

Perbandingan Algoritma Catboost dan Extra Trees Classifier Untuk Prediksi Tingkat Keberhasilan Panen Padi Berdasarkan Faktor Cuaca dan Kondisi Tanah

Rio Fadly

Universitas Muhammadiyah Sumatera Utara, Indonesia

Email: riofadly534@gmail.com

ABSTRAK

Produksi padi merupakan salah satu faktor penting dalam menjaga ketahanan pangan, khususnya di Indonesia yang menjadikan beras sebagai bahan pangan utama. Namun, hasil panen padi sering dipengaruhi oleh berbagai faktor lingkungan seperti curah hujan, suhu, kelembapan, pH tanah, serta kandungan unsur hara tanah seperti nitrogen, fosfor, dan kalium. Ketidakpastian kondisi lingkungan tersebut menyebabkan petani kesulitan dalam memprediksi potensi hasil panen secara akurat. Oleh karena itu, diperlukan suatu sistem yang mampu membantu memprediksi hasil panen padi berdasarkan kondisi lingkungan dan tanah secara cepat dan akurat.

Penelitian ini bertujuan untuk membangun sistem prediksi hasil panen padi berbasis machine learning dengan menggunakan algoritma CatBoost dan ExtraTrees, serta mengimplementasikan Explainable Artificial Intelligence (XAI) menggunakan metode SHAP (SHapley Additive Explanations) untuk memberikan interpretasi terhadap hasil prediksi model. Dataset yang digunakan terdiri dari 1500 data yang memuat variabel curah hujan, suhu, kelembapan, pH tanah, nitrogen, fosfor, dan kalium dengan tiga kelas hasil panen yaitu rendah, sedang, dan tinggi. Data kemudian diproses menggunakan metode train-test split dengan proporsi 80% data pelatihan dan 20% data pengujian.

Hasil penelitian menunjukkan bahwa algoritma CatBoost memiliki performa yang lebih baik dibandingkan ExtraTrees dengan nilai akurasi yang lebih tinggi. Berdasarkan evaluasi menggunakan ROC Curve, model CatBoost memperoleh nilai AUC sebesar 0,99 pada kelas rendah, 0,98 pada kelas sedang, dan 0,99 pada kelas tinggi, yang termasuk dalam kategori excellent classification. Sementara itu, model ExtraTrees memperoleh nilai AUC sebesar 0,96 pada kelas rendah, 0,81 pada kelas sedang, dan 0,93 pada kelas tinggi. Analisis menggunakan metode SHAP menunjukkan bahwa faktor yang paling berpengaruh terhadap hasil prediksi panen adalah nitrogen, pH tanah, dan kelembapan. Sistem yang dibangun juga diimplementasikan dalam bentuk aplikasi web berbasis Laravel yang dapat digunakan untuk melakukan prediksi dan memberikan analisis kondisi lahan secara otomatis.

Dengan adanya sistem ini diharapkan dapat membantu petani atau pengguna dalam mengetahui potensi hasil panen padi serta memahami faktor-faktor lingkungan yang mempengaruhi hasil panen sehingga dapat menjadi alat bantu pengambilan keputusan (decision support system) dalam pengelolaan lahan pertanian.

Kata Kunci: *Machine Learning*, *Prediksi Panen Padi*, *CatBoost*, *ExtraTrees*, *Explainable AI*, *SHAP*.

ABSTRACT

Rice production is one of the important factors in maintaining food security, especially in Indonesia where rice is the main staple food. However, rice yield is often influenced by various environmental factors such as rainfall, temperature, humidity, soil pH, and soil nutrients including nitrogen, phosphorus, and potassium. The uncertainty of these environmental conditions makes it difficult for farmers to accurately predict potential crop yields. Therefore, a system is needed to help predict rice yield based on environmental and soil conditions quickly and accurately.

This study aims to develop a rice yield prediction system based on machine learning using the CatBoost and ExtraTrees algorithms, and to implement Explainable Artificial Intelligence (XAI) using the SHAP (SHapley Additive Explanations) method to provide interpretation of the model predictions. The dataset used in this study consists of 1500 records containing variables of rainfall, temperature, humidity, soil pH, nitrogen, phosphorus, and potassium with three yield classes, namely low, medium, and high. The dataset was processed using the train-test split method with a proportion of 80% training data and 20% testing data.

