



Pemodelan Support Vector Regression Dengan Optimasi Grid Search Untuk Prediksi Curah Hujan Di Kabupaten Berau

Fadilla¹, Nariza Wanti Wulan Sari^{1*}, Meirinda Fauziyah¹, Erlyne Nadhilah Widyaningrum¹, Sri Wahyuningsih¹

¹Statistika, Universitas Mulawarman, Indonesia

*Email Penulis Korespondensi: nariza@fmipa.unmul.ac.id

Abstrak: Fluktuasi curah hujan yang terjadi dari waktu ke waktu dapat meningkatkan potensi bencana hidrometeorologi, seperti banjir dan tanah longsor. Kondisi tersebut menunjukkan pentingnya pengembangan model yang dapat memberikan prediksi curah hujan secara akurat. Penelitian ini bertujuan menerapkan Support Vector Regression (SVR) dengan kernel polynomial dan optimasi grid search untuk memprediksi curah hujan di Kabupaten Berau. Data yang digunakan berupa curah hujan bulanan sebagai variabel respon serta suhu udara, kelembapan udara, tekanan udara, kecepatan angin, dan penyinaran matahari sebagai variabel prediktor. Data mencakup periode Januari 2014 hingga Desember 2025 dan dinormalisasi menggunakan metode min-max normalization. Optimasi grid search dilakukan dengan menguji nilai cost sebesar 0,5; 1; 2; dan 4 serta degree sebesar 1 dan 2. Pembentukan model dilakukan pada tiga skenario pembagian data training dan testing, yaitu 70:30, 80:20, dan 90:10. Evaluasi model menggunakan RMSE dan MAPE. Hasil analisis menunjukkan bahwa model dengan performa terbaik diperoleh pada proporsi 90:10, dengan RMSE training sebesar 85,5421 dan testing sebesar 86,8895. Nilai MAPE yang dihasilkan masing-masing sebesar 45,3880 dan 61,4889. Prediksi model mampu menggambarkan kecenderungan umum curah hujan aktual, sehingga SVR dengan kernel polynomial dan optimasi grid search dapat digunakan sebagai alternatif dalam prediksi curah hujan di Kabupaten Berau.

Kata Kunci: Curah hujan, grid search, kernel polynomial, RMSE, support vector regression.

Abstract: Rainfall fluctuations over time can increase the risk of hydrometeorological disasters, such as floods and landslides. This condition highlights the importance of developing a model capable of accurately predicting rainfall. This study aims to apply Support Vector Regression (SVR) with a polynomial kernel and grid search optimization to predict rainfall in Berau Regency. The data consist of monthly rainfall as the response variable and air temperature, humidity, air pressure, wind speed, and solar radiation as predictor variables. The dataset covers the period from January 2014 to December 2025 and was normalized using the min-max normalization method. Grid search optimization was performed by testing cost values of 0.5, 1, 2, and 4 and degree values of 1 and 2. The model was developed using three training and testing data proportions, namely 70:30, 80:20, and 90:10. Model performance was evaluated using RMSE and MAPE. The results indicate that the best-performing model was obtained using the 90:10 data proportion, with training and testing RMSE values of 85.5421 and 86.8895, respectively. The corresponding MAPE values were 45.3880 and 61.4889. The predicted values were able to represent the general pattern of actual rainfall. Therefore, SVR with a polynomial kernel and grid search optimization can be considered an alternative approach for rainfall prediction in Berau Regency.

Keywords: Rainfall, grid search, polynomial kernel, RMSE, support vector regression.

Pendahuluan

Kemajuan teknologi informasi yang semakin pesat telah memunculkan berbagai inovasi dalam kehidupan, salah satunya adalah kecerdasan buatan (Artificial Intelligence atau AI) (Huda dkk., 2025). AI merupakan pendekatan yang memungkinkan komputer melakukan berbagai tugas secara cerdas serta mengintegrasikan data dalam skala besar dan kompleks (Lazarusli, 2025). Salah satu bidang dalam AI yang banyak digunakan adalah machine learning, yaitu pendekatan

yang memungkinkan sistem mengenali pola berdasarkan data untuk menghasilkan prediksi maupun klasifikasi (Nurhalizah, Ardianto, & Purwono, 2024). Proses pemodelannya melibatkan data training untuk membentuk model serta data testing untuk mengukur kinerjanya pada data yang belum pernah digunakan (Retnoningsih & Pramudita, 2020).

Support Vector Machine (SVM) merupakan salah satu algoritma supervised learning yang digunakan dalam machine learning (Hasibuan & Karim, 2022). SVM menggunakan hyperplane optimal untuk memodelkan hubungan antarvariabel, sedangkan penerapannya pada permasalahan regresi dikenal sebagai Support Vector Regression (SVR) (Nurhalizah, Ardianto, Purwono, 2024). SVR digunakan untuk membangun model regresi yang mampu merepresentasikan hubungan data dengan meminimalkan kesalahan prediksi (Kusuma & Kudus, 2022). Dalam menangani hubungan nonlinier, SVR dapat menggunakan beberapa fungsi kernel, seperti linear, polynomial, dan Radial Basis Function (RBF) (Rais, 2022). Kinerja SVR sangat dipengaruhi oleh pemilihan hyperparameter, seperti cost (C), epsilon (ϵ), dan parameter kernel, sehingga pemilihan parameter yang tidak tepat dapat menyebabkan overfitting maupun underfitting (Kamaluddin, 2025). > Untuk menentukan kombinasi hyperparameter yang paling sesuai, dapat digunakan metode grid search dengan menguji berbagai kombinasi parameter secara sistematis (Iriananda, Budiawan, Rahman, & Istiadi, 2024).

Penerapan SVR telah dilakukan pada berbagai bidang. Ruliana, Rais, Marni, & Ahmar (2024) menerapkan SVR dengan kernel RBF dan optimasi grid search untuk memprediksi inflasi bulanan di Kota Makassar dan memperoleh RMSE sebesar 0,029. Mashita, Basuki, dan Hayatin (2020) menggunakan kernel linear, RBF, dan polynomial untuk meramalkan konsumsi listrik dengan proporsi data training dan data testing 80:20, dengan kernel polynomial menghasilkan RMSE terkecil sebesar 149.761,2. Sementara itu, Patiallo, Fathurahman, Prangga, & Nadhilah (2025) menerapkan SVR dengan kernel polynomial, penelitian ini menggunakan data bulanan sebagai dasar dalam melakukan prediksi curah hujan di Kabupaten Berau pada bulan Januari 2014–Desember 2023 dengan proporsi 80:20 dan memperoleh RMSE sebesar 0,1786. Hasil penelitian tersebut menunjukkan bahwa SVR dapat diterapkan untuk berbagai permasalahan prediksi, termasuk prediksi curah hujan.

