraian gugat yang diajukan oleh pihak istri dan telah diputuskan oleh pengadilan. Sementara itu, 110.440 kasus merupakan perceraian talak yang diajukan oleh pihak suami dan juga telah diputuskan oleh pengadilan. Data ini menggambarkan bahwa pada tahun 2021, terjadi lebih dari setengah juta kasus perceraian di Indonesia [2]. Penting untuk dicatat bahwa kasus-kasus perceraian ini terjadi di berbagai daerah, termasuk salah satunya adalah kota Pekanbaru.

Sebagai kota yang terus berkembang, Pekanbaru juga menghadapi peningkatan jumlah perceraian dari tahun ke tahun. Data yang diperoleh dari Situs Direktori Putusan Mahkamah Agung Republik Indonesia untuk Pengadilan Agama Pekanbaru menunjukkan tren ini, berdasarkan data per bulan yang mencakup periode dari 01 Januari 2015 hingga 30 April 2023. Pada tahun 2021, terdapat penurunan signifikan dalam jumlah perceraian, yaitu sebanyak 1700, tetapi pada tahun 2022, angkanya kembali naik menjadi 2202 kasus [3]. Oleh karena itu, diperlukan prediksi mengenai perkiraan jumlah perceraian di Pekanbaru guna memungkinkan perencanaan tindakan yang sesuai untuk mengatasi tren peningkatan ini.

Berbagai metode telah digunakan dalam upaya memprediksi data time series, salah satunya adalah *Support Vector Regression* (SVR). Metode ini merupakan metode regresi yang menggabungkan konsep pembelajaran mesin dan beragam jenis kernel digunakan, termasuk linear, polynomial, dan rbf untuk meramalkan nilai pada titik waktu selanjutnya. Beberapa penelitian terkait penggunaan kernel SVR, yaitu

penelitian Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode *Support Vector Regression* dengan kernel Polynomial menunjukkan hasil yang sangat akurat, dengan MAPE sebesar 7.95%, yang dapat disebut "Sangat Bagus". Sebagai perbandingan, kernel RBF menghasilkan MAPE sebesar 13.35%, yang masih dianggap "Bagus". Penentuan parameter C juga mempengaruhi hasil peramalan, dengan nilai C yang lebih kecil menghasilkan MAPE yang lebih baik [4]. Selanjutnya penelitian Prediksi Laju Pertumbuhan Penduduk Menggunakan Metode *Support Vector Regression* (Studi Kasus: Kota Malang). Hasil penelitian menunjukkan bahwa penggunaan kernel linear menghasilkan nilai MAPE sebesar 0,0985%, sementara penggunaan kernel Gaussian RBF menghasilkan nilai MAPE sebesar 0,38192% [5]. Selain itu, ada beberapa penelitian terkait telah dilakukannya prediksi jumlah perceraian. Sebagai contoh, sebuah penelitian pertama menerapkan Algoritma Regresi Linear Berganda untuk memproyeksikan jumlah perceraian di Pengadilan Agama Simalungun, dan hasilnya menunjukkan bahwa pada tahun 2019, estimasi jumlah perceraian adalah 961 [6]. Penelitian lainnya, seperti Pemodelan *Vector Autoregressive* (VAR) pada Data Jumlah Perceraian di Kota Pekanbaru, menghasilkan tingkat kesalahan prediksi sebesar 47,7648% [7].

Berbagai studi sebelumnya telah melakukan perbandingan antara berbagai metode analisis, seperti yang terlihat dalam penelitian yang mengevaluasi Metode Regresi Linear Berganda dan Metode *Support Vector Regression* (SVR) dalam menganalisis dampak cuaca terhadap produktivitas padi di Indonesia. Hasil penelitian tersebut menunjukkan bahwa SVR memiliki tingkat kesalahan yang lebih rendah daripada metode regresi linear, dengan perbedaan kesalahan (MAE) sekitar 0.7231. Selain itu, SVR juga menghasilkan *Root Mean Square Error* (RMSE) yang lebih rendah, yaitu sebesar 0.6672, dibandingkan dengan metode regresi linear. Hal ini menunjukkan superioritas metode SVR dalam analisis ini [8]. Penelitian lain yang membandingkan Metode *Ridge Regression* dan *Support Vector Regression* (SVR) dalam memprediksi indeks batubara di PT XYZ juga menghasilkan hasil yang serupa. SVR terbukti sebagai metode terbaik dengan RMSE terendah, yakni 0.619, ketika menggunakan kernel-polynomial dengan parameter sigma sebesar 0.100 dan nilai c sebesar 1 [9]. Selanjutnya, dalam penelitian yang fokus pada prediksi harga saham PT. Astra International Tbk menggunakan Metode *Autoregressive Integrated Moving Average* (ARIMA) dan SVR, ditemukan bahwa SVR lebih unggul dalam skema analisis mingguan maupun selama periode tiga bulan, dengan perbedaan (Gap) sekitar 0.013941 dibandingkan dengan Metode ARIMA. Hasil ini menekankan bahwa Metode SVR adalah pilihan yang lebih baik, terutama bagi para investor yang harus mengambil keputusan dalam konteks ini [10].

