

**ALGORITMA MACHINE LEARNING UNTUK PREDIKSI PRODUKSI
TANAMAN PADI DI PULAU SUMATERA**

Muhammad Fikri Hidayat¹, Junadhi²

STMIK AMIK Riau

E-mail: 2110031802062@sar.ac.id¹, junadhi@sar.ac.id²

ABSTRAK

Pulau Sumatera memiliki lebih dari 50 persen lahan pertanian di setiap provinsi dengan komoditas pangan utama yang paling dominan adalah beras, sedangkan komoditas minor lainnya adalah jagung, kacang tanah, dan ubi jalar. Hasil pertanian di Pulau Sumatera sangat rentan terhadap perubahan iklim dan dampak negatifnya dapat mempengaruhi pola tanam, waktu penanaman, produksi, dan kualitas hasil panen. Selain itu, peningkatan suhu bumi akibat dampak pemanasan global akan memengaruhi pola presipitasi, penguapan, aliran air, kelembaban tanah, dan variasi iklim yang sangat fluktuatif secara keseluruhan dapat mengancam keberhasilan produksi pertanian. Berdasarkan konteks ini, diperlukan prediksi produksi tanaman padi menggunakan algoritma machine learning. Dalam penelitian ini, tiga metode regresi yang digunakan adalah Lasso Regression, Random Forest Regression, dan Gradient Boosting Regression. Data yang digunakan ini berjumlah 224 data dan terkait catatan informasi dari tahun 1993 hingga 2020 yang mencakup 8 provinsi di Pulau Sumatera, yaitu Nanggroe Aceh Darussalam, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, dan Lampung. Berdasarkan pengujian yang dilakukan didapatkan bahwa algoritma Gradient Boosting Regression mampu menghasilkan nilai r^2 score terbaik dengan nilai sebesar 86,92%. Hasil dari studi ini diharapkan dapat memberikan informasi yang berguna bagi para petani dalam meningkatkan produktivitas tanaman padi.

Kata Kunci — Gradient Boosting Regression, Lasso Regression, Machine Learning, Optimasi Produksi Padi, Prediksi, Random Forest Regression, Tanaman Padi.

ABSTRACT

Sumatra Island has more than 50 percent of agricultural land in each province, with the most dominant main food commodity being rice, while other minor commodities include corn, peanuts, and sweet potatoes. Agricultural yields in Sumatra Island are highly vulnerable to climate change, and its negative impacts can affect planting patterns, timing, production, and the quality of harvest. Furthermore, the rising global temperatures resulting from the effects of global warming will influence precipitation patterns, evaporation, water flow, soil moisture, and overall fluctuating climate variations, posing a threat to the success of agricultural production. Given this context, the prediction of rice crop production using machine learning algorithms is needed. In this study, three regression methods employed were Lasso Regression, Random Forest Regression, and Gradient Boosting Regression. The data used consisted of 224 records spanning from 1993 to 2020, encompassing eight provinces in Sumatra Island, namely Nanggroe Aceh Darussalam, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, and Lampung. Based on the conducted tests, it was found that the Gradient Boosting Regression algorithm produced the best r^2 score with a value of 86.92%. The findings of this study are expected to provide valuable information for farmers in enhancing rice crop productivity.

Keywords — Gradient Boosting Regression, Lasso Regression, Machine Learning, Prediction, Random Forest Regression, Rice Crop, Rice Production Optimization.