Curah hujan menunjukkan banyaknya air hujan yang jatuh di suatu wilayah selama jangka waktu tertentu dan dinyatakan dalam satuan milimeter (mm) (Kamil, 2021). Kabupaten Berau di Kalimantan Timur termasuk wilayah yang memiliki potensi terdampak bencana hidrometeorologi, khususnya banjir yang dapat dipicu oleh intensitas hujan yang tinggi. Bencana banjir pernah terjadi di beberapa kecamatan di Kabupaten Berau, seperti Segah, Teluk Bayur, dan Tanjung Redeb pada tahun 2020 (Karondia, Fitriani, & Hizkia, 2022). Selain itu, banjir di kawasan DAS Kelay pada 15 Mei 2021 menjadi salah satu kejadian banjir paling parah dalam 20 tahun terakhir (Musnanda, 2022). Oleh karena itu, prediksi curah hujan diperlukan untuk mendukung penyediaan informasi dalam mengantisipasi kemungkinan terjadinya bencana hidrometeorologi, seperti banjir dan tanah longsor. (Setiawan, 2021).

Metodologi

AI dapat dipahami sebagai kemampuan suatu sistem dalam mengolah informasi dari lingkungan, mempelajari pola berdasarkan data, kemudian menyesuaikan hasil pembelajarannya untuk mencapai tujuan tertentu (Siahaan, Jasa, Anderson, Roisana, Lim, & Yudianto, 2020). Dalam perkembangannya, terdapat beberapa pendekatan yang digunakan dalam AI mencakup Fuzzy Logic (FL), Evolutionary Computing (EC), serta machine learning.

Machine learning merupakan bagian dari AI yang memungkinkan komputer mempelajari pola berdasarkan data sehingga dapat menghasilkan keputusan atau prediksi secara otomatis tanpa instruksi pemrograman yang spesifik (Nurhalizah, Ardianto & Purwono, 2024). Berdasarkan proses pembelajarannya, machine learning terdiri atas supervised learning, unsupervised learning, dan reinforcement learning (Wijoyo, Saputtra, Ristanti, Sya'ban, Amalia & Febriansyah, 2024).

Supervised learning adalah pendekatan pembelajaran mesin yang memanfaatkan data berlabel sebagai dasar pelatihan agar model dapat mempelajari hubungan antara input dan hasil yang diharapkan. Perbedaan antara hasil prediksi dan label aktual disebut loss yang digunakan sebagai umpan balik untuk memperbaiki model (Elwirehardja, Suparyanto & Pardamean, 2023). Unsupervised learning merupakan metode machine learning yang bertujuan mengidentifikasi pola

atau struktur tertentu dalam data tanpa menggunakan variabel keluaran yang telah diberi label. (Nurhalizah, Ardianto & Purwono, 2024). Reinforcement learning merupakan metode pembelajaran yang memungkinkan sistem memperoleh pengetahuan melalui proses coba-coba dan interaksi dengan lingkungan yang dinamis.

Support Vector Regression (SVR) adalah pengembangan dari SVM yang diterapkan pada analisis regresi untuk menghasilkan nilai prediksi. Metode ini berupaya membentuk fungsi regresi yang mampu menggambarkan hubungan data dengan kesalahan prediksi yang rendah. (Kusuma & Kudus, 2022). Berbeda dengan SVM yang digunakan untuk klasifikasi, SVR digunakan untuk memprediksi nilai kontinu (Suyono, Kusriani, & Arief, 2022). SVR menggunakan batas toleransi kesalahan yang disebut epsilon (ϵ). Nilai ϵ membentuk suatu tube di sekitar hyperplane, sehingga data yang berada di dalam batas tersebut tidak dikenakan penalti. Model SVR dibentuk melalui proses optimasi dengan mempertimbangkan parameter cost (C), ϵ , serta koefisien Lagrange. Penyelesaian optimasi dapat dilakukan menggunakan fungsi Lagrange dan kondisi Karush-Kuhn-Tucker (KKT), sedangkan koefisien Lagrange dapat diperoleh dengan bantuan software RStudio (Caraka, Yasin, & Basyiruddin, 2017; Patiallo, Fathurahman, Prangga & Nadhilah). Untuk hubungan nonlinear, SVR menggunakan fungsi kernel yang memetakan data dari ruang input ke ruang fitur berdimensi lebih tinggi. Fungsi kernel digunakan dalam SVR untuk menangani hubungan nonlinear tanpa harus melakukan pemetaan data secara eksplisit ke ruang berdimensi tinggi (Saputra, Wigena, & Sartono, 2019; Furi, Jondri, & Saepudin, 2015). Fungsi kernel yang sering digunakan dalam metode SVR meliputi beberapa jenis berikut (Meyer dkk., 2024):

1. Kernel Linear
Digunakan ketika hubungan data dapat direpresentasikan secara linear.
2. Kernel Polynomial
Digunakan untuk memodelkan hubungan nonlinear dengan mempertimbangkan derajat (degree) fungsi polynomial.
3. Kernel Radial Basis Function (RBF)
Digunakan untuk data yang memiliki hubungan nonlinear dan memungkinkan setiap titik data memberikan pengaruh terhadap pembentukan pola prediksi.

Grid search adalah teknik pencarian yang dilakukan dengan menguji berbagai nilai hyperparameter dalam ruang parameter yang telah ditetapkan untuk memperoleh kombinasi yang memberikan performa model terbaik (Safira, Warsito, & Rusgiyono, 2023). Selanjutnya, dalam proses pembentukan model, Data training berfungsi sebagai dasar dalam proses pembentukan model, sementara data testing digunakan untuk mengukur kemampuan model dalam menghasilkan prediksi pada data yang tidak melibatkan saat pelatihan (Utami & Ulama, 2015).

