Jurnal Teknologi Terpadu

https://journal.nurulfikri.ac.id/index.php/jtt
ISSN : 2477-0043 ISSN ONLINE : 2460-7908

KOMPARASI ALGORITMA *MACHINE LEARNING* DALAM MEMPREDIKSI KAPASITAS PRODUKSI POTENSIAL AIR BERSIH DI INDONESIA

Tatang Rohana[1], Hilda Yulia Novita[2], Euis Nurlaelasari[3]

[1,2,3]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang
Karawang, Jawa Barat, Indonesia 41361
tatang.rohana@ubpkarawang.ac.id, hilda.yulia@ubpkarawang.ac.id, euis.nurlaelasari@ubpkarawang.ac.id

Abstract

Clean water availability is a key indicator of sustainable development, particularly in developing countries like Indonesia. Factors such as population growth, climate change, and urbanization contribute to fluctuations in clean water supply. This study aims to estimate the potential for clean water production in Indonesia using various machine learning algorithms, such as Linear Regression, Decision Tree, Random Forest, Multilayer Perceptron, XGBoost (Extreme Gradient Boosting), and Neural Network. Each algorithm was evaluated based on Mean Squared Error (MSE), Mean Absolute Error (MAE), R-squared (R^2), and prediction accuracy. The results show that Linear Regression achieved the lowest MSE (9.31E-18), nearly zero, indicating extremely accurate predictions. Neural Network and Multilayer Perceptron also performed well, with MSE values of 0.00010898 and 0.00018004, respectively. Moreover, Linear Regression and Neural Network achieved R^2 scores of 1 and 0.9905, suggesting they can explain nearly all variability in the target data. These findings highlight the effectiveness of Linear Regression, Neural Network, and Multilayer Perceptron in modeling clean water production capacity. Therefore, these algorithms are recommended as the most reliable approaches for supporting data-driven decisions in clean water resource planning and management in Indonesia.

***Keywords**: Decision Tree, Linear Regression, Machine Learning, Prediction, Random Forest*

Abstrak

Ketersediaan air bersih merupakan indikator penting dalam pembangunan berkelanjutan, terutama di negara berkembang seperti Indonesia. Pertumbuhan penduduk, perubahan iklim, dan urbanisasi menyebabkan fluktuasi dalam ketersediaan air bersih. Penelitian ini bertujuan untuk memperkirakan potensi produksi air bersih di Indonesia menggunakan algoritma pembelajaran mesin, seperti Regresi Linier, *Decision Tree, Random Forest, Multilayer Perceptron, XGBoost (Extreme Gradient Boosting)*, dan *Neural Network*. Setiap algoritma dievaluasi dan dibandingkan berdasarkan *Mean Squared Error* (MSE), koefisien determinasi (R^2), *Mean Absolute Error* (MAE), dan tingkat akurasi prediksi. Berdasarkan hasil evaluasi kinerja dari masing-masing algoritma, maka terlihat Regresi Linier memiliki nilai MSE yang sangat rendah (9.31E-18), hampir mendekati nol, yang menunjukkan bahwa model ini sangat tepat dalam memprediksi target pada dataset ini. *Neural Network* dan *Multilayer Perceptron* juga memiliki MSE yang sangat rendah, yaitu 0.00010898 dan 0.0001800368, yang menunjukkan performa model yang sangat baik dengan *error* yang sangat kecil. Regresi Linier dan *Neural Network* mencapai nilai $R^2 = 1$ dan 0.9905, yang berarti model ini dapat menjelaskan hampir 100% variasi dari data target, menunjukkan prediksi yang sangat akurat. Secara keseluruhan, Regresi Linier, *Neural Network*, dan *Multilayer Perceptron* dapat direkomendasikan sebagai algoritma yang paling efektif untuk prediksi kapasitas produksi air bersih di Indonesia.

Kata kunci : *Decision Tree, Machine Learning, Prediksi, Random Forest,* Regresi Linier

1. PENDAHULUAN

Di semua negara, ketersediaan air bersih merupakan suatu keharusan yang harus dipenuhi untuk memenuhi kebutuhan air bagi warga negaranya, tak terkecuali Indonesia. Dengan populasi yang terus bertambah dan urbanisasi yang semakin pesat, kebutuhan akan air bersih meningkat signifikan setiap tahunnya. Namun, berbagai faktor seperti perubahan iklim, degradasi lingkungan, dan keterbatasan infrastruktur sering kali menghambat kapasitas produksi dan distribusi air bersih [1]. Untuk pengelolaan air bersih secara berkelanjutan diperlukan suatu cara atau strategi yang efektif dan berbasis data dalam pengelolaan ketersediaan air bersih tersebut. Meskipun air menutupi sekitar 70% dari permukaan bumi dengan total sekitar 1,4 miliar kilometer kubik, sayangnya hanya sekitar 0,003% dari jumlah tersebut yang dapat dimanfaatkan dengan baik. Sebagian besar air, yaitu sekitar

Diterima 7 Desember 2024, Direvisi 16 Juni 2025, Diterima untuk publikasi 28 Juli 2025

97%, berada di samudra atau laut dengan kadar garam yang terlalu tinggi untuk sebagian besar kebutuhan. Dari sisa 3% air tawar, hampir seluruhnya, kurang lebih 87%, terdapat pada lapisan es di kutub atau jauh di bawah permukaan tanah.

