

Perbandingan Kinerja Logistic Regression dan Decision Tree dalam Memprediksi Produksi Padi di Sumatera

Muhammad Fawwaz Albasith, Ahmad Albani Islami, Rido Rabani Kurniawan, Muhammad Yunas Hidayatullah, Muhammad Aliva Nurramadhan

Universitas Bumigora, Mataram, Indonesia

Correspondence : e-mail: 2201010027@universitasbumigora.ac.id

Abstrak

Penelitian ini membandingkan kinerja dua algoritma klasifikasi, yaitu Logistic Regression dan Decision Tree, dalam memprediksi tingkat produksi padi di delapan provinsi di Sumatera. Data yang digunakan mencakup kurun waktu 1993 hingga 2020 dan melibatkan variabel seperti curah hujan, suhu, kelembapan, dan luas panen. Seluruh data diproses melalui tahap praproses dan transformasi untuk menghasilkan model klasifikasi multi-kelas: Rendah, Sedang, dan Tinggi. Hasil evaluasi menunjukkan bahwa kedua model memiliki akurasi yang sama sebesar 79,41%. Namun, Decision Tree menunjukkan F1-score tertimbang yang sedikit lebih tinggi, yaitu 0,7815 dibandingkan dengan 0,7760 pada Logistic Regression. Hal ini mengindikasikan bahwa Decision Tree lebih efektif dalam mengenali pola data yang kompleks dan tidak seimbang. Temuan ini menunjukkan bahwa pemilihan algoritma yang sesuai sangat penting dalam mendukung keputusan strategis di sektor pertanian. Penelitian ini dapat dikembangkan lebih lanjut dengan melibatkan fitur tambahan dan algoritma prediktif lain untuk meningkatkan akurasi model.

Kata kunci: produksi padi, klasifikasi, logistic regression, decision tree, machine learning

Abstract

This study compared the performance of two classification algorithms, namely Logistic Regression and Decision Tree, in predicting rice production levels across eight provinces in Sumatra. The dataset covered the period from 1993 to 2020 and included variables such as rainfall, temperature, humidity, and harvested area. All data underwent preprocessing and transformation stages to build a multi-class classification model: Low, Medium, and High. The evaluation results showed that both models achieved the same accuracy of 79.41%. However, the Decision Tree yielded a slightly higher weighted F1-score of 0.7815 compared to 0.7760 from Logistic Regression. This indicates that the Decision Tree was more effective in capturing complex and imbalanced data patterns. These findings highlight the importance of selecting appropriate algorithms to support strategic decision-making in the agricultural sector. Future studies may enhance this research by incorporating additional features and more advanced predictive algorithms to improve model performance.

Keywords: rice production, classification, logistic regression, decision tree, machine learning

1. Pendahuluan

Pertanian merupakan sektor strategis dalam mendukung ketahanan pangan nasional, terutama di negara agraris seperti Indonesia. Salah satu komoditas utama dalam sektor ini adalah padi, yang menjadi makanan pokok mayoritas penduduk. Oleh karena itu, kestabilan dan peningkatan produksi padi menjadi prioritas dalam perencanaan pembangunan nasional di bidang pertanian. Namun, realitas di lapangan menunjukkan bahwa produksi padi sering kali mengalami fluktuasi akibat berbagai faktor, baik dari segi iklim maupun non-iklim, seperti curah hujan, suhu, kelembapan udara, serta luas lahan panen [1], [2], [3]. Untuk memitigasi risiko fluktuasi tersebut, diperlukan pendekatan ilmiah yang mampu memprediksi produksi padi secara lebih akurat. *Forecasting* atau peramalan merupakan salah satu teknik penting dalam proses pengambilan keputusan berbasis data. Dalam beberapa tahun

terakhir, metode *machine learning* telah banyak digunakan untuk mendukung proses *forecasting* dalam berbagai bidang, termasuk sektor pertanian. Di antara berbagai metode yang tersedia, *Logistic Regression* dan *Decision Tree* merupakan dua algoritma yang relatif sederhana namun mampu memberikan hasil yang cukup baik dalam konteks klasifikasi dan prediksi [4], [5], [6].

