

PENERAPAN *MACHINE LEARNING* DENGAN ALGORITMA SUPPORT VECTOR MACHINE UNTUK PREDIKSI KELEMBAPAN UDARA RATA-RATA

Indah Dwi Sulistyowati^{1*}, Sunarno Sunarno², dan Djuniadi Djuniadi³

^{1,2}Fakultas Matematika dan Ilmu Pengetahuan, Fisika, Universitas Negeri Semarang,

³ Fakultas Teknik, Teknik Elektro, Universitas Negeri Semarang

*ndahe1549@students.unnes.ac.id

Abstrak

Machine learning dapat digunakan untuk memprediksi suatu data. *Support Vector Machine* merupakan bagian dari teknik data mining yang dipergunakan untuk mengidentifikasi dan memprediksi hubungan antara variabel pada suatu dataset. Metode ini efektif untuk melakukan prediksi baik itu untuk klasifikasi ataupun analisis regresi. Perangkat lunak *Orange Data Mining 3.3.12* digunakan untuk melakukan proses prediksi. Selanjutnya algoritma *Support Vector Machine* digunakan untuk memprediksi Kelembapan udara. Data masukan adalah suhu, kecepatan angin, penyinaran matahari, dan juga Kelembapan udara harian maksimum dan minimum. Data diambil dari Stasiun Meteorologi Jawa Timur di wilayah Malang pada tahun 2015-2023 sebanyak 2922 dataset. Tujuan penelitian ini adalah untuk mendapatkan nilai RMSE, MAE, dan *R-Squared* (R^2). Rasio perbandingan data pelatihan dan data pengujian ditetapkan pada 70:30. Hasil penelitian menunjukkan bahwa hasil akurasi *Root Mean Squared Error* (RMSE) dengan nilai 3,378, *Mean Absolute Error* (MAE) dengan nilai 2,738, dan *R-squared* (R^2) dengan nilai 0,723. Berdasarkan hasil korelasi tersebut menunjukkan bahwa algoritma *Support Vector Machine* ini termasuk memiliki pengaruh kuat terhadap hasil prediksi Kelembapan udara rata-rata harian.

Kata Kunci: *Kelembapan Udara; MAE; Prediksi; RMSE; R-Squared; Support Vector Machine*

Abstract

Machine learning can be used to predict data. Support Vector Machine is a part of data mining technique that is used to identify and predict the relationship between variables in a dataset. This method is effective for making predictions for both classification and regression analysis. Orange Data Mining 3.3.12 software is used to perform the prediction process. Furthermore, the Support Vector Machine algorithm was used to predict air humidity. The input data are temperature, wind speed, solar irradiation, and also maximum and minimum daily air humidity. The data was taken from the East Java Meteorological Station in the Malang area in 2015-2023 totaling 2922 datasets. The reason of this research is to get the RMSE, MAE, and *R-Squared* (R^2) values. The proportion of training data and test data is set at 70:30. The results showed that the Root Mean Squared Error (RMSE) accuracy results with a value of 3.378, Mean Absolute Error (MAE) with a value of 2.738, and *R-squared* (R^2) with a value of 0.723. Based on the correlation results, it shows that the Support Vector Machine algorithm has a strong influence on the prediction results of daily average air humidity.

Keywords: *Air Humidity; MAE; Prediction; RMSE; R-Squared; Support Vector Machine*

1. Pendahuluan

Kondisi cuaca berperan signifikan dalam menunjang aktivitas sehari-hari manusia (Seprima & Defrianto, 2020). Keterkaitan antara cuaca dan aktivitas manusia sangatlah dekat, dimana kondisi cuaca dapat membatasi jenis kegiatan yang dapat dilaksanakan (Pratama et al., 2022; Syaputra et al., 2023; Zulfiani & Fauzi, 2023). Prakiraan cuaca yang tepat sangat membantu dalam pengelolaan kualitas hidup peradaban dengan cara yang lebih efektif dan efisien.

Kelembapan udara merupakan salah satu unsur cuaca yang memungkinkan untuk diprediksi. Kelembapan udara adalah jumlah uap air yang terkandung di atmosfer. Kelembapan udara yang tinggi mengindikasikan banyaknya kandungan uap air dalam kondisi basah. Adapun faktor yang mempengaruhi Kelembapan udara diantaranya angin, suhu udara, sumber uap, ketersediaan air, dan tekanan udara (Swarinoto & Sugiyono, 2011).