Setelah pembagian data dilakukan, tahap berikutnya adalah normalisasi, yaitu penyesuaian nilai data ke skala tertentu agar proses pembelajaran model menjadi lebih stabil dan optimal (Parto & Suhu, 2015). Selanjutnya, untuk mengetahui kinerja model yang telah dibangun, RMSE digunakan untuk mengukur tingkat penyimpangan antara hasil prediksi dengan nilai aktual. Nilai RMSE yang semakin rendah atau mendekati nol menunjukkan kesalahan prediksi yang semakin kecil, sehingga performa model dinilai semakin baik. (Maulana, Setiawan, & Dewi, 2019).

Selain RMSE, Mean Absolute Percentage Error (MAPE) digunakan untuk mengukur besarnya kesalahan prediksi dalam bentuk persentase berdasarkan perbedaan absolut antara hasil prediksi dan nilai aktual (Krislianti dkk., 2023). Nilai MAPE dihitung dengan mengambil rata-rata persentase kesalahan dari seluruh hasil prediksi. Dalam penelitian mengenai prediksi curah hujan, curah hujan merupakan jumlah air hujan yang jatuh pada suatu wilayah dalam periode tertentu (BPS, 2023). Curah hujan dipengaruhi oleh beberapa faktor meteorologi. Dalam penelitian ini, faktor yang digunakan sebagai variabel prediktor meliputi:

1. Suhu udara, yaitu ukuran tingkat panas atau dinginnya udara yang dapat dipengaruhi oleh ketinggian suatu wilayah (BPS, 2023; Purwantara, 2015).

2. Kelembaban udara, yaitu jumlah uap air yang terkandung di udara dan memiliki hubungan erat dengan pembentukan curah hujan (BPS, 2023; Priyahita, Sugianti, & Aliah, 2016).
3. Tekanan udara, yaitu berat kolom udara per satuan luas yang dapat memengaruhi pola angin dan perubahan cuaca, termasuk curah hujan (BPS, 2023; Yukifli, Asrizal, & Ardi, 2016).
4. Kecepatan angin, yaitu kecepatan pergerakan udara secara horizontal yang berkaitan dengan proses penguapan dan perpindahan udara lembab (Simbolon, Ruhiat, & Saefullah, 2022).
5. Penyinaran matahari, yaitu lama matahari bersinar di permukaan bumi yang berperan dalam proses penguapan air dan pembentukan awan hingga terjadinya hujan (Pujiastuti & Harjoko, 2016).

Hasil dan Pemahasan

Hasil

Tahap awal analisis dilakukan dengan melihat karakteristik data melalui statistika deskriptif. Hasil perhitungan untuk variabel curah hujan, suhu udara, kelembapan udara, tekanan udara, kecepatan angin, dan penyinaran matahari ditampilkan pada Tabel 1.

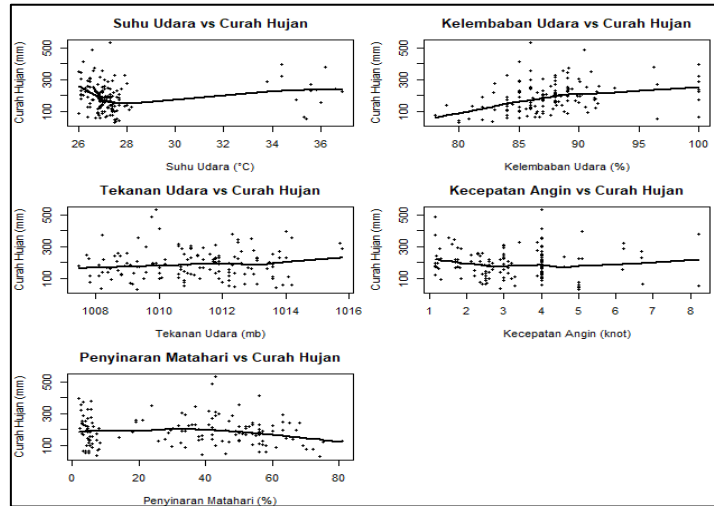
Tabel 1. Statistika Deskriptif

Variabel	Rata-rata	Simpangan Baku	Minimum	Maksimum
Y	192,0743	92,0384	27	535
X_1	27,7827	2,4846	26	36,9000
X_2	87,7737	4,1410	78	100
X_3	1.011,0710	1,8682	1.007,4500	1.015,8000
X_4	3,3106	1,3846	1,1400	8,2300
X_5	33,1319	23,0805	1,9000	81

Berdasarkan Tabel 1, curah hujan mempunyai nilai rata-rata sebesar 192,0743 mm, dengan nilai terendah 27 mm dan tertinggi 535 mm. Simpangan baku yang diperoleh sebesar 92,0384 mm, yang menunjukkan adanya variasi nilai curah hujan selama periode pengamatan. Suhu udara memiliki rata-rata 27,7827°C dengan simpangan baku 2,4846°C. Kelembapan udara mempunyai rata-rata 87,7737% dan simpangan baku 4,1410%.

Pada variabel tekanan udara, diperoleh rata-rata sebesar 1.011,0710 mb dengan simpangan baku 1,8682 mb. Kecepatan angin memiliki rata-rata 3,3106 knot dan simpangan baku 1,3846 knot. Sementara itu, penyinaran matahari memiliki rata-rata 33,1319% dengan simpangan baku 23,0805%. Hasil tersebut memperlihatkan bahwa setiap variabel memiliki rentang dan skala pengukuran yang berbeda.

Karakteristik perubahan setiap variabel selama periode pengamatan kemudian ditampilkan dalam bentuk grafik pada Gambar 1. Grafik tersebut memperlihatkan adanya perubahan nilai pada seluruh variabel dari satu periode ke periode lainnya.



Gambar 1. Plot Data Variabel Penelitian

Curah hujan menunjukkan pola yang berfluktuasi, dengan beberapa periode memiliki nilai yang relatif rendah dan periode lainnya menunjukkan peningkatan yang cukup tinggi. Variabel meteorologis lainnya juga mengalami perubahan sepanjang waktu. Pola tersebut menunjukkan bahwa kondisi meteorologis yang digunakan sebagai prediktor tidak bersifat tetap selama periode penelitian.