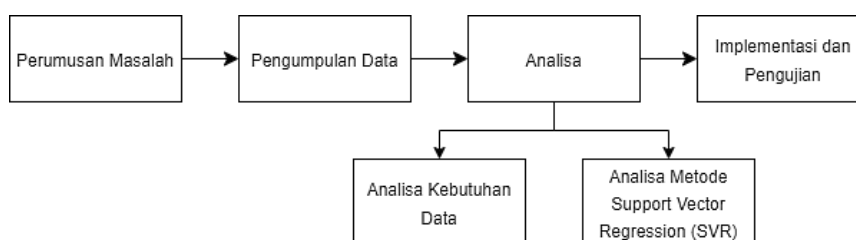
Selain itu, dalam beberapa penelitian, metode *Support Vector Regression* (SVR) telah digunakan. Sebagai contoh, dalam sebuah penelitian tentang Prediksi Harga Batu Bara menggunakan SVR, ditemukan bahwa penggunaan kernel Gaussian RBF menghasilkan rata-rata kesalahan persentase mutlak (MAPE) sekitar 9,64%, sementara kernel ANOVA menghasilkan sekitar 8,38%. Prestasi ini dianggap sangat baik karena diuji pada 48 dataset pelatihan dan diuji pada 12 dataset uji dengan parameter optimal seperti ϵ sebesar 0,00001, cLR sebesar 0,01, C sebesar 0,5, dan λ sebesar 0,5 pada kernel Gaussian, serta 1 pada kernel ANOVA [11]. Penelitian lain yang berkaitan dengan Prediksi Nilai Tukar Dolar Amerika Serikat terhadap Rupiah menggunakan SVR menghasilkan parameter SVR yang cocok dengan data pelatihan, dengan nilai R-squared (R²) sekitar 0,9223 dan *Root Mean Square Error* (RMSE) sekitar 54,3156. Tambahan, estimasi parameter model SVR mampu memprediksi nilai tukar jual harian Dolar Amerika terhadap Rupiah jika dibandingkan dengan data pengujian, dengan R² sekitar 0,5397 dan RMSE sekitar 66,8015 [12].

Penelitian ini dilakukan untuk menggunakan metode *Support Vector Regression* (SVR) dalam memproyeksikan angka perceraian dengan fokus pada menemukan kernel terbaik. Dengan harapan hasil penelitian ini akan bermanfaat, terutama dalam konteks penelitian sosial, untuk memahami tren sosial dan dampaknya pada masyarakat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah serangkaian proses yang harus diikuti dalam menjalankan sebuah penelitian. Proses ini mencakup berbagai tahapan yang dilakukan secara terstruktur, yang dimulai dari merumuskan masalah, mengumpulkan data, menganalisis data, hingga menyajikan hasil penelitian. Anda bisa melihat gambaran keseluruhan tahapan penelitian ini dalam Ilustrasi pada Gambar 1.



Gambar 1. Tahapan Penelitian

a. Perumusan Masalah

Langkah awal dalam melakukan penelitian melibatkan pembuatan rumusan masalah. Pada tahap ini, peneliti akan mencari permasalahan yang perlu diselidiki melalui berbagai sumber seperti observasi lingkungan, literatur, situs web, dan sebagainya. Setelah mengidentifikasi permasalahan, peneliti akan mendalami berbagai aspek yang berkaitan dengan masalah tersebut dan mencari cara untuk mengatasi masalah tersebut. Dalam konteks penelitian ini, rumusan masalahnya adalah “Bagaimana bentuk model *Support Vector Regression* (SVR) untuk menentukan serta memperoleh hasil prediksi terbaik dan memiliki nilai kesalahan terkecil”.

b. Pengumpulan Data

Dalam penelitian ini, tahap awal dalam pengumpulan data dimulai dengan eksplorasi sumber-sumber literatur, yang mencakup pencarian informasi dan data yang relevan dari berbagai sumber seperti buku, publikasi ilmiah, dan jurnal yang mendukung kerangka penelitian yang sedang dilaksanakan. Kemudian, data dikumpulkan secara manual dan diinput ke dalam format excel. Data ini diperoleh dari Website Direktori Putusan Mahkamah Agung Republik Indonesia untuk Pengadilan Agama Pekanbaru yang tersedia secara publik. Data yang berhasil dikumpulkan mencakup catatan harian dari 05 Januari 2015 hingga 27 April 2023, dengan total sebanyak 1142 data time series perceraian.

c. Analisa Kebutuhan Data

Analisis kebutuhan data adalah tahapan dalam suatu proyek atau sistem yang bertujuan untuk mengenali, mendefinisikan, dan memahami kebutuhan yang berkaitan dengan data. Tahap analisis ini bertujuan untuk memastikan bahwa data yang diperlukan untuk mencapai tujuan proyek atau sistem telah diidentifikasi dengan tepat, struktur data telah dirancang dengan benar, dan data tersebut dapat diakses, dikelola, dan dimanfaatkan secara efisien. Proses ini mencakup berbagai tahapan yang dilakukan secara terstruktur, yang dimulai dari *cleaning* data, data time series perbulan, menentukan fitur, normalisasi data, pembagian data.

1. *Cleaning* Data

Cleaning data adalah proses yang penting dalam pengelolaan dan analisis data. Tujuannya adalah untuk meningkatkan kualitas, akurasi, dan keandalan data dengan mengidentifikasi, memperbaiki, atau menghapus kesalahan dan inkonsistensi. Manfaatnya mencakup peningkatan kualitas data, akurasi analisis, pengurangan risiko keputusan yang salah, peningkatan efisiensi operasional, dan penyediaan dasar yang kuat untuk analisis lanjutan.

2. Pola Data Time Series

Pola data time series mengacu pada pendekatan yang digunakan untuk mengatur serta menganalisis data yang diukur atau dicatat pada interval waktu tertentu, seperti harian, dengan tujuan memvisualisasikan dan memahami perubahan data dalam skala waktu yang lebih besar, seperti bulanan. Time series memiliki beberapa karakteristik khusus yaitu tren, siklus, musiman, dan fluktuasi acak [13]. Dengan menggunakan pola data time series, kita dapat mengungkap hubungan yang mungkin tersembunyi dalam data, mengidentifikasi perubahan signifikan, serta membuat prediksi berdasarkan pola-pola yang telah diamati dalam rentang waktu tertentu.