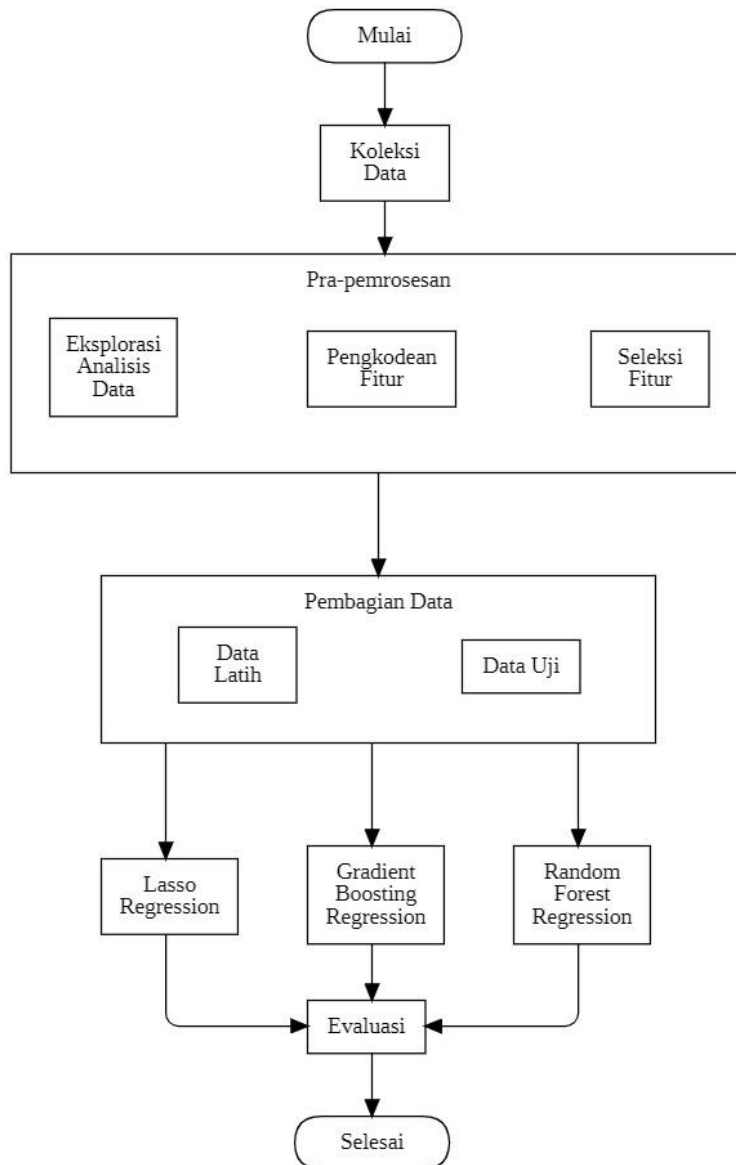
1. PENDAHULUAN

Sumatera adalah salah satu pulau di Indonesia yang memiliki luas lahan pertanian yang cukup luas dan beragam [1]. Sumatera mempunyai bermacam tumbuhan pangan dalam zona pertanian salah satunya yakni padi [2]. Padi merupakan tanaman pangan penting yang ditanam di seluruh dunia [3]. Produksi padi di Sumatera cukup tinggi, dengan luas lahan pertanaman yang cukup luas dan kondisi iklim yang cocok untuk pertumbuhan padi. Produksi padi dipengaruhi oleh berbagai faktor, seperti kondisi lingkungan, teknologi pertanian, dan faktor sosial ekonomi [4]. Beberapa daerah di Sumatera yang dikenal sebagai sentra produksi padi adalah Aceh, Jambi, Riau, Sumatera Utara, dan Sumatera Barat. Penelitian menunjukkan bahwa produksi padi dapat ditingkatkan melalui penerapan teknologi pertanian yang baik, seperti sistem irigasi, pemupukan, dan pengendalian hama dan penyakit [5]. Namun, produktivitas padi di Sumatera masih di bawah rata-rata nasional, hal ini disebabkan karena masih banyaknya penggunaan teknologi yang tidak sesuai dan masalah lingkungan yang masih belum teratasi. Terdapat sebagian riset yang menggunakan algoritma machine learning prediksi serta jadi acuan peneliti dalam melaksanakan penelitian ini. Penelitian [6] yang menggunakan algoritma Random Forest Regression dalam memprediksi harga sewa apartemen di DKI Jakarta pada website mamikos.com mendapati hasil performa MAE sebesar 520686,33 setelah dilakukan tuning parameter menggunakan model searchgridcv dari hasil performa MAE sebesar 568424,38. Kemudian, penelitian [7] yang menggunakan algoritma Linear Regression, Lasso Regression, Ridge Regression, Random Forest Regression, dan Support Vector Regression dalam memprediksi harga pembelian mobil mendapati hasil performa R^2 sebesar 0,99958 dan MAE sebesar 2284865,29.