Teknologi *Mechine Learning* (ML telah berkembang menjadi alat yang potensial dalam berbagai sektor, termasuk manajemen sumber daya air. Algoritma ML dapat digunakan untuk memprediksi berbagai variabel seperti dalam penelitian [2] [3] [4] [5] [6] [7] dalam berbagai topik penelitian, termasuk prediksi kapasitas produksi air bersih di masa depan. Prediksi yang akurat tentang kapasitas produksi air bersih sangat penting untuk mendukung perencanaan, pengambilan keputusan, serta pengelolaan infrastruktur air agar mampu memenuhi kebutuhan masyarakat [1]. Namun, salah satu tantangan utama dalam penerapan ML pada prediksi kapasitas air bersih adalah pemilihan algoritma yang paling tepat [8]. Berbagai algoritma ML memiliki kinerja yang bervariasi tergantung pada karakteristik dataset dan permasalahan yang ada. Karenanya dibutuhkan suatu terobosan proses komparasi untuk mengidentifikasi algoritma yang paling efektif dan akurat dalam memprediksi kapasitas produksi air bersih di Indonesia.

Berdasarkan statistik air bersih tahun 2023 oleh Badan Pusat Statistik (BPS), kebutuhan air bersih semakin meningkat sehingga perlu ada kajian atau studi tentang persediaan air bersih dalam bentuk klasifikasi maupun prediksi ketersediaan air bersih. Penelitian–penelitian yang sudah dilakukan tentang ketersediaan air di antaranya, [1] Melakukan penelitian tentang analisis kebutuhan dan ketersediaan air bersih di kota Bandung. Pada kajian ini ditunjukkan bahwa pada tahun 2029, Kecamatan Kiaracondong akan membutuhkan sekitar 21,276 liter air bersih per detik. Penelitian lainnya [2] menghasilkan akurasi dalam menentukan apakah kualitas air layak dikonsumsi atau tidak. Tingkat akurasi yang diperoleh untuk masing-masing metode adalah 60,19% dengan metode *Decision Tree*, 62,80% dengan metode *Logistic Regression*, 68,59% dengan metode *Support Vector Machine* (SVM), dan 69,54% dengan metode ANN. Prediksi tentang kebutuhan pemakaian air [5], objek penelitian ini adalah pemakaian atau kebutuhan air di PDAM kota Malang. Hasilnya akurasi prediksi pemakaian ini cukup baik. Penelitian tentang prediksi kualitas air Ciliwung [9] menghasilkan akurasi metode SVM di atas 89%. Penelitian lain yang sudah dilakukan tentang prediksi di antaranya [3] [10] [11] [12] [13] .

Beberapa algoritma *ML* seperti *Decision Tree*, *Random Forest*, SVM, *Neural Network*, *XGBoost*, *Gradient Boosting*, dan *Multilayer Perceptron* sudah digunakan pada penelitian [14] [15] [7] [16]. Pada penelitian ini, model algoritma yang akan dibandingkan terdiri dari 6 algoritma. Harapannya, penelitian ini bisa memberikan kontribusi signifikan terhadap perencanaan sumber daya air yang lebih baik dan berkelanjutan di Indonesia.

2. METODE PENELITIAN

2.1. Objek Penelitian

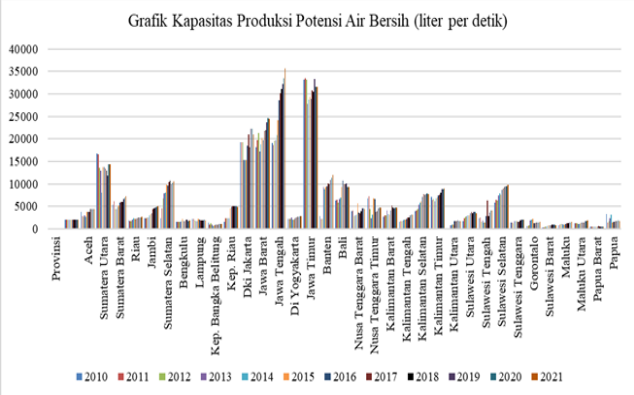
Kapasitas produksi potensial air bersih yang dihasilkan oleh perusahaan air bersih di seluruh Indonesia dari tahun 2010 sampai dengan tahun 2021 pada tabel 1 merupakan objek pada penelitian ini.

Tabel 1. Jumlah produksi Perusahaan Air Bersih Menurut Provinsi (liter per detik)

No	Provinsi	Data Kapasitas Produksi Air Bersih				
		2010	2011	…….	2020	2021
1	Aceh	3750	2803		4441	4394
2	Sumatera Utara	16700	16610		14336	14270
3	Sumatera Barat	5269	6060		6960	7171
4	Riau	1903	1672		2512	2583
5	Jambi	2307	2409		4855	4999
6	Sumatera Selatan	2303	4484		10193	10513
7	Bengkulu	1520	1535		1779	1899
8	Lampung	2125	2146		2043	1892
9	Kep. Bangka Belitung	1350	895		1118	1118
10	Kep. Riau	1558	2316		5042	4847
11	DKI Jakarta	19300	19300		22260	20967
12	Jawa Barat	18198	19720		24570	24428
13	Jawa Tengah	19053	18614		33405	35605
14	Di Yogyakarta	2235	2117		2840	2793
15	Jawa Timur	33215	33416		31543	31495
16	Banten	2852	2249		11376	11954
17	Bali	6311	6474		9343	9254
18	Nusa Tenggara Barat	3899	4086		4607	4453
19	Nusa Tenggara Timur	6744	7258		4648	4711
20	Kalimantan Barat	2684	2973		4685	4736
21	Kalimantan Tengah	1568	1634		3131	3075
22	Kalimantan Selatan	3868	4144		7938	7750
23	Kalimantan Timur	7112	6440		8786	8951
24	Kalimantan Utara	0	0		1655	1655
25	Sulawesi Utara	1655	2296		3824	3373
26	Sulawesi Tengah	2401	2500		4063	4000
27	Sulawesi Selatan	5839	6494		9503	9769
28	Sulawesi Tenggara	1398	1435		1981	2067
29	Gorontalo	675	666		1405	1483
30	Sulawesi Barat	328	348		959	811
31	Maluku	784	880		1399	1540
32	Maluku Utara	1211	1226		1745	1808
33	Papua Barat	417	510		506	511

