BRAHMANA: Jurnal Penerapan Kecerdasan Buatan
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Desember (2023), pp. 77-84

Pemanfaatan *Machine Learning* untuk Memprediksi Kandungan *Dissolved Oxygen* (DO) pada Air Sungai Menggunakan Metode *Decision Tree Regressor* (DTR) dan *Support Vector Regressor* (SVR)

Ikhwanussafa Sadidan[1], Gina Lova Sari[2], Edmund Uncok Armin[3], Fakhri Ikhwanul Alifin[4], Venny Ulya Bunga[5]
[1,2,5]Program Studi Teknik Lingkungan, Fakultas Teknik, Universitas Singaperbangsa Karawang, Indonesia
[3]Program Studi Teknik Elektro, Fakultas Teknik, Universitas Singaperbangsa Karawang, Indonesia
[4]Program Studi Teknik Lingkungan, Fakultas Teknik, Universitas Singaperbangsa Karawang, Indonesia
[4]Program Studi Teknik Industri, Fakultas Teknik, Universitas Singaperbangsa Karawang, Indonesia
E-mail: [1]ikhwanussafa.sadidan@ft.unsika.ac.id, [2]ginalovasari@unsika.ac.id, [3]edmund.ucok@ft.unsika.ac.id, [4]fakhri.ikhwan@ft.unsika.ac.id, [5]venny.ulya @ft.unsika.ac.id

Abstract

Water quality is a key factor in maintaining a healthy river ecosystem and supporting the life of aquatic organisms. The measurement of Dissolved Oxygen (DO) content in water is one of the crucial parameters that assess the level of dissolved oxygen, directly influencing the life of aquatic organisms. This study aims to predict the Dissolved Oxygen (DO) content in the Citarum River Irrigation Area by employing Support Vector Regression (SVR) analysis and Decision Tree Regressor. The predictive model was developed by analyzing the relationship between other water quality parameters such as Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), and temperature. The analysis results indicate that the accuracy score of the Decision Tree Regressor analysis is superior to that of the Support Vector Regression (SVR) analysis.

Keywords: *Dissolved Oxygen, Citarum River, Decision Tree Regressor, Support Vector Regression*

Abstrak

Kualitas air yang baik adalah faktor kunci dalam menjaga ekosistem sungai yang sehat dan mendukung kehidupan organisme air. Kandungan Dissolved Oxygen (DO) dalam air adalah salah satu parameter penting yang mengukur tingkat oksigen terlarut, yang secara langsung mempengaruhi kehidupan makhluk hidup di dalam air. Studi ini bertujuan untuk memprediksi kandungan Dissolved Oxygen (DO) dalam air di Kawasan Irigasi Sungai Citarum dengan menggunakan analisis Support Vector Regression (SVR) dan Decision Tree Regressor. Model prediksi ini dilakukan dengan cara menganalisis hubungan antara parameter kualitas air lain seperti Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), dan temperatur. Hasil analisis menunjukkan bahwa skor akurasi analisis Decision Tree Regressor lebih akurat dibandingkan dengan hasil analisis Support Vector Regression (SVR).