3. Pemilihan Fitur

Pemilihan fitur adalah salah satu tahap penting dalam proses analisis data dan pembangunan model machine learning. Ini melibatkan pemilihan subset dari fitur atau variabel yang ada dalam dataset yang akan digunakan untuk melatih model atau melakukan analisis. Pemilihan fitur memiliki dampak signifikan pada kinerja model dan dapat membantu mengurangi kompleksitas, menghemat waktu komputasi, dan menghindari overfitting. Fitur yang digunakan dalam penelitian ini x1 hingga x5 sebagai jumlah perceraian bulan sebelumnya dan y sebagai nilai aktual.

4. Normalisasi Data

Normalisasi dilakukan untuk menyederhanakan perhitungan dan mengurangi tingkat kesalahan dalam proses komputasi. Dalam penelitian ini, normalisasi dapat dilakukan dengan menggunakan metode seperti Min-Max Scaling. Berikut adalah rumus perhitungan normalisasi data [14] :

$$\text{Normalisasi} = \frac{(x-x_{min})}{x_{max}-x_{min}} \quad (1)$$

Dimana x adalah nilai data yang akan dinormalisasi, x_{max} adalah nilai maksimum dari dataset, dan x_{min} adalah nilai minimum dari dataset.

5. Pembagian Data

Pembagian data adalah proses dalam analisis data dan pemodelan yang melibatkan pemisahan dataset menjadi subset yang berbeda. Subset ini mencakup set pelatihan dan set pengujian. Set pelatihan digunakan untuk melatih model sedangkan set pengujian digunakan untuk menguji kinerja model secara akhir.

d. Analisa Metode SVR

Proses ini mencakup berbagai tahapan yang dilakukan secara terstruktur, yang dimulai dari pemilihan kernel, pemilihan parameter kernel, pemodelan SVR, prediksi, evaluasi model. Proses tersebut diselesaikan menggunakan bahasa pemrograman Python pada Google Colab.

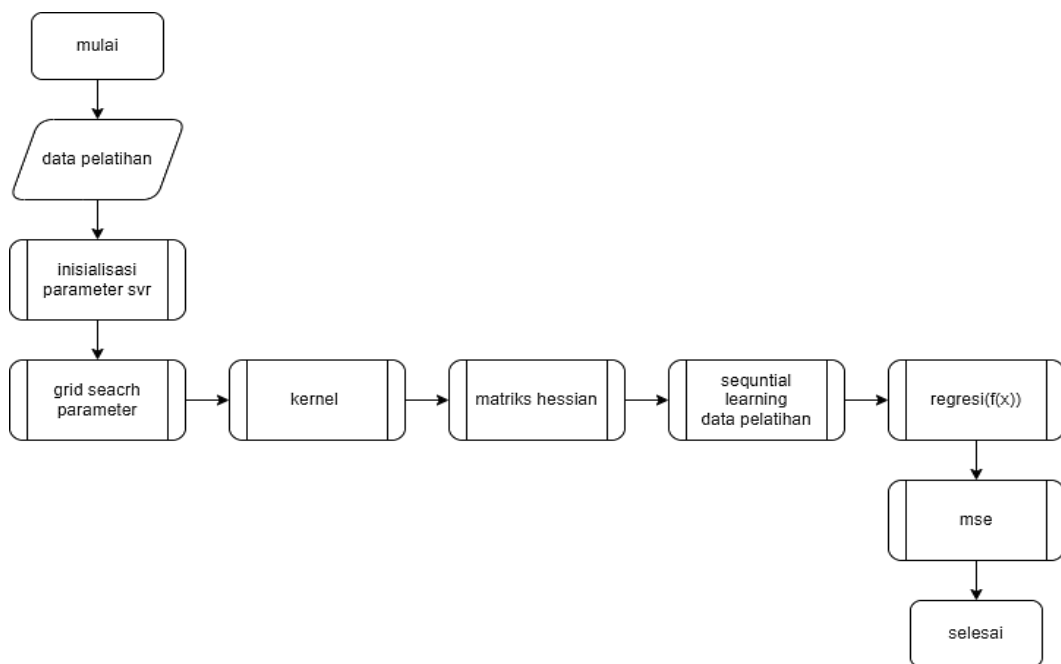
2.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) adalah sebuah turunan dari Support Vector Machine (SVM) yang digunakan dalam konteks regresi. Tujuannya yaitu memberikan prediksi yang sangat akurat dengan mengatasi masalah-masalah dalam regresi. Berbeda dengan SVM yang biasanya menghasilkan nilai bulat dalam kasus klasifikasi, Support Vector Regression (SVR) difokuskan pada penyelesaian kasus regresi dengan output berupa bilangan riil [15]. Konsep SVR memiliki potensi menghasilkan estimasi yang akurat karena mampu mengatasi masalah overfitting, yaitu kecenderungan di mana model terlalu sesuai dengan data pelatihan sehingga akurasi prediksi pada data baru dapat menurun. Overfitting terjadi ketika data selama proses pelatihan memberikan prediksi dengan akurasi yang hampir sempurna [16]. Sasaran utama algoritme SVR adalah menemukan Hyperplane terbaik, yang merupakan garis pemisah optimal. Hyperplane terbaik ditemukan dengan mengukur margin, yaitu jarak dari hyperplane ke data terdekat. Data yang paling dekat dengan margin disebut sebagai support vector.