No	Provinsi	Data Kapasitas Produksi Air Bersih				
		2010	2011	…….	2020	2021
34	Papua	3222	1355		1878	1660

Data kapasitas potensi air bersih yang digunakan dalam penelitian ini terdiri dari hasil produksi air bersih dari setiap provinsi di seluruh Indonesia dari tahun 2010 sampai dengan tahun 2021 yang dapat dilihat pada gambar 1.



Gambar 1. Grafik Kapasitas Produksi Potensi Air Bersih (liter per detik)

2.2 Pengumpulan Data

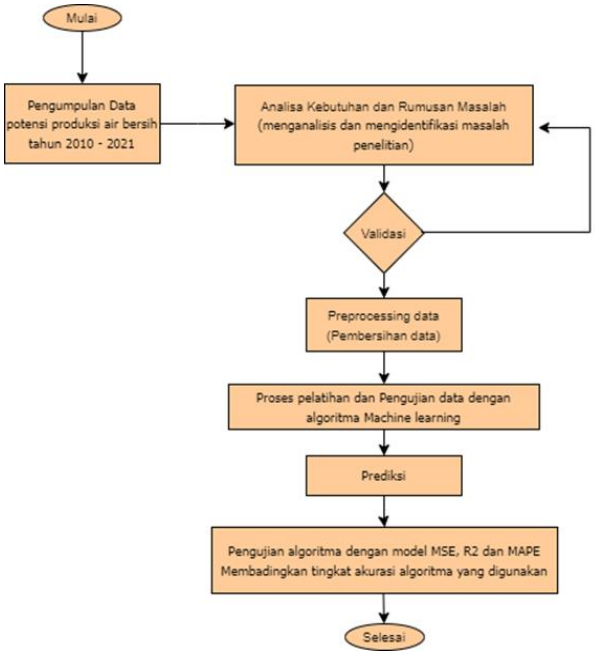
Data adalah unit informasi yang tercatat pada media tertentu, memiliki ciri khas yang membedakannya dari data lain, serta dapat dianalisis dan relevan untuk tujuan tertentu. Pengumpulan data dilakukan melalui prosedur yang sistematis dan terstandarisasi untuk memperoleh data yang diperlukan. Dalam penelitian ini, penulis menggunakan dua metode pengumpulan data, yaitu observasi dan dokumentasi. Observasi dilakukan dengan meninjau data sekunder dari Badan Pusat Statistik (BPS) Indonesia.

2.3 Analisis Data

Teknik analisis data yang diterapkan melibatkan pembagian data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data pelatihan berfungsi untuk melatih model, sementara data pengujian digunakan untuk mengevaluasi kinerja model. Kesimpulan dari hasil pengujian model kemudian akan diverifikasi melalui diagnosis pada data pengujian. Analisis data pada penelitian ini bertujuan untuk mengetahui seberapa akurat model ML (ML) dalam memprediksi kapasitas produksi potensial air bersih yang dihasilkan oleh perusahaan air bersih di seluruh Indonesia. Proses pengukuran kinerja algoritma dilakukan dengan menggunakan uji nilai hasil *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *R-squared* (R^2).

2.4 Tahapan Penelitian

Pada pelaksanaan penelitian ini maka gambar di bawah ini merupakan tahapan yang menjelaskan proses penelitian.



Gambar 2. Tahapan Penelitian

Gambar 2 merupakan tahapan–tahapan dalam penelitian ini, dapat dijelaskan sebagai berikut:

1. Pengumpulan Data dan Menentukan Objek Penelitian

   Penelitian ini dilakukan secara berkala selama sekitar 9 bulan, mulai dari Juni 2024 hingga Desember 2024. Data yang akan digunakan dalam penelitian ini adalah data mengenai potensi produksi air bersih di seluruh Indonesia dari tahun 2010 hingga 2021.

2. Analisa Kebutuhan Data Penelitian

   Untuk memperkuat referensi dalam penelitian ini, juga digunakan data berupa materi pendukung atau jurnal yang dapat membantu dalam proses penelitian. Beberapa jurnal diambil dari koleksi jurnal resmi yang tersedia di situs resmi yang dapat diakses di internet.

3. Merumuskan Masalah Penelitian

   Dalam penelitian ini, dirumuskan masalah yang berkaitan dengan data kelulusan mahasiswa tepat waktu meliputi:

   a. Bagaimana menerapkan algoritma ML untuk memprediksi kapasitas potensi produksi air bersih di Indonesia.
   b. Bagaimana tingkat akurasi dan sebaran data yang terbentuk pada ML dari hasil uji data yang dilakukan.
   c. Bagaimana cara menganalisis hasil data uji dan *training* dari algoritma ML.

4. Validasi Data Penelitian

   Validasi data merupakan langkah penting dalam penelitian ini, karena bertujuan untuk memastikan

bahwa data yang digunakan memenuhi kriteria yang telah ditetapkan, serta dapat dipertanggungjawabkan sumber dan kebenarannya.