Seiring berkembangnya teknologi, penerapan metode *data mining* dapat dilakukan untuk melakukan prediksi cuaca sehingga mendapatkan nilai akurasi yang baik. *Data mining* merupakan rangkaian proses pengumpulan data dalam skala besar. Inti dari *data mining* untuk mempermudah proses pengambilan keputusan dengan prediksi yang akurat. Manfaat dari metode ini terletak pada kemudahannya dalam menganalisis dataset yang luas (Arif & Nurmaleni, 2023; Jayadianti et al., 2020; Nurul Salsabila Syam et al., 2022). Salah satu metode yang digunakan dalam data mining adalah *machine learning*.

Penggunaan *machine learning* saat ini menjadi alat penting untuk analisis prediktif dalam beberapa tahun belakangan ini. *Machine learning* bisa diartikan sebagai serangkaian algoritma yang dirancang untuk mengidentifikasi pola dalam kumpulan data (Chen et al., 2018). Perkembangan teknologi pada *machine learning* telah memungkinkan prediksi yang lebih efisien dan tepat.

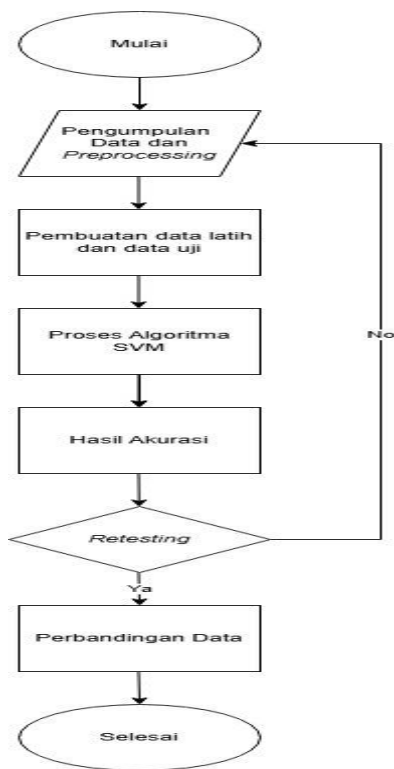
Ketepatan dalam memprediksi cuaca bergantung pada pemilihan algoritma yang akurat dan sesuai (Intan et al., 2021). Algoritma *Support Vector Machine* merupakan model pengklasifikasian yang menggunakan model biner atau diskriminatif serta juga digunakan untuk mencari hyperplane paling baik dengan mengoptimalkan jarak antara kelas (Risanti, 2024).

Perangkat Lunak *Orange Data Mining* sebagai salah satu tools dalam pengolahan *machine learning* dan *data mining*. *Orange* memungkinkan penggunaan pada lingkup visual programming serta eksplorasi data analitik dengan bahasa pemrograman Python. Perangkat lunak ini terdiri atas berbagai modul yang dikenal sebagai *widget*. *Orange Data Mining* kompatibel dengan bermacam-macam OS, termasuk mac OS, Windows, dan Linux (Kukavadiya & Divecha, 2020).

Penelitian ini bertujuan untuk melakukan prediksi Kelembapan udara menerapkan model *machine learning* dengan algoritma *Support Vector Machine*. Data yang diterapkan meliputi berbagai pengukuran parameter suhu, kecepatan angin, penyinaran matahari, dan juga Kelembapan udara harian maksimum dan minimum. Data diambil dari Stasiun Meteorologi Jawa Timur di wilayah Malang pada rentang waktu tahun 2015-2023. Selanjutnya dari data tersebut, akan di-training dan di-test hingga didapatkan nilai akurasi yang dinyatakan dalam nilai RMSE, MAE, dan *R-Squared* (R^2).

2. Metode Penelitian

Penelitian ini menerapkan algoritma *Support Vector Machine* untuk melakukan prediksi keakuratan Kelembapan udara. Perangkat lunak *Orange Data Mining* 3.3.12 digunakan untuk menerapkan algoritma *Support Vector Machine*. Tahapan penelitian dapat terlihat pada gambar 1. Penjelasan setiap tahap disampaikan pada alinea berikutnya.



Gambar. 1 Tahapan Penelitian.