Kata Kunci: *microsleep, Deep Learning, YOLOv8*

ISSN: 2715-9906

1. Pendahuluan

Air merupakan salah satu sumber daya alam yang memainkan peran penting dalam kehidupan dan penghidupan manusia [1]. Eksploitasi dan pemanfaatan sumber daya air secara berlebihan telah menimbulkan serangkaian permasalahan, seperti penurunan kualitas air, rusaknya wilayah perairan, dan degradasi struktur ekosistem sungai. Hal ini sangat membahayakan pembangunan sosial dan ekonomi serta keselamatan masyarakat [2]. Sungai Citarum dikenal sebagai sungai terbesar dan terpanjang di Jawa Barat, Indonesia, dengan luas mencapai 661.015 hektar dan panjang 297 km [3]. Aktivitas masyarakat di sepanjang DAS menjadikan DAS Citarum sebagai salah satu sungai paling tercemar di dunia karena limbah dan air limbah sering kali dibuang ke badan air tanpa pengolahan yang baik, sehingga mengakibatkan penurunan kualitas air dan ancaman terhadap pemanfaatannya [4]. Prediksi kualitas air sangat penting dilakukan untuk mencegah dan menangani masalah pencemaran air. Prediksi ini dapat membantu memahami sepenuhnya tren dinamis lingkungan ekologi air dan memperingatkan kemungkinan terjadinya pencemaran [2]. Fokus pengendalian pencemaran air telah bergeser dari perbaikan ke pencegahan. Untuk mengurangi polusi air secara efektif, penting untuk memprediksi tren kualitas air di masa depan secara akurat. Peringatan dini ini akan mendorong pengelolaan sumber daya air secara ilmiah, menjaga keberlanjutan ekosistem, dan melindungi kesehatan manusia [1].

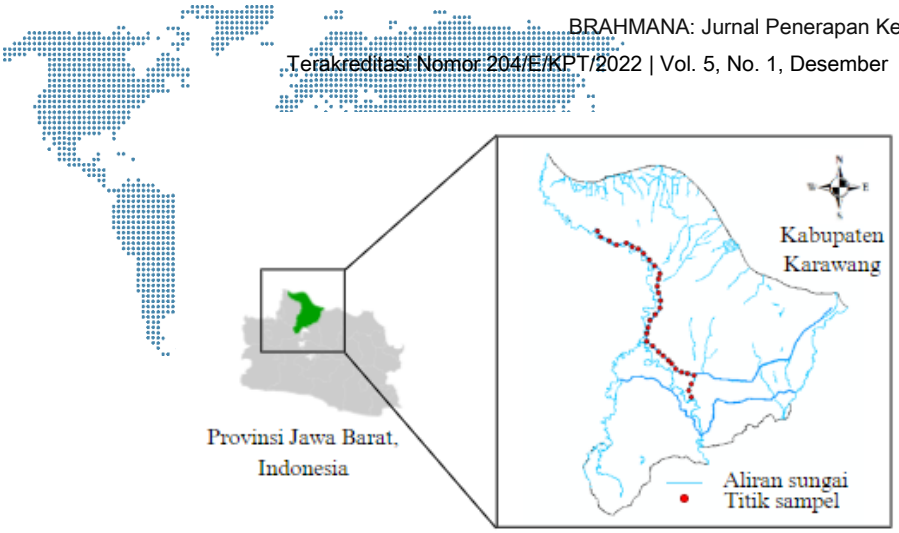
Saat ini, kecerdasan buatan telah diaplikasikan dalam berbagai bidang keilmuan [5], teknik *machine learning* atau pembelajaran mesin pun sudah diperkenalkan untuk memprediksi kualitas air [6]. Metode *machine learning* dapat digunakan pada pemetaan hubungan antara data *input* dan respons *output* yang diinginkan melalui mekanisme proses yang kompleks. Penggunaan data historis yang melimpah, termasuk perubahan data yang dinamis, hubungan *non-linier* tertentu dapat diidentifikasi secara tepat [7].

2. Metodologi Penelitian

Metode yang biasa digunakan pada proses klasifikasi atau pun prediksi adalah *Support Vector Machine (SVM)*, *Backpropagation*, *Radial Basis Function (RBF)*, K- Nearest Network, Analisis Diskriminan, *Simple Logistic Classifier (SLC)*, *Fuzzy*, dan lain-lain [8]. Metode *Support Vector Machine (SVM)* mampu menyesuaikan jenis data yang digunakan untuk berbagai input dengan jumlah kesalahan minimal. Tingkat efisiensi dan akurasi yang dihasilkan oleh metode ini pun terbilang baik [9]. Untuk penyelesaian masalah regresi, diterapkan metode *Support Vector Regression (SVR)* yang memiliki algoritma khusus untuk kasus regresi yang menghasilkan luaran berupa bilangan riil. Konsep algoritma SVR dapat menghasilkan nilai prediksi yang baik karena SVR mempunyai kemampuan untuk menyelesaikan masalah overfitting, yaitu saat data testing atau training menghasilkan akurasi prediksi hampir sempurna [10].