The results show that the CatBoost algorithm performs better than ExtraTrees with higher accuracy performance. Based on evaluation using the ROC Curve, the CatBoost model achieved AUC values of 0.99 for the low class, 0.98 for the medium class, and 0.99 for the high class, which are categorized as excellent classification performance. Meanwhile, the ExtraTrees model achieved AUC values of 0.96 for the low class, 0.81 for the medium class, and 0.93 for the high class. The SHAP analysis indicates that the most influential factors affecting the prediction results are nitrogen, soil pH, and humidity. The developed system is also implemented in a web-based application using the Laravel framework, which allows users to perform predictions and obtain automatic analysis of land conditions.

This system is expected to assist farmers or users in identifying potential rice yields and understanding environmental factors that influence crop production. Therefore, it can serve as a decision support system in agricultural land management.

Keywords: Machine Learning, Rice Yield Prediction, CatBoost, ExtraTrees, Explainable AI, SHAP.

PENDAHULUAN

Sektor pertanian merupakan salah satu pilar utama dalam menggerakkan perekonomian Indonesia, sejalan dengan karakteristik Indonesia sebagai negara agraris. Di antara berbagai komoditas pertanian, padi memiliki peranan yang sangat strategis karena menjadi sumber pangan pokok bagi lebih dari 280 juta penduduk Indonesia (Nizami et al., 2025). Namun, dalam beberapa dekade terakhir, sektor pertanian menghadapi tantangan yang semakin kompleks, terutama berkaitan dengan perubahan iklim, degradasi kualitas tanah, serta keterbatasan dalam memprediksi hasil panen secara akurat. Curah hujan yang tidak menentu, peningkatan suhu udara, perubahan pola musim, serta variasi kondisi tanah sering kali menyebabkan produktivitas padi mengalami fluktuasi di berbagai wilayah. Kondisi ini menuntut adanya pendekatan baru yang mampu membantu petani maupun pemerintah dalam memprediksi keberhasilan panen secara lebih tepat dan berbasis data.

Selama ini, prediksi keberhasilan panen padi umumnya masih mengandalkan metode manual maupun pendekatan statistik sederhana yang belum mampu menangkap hubungan kompleks antarvariabel lingkungan. Padahal, keberhasilan panen dipengaruhi oleh banyak faktor, seperti curah hujan, suhu udara, kelembapan,

intensitas penyinaran, pH tanah, kandungan unsur hara (N, P, K), tekstur tanah, jenis varietas, hingga pola pemupukan. Interaksi antarvariabel tersebut membutuhkan metode pengolahan data yang lebih canggih agar hasil prediksi dapat dilakukan secara lebih akurat dan konsisten.

Machine learning merupakan proses memperoleh pengetahuan dari data yang menggabungkan konsep statistik, kecerdasan buatan, dan ilmu komputer. Bidang ini juga dikenal sebagai predictive analytics atau statistical learning (Shawon et al., 2025). Dengan memanfaatkan data sekunder, machine learning mampu menghasilkan model prediksi panen yang lebih akurat. Akan tetapi, sebagian besar penelitian sebelumnya di Indonesia masih menggunakan algoritma umum seperti Linear Regression, Decision Tree, Random Forest, maupun Support Vector Machine (SVM). Sementara itu, penerapan algoritma modern seperti CatBoost dan Extra Trees Classifier masih relatif terbatas, khususnya pada penelitian pertanian yang memanfaatkan faktor cuaca dan kondisi tanah.

CatBoost merupakan algoritma boosting modern yang dikembangkan oleh Yandex dengan kemampuan unggul dalam menangani fitur kategorikal dan data nonlinier tanpa memerlukan proses encoding yang rumit. Algoritma ini mampu memberikan performa tinggi pada dataset yang terdiri atas kombinasi variabel numerik dan kategorikal, sehingga sangat sesuai diterapkan pada data pertanian yang bersifat heterogen. Di sisi lain, Extra Trees Classifier merupakan salah satu metode ensemble berbasis pohon keputusan yang memiliki tingkat randomisasi lebih tinggi dibandingkan Random Forest. Melalui pengacakan dalam pemilihan fitur dan ambang batas, algoritma ini mampu menghasilkan generalisasi model yang lebih baik pada dataset berukuran besar.