Berdasarkan hal di atas, pada penelitian ini akan dilakukan perbandingan sebagian algoritma machine learning prediksi diantaranya Lasso Regression, Random Forest Regression, dan Gradient Boosting Regression terhadap dataset produksi tanaman padi di Sumatera dalam kurun tahun 1993-2020 yang dibagi menjadi 70% data latih serta 30% data uji. Dataset yang digunakan dalam penelitian berasal dari Kaggle Open Datasets [8] dengan ukuran 224 baris dan 7 kolom. Penelitian ini diharapkan dapat memberikan rekomendasi rentang target produksi per tahun mendatang untuk setiap provinsi yang ada di Sumatera.

2. METODE PENELITIAN

Berdasarkan alur metode penelitian yang digambarkan pada Gambar 1, tahapan proses penelitian yang digunakan dalam penelitian ini dijelaskan secara rinci sebagai berikut:



Gambar 1. Metodologi Penelitian

3. HASIL DAN PEMBAHASAN

Koleksi Data

Koleksi data yang akan dipergunakan dalam studi ini adalah data terbuka yang bersumber dari

Kaggle Open Datasets. Dataset tersebut merupakan data produksi tanaman padi di pulau sumatera yang berjumlah 224 data.

Tabel 1. Koleksi data yang digunakan

No	Provinsi	Tahun	Produksi	Luas Panen	Curah Hujan	Kelembapan	Suhu rata-rata
1	Aceh	1993	1329536.00	323589.00	1627.0	82.00	26.06
2	Aceh	1994	1299699.00	329041.00	1521.0	82.12	26.92
3	Aceh	1995	1382905.00	339253.00	1476.0	82.72	26.27
...	...	...	...	...	...	...	...
223	Lampung	2019	2164089.33	464103.42	1706.4	78.03	27.23
224	Lampung	2020	2604913.29	545149.05	2211.3	75.80	24.58

Pra-pemrosesan

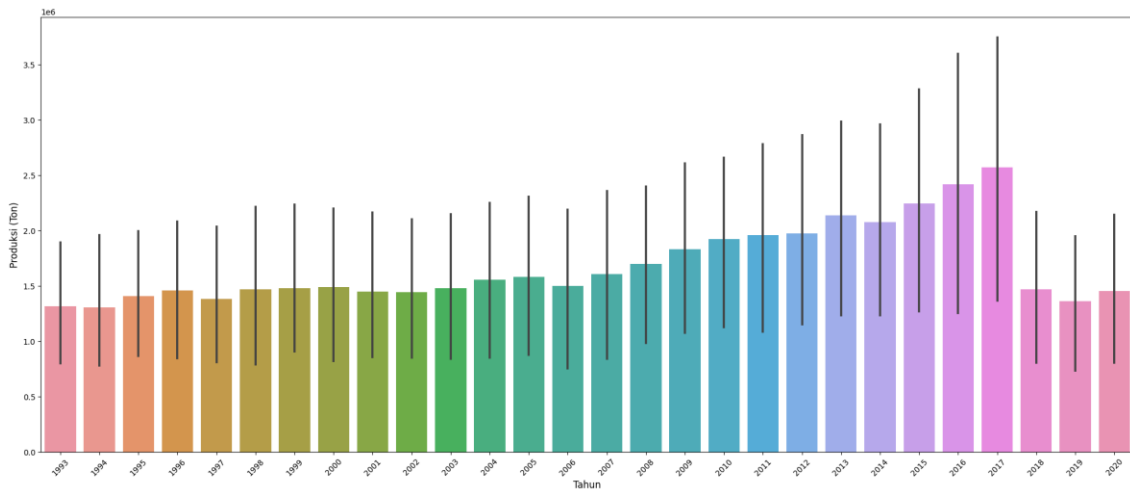
Eksplorasi Analisis Data

Dari data yang sudah didapat sebelumnya, selanjutnya data tersebut akan dianalisis lebih lanjut untuk mengetahui informasi dari produksi tanaman padi di pulau Sumatera. Berikut bisa kita ketahui bahwa produksi tertinggi selama tahun 1993 hingga 2020 ialah provinsi Sumatera Utara dengan jumlah produksi sebanyak 93 juta ton.