Konsep dasar dari algoritme Support Vector untuk regresi ini melibatkan perhitungan nilai fungsi linear dengan mempertimbangkan Lagrange multiplier yang non-negatif, α_i , dan α_i^* . Solusi masalah ini biasanya ditemukan melalui pendekatan pemrograman kuadratik. Untuk mencapai aproksimasi yang optimal, digunakan formulasi yang telah dimodifikasi [17]. Selain itu, Support Vector Regression (SVR) juga mampu mengatasi masalah overfitting dan underfitting yang sering terjadi pada model regresi tradisional. Selain itu, SVR juga mampu menangani masalah data yang tidak teratur (nonlinear) dan noise pada data [18]. Nonlinearitas dalam pemodelan SVR ditunjukkan dalam persamaan 2.

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot K(x, x_i) + b \tag{2}$$

Dimana $f(x)$ adalah fungsi prediksi, n adalah jumlah pelatihan. Koefisien Lagrange direpresentasikan sebagai α_i^* dan α_i . Fungsi kernel $K(x_i, x)$ adalah fungsi kernel digunakan untuk mengukur kemiripan antara data pelatihan x_i dan data uji x . Selain itu, b merupakan nilai bias dalam model. Diagram alir penyelesaian masalah prediksi jumlah perceraian menggunakan algoritme SVR secara garis besar digambarkan pada Gambar 2.



Gambar 2. Diagram Alir SVR

Penyelesaian bagi regresi non-linear dapat diperoleh dengan lebih efisien dan hasil yang lebih optimal jika dibandingkan dengan penerapan metode yang lebih sederhana [19]. Algoritma ini melibatkan beberapa langkah:

a. Inisialisasi parameter

Dalam metode Support Vector Regression (SVR), terdapat beberapa parameter penting, seperti C dan ϵ (epsilon) yang berpengaruh dalam menentukan toleransi kesalahan.

b. Perhitungan matriks hessian

Berikut persamaan untuk menghitung matriks hessian

$$H_{ij} = K(x_i, x_j) + \lambda^2, \text{ untuk } i, j = 1, \dots, n \tag{3}$$

Dimana H_{ij} adalah matriks Hessian, sedangkan $K(x_i, x_j)$ adalah fungsi kernel. Selain itu, λ^2 diwakili sebagai variabel skalar.

c. Iterasi pada metode SVR

Perhitungan Kesalahan SVR dilakukan dengan menggunakan rumus berikut:

$$E_i = y_i - \sum_{j=1} (\alpha_j^* \alpha_j) R_{ij} \quad (4)$$

Dimana E_i adalah nilai kesalahan pada posisi ke- i , yang mengukur seberapa besar perbedaan antara nilai prediksi dan nilai aktual (y_i) pada data latih ke- i . α_j^*, α_j adalah nilai Lagrange Multiplier terkini, yang merupakan bagian dari metode optimasi untuk menemukan nilai optimum. Perubahan Lagrange Multiplier pada SVR dihitung menggunakan rumus:

$$\delta \alpha_i^* = \min\{\max(\gamma(E_i - \varepsilon), -\alpha_i^*), C - \alpha_i^*\} \quad (5)$$

$$\delta \alpha_i = \min\{\max(\gamma(E_i - \varepsilon), -\alpha_i), C - \alpha_i\} \quad (6)$$

Perubahan pada nilai α_i^* direpresentasikan sebagai $\delta \alpha_i^*$, sedangkan perubahan pada nilai α_i dirumuskan sebagai $\delta \alpha_i$. Parameter γ mengindikasikan nilai laju pembelajaran, sementara ε merupakan toleransi kesalahan. C digunakan untuk menentukan nilai kompleksitas. Perbaharuan Lagrange Multiplier pada SVR diberikan oleh rumus:

$$\alpha_i^* = \delta \alpha_i^* + \alpha_i^* \quad (7)$$

$$\alpha_i = \delta \alpha_i + \alpha_i \quad (8)$$

- d. Proses ini diulang secara iteratif, kembali ke langkah kedua, hingga mencapai kondisi iterasi maksimum atau $\max(|\alpha_i|) < \varepsilon$ dan $\max(|\alpha_i^*|) < \varepsilon$
- e. Perhitungan hasil peramalan pada SVR diperoleh setelah melakukan perhitungan dari persamaan regresi sebagaimana diformulasikan dalam persamaan (2).

2.2.1 Fungsi Kernel

Terdapat beberapa jenis fungsi kernel yang umum digunakan pada SVR, seperti kernel linear, kernel polinomial, dan kernel *Radial Basis Function* (RBF). Berikut ini adalah rumus dari beberapa varian kernel yang umum digunakan [20].

- a. Kernel linear

$$K(x, x_i) = x \cdot x_i \quad (9)$$

Di mana x dan x_i adalah data.

- b. Kernel polynomial

$$K(x, x_i) = (x \cdot x_i + c)^d \quad (10)$$

Di mana c (coef0) adalah konstanta dan d (degree) adalah derajat polynomial.

- c. Kernel Radial Basis Function (RBF)

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (11)$$

Di mana σ adalah parameter lebar fungsi kernel (gamma).

2.2.2 Grid search Parameter

Algoritma grid search adalah algoritma yang digunakan untuk menentukan dan mencari parameter optimal pada model SVR. Dalam algoritma ini, jangkauan parameter yang akan dioptimalkan dibagi ke dalam grid, dan selanjutnya, algoritma akan melintasi semua titik pada grid tersebut untuk mendapatkan parameter optimal. [21]. Proses *grid search* melibatkan pemeriksaan satu persatu terhadap kombinasi parameter, dengan membandingkan nilai kinerja terbaik yang dihasilkan oleh setiap parameter.