Data penelitian selanjutnya dibagi menjadi data training dan testing berdasarkan urutan waktu. Tiga skenario pembagian yang digunakan adalah 70:30, 80:20, dan 90:10. Jumlah data training dan data testing ditentukan menggunakan Persamaan berikut:

$$\sum training = \text{Proporsi data training} \times n \quad (1)$$

$$\sum testing = n - \text{jumlah data training} \quad (2)$$

Pada skenario 70:30, sebanyak 101 observasi pertama digunakan sebagai data training, sedangkan 43 observasi berikutnya menjadi data testing. Skenario 80:20 terdiri atas 115 observasi training dan 29 observasi testing. Adapun skenario 90:10 menggunakan 130 observasi training dan 14 observasi testing.

Pembagian tersebut dilakukan secara berurutan, sehingga data yang digunakan untuk pengujian berasal dari periode setelah data yang digunakan dalam proses pelatihan. Dengan cara ini, kemampuan model dalam menghasilkan prediksi terhadap data pada periode berikutnya dapat diamati.

Sebelum model SVR dibentuk, seluruh variabel dilakukan normalisasi menggunakan metode min-max normalization. Proses ini diterapkan pada seluruh 144 observasi dengan tujuan menyetarakan skala nilai dari masing-masing variabel. Adapun perhitungan normalisasi data menggunakan metode min-max normalization berdasarkan persamaan sebagai berikut.

$$x_{ig}^* = \frac{x_{ig} - \min(x_g)}{\max(x_g) - \min(x_g)} \quad (3)$$

Pemodelan dilakukan menggunakan metode Support Vector Regression (SVR) dengan kernel polynomial. Model dibentuk berdasarkan data training yang telah dinormalisasi. Fungsi kernel digunakan untuk memperoleh hubungan antara variabel prediktor dan curah hujan dalam proses pembentukan fungsi regresi. Model SVR secara umum yang dituliskan pada Persamaan berikut

$$f(x_i) = \sum_{i=1}^n (a_i - a_i^*) (\gamma(x_i \cdot x_j) + c)^d + b \quad (4)$$

Untuk menentukan model yang memberikan hasil paling baik, dilakukan optimasi parameter menggunakan metode grid search. Proses Grid Search dilakukan untuk memperoleh kombinasi parameter terbaik berdasarkan nilai RMSE terkecil. Hasil pengujian optimasi hyperparameter dilakukan dengan menguji beberapa kombinasi nilai cost (C) dan degree (d).

Nilai C yang digunakan adalah 0,5; 1; 2; dan 4, sedangkan nilai degree yang diuji adalah 1 dan 2.

Tabel 2. Hasil Grid Search

Proporsi	Parameter		Nilai Error
	C	d	RMSE
70:30	0,5	1	0,3545
80:20	0,5	1	0,1933
90:10	4,0	1	0,1710

Hasil pencarian parameter terbaik berdasarkan RMSE terkecil disajikan pada Tabel 2. Untuk proporsi 70:30, kombinasi optimal diperoleh pada C = 0,5 dan degree = 1 dengan RMSE sebesar 0,3545. Pada proporsi 80:20, kombinasi yang memberikan nilai terbaik adalah C = 0,5 dan degree = 1, dengan RMSE sebesar 0,1933. Sementara itu, proporsi 90:10 menghasilkan kombinasi terbaik pada C = 4 dan degree = 1, dengan RMSE sebesar 0,1710.

Dari hasil tersebut terlihat bahwa degree = 1 menjadi parameter terpilih pada seluruh skenario pembagian data. Nilai RMSE terendah dari ketiga skenario diperoleh pada proporsi 90:10, sehingga parameter C = 4 dan degree = 1 digunakan untuk membentuk model pada skenario tersebut.

Setelah kombinasi parameter diperoleh, proses selanjutnya adalah membentuk model SVR pada masing-masing proporsi. Hasil pelatihan menghasilkan sejumlah support vector yang digunakan dalam pembentukan fungsi regresi. Pada proporsi 70:30 diperoleh 50 support vector. Jumlah tersebut menunjukkan observasi yang berperan dalam pembentukan model SVR pada masing-masing skenario.

Tabel 3. Support Vector Model SVR dengan Proporsi Data 70:30

No	X_1	X_2	X_3	X_4	X_5
1	0,3182	0,4667	1,0000	1,0000	0,6817
2	0,4545	0,6000	0,9111	0,7409	0,7072
3	0,4091	0,4000	0,9556	0,7409	0,7072
⋮	⋮	⋮	⋮	⋮	⋮
48	0,0455	0,7333	0,7481	0,1321	0,4398
49	0,5455	0,4667	0,5259	0,1839	0,5798
50	0,6364	0,4667	0,5852	0,1580	0,7708

Perhitungan kernel dilanjutkan pada seluruh support vector sehingga terbentuk matriks kernel polynomial berukuran 50×50 untuk proporsi 70:30 sebagai berikut.

$$K(x_i, x_j) = \begin{bmatrix} 0,5667 & 0,5117 & \dots & 0,3378 \\ 0,5117 & 0,4892 & \dots & 0,3529 \\ \vdots & \vdots & \ddots & \vdots \\ 0,3378 & 0,3529 & \dots & 0,3169 \end{bmatrix}$$

Setelah matriks kernel polynomial diperoleh, tahap berikutnya adalah menentukan parameter model yang diperlukan dalam membentuk fungsi regresi SVR. Parameter tersebut diperoleh berdasarkan hasil pelatihan menggunakan fungsi svm pada Software R. Hasil pemodelan selanjutnya digunakan untuk menyusun fungsi regresi SVR pada proporsi data 70:30.

Berdasarkan parameter yang diperoleh dari hasil pemodelan, nilai tersebut disubstitusikan ke dalam fungsi regresi sehingga menghasilkan model SVR dengan kernel polynomial untuk proporsi data 70:30 seperti yang ditunjukkan pada Persamaan (5).

$$\begin{aligned}
 f(x_i) &= \sum_{i=1}^n (a_i - a_i^*)(\gamma(x_i \cdot x_j) + c)^d + b \\
 &= \sum_{i=1}^{50} (a_i - a_i^*)(\gamma(x_i \cdot x_j) + d)^1 + 0.2790
 \end{aligned} \tag{5}$$

Pada proporsi data 80:20, proses pembentukan model dilakukan dengan tahapan yang sama. Data training yang telah dinormalisasi digunakan untuk memperoleh support vector melalui proses pelatihan SVR. Berdasarkan hasil pemodelan, diperoleh 58 support vector yang digunakan dalam pembentukan fungsi regresi yang disajikan pada Tabel 4.