Gambar 2. Produksi padi berdasarkan provinsi

Adapun informasi yang didapatkan ialah total produksi padi di pulau Sumatera berdasarkan tahunnya dan diketahui bahwa produksi tanaman padi tertinggi di pulau Sumatera terjadi pada tahun 2017 dengan jumlah produksi sebanyak 27 juta ton. Informasi ini dapat terlihat pada grafik berikut ini.



Gambar 3. Produksi padi berdasarkan tahun

Pengkodean Fitur

Pada tahapan ini, diperlukannya pengkodean fitur terhadap data yang digunakan dalam penelitian. Dimana fitur provinsi bernilai kategorikal dan harus diubah menjadi label numerik, hasil pengkodean fitur terlihat pada tabel berikut.

Tabel 2. Pengkodean fitur

No	Provinsi	Tahun	Produksi	Luas Panen	Curah Hujan	Kelembapan	Suhu rata-rata
1	0	1993	1329536.00	323589.00	1627.0	82.00	26.06
2	0	1994	1299699.00	329041.00	1521.0	82.12	26.92
3	0	1995	1382905.00	339253.00	1476.0	82.72	26.27
...	...	...	...	...	...	...	...
223	3	2019	2164089.33	464103.42	1706.4	78.03	27.23
224	3	2020	2604913.29	545149.05	2211.3	75.80	24.58

Seleksi Fitur

Pada tahapan ini, fitur-fitur yang tidak relevan untuk prediksi akan dihapus agar meningkatkan hasil evaluasi dari algoritma-algoritma yang akan diujikan. Fitur-fitur yang dihapuskan ialah fitur No dan Tahun, hasil seleksi fitur terlihat pada tabel berikut

Tabel 3. Seleksi fitur

Provinsi	Produksi	Luas Panen	Curah Hujan	Kelembapan	Suhu rata-rata
0	1329536.00	323589.00	1627.0	82.00	26.06
0	1299699.00	329041.00	1521.0	82.12	26.92
0	1382905.00	339253.00	1476.0	82.72	26.27
...	...	...	...	...	...
3	2164089.33	464103.42	1706.4	78.03	27.23
3	2604913.29	545149.05	2211.3	75.80	24.58

Pembagian Data

Pada tahap ini, dilakukan percobaan pembagian data dengan rasio 70:30 dan data dibagi menjadi dua subset yang berbeda, yaitu data uji (test data) dan data latih (train data). Berikut tabel yang menunjukkan jumlah pembagian data latih dan data uji dengan rasio tersebut.

Tabel 4. Data latih dan data uji

Rasio Pembagian		Jumlah Data	
Data Latih	Data Uji	Data Latih	Data Uji
70%	30%	156	68

Lasso Regression

Algoritma Lasso Regression yang dibangun menggunakan parameter alpha dan bernilai 0,1. Selanjutnya data dilatih menggunakan data latih terhadap algoritma Lasso Regression dan melakukan prediksi menggunakan data uji.

Gradient Boosting Regression

Algoritma Gradient Boosting Regression yang dibangun menggunakan parameter subsample yang bernilai 0,1, n_estimators yang bernilai 1000, max_depth yang bernilai 7, min_samples_leaf yang bernilai 2, min_samples_split yang bernilai 2, dan learning_rate yang bernilai 0,01. Selanjutnya data dilatih menggunakan data latih terhadap algoritma Gradient Boosting Regression dan melakukan prediksi menggunakan data uji.

Random Forest Regression

Algoritma Random Forest Regression yang dibangun menggunakan parameter n_estimators yang bernilai 365, max_depth yang bernilai 3, min_samples_leaf yang bernilai 10, dan min_samples_split yang bernilai 4. Selanjutnya data dilatih menggunakan data latih terhadap algoritma Random Forest Regression dan melakukan prediksi menggunakan data uji.

Evaluasi

Pada tahapan ini merupakan tahapan akhir dari proses prediksi terhadap algoritma machine learning yang digunakan. Algoritma machine learning tersebut akan diukur keakuratannya berdasarkan metrik evaluasi prediksi antara lain R2-Score, Mean Absolute Error, dan Mean Squared Error. Berikut hasil dari evaluasi algoritma machine learning yang diujikan.