2.2.3 Evaluasi Model

Evaluasi dapat dilakukan dengan beberapa metrik evaluasi yang umum digunakan dalam regresi. Berikut salah satu metrik evaluasi yang sering digunakan pada SVR adalah:

- a. *Mean Squared Error* (MSE)

MSE adalah metode untuk mengestimasi rata-rata perbedaan kuadrat antara nilai aktual dan nilai yang diprediksi, yang dapat dihitung dengan menggunakan rumus evaluasi MSE berikut :

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_i - A_i)^2 \quad (12)$$

Di mana n merepresentasikan jumlah titik data. Setiap data diindeks P_i sebagai untuk nilai prediksi dan A_i untuk nilai aktual pada titik data ke- i . Semakin rendah nilai MSE, semakin tepat model dalam memprediksi data. Meskipun demikian, seperti halnya RMSE, MSE juga memiliki kekurangan dalam hal sensitivitas terhadap nilai-nilai ekstrem [14].

3. HASIL DAN PEMBAHASAN

3.1 Cleaning Data

Proses *cleaning* pada data time series perhari dilakukan untuk mengatasi nilai data yang hilang. Sehingga data awal yang berjumlah 3042 data berubah menjadi 1042 data.

3.2 Pola Data Time Series

Selanjutnya, data perhari diolah menjadi bentuk pola data time series bulanan. Maka, jumlah data tentang perceraian yang awalnya berjumlah 1042 data dapat diubah menjadi pola data bulanan dengan total 100 data.

3.3 Pemilihan Fitur

Tahapan berikutnya adalah melakukan pemilihan fitur dengan menggunakan fitur X yaitu X1 hingga X5 sebagai pengukuran jumlah perceraian dalam bulan sebelumnya, sedangkan Y mewakili nilai aktual. Dalam proses penamaan fitur ini, X1 menunjukkan data satu bulan sebelum nilai aktual, X2 menunjukkan data dua bulan sebelumnya, dan seterusnya hingga X5 yang menggambarkan data lima bulan sebelum nilai aktual. Variabel Y digunakan untuk merepresentasikan nilai aktual. Informasi mengenai hasil pemilihan fitur dapat ditemukan pada Tabel 1 di bawah.

Tabel 1. Pemilihan Fitur

No	X1	X2	X3	X4	X5	Y
1	85	23	162	110	153	142
2	23	162	110	153	142	66
3	162	110	153	142	66	148
...
98	169	179	96	0	0	0
99	179	96	0	0	0	0
100	96	0	0	0	0	0

3.4 Normalisasi Data

Dalam tahapan ini, nilai x1 dinyatakan dengan $x_{min} = 0$ dan $x_{max} = 655$. Normalisasi dilakukan pada nilai tersebut hingga proses perhitungan jumlah perceraian selesai, dan data hasilnya disertakan dalam Tabel 2.

Tabel 2. Normalisasi Data

No	X1	X2	X3	X4	X5	Y
1	0.129771	0.035115	0.247328	0.167939	0.233588	0.216794
2	0.035115	0.247328	0.167939	0.233588	0.216794	0.100763
3	0.247328	0.167939	0.233588	0.216794	0.100763	0.225954
...
98	0.258015	0.273282	0.146565	0.000000	0.000000	0.000000
99	0.273282	0.146565	0.000000	0.000000	0.000000	0.000000
100	0.146565	0.000000	0.000000	0.000000	0.000000	0.000000

3.5 Pembagian Data

Setelah proses normalisasi selesai, data tersebut dibagi menjadi dua kategori, yakni data pelatihan dan data pengujian. Pembagian ini dilakukan dengan alokasi 80% data untuk data pelatihan dan 20% data untuk data pengujian. Informasi terkait pembagian data ini terdapat pada tabel 3 dan 3.

Tabel 3. Data Pelatihan

No	X1	X2	X3	X4	X5	Y
1	0.129771	0.035115	0.247328	0.167939	0.233588	0.216794
2	0.035115	0.247328	0.167939	0.233588	0.216794	0.100763
3	0.247328	0.167939	0.233588	0.216794	0.100763	0.225954
...
78	0.138931	0.172519	0.204580	0.219847	0.209160	0.299237
79	0.172519	0.204580	0.219847	0.209160	0.299237	0.316031
80	0.204580	0.219847	0.209160	0.299237	0.316031	0.636641

Tabel 4. Data Pengujian

No	X1	X2	X3	X4	X5	Y
81	0.219847	0.209160	0.299237	0.316031	0.636641	0.247328
82	0.209160	0.299237	0.316031	0.636641	0.247328	0.001527
83	0.299237	0.316031	0.636641	0.247328	0.001527	0.175573
...
98	0.258015	0.273282	0.146565	0.000000	0.000000	0.000000
99	0.273282	0.146565	0.000000	0.000000	0.000000	0.000000
100	0.146565	0.000000	0.000000	0.000000	0.000000	0.000000

3.6 Hasil Parameter SVR

Langkah pertama yang diperlukan dalam melakukan prediksi menggunakan metode Support Vector Regression (SVR) yaitu menetapkan nilai parameter. Dalam kernel linear, parameter yang relevan adalah ϵ (Epsilon) dengan range nilai 0.01, 0.1, 0.2. Sedangkan untuk range nilai parameter C (Cost) yaitu 0.1, 1, 10, 100. Nilai-nilai ini ditentukan melalui algoritma grid search menggunakan python di google collab. Parameter yang dipilih adalah yang menghasilkan hyperplane terbaik. Hasil grid search yaitu $\epsilon = 0.1$ dan $C = 0.1$. Berikut proses grid search parameter SVR pada google collab yang dapat di lihat pada gambar 3.

```
[54] from sklearn.svm import SVR
      from sklearn.model_selection import GridSearchCV

      # Data pelatihan
      X_train = data[['X1', 'X2', 'X3', 'X4', 'X5']].values
      y_train = data['Y'].values
      # Menentukan parameter yang akan diuji
      param_grid = {'C': [0.1, 1, 10, 100], 'epsilon': [0.1, 0.2, 0.5]}
      # Membuat model SVR
      svr = SVR()
      # Menggunakan GridSearchCV untuk mencari kombinasi parameter terbaik
      grid_search = GridSearchCV(svr, param_grid, cv=5)
      grid_search.fit(X_train, y_train)
      # Menampilkan parameter terbaik
      best_params = grid_search.best_params_
      print("Parameter terbaik:", best_params)

Parameter terbaik: {'C': 0.1, 'epsilon': 0.1}
```