2.1 Pengumpulan Dataset dan Preprocessing

Proses pengumpulan data dilakukan untuk menghimpun dataset yang akan diaplikasikan dalam pembuatan model klasifikasi *Support Vector Machine*. Sumber data berasal dari hasil pengamatan Stasiun Klimatologi Malang, provinsi Jawa Timur. Data diambil dari aplikasi BMKGSoft dengan rentang waktu tahun 2015-2023. Total dataset untuk setiap parameter adalah 2922 data. Atribut yang digunakan pada penelitian ini meliputi:

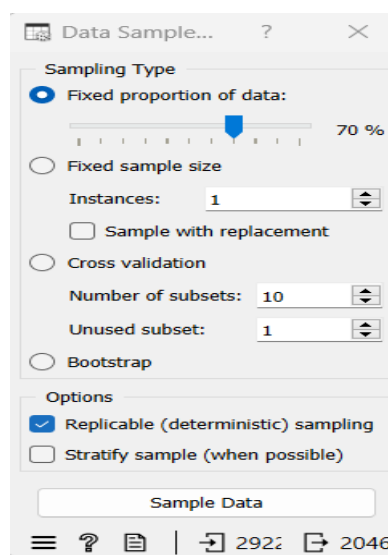
- 1) suhu rata-rata maksimum : nilai suhu rata-rata tertinggi harian dalam °C
- 2) suhu rata-rata minimum : nilai suhu rata-rata terendah harian dalam °C
- 3) kecepatan angin: nilai kecepatan angin harian dalam knots
- 4) lamanya penyinaran matahari : nilai lamanya penyinaran matahari harian dalam jam
- 5) Kelembapan udara maksimum : nilai Kelembapan udara tertinggi harian dalam persen

- 6) Kelembapan udara minimum : nilai Kelembapan udara terendah harian dalam persen.

Preprocessing data adalah langkah penting yang harus dijalankan sebelum data dimasukkan untuk klasifikasi. Tahap ini merupakan proses mengubah data mentah menjadi data yang bebas dari *noise* sehingga siap digunakan sebagai input ke dalam *machine learning* (Djuniadi et al., 2022). *Preprocessing* memastikan data yang akan diklasifikasikan bebas dari atribut kosong. Atribut kosong dapat berpengaruh pada hasil klasifikasi dan akurasi (Bertalya et al., 2021). Pada penelitian ini, *preprocessing* data yang dilakukan adalah dengan penghapusan (*drop*) atribut kosong agar tidak menghambat proses prediksi nantinya

2.2 Pembuatan Data Pelatihan dan Data Pengujian

Langkah berikutnya yaitu melibatkan pemisahan data yaitu data training dan data test. Hasil yang diperoleh berasal dari proses yang dijalankan yaitu dari data set yang tersedia dengan mengeliminasi kesalahan data atau data yang tidak bernilai. Data yang telah dikumpulkan kemudian dibagi secara random 70% atau 2046 data pelatihan, dan 30% yang berjumlah 876 data akan digunakan sebagai data untuk diuji.



Gambar. 2 Pengaturan Rasio perbandingan Data Latih dan Data Uji

2.3 Proses prediksi Kelembapan udara

Tahapan selanjutnya adalah proses prediksi yang dilakukan menggunakan perangkat lunak *orange data mining* dalam melakukan prediksi Kelembapan udara menggunakan algoritma *Support Vector Machine*. Metode algoritma tersebut akan diperoleh nilai RMSE dan nilai MAE yang dibutuhkan untuk mengevaluasi efektivitas model. Nilai yang mendekati 0 menunjukkan prediksi yang lebih akurat, sementara nilai R-Squared jika nilainya mendekati 1 menandakan adanya korelasi kuat antara prediksi dan observasi. Nilai ini terlihat pada widget *Test and Score* di perangkat lunak *Orange tools*.

2.3.1 Root Mean Square Error (RMSE)

Evaluasi keakuratan prediksi memerlukan perbandingan data aktual. Beberapa metode umum digunakan untuk memperoleh nilai kesalahan prediksi, diantaranya adalah RMSE. Metode ini merupakan metode populer pada penilaian hasil prediksi. RMSE menghitung selisih nilai yang diprediksi dan nilai yang sebenarnya terjadi oleh suatu model (Colanus et al., 2017). Nilai RMSE yang rendah, mendekati 0, menunjukkan tingkat akurasi yang tinggi (Pratama W, 2023).