Tabel 4. Support Vector Model SVR dengan Proporsi Data 80:20

No	X_1	X_2	X_3	X_4	X_5
1	0,0660	0,4667	1,0000	1,0000	0,6817
2	0,0943	0,6000	0,9111	0,7409	0,7072
⋮	⋮	⋮	⋮	⋮	⋮
57	0,1415	0,6000	0,6593	0,5337	0,0374
58	0,1321	0,5333	0,7037	0,7150	0,0336

Nilai support vector yang diperoleh selanjutnya digunakan dalam perhitungan kernel polynomial. Perhitungan tersebut dilakukan terhadap seluruh support vector sehingga menghasilkan matriks kernel polynomial berukuran sesuai dengan jumlah support vector pada proporsi 80:20.

$$K(x_i, x_j) = \begin{bmatrix} 0,5374 & 0,4841 & \dots & 0,3399 \\ 0,4841 & 0,4496 & \dots & 0,3054 \\ \vdots & \vdots & \ddots & \vdots \\ 0,3399 & 0,3054 & \dots & 0,2619 \end{bmatrix}$$

Setelah matriks kernel polynomial diperoleh, tahap berikutnya adalah menentukan parameter model yang diperlukan dalam membentuk fungsi regresi SVR. Parameter tersebut diperoleh berdasarkan hasil pelatihan menggunakan fungsi svm pada Software R. Hasil pemodelan selanjutnya digunakan untuk menyusun fungsi regresi SVR pada proporsi data 70:30.

Tahap berikutnya adalah menentukan parameter model berdasarkan hasil pelatihan menggunakan fungsi svm pada Software R. Parameter yang diperoleh kemudian digunakan dalam penyusunan fungsi regresi. Dengan memasukkan parameter tersebut ke dalam persamaan SVR, diperoleh model SVR dengan kernel polynomial untuk proporsi 80:20 yang disajikan pada Persamaan (6).

$$\begin{aligned}
 f(x_i) &= \sum_{i=1}^n (a_i - a_i^*)(\gamma(x_i \cdot x_j) + c)^d + b \\
 &= \sum_{i=1}^{58} (a_i - a_i^*)(\gamma(x_i \cdot x_j) + d)^1 + 0.2059
 \end{aligned} \tag{6}$$

Proses pembentukan model pada proporsi 90:10 dilakukan dengan tahapan serupa. Hasil pelatihan menghasilkan 69 support vector, sebagaimana ditampilkan pada Tabel 5. Keberadaan support vector tersebut menjadi bagian penting dalam pembentukan fungsi regresi SVR yang disajikan pada Tabel 5.

Tabel 5. Support Vector Model SVR dengan Proporsi Data 80:20

No	X_1	X_2	X_3	X_4	X_5
1	0,0642	0,3182	0,8084	0,5444	0,6839

2	0,0917	0,4091	0,7365	0,4034	0,7092
⋮	⋮	⋮	⋮	⋮	⋮
68	0,8624	0,8409	0,5689	1,0000	0,0253
69	0,9358	0,8295	0,5808	1,0000	0,0480

Seluruh support vector kemudian digunakan dalam perhitungan fungsi kernel polynomial sehingga diperoleh matriks kernel untuk proporsi 90:10. Setelah matriks kernel terbentuk, parameter model diperoleh melalui hasil pelatihan menggunakan fungsi svm pada Software R.

$$K(x_i, x_j) = \begin{bmatrix} 0,3046 & 0,2872 & \cdots & 0,2742 \\ 0,2872 & 0,2768 & \cdots & 0,2581 \\ \vdots & \vdots & \ddots & \vdots \\ 0,2742 & 0,2581 & \cdots & 0,5807 \end{bmatrix}$$

Parameter yang dihasilkan selanjutnya digunakan untuk membentuk fungsi regresi SVR. Hasil substitusi parameter ke dalam persamaan menghasilkan model SVR dengan kernel polynomial untuk proporsi 90:10 yang dituliskan pada Persamaan (7).

$$\begin{aligned} f(x_i) &= \sum_{i=1}^n (a_i - a_i^*) (\gamma(x_i \cdot x_j) + c)^d + b \\ &= \sum_{i=1}^{69} (a_i - a_i^*) (\gamma(x_i \cdot x_j) + c)^1 + 0,1309 \end{aligned} \quad (7)$$

Model yang telah terbentuk selanjutnya digunakan untuk menghasilkan prediksi pada data testing. Berdasarkan hasil optimasi dan evaluasi model sebelumnya, proporsi 90:10 dipilih sebagai skenario dengan kinerja terbaik. Data testing pada skenario tersebut terdiri atas observasi ke-131 sampai observasi ke-144. Hasil prediksi pada skala normalisasi untuk ke-14 observasi tersebut disajikan pada Tabel 6. Nilai prediksi yang diperoleh menunjukkan adanya variasi antarperiode, dengan nilai tertinggi sebesar 0,3502 pada observasi ke-131 dan nilai terendah sebesar 0,1769 pada observasi ke-139.

Tabel 6. Hasil Prediksi Data Testing dengan Proporsi 90:10

No	Hasil Prediksi	No	Hasil Prediksi
131	0,3502	138	0,2665
132	0,3337	139	0,1769
133	0,3268	140	0,2670
134	0,3303	141	0,2527
135	0,3108	142	0,2769
136	0,2756	143	0,2319
137	0,2476	144	0,2679

Nilai prediksi yang masih berada dalam skala normalisasi belum dapat langsung dibandingkan dengan curah hujan aktual. Oleh karena itu, dilakukan denormalisasi untuk mengubah kembali hasil prediksi ke satuan awal, yaitu milimeter (mm).