Gambar 3. Hasil Grid search Parameter SVR

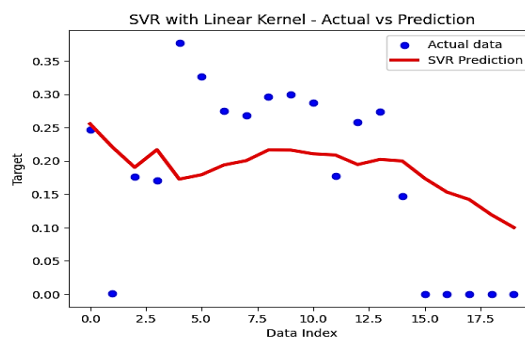
3.7 Hasil Uji Kernel Linear

Pada pengujian kernel linear menggunakan C yaitu 0.1 dan ϵ yaitu 0.1 yang didapat dari proses grid search parameter. Tahap selanjutnya yaitu menghitung nilai alpha dan bias. Hasil bias yang didapatkan yaitu 0.09664053. Sedangkan hasil alpha dapat di lihat pada gambar 4 berikut.

```
Koefisien Lagrange (alpha):
[[-0.1      -0.1      -0.1      -0.1      -0.1      -0.1
  0.1       0.1       0.1       0.1       0.1       0.1
  0.1      -0.1      -0.1      -0.1      -0.1      -0.1
  0.1       0.1       0.1       0.1       0.1       0.1
  0.1       0.1       0.1       0.1       0.1      -0.1
 -0.00660091 0.1       0.1       0.1       0.1       0.1
  0.1       0.1      -0.1      -0.1      -0.1      -0.1
 -0.09339909]]
```

Gambar 4. Hasil Alpha Kernel Linear

Setelah mengidentifikasi parameter optimal dan nilai alpha serta b, prediksi pada data uji dapat dilakukan. Hal ini tergambar secara visual dalam plot pada gambar 5 di bawah.



Gambar 5. Plot Data Prediksi Model SVR Kernel Linear

Dalam Gambar 3, terlihat bahwa pola data hasil prediksi hampir sejajar dengan data aktual. Hal ini mengindikasikan bahwa model SVR dengan kernel linear berhasil mengatasi overfitting dalam memproyeksikan jumlah perceraian. Evaluasi hasil model tersebut menunjukkan Mean Squared Error (MSE) sebesar 0.012767.

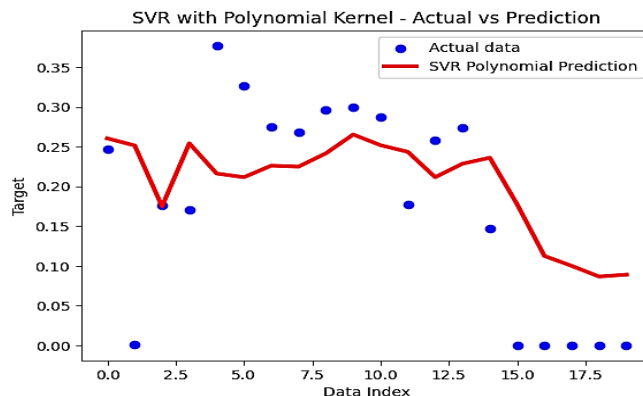
3.8 Hasil Uji Kernel polynomial

Proses selanjutnya sama dengan proses pembentukan SVR dengan kernel linear, langkah awal yang perlu dilakukan ketika menggunakan kernel polynomial adalah menetapkan nilai parameter tertentu. Parameter yang diperlukan untuk SVR dengan kernel ini melibatkan parameter ϵ , C, degree, dan coef0. Pada kernel ini nilai parameter yang di gunakan yaitu $\epsilon = 0.1$, C = 0.1. Sedangkan untuk degree dengan range nilai 2, 3, 4 dan coef0 dengan range nilai 0.1, 0.5, 1.0 yang akan di dipilih melalui grid search parameter. Hasil pemilihan untuk parameter degree yaitu 2 dan coef0 yaitu 0.5. Selanjutnya, menghitung nilai alpha dan bias. Untuk hasil bias yang di dapatkan yaitu 0.10017553. Hasil alpha dapat di lihat pada gambar 6 berikut.

```
Koefisien Lagrange (alpha):
[[-0.1      -0.03635971 -0.1      -0.1      -0.1      -0.09779646
  0.1       0.1       0.1       -0.05706732 -0.03346136 -0.04760931
 -0.02274515 -0.1      -0.1      0.1       -0.07263965 -0.01074043
 -0.03412483 -0.1      0.1       0.1       0.03100696 0.1
 -0.01062895 0.1      -0.03796417 0.1       0.1       0.1
 0.04273078 0.02351732 -0.1      -0.1      0.1       0.07667344
 0.1      -0.1      0.1       0.1       0.1      -0.1
 -0.01279116]]
```

Gambar 6. Hasil Alpha Kernel Polynomial

Setelah parameter optimal dan nilai alpha serta b diidentifikasi, langkah selanjutnya adalah melakukan prediksi pada data uji. Proses ini dapat dilihat secara visual melalui plot yang disajikan pada gambar 7 di bawah .