RMSE merupakan total dari selisih kuadrat atau perbedaan dari nilai aktual dengan nilai prediksi yang dinyatakan pada persamaan:

$$RMSE = \sqrt{\frac{(y' - y)^2}{n}} \tag{1}$$

- n = Banyaknya data
- y' = Data Prediksi
- y = Data Sebenarnya

2.3.2 Mean Absolute Error (MAE)

MAE merupakan teknik yang sering diaplikasikan dalam menilai keakuratan model prediksi. MAE menghitung nilai dari rerata kesalahan mutlak antara nilai sebenarnya dengan prediksi. Tingkat keakuratan prediksi yang dihasilkan dari metode MAE dimanfaatkan untuk mengukur tingkat keakuratan suatu model prediksi (Suryanto, 2019).

MAE merupakan hasil perhitungan dari rerata perbedaan mutlak nilai aktual terhadap nilai prediksi dinyatakan pada persamaan:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x| \tag{2}$$

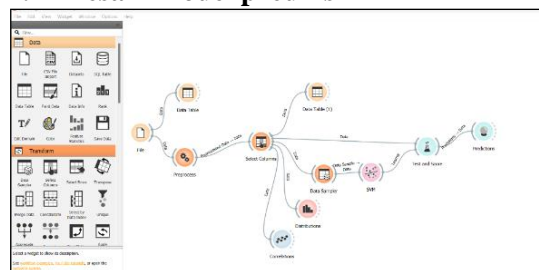
- n = Banyaknya jumlah data
- x_i = Data sebenarnya
- x = Data hasil prediksi

2.3.3 R- square (R²)

R square merupakan ukuran yang menunjukkan pengaruh antara nilai sebenarnya terhadap nilai prediksi dalam suatu model. Dengan rumus:

$$R^2 = 1 - \frac{SS\ Error}{SS\ Total} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{3}$$

2.4 Desain model prediksi



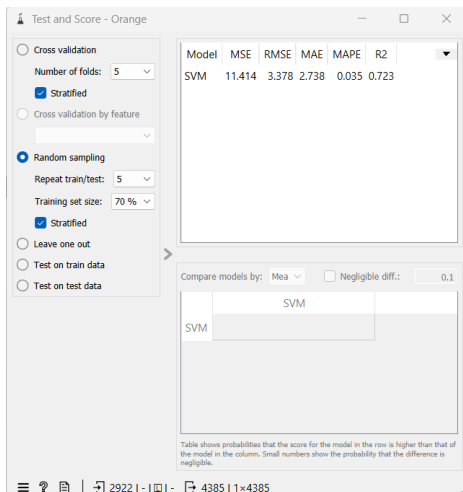
Gambar. 3
Desain model prediksi

Gambar 3. menunjukkan alur dalam pengujian dari model prediksi. Model ini untuk menghitung keberhasilan prediksi. Model ini juga untuk mengevaluasi efektivitas dari hasil prediksi pada orange tool menggunakan widget *Test and Score* untuk pengujian dan penilaian. Hasil prediksi terlihat melalui widget *Predictions*.

3. Hasil dan Pembahasan

Proses ini dilakukan yaitu dari dataset yang tersedia setelah tahap preprocessing. Data yang terhimpun kemudian secara acak dibagi dengan 70% untuk pelatihan dan 30% untuk pengujian serta validasi menggunakan aplikasi orange, dimana 1 atribut sebagai target. Evaluasi kinerja model dilakukan dengan menggunakan model statistic, dengan

memperhatikan beberapa indikator yaitu RMSE, MAE dan R squared. Simulasi menggunakan aplikasi orange dan hasilnya ditunjukkan pada gambar 4.