Proses tersebut dilakukan dengan menggunakan persamaan denormalisasi seperti pada Persamaan (8). Hasil yang diperoleh selanjutnya digunakan sebagai dasar dalam melakukan evaluasi kinerja model.

$$x_{ig} = x_{ig}^* (\max(x_g) - \min(x_g)) + \min(x_g) \quad (8)$$

Pengukuran kinerja dilakukan dengan membandingkan hasil prediksi terhadap nilai aktual menggunakan RMSE dan MAPE. Kedua ukuran tersebut dihitung pada data training dan testing untuk masing-masing skenario pembagian data dengan menggunakan persamaan berikut

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}} \quad (9)$$

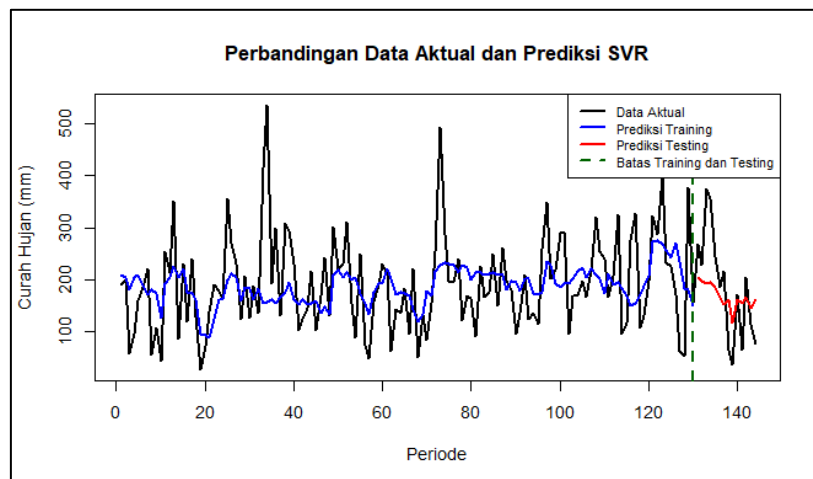
$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \quad (10)$$

Perhitungan nilai RMSE dan MAPE dilakukan dengan menggunakan bantuan *Software R* yang dapat dilihat pada Tabel 7.

Tabel 7. Ukuran Keباikan Model Prediksi Curah Hujan

No	Model Prediksi	Proporsi Data	RMSE Training	RMSE Testing	MAPE Training	MAPE Testing
1	$f(x) = \sum_{i=1}^{50} (a_i - a_i^*) (\gamma(x_i \cdot x_j) + c)^1 + 0,2790$	70:30	83,1917	180,0658	46,2822	75,3942
2	$f(x) = \sum_{i=1}^{58} (a_i - a_i^*) (\gamma(x_i \cdot x_j) + c)^1 + 0,2059$	80:20	82,7451	98,1892	44,6395	65,6183
3	$f(x) = \sum_{i=1}^{69} (a_i - a_i^*) (\gamma(x_i \cdot x_j) + c)^1 + 0,1309$	90:10	85,5421	86,8895	45,3880	61,4889

Berdasarkan Tabel 7, skenario 70:30 menghasilkan RMSE sebesar 83,1917 pada data training dan 180,0658 pada data testing. Nilai MAPE yang diperoleh masing-masing sebesar 46,2822% dan 75,3942%. Pada skenario 80:20, RMSE training tercatat sebesar 82,7451, sedangkan RMSE testing sebesar 98,1892. MAPE yang dihasilkan sebesar 44,6395% untuk training dan 65,6183% untuk testing. Sementara itu, skenario 90:10 memperoleh RMSE training sebesar 85,5421 dan RMSE testing sebesar 86,8895. Nilai MAPE training sebesar 45,3880%, sedangkan MAPE testing sebesar 61,4889%. Jika dilihat dari hasil pada data testing, skenario 90:10 memiliki nilai RMSE dan MAPE paling rendah dibandingkan dua skenario lainnya. Oleh karena itu, model dengan proporsi 90:10 ditetapkan sebagai model terbaik dalam penelitian ini. Model ini selanjutnya divisualisasikan melalui perbandingan data aktual dan hasil prediksi pada Gambar 2.



Gambar 2. Plot Perbandingan Data Aktual dengan Data Prediksi menggunakan Model SVR Terbaik

Kinerja model terbaik kemudian ditampilkan secara visual melalui perbandingan antara data aktual dan hasil prediksi pada Gambar 2. Grafik tersebut memperlihatkan bahwa hasil prediksi secara umum mengikuti arah perubahan data aktual. Namun, garis prediksi tampak lebih

stabil dibandingkan dengan data aktual yang memiliki perubahan lebih tajam. Perbedaan tersebut menunjukkan bahwa model mampu menangkap kecenderungan umum curah hujan, tetapi belum sepenuhnya mengikuti setiap perubahan nilai aktual, terutama pada periode dengan fluktuasi yang lebih besar.

Pembahasan

Berdasarkan hasil evaluasi seluruh skenario, proporsi 90:10 merupakan konfigurasi terbaik. Skenario tersebut menggunakan 130 observasi untuk proses training dan 14 observasi untuk testing. Parameter yang diperoleh melalui grid search adalah $C = 4$ dan $\text{degree} = 1$, dengan nilai RMSE grid search sebesar 0,1710.

Pemilihan model terbaik didasarkan terutama pada performa data testing. Proporsi 90:10 menghasilkan RMSE testing sebesar 86,8895 dan MAPE testing sebesar 61,4889%, yang merupakan nilai terendah dibandingkan dua skenario lainnya. Walaupun RMSE training pada proporsi tersebut sedikit lebih tinggi daripada proporsi 80:20, performa pada data testing justru lebih baik. Kondisi ini menunjukkan bahwa penggunaan data training yang lebih banyak memberikan informasi yang lebih luas bagi model dalam mengenali pola data.

Hasil visualisasi juga memperlihatkan bahwa prediksi model cenderung lebih halus dibandingkan data aktual. Hal tersebut dapat dikaitkan dengan karakteristik SVR yang menggunakan konsep ϵ -insensitive, sehingga model tidak dituntut mengikuti setiap perubahan observasi secara persis. Akibatnya, ketika curah hujan mengalami perubahan yang tajam, nilai prediksi cenderung tidak mengalami perubahan sebesar data aktual.