Penelitian sebelumnya telah banyak menggunakan pendekatan statistik tradisional dalam menganalisis produksi pertanian, namun penggunaan metode *machine learning* berbasis data historis dengan mempertimbangkan variabel lingkungan masih tergolong terbatas, khususnya pada wilayah Sumatera. Padahal, wilayah ini memiliki karakteristik geografis dan iklim yang sangat bervariasi sehingga menimbulkan tantangan tersendiri dalam proses peramalan produksi pertanian [7], [8], [9]. Dalam penelitian ini, penulis mengusulkan pendekatan prediktif dengan membandingkan dua model pembelajaran mesin, yaitu *Logistic Regression* dan *Decision Tree*, dalam melakukan peramalan produksi padi berdasarkan data historis dari delapan provinsi di Sumatera selama periode 1993 hingga 2020. Tujuan utama dari penelitian ini adalah untuk mengevaluasi kinerja kedua model dalam memetakan pola hubungan antara faktor-faktor lingkungan terhadap hasil produksi padi serta memberikan gambaran model mana yang lebih efektif digunakan dalam konteks kebijakan pertanian [10], [11], [12].

Kontribusi utama dari penelitian ini terletak pada penerapan dua metode *machine learning* tersebut dalam konteks agrikultur di Indonesia, khususnya pada komoditas padi, serta penyediaan analisis perbandingan yang dapat digunakan oleh para pengambil kebijakan dan pelaku sektor pertanian sebagai dasar dalam pengambilan keputusan yang lebih akurat dan berbasis data [13], [14], [15].

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional untuk membandingkan performa algoritma *Logistic Regression* dan *Decision Tree* dalam memprediksi produksi padi di wilayah Sumatera. Proses penelitian dilakukan melalui lima tahapan utama: pengumpulan data, pra-proses, transformasi target variabel, pembangunan model, dan evaluasi performa. Data yang digunakan merupakan data sekunder produksi padi tahunan dari delapan provinsi di Sumatera selama tahun 1993–2020. Variabel yang dianalisis meliputi produksi padi, luas panen, curah hujan, kelembapan, dan suhu rata-rata. Data dibersihkan dari duplikasi dan nilai kosong, kemudian dinormalisasi agar sesuai dengan standar input model.

Karena *Logistic Regression* merupakan algoritma klasifikasi, maka variabel target dikonversi menjadi dua kelas: **Produksi Tinggi** dan **Produksi Rendah**, berdasarkan nilai median. Selanjutnya, model *Logistic Regression* dan *Decision Tree* dibangun menggunakan pustaka *scikit-learn* di Python. Data dibagi menjadi 80% data latih dan 20% data uji. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, serta validasi silang sebanyak 5 lipatan (*5-fold cross-validation*). Salah satu rumus evaluasi utama adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Persamaan 1 digunakan untuk menghitung akurasi, yaitu proporsi jumlah prediksi yang benar terhadap seluruh prediksi yang dilakukan oleh model. Dalam hal ini, prediksi yang benar mencakup dua komponen, yaitu *True Positive* (TP) dan *True Negative* (TN). TP adalah jumlah data yang diklasifikasikan dengan benar sebagai produksi tinggi, sedangkan TN adalah jumlah data yang diklasifikasikan dengan benar sebagai produksi rendah. Di sisi lain, *False Positive* (FP) menunjukkan jumlah data yang sebenarnya rendah namun salah diprediksi sebagai tinggi, dan *False Negative* (FN) menunjukkan jumlah data yang sebenarnya tinggi namun salah diprediksi sebagai rendah. Akurasi dihitung dengan menjumlahkan TP dan TN, kemudian dibagi dengan total jumlah prediksi, yaitu $TP + TN + FP + FN$. Nilai akurasi yang tinggi menunjukkan bahwa model mampu melakukan klasifikasi dengan baik secara keseluruhan. Namun demikian, apabila distribusi kelas tidak seimbang, akurasi perlu dilengkapi dengan metrik lain seperti precision, recall, dan F1-score agar evaluasi performa model menjadi lebih komprehensif.