Gambar 4. Tampilan Hasil
Sumber : Peneliti (2024)

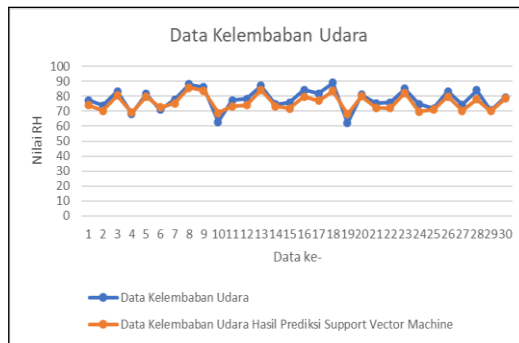
Gambar 4. menampilkan output dari widget *Test and Score* di *Orange Tools*. Dalam tulisan ini hanya memperhatikan nilai RMSE, MAE, dan R² untuk menilai efektivitas model prediksi yang dianalisa. Hasil evaluasi mengindikasikan korelasi antara data actual dan data prediksi dari hasil pengolahan, mempunyai nilai RMSE sebesar 3,378, MAE sebesar 2,738 dan R² mempunyai nilai sebesar 0,723.

Tabel 1. Hasil Prediksi menggunakan Metode *Support Vector Machine*

Data	Data RH	Data RH Hasil Prediksi SVM	Data	Data RH	Data RH Hasil Prediksi SVM	Data	Data RH	Data RH Hasil Prediksi SVM
1	77,35	74,0421	11	77,3	73,2665	21	75,325	72,2494
2	73,75	70,3644	12	78,25	74,0392	22	75,75	72,2079
3	83,25	80,5592	13	87,15	84,3182	23	85,1	82,3516
4	68,275	69,1972	14	74,525	73,2278	24	74,75	69,7963
5	81,5	79,9278	15	75,75	71,9341	25	71,75	71,0247
6	71,05	72,4803	16	84,2	79,9142	26	83,25	79,709
7	77,75	75,437	17	81,75	77,0764	27	74,45	70,3263
8	88	85,7487	18	89	83,6264	28	84	78,0256
9	86	83,8854	19	62	68,1617	29	70,55	70,1076
10	62,75	68,8113	20	81	80,1751	30	79,3	78,6625

Tabel 1 diatas memperlihatkan bahwa sampling data Kelembapan udara dan hasil prediksi data Kelembapan udara menggunakan metode *Support Vector Machine*. Dikarenakan banyaknya data yang digunakan dalam pengolahan ini, penulis hanya menampilkan 30 data. Selisih dari nilai Kelembapan udara tidak terlalu signifikan antara data Kelembapan udara sebenarnya

dengan hasil prediksi dari metode *Support Vector Machine*.



Gambar 5. Hasil Perbandingan Kelembapan Udara aktual dan prediksi
Sumber: Peneliti (2024)

4. Kesimpulan

Penelitian ini menggunakan dataset yang bersumber dari hasil pengamatan Stasiun Klimatologi Malang, provinsi Jawa Timur. Data diambil dari aplikasi BMKGSoft dengan rentang waktu tahun 2015-2023. Total dataset untuk setiap parameter adalah 2922 data. jumlah atribut yang digunakan sebanyak 6 buah. Perangkat lunak yang digunakan adalah Orange. Hasil yang diperoleh, nilai RMSE sebesar 3,378, MAE sebesar 2,738 dan R² mempunyai nilai 0,723. Berdasarkan hasil dapat disimpulkan bahwa model yang dibangun menunjukkan tingkat hubungan kuat, sehingga dapat digunakan untuk melakukan prediksi Kelembapan udara.

Saran untuk penelitian selanjutnya adalah dapat melakukan implementasi metode *machine learning* selain *Support Vector Machine* agar mendapatkan perbandingan yang lebih komperhensif. Selain itu, dapat juga menggunakan tools lainnya, *Orange Data Mining* sebagai alternatif

Daftar Pustaka

Arif, A., & Nurmaleni. (2023). Komparasi Regresi Linear Berganda Dan Support Vector Machine Untuk Prediksi Kopi Robusta Pagar Alam. *Jurnal Ilmiah Betrik*, 14(03), 713–726.

Bertalya, Prihandoko, Setyowati, L., Irawan, F. I., & Irlianti, S. R. (2021). Formulation of