Jika dibandingkan dengan penelitian sebelumnya yang dilakukan oleh Monalisa Randang Patiallo, M. Fathurahman, Surya Prangga, dan Erlyne Nadhilah Widyaningrum, penelitian ini memiliki kesamaan dalam penggunaan metode Support Vector Regression (SVR) dengan kernel polynomial untuk memprediksi curah hujan di Kabupaten Berau. Kesamaan lainnya terletak pada penggunaan variabel prediktor, yaitu suhu udara, kelembapan udara, tekanan udara, kecepatan angin, dan penyinaran matahari. Penelitian sebelumnya menggunakan proporsi data training dan testing sebesar 70:30 dan 80:20, dengan proporsi 80:20 menghasilkan kinerja terbaik. Sementara itu, penelitian ini menggunakan tiga proporsi pembagian data, yaitu 70:30, 80:20, dan 90:10, dengan proporsi 90:10 memberikan hasil terbaik berdasarkan nilai RMSE dan MAPE pada data testing. Perbedaan hasil tersebut menunjukkan bahwa proporsi data yang optimal dapat berbeda pada masing-masing penelitian, meskipun menggunakan metode, kernel, dan variabel prediktor yang relatif sama. Penambahan proporsi data training juga tidak secara langsung menjamin bahwa model akan menghasilkan nilai kesalahan yang lebih kecil, karena kinerja model dipengaruhi oleh kemampuan model dalam melakukan generalisasi terhadap data testing. Dengan demikian, penggunaan proporsi 90:10 pada penelitian ini memberikan kinerja prediksi yang lebih baik dibandingkan proporsi lainnya, sedangkan penelitian sebelumnya memperoleh kinerja terbaik pada proporsi 80:20.

Meskipun model terbaik telah diperoleh, nilai MAPE testing sebesar 61,4889% menunjukkan bahwa kesalahan prediksi masih cukup besar. Dengan demikian, model belum mampu menghasilkan estimasi yang sangat dekat dengan nilai aktual pada seluruh periode. Salah satu kondisi yang terlihat adalah adanya perbedaan yang lebih besar ketika curah hujan mengalami perubahan ekstrem. Hal tersebut sejalan dengan grafik yang menunjukkan bahwa prediksi lebih stabil dibandingkan fluktuasi aktual.

Dengan demikian, hasil penelitian memperlihatkan bahwa SVR kernel polynomial dengan optimasi grid search pada proporsi 90:10 merupakan model yang paling sesuai di antara skenario yang diuji. Model tersebut dapat digunakan sebagai salah satu pendekatan dalam memprediksi curah hujan Kabupaten Berau menggunakan suhu udara, kelembapan udara, tekanan udara, kecepatan angin, dan penyinaran matahari sebagai variabel prediktor. Namun, hasil prediksi tetap perlu dipertimbangkan sebagai estimasi karena model masih memiliki keterbatasan dalam menangkap perubahan curah hujan yang sangat tinggi maupun rendah.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa metode Support Vector Regression (SVR) dengan kernel polynomial dapat digunakan untuk memprediksi curah hujan di Kabupaten Berau dengan menggunakan suhu udara,

kelembaban udara, tekanan udara, kecepatan angin, dan penyinaran matahari sebagai variabel prediktor. Berdasarkan hasil evaluasi menggunakan Root Mean Square Error (RMSE) dan Mean Absolute Percentage Error (MAPE), model terbaik diperoleh pada proporsi data training dan testing sebesar 90:10 dengan kombinasi parameter $\text{cost} = 4$ dan $\text{degree} = 1$ melalui optimasi grid search. Model tersebut menghasilkan nilai RMSE testing sebesar 86,8896 dan MAPE testing sebesar 61,4889%, yang merupakan tingkat kesalahan prediksi paling rendah dibandingkan proporsi data 70:30 dan 80:20. Hasil tersebut menunjukkan bahwa penggunaan proporsi data 90:10 dengan kombinasi parameter tersebut memberikan kinerja prediksi terbaik dalam penelitian ini. Meskipun demikian, nilai MAPE yang masih cukup tinggi menunjukkan bahwa model masih memiliki keterbatasan dalam menghasilkan prediksi yang akurat, sehingga penelitian selanjutnya disarankan untuk membandingkan penggunaan kernel polynomial dengan kernel lainnya, seperti Radial Basis Function (RBF) atau Sigmoid, serta menggunakan metode optimasi hyperparameter lainnya, seperti Genetic Algorithm (GA), Particle Swarm Optimization (PSO), atau Bayesian Optimization. Selain itu, penelitian selanjutnya dapat mengembangkan model hibrida dengan menggabungkan SVR dengan metode lain, seperti Artificial Neural Network (ANN) atau ARIMA, serta mempertimbangkan penambahan variabel prediktor dan periode data yang lebih panjang untuk memperoleh model yang mampu menangkap karakteristik curah hujan dengan lebih baik dan meningkatkan akurasi prediksi.

Daftar Pustaka

- [1] Caraka, R. E., Yasin, H., & Basyiruddin, A. W. (2017). Peramalan crude palm oil (CPO) menggunakan Support Vector Regression kernel radial basis. *Jurnal Matematika*, 7(1), 43–57.
- [2] Elwirehardja, G. N., Suparyanto, T., & Pardamean, B. (2023). Pengenalan konsep machine learning untuk pemula. Instiper Press.
- [3] Furi, R. P., Jondri, & Saepudin, D. (2015). Peramalan financial time series menggunakan Independent Component Analysis dan Support Vector Regression. *e-Proceeding of Engineering*, 2.
- [4] Hasibuan, E., & Karim, A. (2022). Implementasi machine learning untuk prediksi harga mobil bekas dengan algoritma regresi linear berbasis web. *Jurnal Ilmiah Komputasi*, 21(4), 595–602.
- [5] Hermawan, B. M., Hakim, M. A., Arifin, R., & Puspitasari, N. (2024). Pemanfaatan Artificial Intelligence, khususnya Machine Learning dan Deep Learning System dalam pendidikan. *Prosiding Seminar Nasional Amikom Surakarta*, 2, 345–354.
- [6] Huda, S., Azhar, M. A. Z., Putri, N. R., Damayanti, A., Zahrani, F., Sari, D. A., ... & Qalbiyah, F. (2025). Kreasi dakwah di era AI: Optimalisasi penyiaran Islam dengan kecerdasan buatan.
- [7] Iriananda, S. W., Budiawan, R. W., Rahman, A. Y., & Istiadi, I. (2024). Optimasi klasifikasi sentimen komentar pengguna game bergerak menggunakan SVM, grid search dan kombinasi N-Gram. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(4), 743–752.
- [8] Isnaeni, R., Sudarmin, S., & Rais, Z. (2022). Analisis Support Vector Regression (SVR) dengan kernel Radial Basis Function (RBF) untuk memprediksi laju inflasi di Indonesia. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(1), 30–38.
- [9] Kamaluddin, R. (2025). Optimasi model Support Vector Regression (SVR) dengan GridSearchCV untuk prediksi harga saham Bank BRI. *JSR: Jaringan Sistem Informasi Robotik*, 9(1), 88–97.
- [10] Karondia, L. A., Fitriani, R., & Hizkia. (2022). Pemetaan zonasi kerawanan banjir berbasis sistem informasi geografis (studi kasus: Kabupaten Berau, Kalimantan Timur). *Journal of Geodesy and Geomatics*, 18(1), 57–68.
- [11] Kamil, M. F. I. (2021). Analisis data curah hujan yang hilang dengan menggunakan metode rata-rata aljabar dan metode resipokal di stasiun Cibeureum Kertasari, Cicalengka, dan Ciherang (Doctoral dissertation, Institut Teknologi Nasional).
- [12] Krislianti, D. E., Zukhronah, E., & Susanti, Y. (2023). Peramalan harga minyak menggunakan Autoregressive Integrated Moving Average dan Support Vector Regression. *Prosiding Seminar Pendidikan Matematika dan Matematika*, 7.