Gambar 7. Plot Data Prediksi Model SVR Kernel Polynomial

Pada Gambar 4 menunjukkan bahwa pola data hasil prediksi dengan model SVR menggunakan kernel polynomial juga mendekati nilai aktual, menunjukkan kemampuan model tersebut dalam mengatasi masalah overfitting dalam memprediksi jumlah perceraian. Hasil evaluasi model SVR dengan kernel linear mendapatkan MSE sebesar 0.010228.

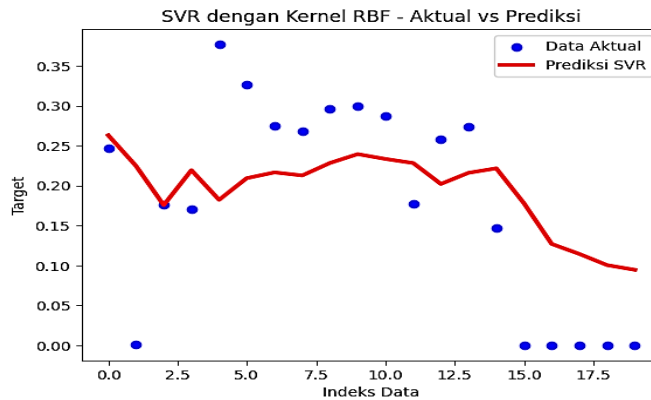
3.9 Hasil Uji Kernel Radial Basis Function (RBF)

Proses yang terjadi pada kernel RBF serupa dengan proses pembentukan SVR menggunakan kernel linear dan kernel polynomial. Langkah pertama yang perlu diambil saat menggunakan fungsi basis radial (RBF) adalah menetapkan nilai parameter tertentu, yang melibatkan parameter ϵ , C, dan gamma dalam konteks SVR dengan kernel RBF. Pada kernel ini nilai parameter yang di gunakan yaitu $\epsilon = 0.1$, C = 0.1. Sedangkan untuk gamma dengan range nilai 0.01, 0.1, 1, 10 yang akan di dipilih melalui *grid search* parameter. Hasil pemilihan untuk parameter gamma yaitu 1. Selanjutnya, menghitung nilai alpha dan bias. Untuk hasil bias yang di dapatkan yaitu 0.15818769. Sedangkan untuk hasil alpha dapat di lihata pada gambar 6 berikut.

```
Koefisien Lagrange (alpha):
[[-0.1      -0.1      -0.1      -0.1      -0.1      -0.01442628
  0.1       0.1       0.1       -0.1      -0.1      -0.1
 -0.1      0.1       0.1      -0.1      -0.1      -0.1
 -0.1      0.1       0.1       0.1       0.1       0.1
 0.1       0.1       0.1       0.1       -0.05518653 0.04669327
 0.1      -0.01763712 0.1      -0.1      0.1       0.1
 0.1      -0.1      0.1       0.1       0.1      -0.1
 -0.1      -0.05944335]]
```

Gambar 8. Hasil Alpha Kernel RBF

Setelah mengidentifikasi parameter optimal dan nilai alpha serta b, langkah selanjutnya adalah melakukan prediksi pada data uji. Proses ini dapat diilustrasikan melalui plot yang terdapat pada gambar 9 di bawah ini.



Gambar 9. Plot Data Prediksi Model SVR Kernel RBF

Pada Gambar 5 menunjukkan bahwa data aktual dibandingkan dengan hasil prediksi menggunakan model SVR dengan kernel RBF juga menunjukkan pola yang mendekati nilai aktual, memberikan indikasi bahwa model tersebut dapat mengatasi masalah *overfitting* dalam memprediksi jumlah perceraian. Hasil evaluasi model SVR dengan kernel linear mendapatkan MSE sebesar 0.010812.

Pada penelitian ini hasil evaluasi dari ketiga kernel akan dibandingkan dapat dilihat dalam tabel 6 di bawah.

Tabel 6. Hasil MSE

Kernel	MSE
Linear	0.012767
Polynomial	0.010228
Radial Basis Functtion (RBF)	0.010812

Dilihat dari Tabel 6, dapat disimpulkan bahwa Model SVR dengan kernel polinomial menunjukkan tingkat akurasi yang lebih baik dalam memprediksi jumlah perceraian. Selain itu, nilai Mean Squared Error (MSE) yang dihasilkan oleh model ini juga lebih rendah dibandingkan dengan Model SVR kernel linear dan Model SVR kernel radial basis function (RBF).

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, metode *Support Vector Regression* (SVR) telah digunakan untuk memprediksikan jumlah perceraian di Kota Pekanbaru. Data perceraian telah diubah menjadi rangkaian data waktu bulanan, dilakukan pemilihan fitur menggunakan data perceraian bulan sebelumnya, dan normalisasi data telah dilakukan. Selanjutnya, data telah dibagi menjadi data untuk pelatihan dan pengujian. Proses pemilihan parameter menggunakan *grid search* telah diterapkan untuk memilih model terbaik. Selajutnya pelatihan SVR terhadap data pelatihan dan melakukan prediksi pada data pengujian. Evaluasi menggunakan *Mean squared error* (MSE) menunjukkan bahwa kernel linear memiliki tingkat kesalahan yang lebih tinggi dibandingkan dengan kernel polynomial dan rbf. Kernel polynomial dan rbf menunjukkan kemampuan yang lebih baik, tetapi kernel polynomial memberikan prediksi yang lebih akurat dengan MSE yang lebih rendah yaitu 0.010228. Sedangkan hasil MSE kernel rbf sebesar 0.010812. Maka, ini menunjukkan bahwa model Polynomial memiliki kualitas prediksi yang lebih baik dibandingkan dengan model linear dan rbf. Penelitian berikutnya dapat menggunakan data selain data time series misalnya faktor-faktor yang mempengaruhi kasus perceraian seperti faktor ekonomi dan pendidikan.