3. Hasil dan Pembahasan

Penelitian ini menghasilkan dua model prediktif menggunakan algoritma *Logistic Regression* dan *Decision Tree Classifier*. Kedua model telah dievaluasi dengan menggunakan metrik akurasi dan F1-score berdasarkan hasil klasifikasi terhadap tiga kelas produksi padi: **Rendah**, **Sedang**, dan **Tinggi**.

3.1. Hasil Evaluasi Model

Tabel berikut menampilkan hasil perbandingan kinerja antara *Logistic Regression* dan *Decision Tree Classifier*:

Tabel 1 Evaluasi Model

Model	Akurasi	F1-score (Weighted Avg)
Logistic Regression	0.7941	0.7760
Decision Tree Classifier	0.7941	0.7815

Berdasarkan Tabel 1, kedua model memiliki tingkat akurasi yang sama, yaitu sebesar **79,41%**. Namun, jika dilihat dari F1-score rata-rata tertimbang, *Decision Tree Classifier* memiliki nilai **0.7815**, sedikit lebih tinggi dibandingkan *Logistic Regression* yang memiliki nilai **0.7760**. Hal ini menunjukkan bahwa meskipun akurasinya sama, *Decision Tree* memberikan keseimbangan yang lebih baik antara precision dan recall pada semua kelas.

3.2. Analisis Kinerja Berdasarkan Kelas Produksi

Secara lebih rinci, model *Logistic Regression* memiliki performa yang sangat baik dalam mengklasifikasikan kelas **Rendah** dan **Sedang**, namun kurang akurat pada kelas **Tinggi**. Hal ini terlihat dari nilai recall kelas Tinggi yang hanya sebesar **0.41**, meskipun precision-nya sempurna (**1.00**), yang berarti model cenderung sangat hati-hati dalam memprediksi kelas tersebut, tetapi tidak mampu menangkap banyak kasus sebenarnya. Sebaliknya, model *Decision Tree Classifier* menunjukkan peningkatan pada kelas **Tinggi**, dengan recall meningkat menjadi **0.47** dan precision sebesar **0.89**. Ini menunjukkan bahwa *Decision Tree* lebih mampu mengenali pola pada kelas produksi yang lebih tinggi, meskipun masih belum optimal.

3.2.1. Implikasi dan Pembahasan

Meskipun perbedaannya tipis, hasil ini mengindikasikan bahwa *Decision Tree* lebih adaptif dalam menangani variasi antar kelas, terutama ketika data tidak seimbang. Dalam konteks pertanian, kemampuan untuk mendeteksi secara tepat kelas produksi **Tinggi** sangat penting, karena berkaitan langsung dengan pengambilan keputusan distribusi dan logistik pangan. Maka, meskipun akurasinya setara, *Decision Tree* lebih unggul dari segi ketepatan prediksi secara menyeluruh (*overall class-wise balance*), yang tercermin dalam F1-score tertimbang.

4. Kesimpulan

Penelitian ini bertujuan untuk membandingkan performa algoritma *Logistic Regression* dan *Decision Tree Classifier* dalam memprediksi produksi padi berdasarkan data iklim dan pertanian di wilayah Sumatera. Berdasarkan hasil evaluasi model, diketahui bahwa kedua algoritma memiliki akurasi yang sama sebesar 79,41%. Namun, dari sisi F1-score tertimbang, *Decision Tree* menunjukkan kinerja yang sedikit lebih baik dibandingkan *Logistic Regression*, terutama dalam mengenali kelas produksi tinggi. Hal ini menunjukkan bahwa struktur pohon keputusan mampu menangkap pola yang lebih kompleks antar variabel, sehingga lebih sesuai digunakan dalam kasus klasifikasi produksi pertanian yang bersifat multi-kelas dan tidak seimbang. Dengan demikian, *Decision Tree Classifier* direkomendasikan sebagai model yang lebih optimal untuk kasus prediksi produksi padi berbasis data historis. Ke depannya, penelitian ini dapat dikembangkan dengan mempertimbangkan algoritma lain seperti *Random Forest* atau *Gradient Boosting*, serta menambahkan variabel eksternal seperti kondisi ekonomi atau jenis varietas padi untuk meningkatkan akurasi dan cakupan prediksi. Selain itu, implementasi sistem prediksi berbasis web atau dashboard interaktif dapat menjadi langkah lanjutan yang mendukung pengambilan keputusan dalam sektor pertanian secara real-time.