- [13] Kusuma, M. P., & Kudus, A. (2022). Penerapan metode Support Vector Regression (SVR) pada data survival KPR PT. Bank ABC, Tbk. *Bandung Conference Series: Statistics*, 2(2), 167–172.
- [14] Lazarusli, I. A. (2025). Pengembangan model prediktif dengan Artificial Intelligence untuk menilai dampak perubahan iklim pada keanekaragaman hayati. *Prosiding Seminar Nasional Sains dan Teknologi Seri III Fakultas Sains dan Teknologi*, 2(1).
- [15] Maulana, N. D., Setiawan, B. D., & Dewi, C. (2019). Implementasi metode Support Vector Regression (SVR) dalam peramalan penjualan roti (studi kasus: Harum Bakery). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2986–2995.
- [16] Mashita, R. M., Basuki, S., & Hayatin, N. (2020). Prediksi pemakaian kWh listrik menggunakan metode Support Vector Regression (SVR) (studi kasus: PT. PLN (Persero) Rayon Seririt). *Jurnal Repositor*, 2(4), 525–540.
- [17] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., & Lin, C. (2024). Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.
- [18] Musnanda. (2022). Kajian banjir Mei 2021 di Kabupaten Berau.
- [19] Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis supervised dan unsupervised learning pada machine learning: Systematic literature review. *Jurnal Ilmu Komputer dan Informatika*, 4(1), 61–72.
- [20] Patiallo, M. R., Fathurahman, M., Prangga, S., & Nadhilah, E. (2025). Prediksi curah hujan di Kabupaten Berau menggunakan Support Vector Regression.
- [21] Parto, S. K., & Sahu, K. (2015). Normalization: A preprocessing stage. *Department of CSE & IT, Burla, Odisha, India*.
- [22] Pujiastuti, A., & Harjoko, A. (2016). Sistem perhitungan lama penyinaran matahari dengan metode Otsu Threshold (studi kasus: ST. Klimatologi Barongan). *Jurnal Compiler*, 5(2), 11–20.
- [23] Rais, Z. (2022). Analisis Support Vector Regression (SVR) dengan kernel Radial Basis Function (RBF) untuk memprediksi laju inflasi di Indonesia. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(1), 30–38.
- [24] Retnoningsih, E., & Pramudita, R. (2020). Mengenal machine learning dengan teknik supervised dan unsupervised learning menggunakan Python. *BINA INSANI ICT Journal*, 7(2), 156–165.
- [25] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan machine learning dalam berbagai bidang. *Jurnal Khatulistiwa Informatika*, 5(1), 490845.
- [26] Ruliana, R., Rais, Z., Marni, M., & Ahmar, A. S. (2024). Implementation of the Support Vector Regression (SVR) method in inflation prediction in Makassar City. *ARRUS Journal of Mathematics and Applied Science*, 4(1), 28–35.
- [27] Safira, A. N., Warsito, B., & Rusgiyono, A. (2023). Analisis Support Vector Regression (SVR) dengan algoritma grid search time series cross validation untuk prediksi jumlah kasus terkonfirmasi COVID-19 di Indonesia. *Jurnal Gaussian*, 11(4), 512–521.
- [28] Saputra, G. H., Wigena, A. H., & Sartono, B. (2019). Penggunaan Support Vector Regression dalam pemodelan indeks saham syariah Indonesia dengan algoritma grid search. *Indonesian Journal of Statistics and Its Applications*, 3(2), 148–160.
- [29] Setiawan, D. (2021). Analisis curah hujan di Indonesia untuk memetakan daerah potensi banjir dan tanah longsor dengan metode Cluster Fuzzy C-Means dan Singular Value Decomposition (SVD). *Engineering, Mathematics and Computer Science Journal (EMACS)*, 3(3), 115–120.
- [30] Siahaan, M., Jasa, C. H., Anderson, K., Rosiana, M. V., Lim, S., & Yudianto, W. (2020). Penerapan Artificial Intelligence (AI) terhadap seorang penyandang disabilitas tunanetra. *Journal of Information System and Technology (JOINT)*, 1(2), 186–193.
- [31] Simbolon, C. D. L., Ruhiat, Y., & Saefullah, A. (2022). Analisis arah dan kecepatan angin terhadap sebaran curah hujan di wilayah Kabupaten Tangerang. *Jurnal Teori dan Aplikasi Fisika*, 10(1), 113–120.
- [32] Suyono, A. A., Kusrini, & Arief, R. (2022). Prediksi indeks harga konsumen komoditas makanan di Kota Surabaya menggunakan Support Vector Regression. *Metik Jurnal*, 6(1), 46–48.

- [33] Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'ban, S., Amalia, M., & Febriansyah, R. (2024). Pembelajaran machine learning. OKTAL: Jurnal Ilmu Komputer dan Science, 3(2), 375–380.