2.1. Pengambilan Sampel

Pengambilan sampel dilakukan dengan menggunakan prinsip *purposive sampling*, yaitu pengambilan sampel dengan menggunakan beberapa pertimbangan tertentu sesuai dengan kriteria yang diinginkan [11]. Sampel air dikumpulkan di 33 titik sepanjang saluran irigasi sepanjang kurang lebih 50 km dan melewati 31 desa di Kabupaten Karawang (lihat Gambar 1). Setiap sampel air dikumpulkan dalam rangkap dua pada jarak sekitar 1,5 km dari setiap lokasi, dikompositkan secara menyeluruh, dipindahkan ke botol kaca steril, dan disimpan pada suhu 4°C.



Gambar 1. Lokasi Pengambilan Sampel

Setelah sampel air terkumpul, dilakukan uji kualitas air di laboratorium untuk mengetahui kandungannya. Adapun parameter yang diukur pada penelitian ini, yaitu *Dissolved Oxygen (DO)*, *Chemical Oxygen Demand (COD)*, *Biologycal Oxygen Demand (BOD)*, dan temperatur. Hasil dari uji laboratorium dapat dilihat pada Tabel 1.

Tabel 1. Hasil Uji Laboratorium

Nomor Sampel	DO (mg/L)	BOD (mg/L)	COD (mg/L)	Temperatur (^{o}C)
1	0.9	12.5	17	29
2	1.2	11.7	19.5	30
3	4	20.9	39.3	38.5
4	3	19.7	28	30
5	4.4	22.2	74.6	31.5
6	3.9	21	33.3	31
7	2.3	21.4	32.8	30
8	7.8	22.9	77.2	29.5
9	6.7	21.3	33.3	30
10	6.3	13.5	20.6	31
11	6.4	22.7	30.8	31.5
12	6.1	23.3	30.6	31.5
13	7.9	11.7	17.7	32
14	6.1	19.7	36.2	32
15	8	17	25	32
16	7	23.2	28.1	30
17	8.6	17.7	25.4	30
18	8.7	17.5	23.5	31
19	8.1	14.2	18.6	33
20	6.8	11.7	20.7	32
21	8.2	14.9	19.7	33
22	7.2	12.4	21.2	30
23	0.7	12.8	24.3	29.4
24	1.5	13.1	21.8	31.1
25	0.9	13.5	48.6	31
26	1	13.1	71.7	31.5
27	7.7	6.2	8.2	31.5
28	7.6	2.4	9	32
29	7.4	2.8	9.7	31.6
30	6.9	2.5	8.9	31.8
31	7.7	5.3	12.5	31.7
32	8.1	4.7	8.1	31.8
33	7.8	1.8	88.8	31.4

Semua parameter yang diuji ini selanjutnya akan dijadikan sebagai variabel untuk memprediksi hubungan antara satu dan yang lainnya. Studi ini menjadikan *Dissolved Oxygen (DO)* sebagai variabel dependen, dan akan dianalisis keterkaitan

kandungannya dengan *Chemical Oxygen Demand (COD)*, *Biologycal Oxygen Demand (BOD)*, dan temperatur menggunakan bantuan *machine learning.*