Meskipun kedua algoritma tersebut menawarkan berbagai keunggulan, penelitian yang secara langsung membandingkan performanya dalam konteks prediksi keberhasilan panen padi masih sangat terbatas. Oleh karena itu, penelitian ini penting dilakukan untuk mengetahui sejauh mana kemampuan CatBoost dan Extra Trees Classifier dalam memprediksi produktivitas padi berdasarkan faktor cuaca dan kondisi tanah. Perbandingan kinerja kedua model diharapkan dapat memberikan wawasan baru terkait algoritma yang paling efektif untuk diterapkan pada sistem prediksi pertanian modern.

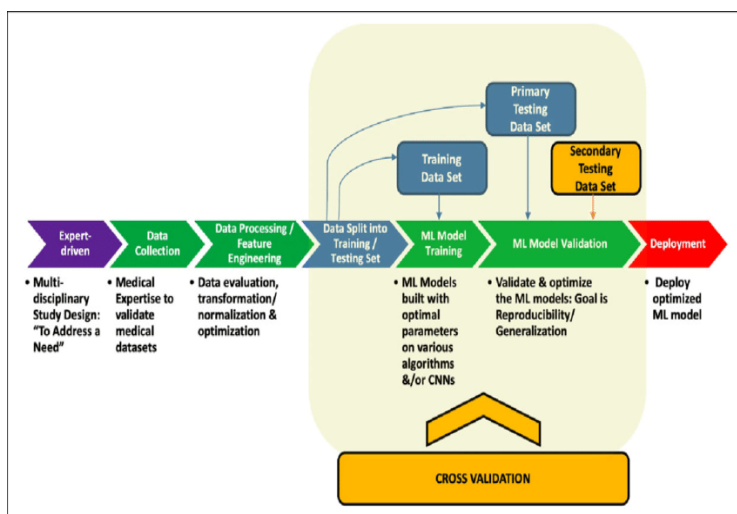
Selain memberikan kontribusi akademik, hasil penelitian ini juga berpotensi memberikan manfaat praktis bagi sektor pertanian Indonesia. Prediksi tingkat keberhasilan panen yang lebih akurat dapat membantu petani, penyuluh pertanian, maupun dinas terkait dalam merencanakan jadwal tanam, strategi pemupukan, pengelolaan irigasi, serta estimasi ketersediaan pangan daerah. Dengan demikian, penelitian ini turut mendukung pengembangan pertanian presisi sebagai salah satu solusi dalam menghadapi tantangan perubahan iklim.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada penerapan dan perbandingan dua algoritma machine learning yang masih jarang digunakan, yaitu CatBoost dan Extra Trees Classifier, untuk memprediksi tingkat keberhasilan

panen padi berdasarkan faktor cuaca dan kondisi tanah menggunakan dataset berukuran besar. Tujuan utama penelitian ini adalah menghasilkan model prediksi yang akurat, stabil, dan mampu mendukung pengambilan keputusan yang efektif dalam sektor pertanian berbasis teknologi informasi.

METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data-driven yang memanfaatkan supervised machine learning untuk memprediksi tingkat keberhasilan panen padi berdasarkan faktor cuaca dan kondisi tanah. Desain penelitian difokuskan pada pembangunan serta pengujian model klasifikasi, kemudian dilanjutkan dengan perbandingan performa dua algoritma ensemble yang memiliki pendekatan berbeda, yaitu CatBoost dan Extra Trees Classifier. Pendekatan tersebut dipilih karena prediksi hasil panen dipengaruhi oleh hubungan nonlinier antar faktor lingkungan, sehingga diperlukan model yang mampu mempelajari pola-pola kompleks dari data historis. Selain itu, proses validasi dan evaluasi model dilakukan secara sistematis agar hasil penelitian dapat dipertanggungjawabkan secara ilmiah serta relevan untuk diterapkan dalam konteks pertanian berbasis data (Filippi et al., 2025).