REFERENCES

- [1] F. Jenz and N. C. Apsari, "Dampak Perceraian Orang Tua Pada Prestasi Anak Remaja," *J. Penelit. dan Pengabd. Kpd. Masy.*, vol. 2, no. 1, p. 1, 2021, doi: 10.24198/jppm.v2i1.33430.
- [2] C. M. Annur, "Kasus Perceraian Meningkat 53%, Mayoritas karena Pertengkaran," *www.katadata.co.id*, 2022. <https://databoks.katadata.co.id/datapublish/2022/02/28/kasus-perceraian-meningkat-53-mayoritas-karena-pertengkaran> (accessed Nov. 30, 2023).
- [3] Direktori Putusan Mahkamah Agung Republik Indonesia, "PUTUSAN UPLOAD PER TAHUN PA PEKANBARU PERCERAIAN," 2022. <https://putusan3.mahkamahagung.go.id/direktori/periode/tahunjenis/upload/pengadilan/pa->

- pekanbaru/kategori/perceraian.html (accessed Nov. 30, 2023).
- [4] N. P. R. Apriyanti, I. K. G. D. Putra, and I. M. S. Putra, "Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 2, p. 72, 2020, doi: 10.24843/jim.2020.v08.i02.p01.
- [5] A. K. Dewi, M. T. Furqon, and R. C. Wihandika, "Prediksi Laju Pertumbuhan Penduduk Menggunakan Metode Support Vector Regression (Studi Kasus: Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 1, pp. 421–427, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] R. A. Samosir, M. F. Rozy, and A. P. Windarto, "Penerapan Algoritma Regresi Linier Berganda dalam Mengestimasi Jumlah Perceraian di Pengadilan Agama Simalungun," *TIN Terap. Inform. Nusant.*, vol. 2, no. 1, pp. 16–20, 2021, [Online]. Available: <https://ejournal.seminar-id.com/index.php/tin>
- [7] A. P. Desvina, "Pemodelan Vector Autoregressive (Var) untuk Data Jumlah Perceraian di Kota Pekanbaru," *J. Sains Mat. dan Stat.*, vol. 7, no. 2, pp. 97–107, 2021, doi: 10.24014/jsms.v7i2.13765.
- [8] R. et all Yulistiani, "Metode Regresi Linier Berganda dan SVR dalam Menentukan Tingkat Pengaruh Cuaca Terhadap Produktivitas Padi di Indonesia," *Semin. Nas. Inform. Sist. Inf. Dan Keamanan Siber*, no. 2017, pp. 29–34, 2018.
- [9] 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," *J. Sains dan Seni ITS*, vol. 9, no. 1, pp. 64–71, 2020, doi: 10.12962/j23373520.v9i1.51021.
- [10] 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)," *J. Ilm. Komputasi*, vol. 20, no. 3, pp. 417–423, 2021, doi: 10.32409/jikstik.20.3.2732.
- [11] O. Bonita, "Prediksi Harga Batu Bara Menggunakan Support Vector Regression (SVR)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6603–6609, 2018.
- [12] S. S. M. Evy Sulistianingsih, "Prediksi Nilai Tukar Dolar Amerika Serikat Terhadap Rupiah Dengan Metode Support Vector Regression (Svr)," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.26418/bbimst.v8i1.30503.
- [13] S. C. Hillmer and W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods.*, vol. 86, no. 413. 1991. doi: 10.2307/2289741.
- [14] I. Cholissodin and A. A. Soebroto, "AI , MACHINE LEARNING & DEEP LEARNING (Teori & Implementasi)," no. July 2019, 2021.
- [15] Trivusi, "Algoritma Support Vector Regression (SVR): Jenis SVM untuk Regresi," 2022. <https://www.trivusi.web.id/2022/08/algoritma-svr.html>
- [16] N. D. Maulana, B. D. Setiawan, and C. Dewi, *Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery)*, vol. 3, no. 3. 2019.
- [17] T. M. N. Utami, D. C. R. Novitasari, F. Setiawan, N. Ulinuha, and Y. Farida, "Tide Prediction in Prigi Beach using Support Vector Regression (SVR) Method," *Sci. J. Informatics*, vol. 8, no. 2, pp. 194–201, 2021, doi: 10.15294/sji.v8i2.28906.
- [18] D. I. Ridwan, C. Setianingsih, and M. A. Murty, "Prediksi Penggunaan Energi Listrik Dengan Menggunakan Metode Support Vector Regression," vol. 8, no. 6, pp. 2–3, 2021.
- [19] W. Anugerah, "Perbedaan SVM Linear dan Non Linear: Mana yang Lebih Baik?," 2023. <https://www.localstartupfest.id/faq/perbedaan-svm-linear-dan-non-linear/>
- [20] M. L. Subiyanto, Y. Amanda, and M. N. Fachrian, "Peramalan Kasus Harian Monkeypox Dunia Dengan Pendekatan Support Vector Regression," pp. 27–36, 2022.
- [21] A. N. Safira, B. Warsito, and A. Rusgiyono, "Analisis Support Vector Regression (Svr) Dengan Algoritma Grid Search Time Series Cross Validation Untuk Prediksi Jumlah Kasus Terkonfirmasi Covid-19 Di Indonesia," *J. Gaussian*, vol. 11, no. 4, pp. 512–521, 2023, doi: 10.14710/j.gauss.11.4.512-521.