2.2. Analisis Data

1. Analisis Regresi

Metode regresi merupakan suatu metode analisis data dalam statistik yang digunakan untuk melakukan peramalan dan mempelajari hubungan antar variable [12]. Dalam analisis regresi atau biasa disebut regresi linier dibedakan menjadi dua yaitu regresi linier sederhana dan regresi linier berganda. Regresi linier sederhana digunakan untuk menghasilkan model yang menggambarkan hubungan antara satu variabel bebas dengan satu variabel terikat. Bentuk umum persamaan regresi linier sederhana adalah:

$$Y = \beta_0 + \beta_1 X_1 \quad (1)$$

Sedangkan model regresi linier berganda merupakan pengembangan dari model regresi linier sederhana. Jika model regresi linier sederhana hanya terdiri dari satu variabel bebas dan satu variabel terikat, maka pada regresi linier berganda variabelnya terdiri lebih dari satu variabel bebas dan satu variabel terikat. Dengan memperbanyak variabel independen, maka bentuk umum persamaan regresi linier berganda adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (2)$$

2. Metode Support Vector Regression (SVR)

SVR, pengembangan dari SVM untuk tujuan regresi, bertujuan mencari fungsi hiperplane yang merupakan fungsi regresi terbaik yang sesuai dengan seluruh data input, dengan kesalahan minimal [13]. Dalam hal ini, SVR berupaya menemukan fungsi yang memiliki deviasi maksimum dari target sebenarnya untuk setiap data training. Jika deviasi tersebut sama dengan nol, itu menunjukkan adanya persamaan regresi yang ideal. Dapat dikatakan bahwa Support Vector (SV) menerapkan konsep mirip dengan jaringan saraf, dengan menggunakan fungsi radial basis sebagai bentuk kasus khusus dari jaringan tersebut. Dalam kasus RBF, algoritma SV secara otomatis menentukan pusat, bobot, dan ambang batas untuk meminimalkan kesalahan pengujian yang diharapkan. Meskipun begitu, keunggulan hasil telah terbukti dalam memecahkan masalah seri SVR dalam beberapa kasus [14].

3. Metode *Decision Tree Regression*

Decision Trees (DT) adalah model non-parametrik dari pembelajaran terpantau yang digunakan untuk analisis klasifikasi dan regresi. Model ini didasarkan pada pohon biner yang membagi satu atau lebih simpul untuk membentuk sebuah pohon keputusan [15]. Algoritma decision trees membagi dataset menjadi kelas-kelas yang lebih kecil dan merepresentasikan hasilnya dalam sebuah simpul daun. Pada dasarnya, pohon keputusan melatih dataset dalam bentuk struktur pohon untuk prediksi. Itulah sebabnya mengapa terkadang disebut sebagai regresi struktur pohon. DT memiliki tiga jenis simpul yang berbeda, yaitu simpul akar, simpul interior, dan simpul daun. Simpul akar adalah simpul pertama yang dibagi menjadi lebih banyak simpul, yang disebut simpul interior. Simpul interior mewakili fitur data model dan aturan keputusan, sementara simpul daun mewakili hasil akhir dari keputusan.

3. Hasil dan Pembahasan

Penelitian ini menggunakan framework *Scikit-learn* pada bahasa pemograman python untuk membuat sebuah model regressor. Penelitian ini menggunakan *Personal Computer* (PC) dengan spesifikasi Intel Core-i5 Gen 9 dan Ram 8gb untuk melakukan training dan

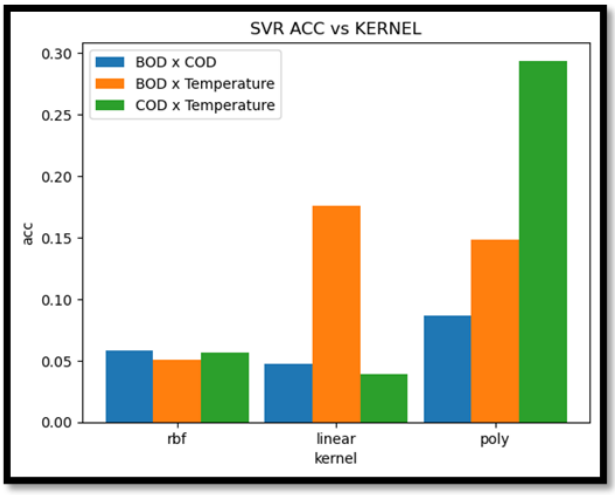
validasi pada model regressor yang telah dibuat. Peneliti menggunakan dua metode machine learning untuk membuat model regressor pada dataset yang digunakan, yaitu *Support Vector Regressor* dan *Decicion Tree Regressor*. Peneliti merancang model regressor dengan menggunakan 2 input data untuk memprediksi DO.