- city health development index using data mining. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(1), 362–369.
<https://doi.org/10.11591/ijeecs.v23.i1.pp362-369>
- Chen, S., Xue, M., Fan, L., Hao, S., Xu, L., Zhu, H., & Li, B. (2018). Automated poisoning attacks and defenses in malware detection systems: An adversarial machine learning approach. *Computers and Security*, 73(B0 Li), 326–344.
<https://doi.org/10.1016/j.cose.2017.11.007>
- Colanus, I., Drajana, R., & Selection, F. (2017). 258730-Metode-Support-Vector-Machine-Dan-Forwar-C588Ca21. 9, 116–123.
- Djuniadi, Iksan, N., Hastawan, A. F., Suni, A. F., Firdaus, A. A., Hidayati, L., Setiawan, D. K., Udayanti, E. D., & Prasetyanto, F. (2022). Face mask detection services of Covid19 monitoring system to maintain a safe environment using deep learning method. *IOP Conference Series: Earth and Environmental Science*, 969(1).
<https://doi.org/10.1088/1755-1315/969/1/012016>
- Intan, I., Aminah Dinayati Ghani, S., Koswara, A. T., Dipa Makassar, U., Arsip Nasional Republik Indonesia, K., & Kemerdekaan, J. P. (2021). Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning Performance Analysis of Weather Forecasting using Machine Learning Algorithms. *Jurnal_Pekommas_Vol._6_No_2*, 2, 1–8.
<https://doi.org/10.30818/jpkm.2021.2060221>
- Jayadianti, H., Cahyadi, T. A., Amri, N. A., & Pitayandanu, M. F. (2020). Metode Komparasi Artificial Neural Network Pada Prediksi Curah Hujan - Literature Review. *Jurnal Tekno Insentif*, 14(2), 48–53.
<https://doi.org/10.36787/jti.v14i2.150>
- Kukasvadiya, M. S., & Divecha, N. H. (2020). Analysis of Data Using Data Mining tool Orange. *International Journal of Engineering Development and Research*, 9(2), 1836–1840. www.ijedr.org
- Nurul Salsabila Syam, Valian Yoga Pudya Ardhana, Eliyah A M Sampetoding, Muh. Sulthan Nazhim, A. Muhammad Risqullah, Muh. Gifari Sakawati, A. Muh. Yusril, & M. Dermawan Mulyodiputo. (2022). Model Support Vector Machine untuk Prediksi pada Penggunaan Energi Listrik di Rumah Hemat Energi. *Jurnal Informatika*, 1(2), 56–59.
<https://doi.org/10.57094/ji.v1i2.360>
- Pratama, A. R. I., Latipah, S. A., & Sari, B. N. (2022). Optimasi Klasifikasi Curah Hujan Menggunakan Support Vector Machine (Svm) Dan Recursive Feature Elimination (Rfe). *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(2), 314–324.
<https://doi.org/10.29100/jupi.v7i2.2675>
- Pratama W, B. S. D. (2023). Prediksi Pemeliharaan Transformator Distribusi Berbasis Artificial Neural Network. *Snestik*, 112–120. <https://ejurnal.itats.ac.id/snestik>
- Risanti, R. (2024). Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, Dan Decision Tree. *XII*, 119–128.
<https://doi.org/10.21009/03.1201.fa18>
- Seprima, M., & Defrianto, D. (2020). Prediksi Curah Hujan Dan Kelembapan Udara Kota Pekanbaru Menggunakan Metode Monte Carlo. *Komunikasi Fisika Indonesia*, 17(3), 134. <https://doi.org/10.31258/jkfi.17.3.134-138>
- Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, 11(1), 78–83.
<https://doi.org/10.32764/saintekbu.v11i1.298>
- Swarinoto, Y. S., & Sugiyono, S. (2011). Pemanfaatan Suhu Udara Dan Kelembapan Udara Dalam Persamaan Regresi Untuk Simulasi Prediksi Total Hujan Bulanan Di Bandar Lampung. *Jurnal Meteorologi Dan Geofisika*, 12(3), 271–281.
<https://doi.org/10.31172/jmg.v12i3.109>
- Syaputra, A., Muslim, B., Prawira, N. S., & Edowinsyah, E. (2023). Implementasi Metode Support Vector Machine Dengan Algoritma Genetika Pada Prediksi Konsumsi Energi Untuk Gedung Beton Bertulang. *Faktor Exacta*, 16(2), 142–153.
<https://doi.org/10.30998/faktorexacta.v16i2.1>

6657

Zulfiani, A., & Fauzi, C. (2023). Penerapan Algoritma Backpropagation Untuk Prakiraan Cuaca Harian Dibandingkan Dengan Support Vector Machine dan Logistic Regression. *Jurnal Media Informatika Budidarma*, 7(3), 1229–1237.
<https://doi.org/10.30865/mib.v7i3.6173>