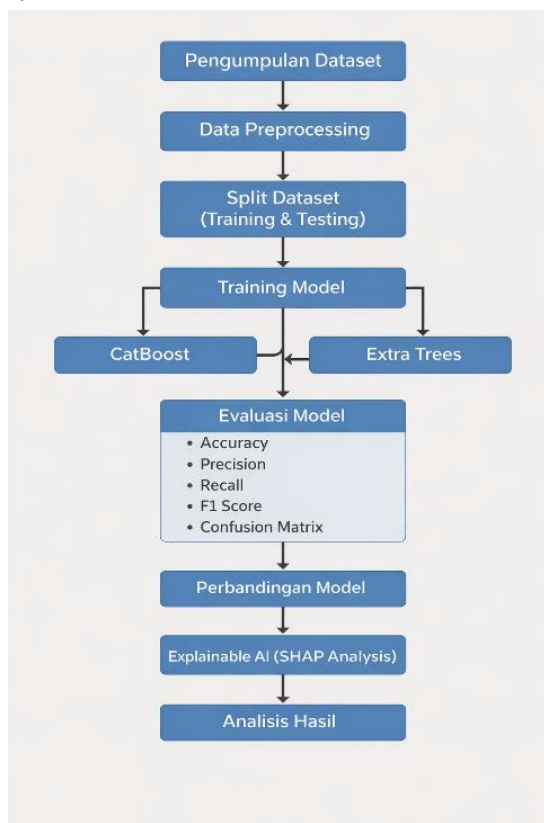
Gambar 1 *Supervised Machine Learning*

Gambar tersebut menunjukkan desain penelitian kuantitatif dengan pendekatan supervised machine learning yang digunakan dalam penelitian ini. Penelitian dirancang untuk membangun serta membandingkan dua algoritma ensemble, yaitu CatBoost dan Extra Trees Classifier, dalam memprediksi tingkat keberhasilan panen padi. Desain penelitian ini menitikberatkan pada penggunaan data historis sebagai dasar pembelajaran model, serta proses evaluasi menggunakan berbagai metrik klasifikasi. Pendekatan komparatif dipilih untuk mengidentifikasi

algoritma yang memiliki performa terbaik dalam menangani data cuaca dan kondisi tanah yang kompleks serta bersifat nonlinier.

Objek penelitian ini adalah prediksi tingkat keberhasilan panen padi sebagai variabel target dengan memanfaatkan variabel lingkungan sebagai fitur, yang meliputi faktor cuaca dan kondisi tanah. Penelitian berfokus pada data historis yang merepresentasikan kondisi lapangan pada suatu wilayah atau area budidaya padi yang tercakup dalam dataset. Adapun pelaksanaan penelitian mencakup beberapa tahapan, mulai dari pengumpulan data, prapemrosesan data, pelatihan model, pengujian, evaluasi, hingga analisis hasil dan penyusunan laporan penelitian.

Tahapan penelitian meliputi: (1) pengumpulan dataset; (2) pemahaman data (data understanding) serta eksplorasi awal; (3) pembagian data menjadi data latih dan data uji; (4) pelatihan model menggunakan algoritma CatBoost dan Extra Trees Classifier; (5) evaluasi model; (6) perbandingan performa kedua model; (7) analisis SHAP (SHapley Additive Explanations); serta (8) analisis perbandingan dan interpretasi hasil penelitian.



Gambar 2. Tahapan Penelitian

Gambar tersebut menunjukkan tahapan penelitian yang dilakukan secara sistematis, dimulai dari proses pengumpulan data, eksplorasi dan prapemrosesan data, pembagian data latih dan data uji, pelatihan model menggunakan algoritma CatBoost dan Extra Trees Classifier, hingga tahap evaluasi serta analisis hasil. Setiap tahapan saling berkaitan dan dirancang untuk memastikan bahwa proses penelitian dapat direplikasi serta menghasilkan temuan yang objektif dan dapat dipertanggungjawabkan.

HASIL DAN PEMBAHASAN

Hasil Preprocessing Data

Tahap preprocessing data merupakan salah satu tahapan penting dalam proses pengembangan model machine learning karena kualitas data yang digunakan akan sangat memengaruhi performa model yang dihasilkan. Pada tahap ini dilakukan beberapa proses pengolahan data untuk memastikan bahwa dataset yang digunakan berada dalam kondisi optimal sebelum digunakan dalam proses pelatihan model.

Namun, pada penelitian ini dataset yang digunakan merupakan dataset rule-based yang telah disiapkan sebelumnya sehingga tidak ditemukan data duplikat, nilai hilang (missing value), maupun outlier yang signifikan. Oleh karena itu, tahapan preprocessing dilakukan secara minimal sebelum proses pelatihan model machine learning.