5. *Pre-Processing*

   Pada tahap ini, data penelitian terlebih dahulu dinormalisasi untuk menghindari ketidakakuratan dan inkonsistensi. Setelah itu, dilakukan proses transformasi data ke dalam bentuk *Min-Max*.

6. Proses Pelatihan dan Pengujian Data

   Untuk dapat menghasilkan sistem yang dapat memprediksi kapasitas potensi produksi air bersih, maka dilakukan proses pelatihan dan pengujian menggunakan algoritma ML yang digunakan.

7. Proses Prediksi

   Langkah selanjutnya adalah proses prediksi, hal ini dilakukan untuk memprediksi kapasitas potensi produksi air bersih di Indonesia untuk masa yang akan datang.

3. HASIL DAN PEMBAHASAN

3.1. *Pre-processing* Data

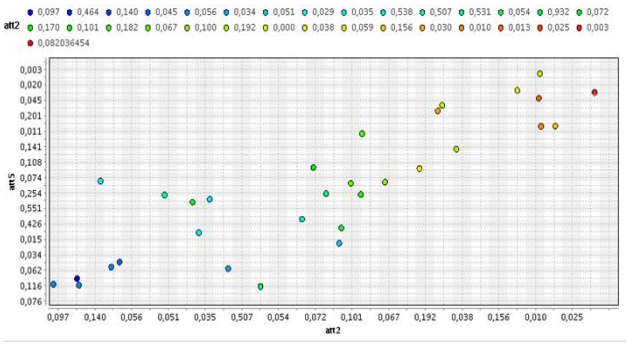
Awal pemrosesan yang dilakukan pada penelitian ini, dengan cara melakukan pra-pengolahan data terhadap data potensi produksi air bersih di seluruh Indonesia tahun 2010 sampai tahun 2021. Terkadang, data tersebut mengandung berbagai masalah yang dapat mempengaruhi hasil dari proses tersebut, seperti *missing value*, data redundan, *outliers*, atau format data yang tidak sesuai dengan sistem. Proses pra-pemrosesan data ini mencakup hal-hal berikut :

1. Pembersihan data (*Data cleaning*)

   Pembersihan data pada gambar 3 dilakukan untuk menghapus data yang tidak efisien dan mengandung kesalahan. Proses ini dilakukan dengan menggunakan aplikasi Rapidminer.

Gambar 3. Proses Pembersihan Data Produksi Air Bersih

Berdasarkan hasil pembersihan data, menunjukkan tidak ditemukan adanya kesalahan atau *missing* pada data yang digunakan pada penelitian ini.



Gambar 4. Plot Data Produksi Air Bersih Tahun 2010

Gambar 4 menunjukkan sebaran data potensi produksi air bersih di semua provinsi yang ada di Indonesia tahun 2010.

2. Normalisasi data ke bentuk Min-Max

   Sebelum data dimasukkan ke dalam jaringan, data terlebih dahulu ditransformasikan ke dalam bentuk data interval (normalisasi). Data tersebut dinormalisasi sehingga berada dalam rentang [0,1]. Tujuan dari normalisasi adalah untuk menyamakan skala nilai setiap data, sehingga setiap data memiliki kontribusi yang proporsional dalam setiap proses. Data hasil konversi *Min-Max* dapat dilihat pada tabel 2.

Tabel 2. Data Hasil Konversi Min-Max

Provinsi	Tahun				
	2010	**2011**	**…….**	**2020**	**2021**
Aceh	0,097	0,07		0,117	0,115
Sumatera Utara	0,464	0,462		0,397	0,395
Sumatera Barat	0,14	0,162		0,188	0,194
Riau	0,045	0,038		0,062	0,064
Jambi	0,056	0,059		0,128	0,132
Sumatera Selatan	0,056	0,118		0,28	0,289
Bengkulu	0,034	0,034		0,041	0,045
Lampung	0,051	0,052		0,049	0,044
Kep. Bangka Belitung	0,029	0,016		0,022	0,022
Kep. Riau	0,035	0,056		0,134	0,128
Dki Jakarta	0,538	0,538		0,622	0,585
Jawa Barat	0,507	0,55		0,687	0,683
Jawa Tengah	0,531	0,518		0,938	1
Di Yogyakarta	0,054	0,051		0,071	0,07
Jawa Timur	0,932	0,938		0,885	0,883
Banten	0,072	0,054		0,313	0,33
Bali	0,17	0,174		0,256	0,253
Nusa Tenggara Barat	0,101	0,107		0,121	0,117
Nusa Tenggara Timur	0,182	0,196		0,122	0,124
Kalimantan Barat	0,067	0,075		0,124	0,125
Kalimantan Tengah	0,035	0,037		0,079	0,078
Kalimantan Selatan	0,1	0,108		0,216	0,21

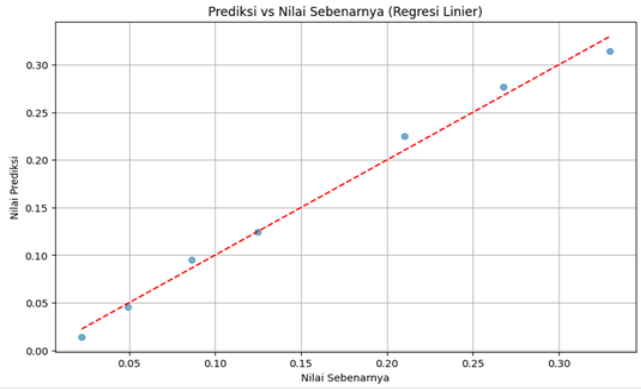
Provinsi	Tahun				
	2010	2011	…….	2020	2021
Kalimantan Timur	0,192	0,173		0,24	0,244
Kalimantan Utara	0	0		0,038	0,038
Sulawesi Utara	0,038	0,056		0,099	0,086
Sulawesi Tengah	0,059	0,062		0,106	0,104
Sulawesi Selatan	0,156	0,175		0,26	0,268
Sulawesi Tenggara	0,03	0,031		0,047	0,049
Gorontalo	0,01	0,01		0,031	0,033
Sulawesi Barat	0	0,001		0,018	0,014
Maluku	0,013	0,016		0,03	0,034
Maluku Utara	0,025	0,025		0,04	0,042
Papua Barat	0,003	0,005		0,005	0,005
Papua	0,082	0,029	0,03	0,044	0,038