Tabel 5. Perbandingan hasil evaluasi algoritma machine learning

Algoritma	R2-Score	Mean Absolute Error	Mean Squared Error
Gradient Boosting Regression	86,92%	242029,711049	159936982035,226868
Lasso Regression	84,39%	243758,563310	190831750516,069946
Random Forest Regression	83,30%	273467,912324	204173509645,281738

4. KESIMPULAN

Berdasarkan hasil penelitian yang membandingkan kinerja tiga algoritma machine learning yaitu, Lasso Regression, Gradient Boosting Regression, dan Random Forest Regression. Didapatkan hasil sebagai berikut:

1. Gradient Boosting Regression memiliki R2 Score tertinggi sebesar 86,92%. R2 Score digunakan untuk mengukur sejauh mana model dapat menjelaskan variasi dalam data target. Semakin tinggi nilai R2 Score, semakin baik model dapat menjelaskan variasi data.
2. Gradient Boosting Regression memiliki Mean Absolute Error (MAE) terendah sebesar 242029711,049. MAE digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MAE, semakin akurat model dalam memprediksi produksi tanaman padi.

3. Gradient Boosting Regression memiliki Mean Squared Error (MSE) terendah sebesar 159936982035,226868. MSE digunakan untuk mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MSE, semakin akurat model dalam memprediksi produksi tanaman padi.

Berdasarkan data produksi padi berdasarkan provinsi, Sumatera Utara memiliki total produksi tertinggi sebesar 93 juta ton. Hal ini menunjukkan bahwa Sumatera Utara adalah provinsi dengan produksi padi yang signifikan.

Selain itu, data menunjukkan bahwa produksi padi tertinggi terjadi pada tahun 2017 sebesar 26 juta ton. Ini menunjukkan bahwa tahun 2017 merupakan tahun dengan produksi padi yang paling tinggi di Pulau Sumatera. Informasi ini dapat berguna untuk perencanaan pertanian dan pengambilan keputusan terkait produksi tanaman padi di pulau Sumatera

DAFTAR PUSTAKA

- [1] B. R. Aprildahani, C. T. H. Permana, and S. T. E. W. Utama, "Kebutuhan Lahan Pertanian Minimum untuk Kesejahteraan Petani di Pulau Sumatera," *J. Sci. Appl. Technol.*, vol. 5, no. 1, pp. 116–125, 2021.
- [2] I. M. Noor and M. P. Ir Muhammad Saleh, *Inovasi teknologi lahan rawa mendukung kedaulatan rawa*. Perpustakaan Balittra, 2021.
- [3] Y. Setiani, U. Unang, and B. Rofatin, "Penentuan Komoditas Unggulan Sub Sektor Tanaman Pangan dan Hortikultura di Setiap Kecamatan Kabupaten Tasikmalaya," *J. Agristan*, vol. 3, no. 2, pp. 149–171, 2021.
- [4] R. Harini, R. D. Ariani, S. Supriyati, and M. C. Satriagasa, "Analisis luas lahan pertanian terhadap produksi padi di Kalimantan Utara," *J. Kawistara*, vol. 9, no. 1, pp. 15–27, 2019.
- [5] T. Soedarto and R. K. Ainiyah, *Teknologi Pertanian Menjadi Petani Inovatif 5.0: Transisi Menuju Pertanian Modern*. Uwais Inspirasi Indonesia, 2022.
- [6] I. L. Mulyahati, "Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python (Studi Kasus: Apartemen di DKI Jakarta Pada Website mamikos. com)," 2020.
- [7] A. M. M. Fattah, A. Voutama, N. Heryana, and N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, pp. 1669–1678, 2022.
- [8] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and R. Rahmaddeni, "Implementasi Algoritma Decision Tree C4. 5 dan Support Vector Regression untuk Prediksi Penyakit Stroke: Implementation of Decision Tree Algorithm C4. 5 and Support Vector Regression for Stroke Disease Prediction," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 61–67, 2022.
- [9] M. Makbul, "Metode pengumpulan data dan instrumen penelitian," 2021.
- [10] S. Sarosa, *Analisis data penelitian kualitatif*. Pt Kanisius, 2021.
- [11] H. Hartatik *et al.*, *DATA SCIENCE FOR BUSINESS: Pengantar & Penerapan Berbagai Sektor*. PT. Sonpedia Publishing Indonesia, 2023.
- [12] Y. Rifa'i, "Analisis Metodologi Penelitian Kulitatif dalam Pengumpulan Data di Penelitian Ilmiah pada Penyusunan Mini Riset," *Cendekia Inov. Dan Berbudaya*, vol. 1, no. 1, pp. 31–37, 2023.
- [13] I. Andriano, "Visualisasi Data Hulu Migas Di Cekungan Sumatera Tengah Di Pusat Penelitian Pengembangan Teknologi Minyak Dan Gas Bumi 'Ppptomg' Lemigas." Univeristas Komputer Indonesia, 2022.
- [14] I. G. I. Sudipa, I. B. G. Sarasvananda, H. Prayitno, I. N. T. A. Putra, R. Darmawan, and D. A. WP, *Teknik Visualisasi Data*. PT. Sonpedia Publishing Indonesia, 2023.
- [15] L. P. I. Kharisma *et al.*, *METODE SPK FAVORIT DI MASA DEPAN: Teori dan Contoh*. PT. Sonpedia Publishing Indonesia, 2023.