3.1. Hasil Training Model Regresi

Proses training dilakukan dengan memvariasikan filter atau metode yang digunakan pada model *regressor*. Peneliti juga mengkombinasikan input dari model *regressor* yaitu "BOD x COD", "BOD x Temperatur", dan "COD x Temperatur".

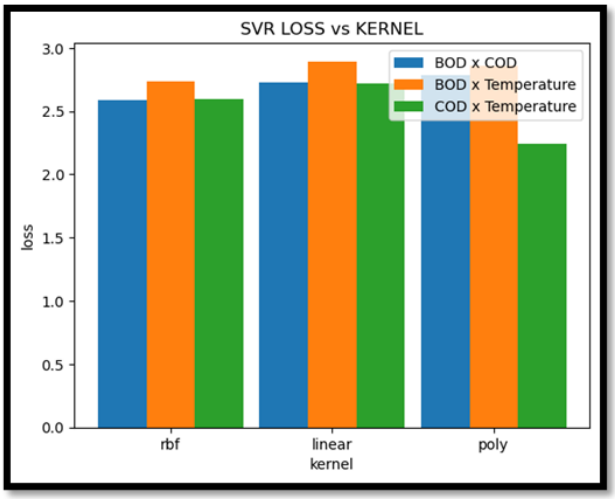
3.2. Hasil Training Support Vector Regression

Pada proses training metode SVR, peneliti mentraining model dengan memvariasikan kernel SVR yaitu RBF, Linear, dan Polynomial pada setiap kombinasi dataset. Dari hasil training, model mendapatkan akurasi tertinggi pada kernel Polynomial dengan menggunakan kombinasi dataset "COD x Temperatur" dengan skor 29.4%. Diagram batang akurasi SVR dapat dilihat pada Gambar 2.



Gambar 2. *Support Vector Regresor Training Accuracy*

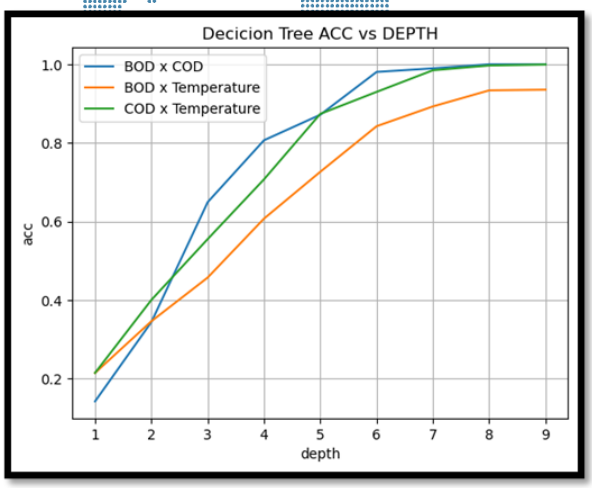
Skor *Root Mean Square Error* terendah didapatkan dengan kernel yang sama yaitu polynomial pada dataset "COD x Temperatur" dengan nilai 2.24 yang dapat dilihat pada Gambar 3.