3.2. Pelatihan Dan Pengujian Data

Data dibagi menjadi dua bagian, yaitu data latih (*Training set*) periode tahun 2010 – 2019 dan data uji (*Testing set*) periode 2020 - 2021.

3.2.1 Pelatihan dan Pengujian Algoritma Regresi Linier

Proses selanjutnya dilakukan proses pelatihan dan pengujian data. Data yang sudah dilakukan proses pelatihan selanjutnya dilakukan pelatihan data. Data selanjutnya dijadikan sumber data untuk proses *training* (pelatihan) dan *testing* (data uji) pada proses analisis Regresi Linier. Untuk proses analisis, data dibagi menjadi dua tahap, yaitu data *training* dan data *testing*.

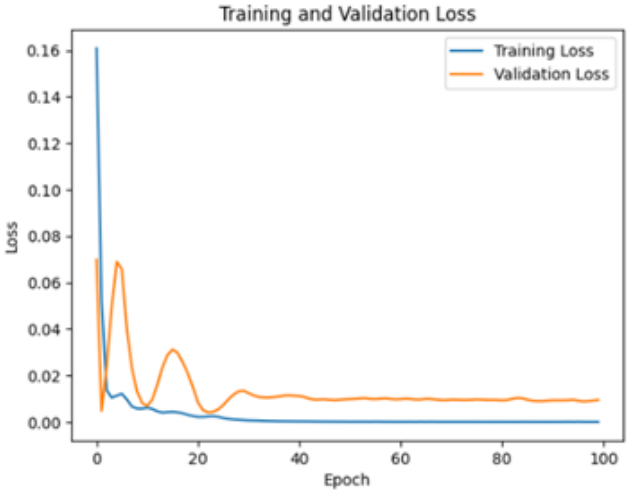


Gambar 5. Perbandingan antara Nilai Prediksi dan Data Aktual

Dari hasil pengujian pada gambar 5 didapatkan nilai MSE sebesar 9.306368624128321e-33, nilai MAE sebesar 7.632783294297951e-17, dan nilai R^2 sebesar 1.0.

3.2.2 Pelatihan dan Pengujian Algoritma *Multilayer Perceptron*

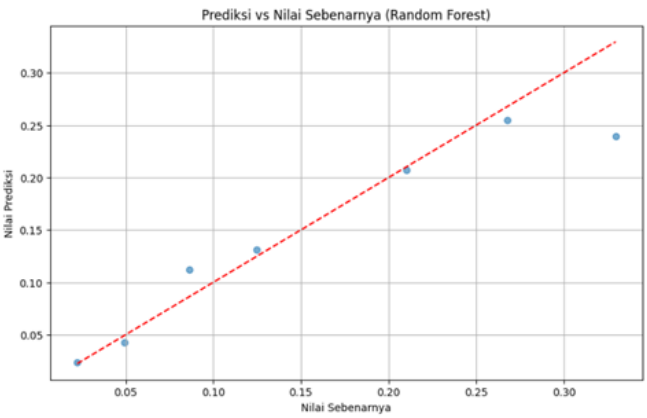
Data yang sudah dibersihkan, selanjutnya dijadikan sumber data untuk proses *training* (pelatihan) dan *testing* (pengujian) pada algoritma *Multilayer Perceptron*. Untuk proses analisis, data dibagi menjadi dua tahap, yaitu data *training* dan data *testing*. Proses pelatihan dilakukan dengan melakukan 100 *epoch*, dan menghasilkan nilai MSE sebesar 0.0001800368, MAE sebesar 0.011356400471, dan R^2 sebesar 0.98436373873 yang dapat dilihat pada gambar 6.



Gambar 6. Grafik *Training and Validation Loss*

3.2.3 Pelatihan dan Pengujian Algoritma *Random Forest*

Pelatihan dan pengujian selanjutnya adalah algoritma *Random Forest*, algoritma ini merupakan metode *ensemble* berbasis pohon keputusan, memiliki keunggulan dalam hal akurasi dan kemampuan generalisasi. Sama dengan pelatihan sebelumnya, untuk proses analisis, data dibagi menjadi dua bagian, yaitu data *training* dan data *testing*.