Daftar Pustaka

- [1] P. C. Saibabu, H. Sai, S. Yadav, and C. R. Srinivasan, "Synthesis of model predictive controller for

- an identified model of MIMO process,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 2, pp. 941–949, 2019, doi: 10.11591/ijeecs.
- [2] M. A. Basir, M. S. Hussin, and Y. Yusof, “Integrated bio-search approaches with multi-objective algorithms for optimization and classification problem,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 5, pp. 2421–2431, 2020, doi: 10.12928/TELKOMNIKA.V18I5.15141.
- [3] I. H. Rahmana, A. R. Febriyani, I. Ranggadara, Suhendra, and I. S. Karima, “Comparative study of extraction features and regression algorithms for predicting drought rates,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 20, no. 3, pp. 638–646, 2022, doi: 10.12928/TELKOMNIKA.v20i3.23156.
- [4] Y. A. Mohammed and E. G. Saleh, “Comparative study of logistic regression and artificial neural networks on predicting breast cancer cytology,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 2, pp. 1113–1120, 2020, doi: 10.11591/ijeecs.v21.i2.pp1113-1120.
- [5] C. I. Imrane, N. Said, B. El Mehdi, E. K. A. Seddiq, and M. Abdelaziz, “Machine learning for decoding linear block codes: Case of multi-class logistic regression model,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, no. 1, pp. 538–547, 2021, doi: 10.11591/ijeecs.v24.i1.pp538-547.
- [6] H. Agusta, E. Santosa, Dulbari, D. Guntoro, and S. Zaman, “Continuous Heavy Rainfall and Wind Velocity During Flowering Affect Rice Production,” *Agrivita*, vol. 44, no. 2, pp. 290–302, 2022, doi: 10.17503/agrivita.v44i2.2539.
- [7] Dulbari *et al.*, “Local adaptation to extreme weather and it’s implication on sustainable rice production in Lampung, Indonesia,” *Agrivita*, vol. 43, no. 1, pp. 125–136, 2021, doi: 10.17503/agrivita.v43i1.2338.
- [8] J. R. Asor, J. L. Lerios, S. B. Sapin, J. O. Padallan, and C. A. C. Buama, “Fire incidents visualization and pattern recognition using machine learning algorithms,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 3, pp. 1427–1435, 2021, doi: 10.11591/ijeecs.v22.i3.pp1427-1435.
- [9] T. Amelia and R. Mohamed, “A decision tree approach based on BOCR for minimizing criteria in requirements prioritization,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 2, pp. 1094–1104, 2023, doi: 10.11591/ijeecs.v32.i2.pp1094-1104.
- [10] H. Elmannai and A. D. Al-Garni, “Classification using semantic feature and machine learning: Land-use case application,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 4, pp. 1242–1250, 2021, doi: 10.12928/TELKOMNIKA.v19i4.18359.
- [11] M. G. Hussain, B. Sultana, M. Rahman, and M. R. Hasan, “Comparison analysis of Bangla news articles classification using support vector machine and logistic regression,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 21, no. 3, pp. 584–591, 2023, doi: 10.12928/TELKOMNIKA.v21i3.23416.
- [12] S. B. Abhi *et al.*, “An intelligent wind turbine with yaw mechanism using machine learning to reduce high-cost sensors quantity,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 31, no. 1, pp. 10–21, 2023, doi: 10.11591/ijeecs.v31.i1.pp10-21.
- [13] H. A. J. H. Almuhan and H. H. Abbas, “Classification of specialities in textual medical reports based on natural language processing and feature selection,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 1, pp. 163–170, 2022, doi: 10.11591/ijeecs.v27.i1.pp163-170.
- [14] N. Agnihotri and S. K. Prasad, “Hybrid logistic regression support vector model to enhance prediction of bipolar disorder,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1294–1300, 2024, doi: 10.11591/ijeecs.v36.i2.pp1294-1300.
- [15] N. Jahan and R. Shahariar, “Predicting fertilizer treatment of maize using decision tree algorithm,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 20, no. 3, pp. 1427–1434, 2020, doi: 10.11591/ijeecs.v20.i3.pp1427-1434.