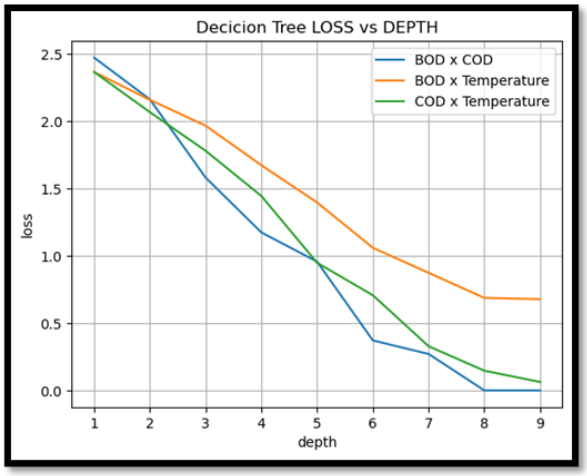
Gambar 3. *Support Vector Regresor Training MSQE*

3.3. Hasil Training Decision Tree Regresion

Pada proses training menggunakan model *Decicion Tree Regressor*, peneliti memvariasikan *depth* yang digunakan dari 1-9 pada setiap kombinasi dataset untuk memprediksi DO. Dari gambar 4, dapat dilihat bahwa skor akurasi pada dataset "BOD x COD".



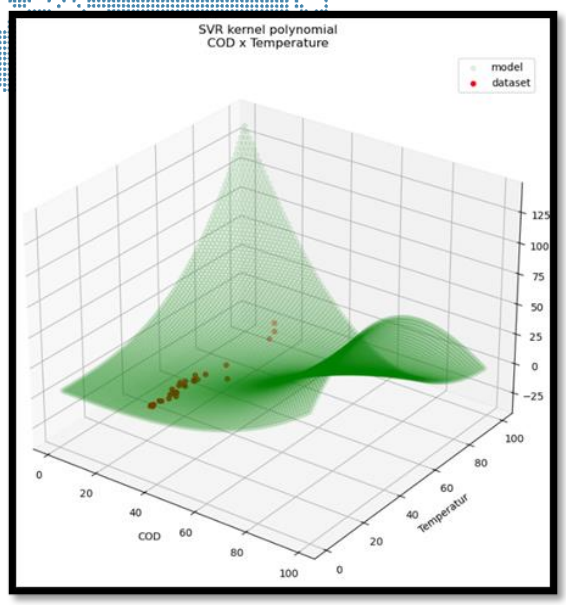
Gambar 4. *Decision Tree Regresor Training Accuracy*



Gambar 5. Decision Tree Regresor Training MSQE

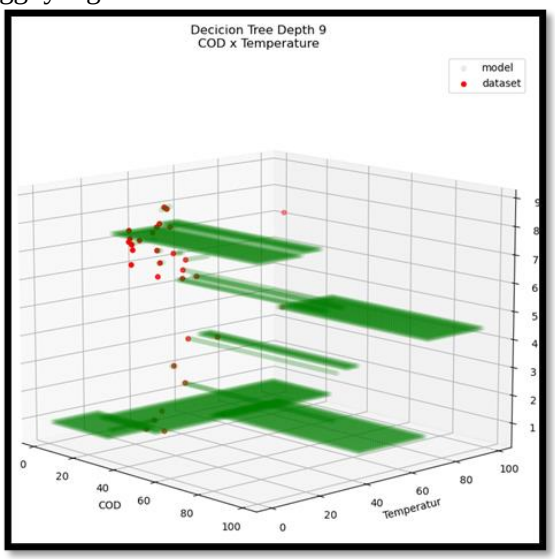
3.4. Visualisasi Model Regresi

Pada tahap ini, peneliti memvariasikan hasil *input* data dengan hasil akurasi model terbaik yang didapatkan pada saat pelatihan model dengan range 0 sampai dengan 100 untuk data Temperatur dan COD. Pada Gambar 6 terlihat bahwa hasil prediksi model SVM membentuk sebuah *surface* yang menghasilkan nilai DO tertinggi melebihi angka 125.