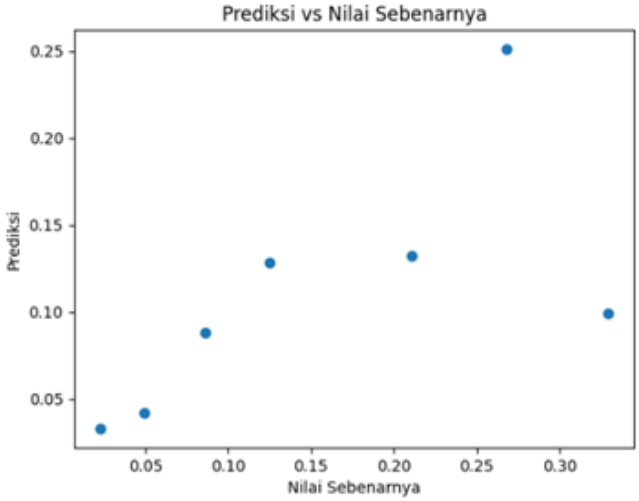
Gambar 7. Perbandingan antara Nilai Aktual dan Prediksi

Berdasarkan gambar 7 hasil pengujian yang dilakukan maka didapatkan nilai MSE 0.0012908747851076011, R^2 sebesar 0.8748804498896507, dan *MAPE* 14.210977728 %.

3.2.4 Pelatihan dan Pengujian Algoritma *XGBoost* (*Extreme Gradient Boosting*)

Algoritma selanjutnya adalah *XGBoost* (*Extreme Gradient Boosting*). Untuk algoritma ini pun sama dilakukan pelatihan dan pengujian, di mana data dibagi dua bagain data *training* dan data *testing*. Dari hasil pelatihan menggunakan algoritma *XGBoost* (*Extreme Gradient Boosting*), dihasilkan nilai MSE sebesar 0.0085409135324,

MAE sebesar 0.049742894000, dan R^2 sebesar 0.2582186178. Nilai MSE ini menunjukkan rata-rata dari kuadrat kesalahan antara nilai prediksi dan nilai aktual. Semakin rendah nilai MSE, semakin baik model dalam hal akurasi prediksi. Nilai MSE ini cukup kecil, yang berarti bahwa pada beberapa bagian model dapat memprediksi dengan cukup baik.

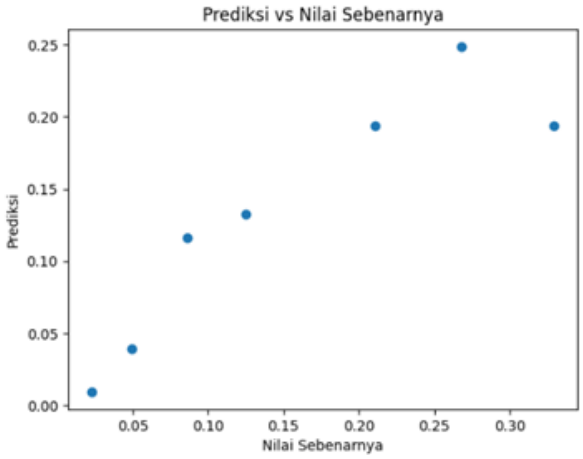


Gambar 8. Perbandingan antara Nilai Aktual dan Prediksi

Berdasarkan gambar 8 nilai MAE menunjukkan rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual. Nilai MAE sebesar 0.0497 menunjukkan bahwa, secara rata-rata, prediksi model meleset sekitar 0.0497 dari nilai aktual. Nilai MAE ini juga terbilang rendah, yang menunjukkan bahwa sebagian besar prediksi cukup dekat dengan nilai aktual. Nilai R^2 sebesar 0.2582 atau sekitar 25.82% menunjukkan bahwa model hanya mampu menjelaskan sekitar 25.82% variabilitas data. Ini adalah nilai yang rendah, yang berarti bahwa model belum mampu menjelaskan sebagian besar variabilitas yang ada dalam data.

3.2.5 Pelatihan dan Pengujian Algoritma *Decision Tree*

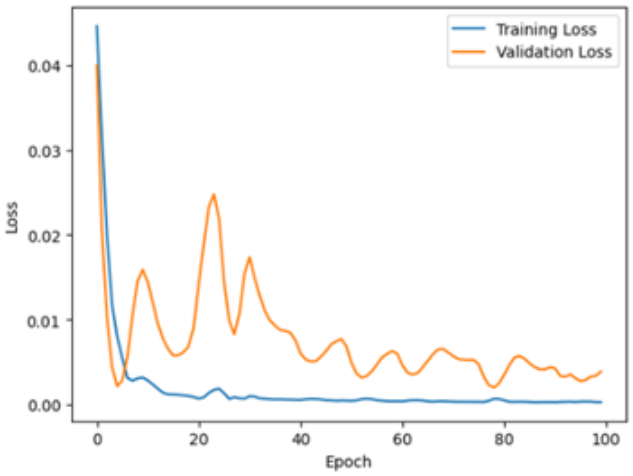
Algoritma *Decision Tree* merupakan algoritma selanjutnya dalam penelitian ini. Data yang sudah dibersihkan, selanjutnya dijadikan sumber data untuk proses *training* (pelatihan) dan *testing* (pengujian). Dari hasil pelatihan dan pengujian menggunakan algoritma *Decision Tree* menghasilkan nilai MSE sebesar 0.002889241878, MAE sebesar 0.03304749061, dan R^2 sebesar 0.7490683138. Secara keseluruhan, hasil ini menunjukkan bahwa model *Decision Tree* memiliki performa yang lebih baik dibandingkan hasil sebelumnya dari model *XGBoost*, khususnya dalam hal R^2, yang jauh lebih tinggi (74.90% dibandingkan 25.82%). Gambar 9 menunjukkan bahwa model *Decision Tree* lebih efektif dalam menjelaskan variabilitas data dan menghasilkan prediksi yang lebih akurat pada dataset ini.