- [16] E. Ardyan *et al.*, *METODE PENELITIAN KUALITATIF DAN KUANTITATIF: Pendekatan Metode Kualitatif dan Kuantitatif di Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2023.
- [17] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naïve Bayes*, vol. 1. Kreatif, 2020.
- [18] I. Junaedi, N. Nuswantari, and V. Yasin, "Perancangan Dan Implementasi Algoritma C4. 5 Untuk Data Mining Analisis Tingkat Risiko Kematian Neonatum Pada Bayi," *JISICOM (Journal Inf. Syst. Informatics Comput.*, vol. 3, no. 1, pp. 29–44, 2019.
- [19] L. FAHADRA, "SISTEM CHATBOT LAYANAN INFORMASI TUGAS AKHIR MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)." UNIVERSITAS ISLAM SULTAN AGUNG, 2023.
- [20] I. D. Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*, vol. 1. Unri Press, 2021.
- [21] E. Edwar, I. G. A. N. R. Semadi, M. Samsudin, and I. K. Dharmendra, "Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini PILKADA DKI 2017)," *INFORMATICS Educ. Prof. J. Informatics*, vol. 8, no. 1, pp. 11–18, 2023.
- [22] D. ARIYOGA, "Perbandingan Metode Seleksi Fitur Filter, Wrapper, Dan Embedded Pada Klasifikasi Data Nirs Mangga Menggunakan Random Forest Dan Support Vector Machine (Svm)," 2022.
- [23] S. C. AISHWARYA, "Optimasi Model Sistem Rekomendasi Film Dengan Neural Network (Studi Kasus: Platform Letterboxd)," 2023.
- [24] D. R. S. Saputro, *WEKA 3.6. 9 (Waikato Environment for Knowledge Analysis): Tools untuk Memahami Machine Learning*. Stiletto Book, 2023.
- [25] I. Setiawan, R. F. A. Cahyani, and I. Sadida, "EXPLORING COMPLEX DECISION TREES: UNVEILING DATA PATTERNS AND OPTIMAL PREDICTIVE POWER," *J. Innov. Futur. Technol.*, vol. 5, no. 2, pp. 112–123, 2023.
- [26] A. C. Muhammad *et al.*, *Dasar-dasar Pembelajaran Mesin:(Foundations of Machine Learning)*. Sada Kurnia Pustaka, 2023.
- [27] S. Hermawan, S. Budi, and S. Kom, "Analisis dan Prediksi Pertempuran Game Of Thrones Menggunakan Algoritma Random Forest dan Logistic Regression," *J. Strateg. Maranatha*, vol. 3, no. 2, pp. 454–459, 2021.
- [28] S. S. Nurashila, F. Hamami, and T. F. Kusumasari, "PERBANDINGAN KINERJA ALGORITMA RECURRENT NEURAL NETWORK (RNN) DAN LONG SHORT-TERM MEMORY (LSTM): STUDI KASUS PREDIKSI KEMACETAN LALU LINTAS JARINGAN PT XYZ," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 3, pp. 864–877, 2023.