Gambar 6. Visualisasi Support Vector Regresor terhadap input COD dan Temperatur

Sementara itu, Gambar 7 menunjukkan hasil prediksi DO dengan variasi nilai 0-100 pada data COD dan temperatur. Hasil prediksi tervisualisasikan menjadi bidang datar dengan nilai DO tertinggi yang dihasilkan oleh model sebesar 4.



Gambar 7. Visualisasi *Decision Tree Regressor* terhadap input COD dan Temperatur

4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan menggunakan metode SVR dan Decicion Tree Regressor dengan mengkombinasikan input dan filter model, didapatkan skor akurasi tertinggi sebesar 100 % pada model Decicion Tree Regressor dengan depth 9 dan 29.4% pada model SVR dengan kernel polynomial dengan menggunakan data input COD dan temperatur. Loss pada model dihitung menggunakan metode RSME dan mendapatkan skor terkecil yaitu 2.24 pada metode SVR dengan kernel polynomial dan 0.0 pada metode Decision Tree Regressor dengan depth 9. Dari hasil kedua model tersebut, dapat disimpulkan metode Decision Tree Regressor dengan depth 9 menghasilkan hasil prediksi yang lebih baik jika dibandingkan dengan metode SVR menggunakan dataset yang peneliti miliki.

Daftar Pustaka

[1] Suwari. 2021. Analysis of water quality status using method of water pollution index: a case study on the Dendeng River. International Journal of Research - GRANTHAALAYAH. 9 (5): 200–218.

[2] Wu, H.; Cheng, S.; Xin, K.; Ma, N.; Chen, J.; Tao, L.; Gao, M. 2022. Water quality prediction based on multi-task learning. Int. J. Environ. Res. Public Health 2022. 19 (15) : 1-19.

[3] Sari, G. L., Hadining, A.F., & H. Sudrajat. 2020. Analisis karakteristik fisik-kimiawi air daerah aliran Sungai Citarum di Waduk Jatiluhur. J. Teknik Lingkungan. 6(1):1–9.

[4] Quay, C. 2018. Water quality impacts of the citarum river on Jakarta and surrounding Bandung Basin. The Ohio State University Libraries.

[5] Min, Hokey. (2010). Artificial intelligence in supply chain management: theory and applications. International Journal of Logistics Research and Applications. 13(1): 13-39.

[6] Liu, M., Lu, J. 2014. Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river. Environ Sci Pollut Res. 21(18) : 1-18.

[7] Deng, T., Chau, K.W., Duan, H.F. 2021. Machine learning based marine water quality prediction for coastal hydro-environment management. J. Environ. Management 284 : 1-14.

[8] Isnaeni, Sudarmin, & Rais, Z. 2022. Analisis Support Vector Regression (SVR) dengan kernel Radial Basis Function (RBF) untuk memprediksi laju inflasi di Indonesia VARIANSI : J. of Statistics and Its Application on Teaching and Research. 4(1) : 30-38.

[9] Widiastuti, N. I., Rainarli, E., & Dewi, K. E. 2017. Peringkasan dan support vector machine pada klasifikasi dokumen. J. Infotel. 9(4) : 416-421.

[10] Furi, R. P., Si, M., & Saepudin, D. 2015. hal. 3608-3617. Prediksi financial time series menggunakan independent component analysis dan support vector regression Studi Kasus : IHSG dan JII. e-Proceeding of Engineering 2(2) Agustus 2015.

[11] Sugiyono. 2011. Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung, Alfabeta.

[12] Kutner, M.H., Nachsteim, C.J., & Neter, J. 2004. Applied Linear Regression Models, 4th penyunt., New York: McGraw-Hill Companies, Inc.

[13] Smola, A. J., & Sch, B. 2004. Statistics and Computing - A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222.

[14] Caraka, R. E. (2017). Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis. 7(1), 43–57.

[15] Kadavi, P. R., Lee, C., & Lee, S. 2019. Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. Environmental Earth Sciences.