Gambar 9. Perbandingan antara Nilai Aktual dan Prediksi

3.2.6 Pelatihan dan Pengujian Algoritma *Neural Network*

Pelatihan dan pengujian algoritma selanjutnya adalah *Neural Network*. Setelah data dibagi dua antara data *training* dan data uji, selanjutnya dilakukan pelatihan dan pengujian, di mana *epoch* yang dilakukan sebanyak 100 kali. Dari hasil pelatihan dan pengujian menggunakan algoritma *Neural Network* pada gambar 10 didapatkan MSE sebesar 0.000108982952, MAE sebesar 0.008737408082, dan R^2 sebesar 0.990534791. Hasil ini menunjukkan bahwa model *Neural Network* memiliki performa yang sangat baik dengan MSE dan MAE yang sangat rendah serta R^2 yang mendekati 1. Ini mengindikasikan bahwa model *Neural Network* sangat akurat dan mampu menjelaskan variabilitas data dengan baik, sehingga model ini sangat efektif untuk dataset yang digunakan.



Gambar 10. Hasil Pelatihan dan Pengujian Model *Neural Network*

3.3. Hasil dan Evaluasi

Berdasarkan hasil evaluasi kinerja dari masing-masing algoritma yang digunakan berdasarkan dengan model MSE, MAE, dan R^2, maka terlihat Regresi Linier memiliki nilai MSE yang sangat rendah (9.31E-18), hampir mendekati nol, yang menunjukkan bahwa model ini sangat tepat dalam memprediksi target pada dataset ini. *Neural Network* dan *Multilayer Perceptron* juga memiliki MSE yang sangat

rendah, yaitu sebesar 0.00010898 dan 0.0001800368, yang menunjukkan performa model yang sangat baik dengan *Error* yang sangat kecil. Sebaliknya, *XGBoost* memiliki MSE yang lebih tinggi sebesar 0.00854091, menunjukkan bahwa model ini kurang tepat dalam memprediksi pada dataset yang digunakan.

Regresi Linier dan *Neural Network* mencapai nilai $R^2 = 1$ dan 0.9905, yang berarti model ini dapat menjelaskan hampir 100% variasi dari data target, menunjukkan prediksi yang sangat akurat. *Multilayer Perceptron* juga menunjukkan R^2 tinggi sebesar 0.9844, menunjukkan bahwa model ini cukup kuat dalam menangkap pola data. *XGBoost*, dengan R^2 sebesar 0.2582, menunjukkan bahwa model ini hanya mampu menjelaskan sekitar 25.8% variasi dari data target, yang mengindikasikan bahwa model ini kurang cocok untuk dataset ini.

Neural Network memiliki nilai MAE terendah sebesar 0.00873740, menunjukkan kesalahan rata-rata yang sangat kecil pada prediksi, yang berarti model ini memberikan prediksi yang cukup akurat. Regresi Linier dan *Multilayer Perceptron* juga memiliki MAE yang rendah, menunjukkan akurasi prediksi yang tinggi. *XGBoost* memiliki nilai MAE yang relatif lebih tinggi, yakni sebesar 0.04974289, yang mengonfirmasi bahwa model ini tidak optimal pada dataset ini.

Regresi Linier, *Neural Network*, dan *Multilayer Perceptron* menunjukkan kinerja yang sangat baik, dengan nilai MSE, MAE, dan R^2 yang menunjukkan kemampuan prediksi yang tinggi. *Random Forest* juga menunjukkan performa yang cukup baik dengan R^2 sebesar 0.8749. *XGBoost* menunjukkan performa yang kurang optimal dalam prediksi kapasitas air pada dataset ini, dengan MSE dan MAE yang lebih tinggi serta R^2 yang jauh lebih rendah dibandingkan dengan algoritma lainnya. Tabel 3 berikut adalah perbandingan hasil evaluasi untuk masing-masing algoritma yang digunakan dalam penelitian ini.

Tabel 3. Hasil Evaluasi Algoritma

No	Algoritma	MSE	R^2 Score	MAE
1	Regresi Linier	9,31E-18	1	7,63E-02
2	*Decision Tree*	0.00288924	0.74906831	0.03304749
3	*Random Forest*	0.00129087	0.87488044	
4	*Multilayer Perceptron*	0.0001800368	0.98436373	0.01135640
5	*XGBoost* (*Extreme Gradient Boosting*)	0.00854091	0.25821861	0.04974289
6	*Neural Network*	0.00010898	0.99053479	0.00873740

4. KESIMPULAN

Berdasarkan hasil evaluasi model menggunakan metrik *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan R-*squared* (R^2), dapat disimpulkan bahwa Regresi Linier, *Neural Network*, dan *Multilayer Perceptron* merupakan algoritma dengan performa terbaik dalam memprediksi kapasitas produksi air bersih pada dataset ini. Hal ini terlihat dari nilai MSE yang sangat rendah, MAE yang kecil, serta nilai R^2 yang mendekati atau mencapai 1, yang menunjukkan bahwa ketiga model ini mampu menangkap pola data dan menghasilkan prediksi yang sangat akurat.

Random Forest juga menunjukkan kinerja yang cukup baik, dengan R^2 sebesar 0.8749, meskipun tidak seunggul Regresi Linier, *Neural Network*, dan *Multilayer Perceptron*. Sementara itu, *XGBoost* menunjukkan performa yang kurang optimal, dengan MSE dan MAE yang lebih tinggi serta R^2 yang rendah (0.2582), yang menunjukkan bahwa algoritma ini kurang cocok untuk dataset ini dan tidak dapat menangkap pola data dengan baik seperti model-model lainnya.

Secara keseluruhan, Regresi Linier, *Neural Network*, dan *Multilayer Perceptron* dapat direkomendasikan sebagai algoritma yang paling efektif untuk prediksi kapasitas produksi air bersih dalam penelitian ini, dengan akurasi tinggi dan kesalahan prediksi yang minimal.

Hasil penelitian ini memiliki implikasi penting bagi perumusan kebijakan pengelolaan sumber daya air bersih di Indonesia, khususnya dalam konteks perencanaan dan pengambilan keputusan berbasis data. Dengan diketahuinya bahwa algoritma Regresi Linier, *Neural Network*, dan *Multilayer Perceptron* mampu menghasilkan prediksi kapasitas produksi air bersih dengan akurasi tinggi, maka instansi terkait, seperti dinas air minum, perencanaan wilayah, dan lembaga pengelola lingkungan, dapat memanfaatkan model-model ini untuk:

1. Perencanaan produksi dan distribusi air bersih.
2. Identifikasi wilayah rawan krisis air.
3. Alokasi anggaran dan investasi infrastruktur.
4. Monitoring dan evaluasi kinerja sistem air bersih.
5. Pengembangan sistem pendukung keputusan (DSS).

Ucapan Terima Kasih

Penulis menyampaikan rasa terima kasih kepada seluruh pihak di Fakultas Ilmu Komputer Universitas Buana Perjuangan Karawang, yang telah memberikan dukungan, masukan, dan motivasi sehingga penulis dapat mempublikasikan hasil penelitian ini.

DAFTAR PUSTAKA

[1] G. M. Geoniti dan F. Yustiana, “Analisis Kebutuhan dan Ketersediaan Air PDAM Tirtawening Wilayah Bandung Timur Kecamatan Kiaracondong Kota Bandung,” *Reka Racana,* vol. xx, pp. 1-7, 2020.

[2] D. Hartanti dan A. I. Pradana, "Komparasi Algoritma *Machine Learning* dalam Identifikasi Kualitas Air," *SMARTICS Journal,* vol. 9, pp. 1-6, 2023.

[3] T. H. Hasibuan dan D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 pada UIN Syarif Hidayatullah Jakarta," *SKANIKA: Sistem Komputer dan Teknik Informatika,* vol. 6, no. 1, pp. 61-74, 2023.

[4] D. A. Putra dan M. Kamayani, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes," *Seminar Nasional TEKNOKA,* vol. 5, pp. 34-40, 2020.

[5] B. Putro, M. T. Furqon dan S. H. Wijoyo, "Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode *Exponential Smoothing,*" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2, pp. 4679-4686, 2018.

[6] T. Rohana, J. Indra dan G. G. Munzi, "Kajian Model *Backpropagation* dan *Hybrid* ANFIS dalam Memprediksi Pertumbuhan Penduduk di Kabupaten Karawang," *JOURNAL OF INFORMATION SYSTEM RESEARCH,* vol. 4, pp. 374-381, 2023.

[7] S. Shabbir, "*Comparing Performance of J48, Multilayer Perceptron*," *Research,* 2015.

[8] M. Thoriq, A. E. Syaputra dan Y. S. Eirlangga, "Perkiraan Kebutuhan Air Bersih Menggunakan Metode Jaringan Syaraf Tiruan *Backpropagation*," *JURNAL FASILKOM,* vol. 13, pp. 438-444, 2023.

[9] M. Haekal dan W. C. Wibowo, "Prediksi Kualitas Air Sungai Menggunakan Metode Pembelajaran Mesin: Studi Kasus Sungai Ciliwung," *Jurnal Teknologi Lingkungan,* vol. 24, no. 2, pp. 273-282, 2023.

[10] T. Rohana, "*Performance Evaluation of Adaptive Neuro-Fuzzy Inference System (ANFIS) In Predicting New Students* (Case Study: UBP Karawang)," *BIT and CS,* vol. 2, no. 2, pp. 31-37, 2021.

[11] T. Rohana, E. Nurlaelasari, E. E. Awal dan H. Y. Novita, "Kajian Model Jaringan Syaraf Tiruan untuk Memprediksi Secara Dini Tingkat Kelulusan Mahasiswa," *Technologia: Jurnal Ilmiah,* vol. 15, no. 4, pp. 629-640, 2024.

[12] U. Riyanto, "Penerapan Algoritma *Multilayer Perceptron* (MLP) Dalam Menentukan Kelayakan Kenaikan Jabatan: Studi Kasus PP. Abc - Jakarta," *Jurnal Teknik Informatika (JIKA),* vol. 2, no. 1, pp. 58-65, 2018.

[13] F. J. Zebua, R. P. B. Manalu dan M. N. K. Nababan, "Prediksi Kelulusan Mahasiswa Menggunakan Perbandingan Algoritma C5.0 Dengan *Regression Linear*," *Jurnal TEKINKOM,* vol. 4, no. 2, pp. 230-238, 2021.

[14] R. Soelistijadi, T. D. Wismarini, S. Eniyati dan S. S., "Pemodelan Prediktif Menggunakan Metode *Ensemble Learning* XGBoost dalam Peningkatan Akurasi Klasifikasi Penyakit Ginjal," *Kesatria,* vol. 5, no. 4, pp. 1866-1875, 2024.

[15] N. Muhammad, Data Mining Untuk Memprediksi Kelulusan Mahasiswa Jurusan Teknik Informatika UIN Syarif Hidayatullah Jakarta Menggunakan Metode Klasifikasi C4.5, Jakarta: UIN Syarif Hidayatullah, 2022.

[16] T. Rohana, "Implementasi Model *Hybrid* Dalam Memprediksi Penyebaran Covid-19 Di Wilayah Jawa Barat," *Seminastika,* vol. 4, no. 2, pp. 124-137, 2021.