Tahap preprocessing juga bertujuan untuk memastikan bahwa setiap variabel dalam dataset memiliki format yang konsisten sehingga dapat diproses dengan baik oleh algoritma machine learning. Menurut Dinh et al. (2025), preprocessing data merupakan tahap yang sangat penting dalam pipeline machine learning karena data yang tidak bersih dapat menyebabkan model menghasilkan prediksi yang tidak akurat.

Pembagian Dataset

Pembagian dataset ini bertujuan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data baru yang belum pernah digunakan dalam proses pelatihan model. Dalam penelitian ini, pembagian dataset dilakukan menggunakan metode *train-test split* dengan rasio 80:20, yaitu 1.200 data training dan 300 data testing. Sebanyak 80% dari dataset digunakan sebagai data latih, sedangkan 20% sisanya digunakan sebagai data uji.

Data latih digunakan untuk melatih model *machine learning* agar dapat mengenali pola hubungan antara variabel input dan variabel target. Sementara itu, data uji digunakan untuk mengukur performa model setelah proses pelatihan selesai dilakukan. Dengan menggunakan metode pembagian dataset ini, performa model dapat diuji secara objektif terhadap data yang tidak digunakan selama proses pelatihan.

Hasil Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa algoritma machine learning dalam melakukan klasifikasi tingkat keberhasilan panen padi. Dalam penelitian ini, evaluasi model dilakukan menggunakan beberapa metrik evaluasi yang umum digunakan dalam tugas klasifikasi, yaitu accuracy, precision, recall, dan F1-score. Metrik-metrik tersebut digunakan untuk

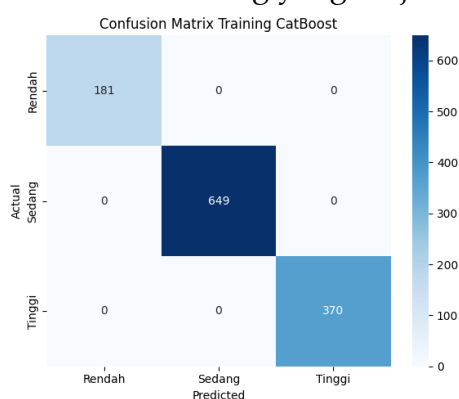
mengukur sejauh mana model mampu melakukan prediksi secara benar terhadap data uji.

Accuracy digunakan untuk mengukur persentase prediksi yang benar terhadap seluruh data uji. Precision digunakan untuk mengukur tingkat ketepatan model dalam memprediksi kelas tertentu. Recall digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data yang termasuk dalam suatu kelas. Sementara itu, F1-score merupakan kombinasi antara precision dan recall yang memberikan gambaran keseimbangan antara kedua metrik tersebut.

Hasil evaluasi model kemudian dibandingkan untuk mengetahui algoritma mana yang memiliki performa terbaik dalam memprediksi tingkat keberhasilan panen padi. Menurut Foody (2023), penggunaan beberapa metrik evaluasi dalam model klasifikasi sangat penting karena setiap metrik memberikan perspektif yang berbeda dalam menilai performa model.

Confusion Matrix dan Kurva ROC CatBoost

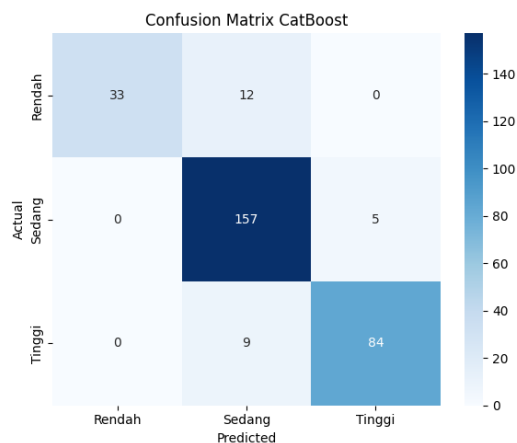
Pengujian dilakukan menggunakan confusion matrix yang terdiri atas accuracy, precision, recall, dan F1-score dengan dataset sebanyak 1.500 data yang diolah menggunakan algoritma CatBoost. Pengujian confusion matrix untuk dataset yang diolah menggunakan CatBoost dapat dilihat pada gambar di bawah ini. Pengujian dilakukan menggunakan data testing yang berjumlah 1.200 data.



Gambar 2 Confusion Matrix Training CatBoost

Setelah melakukan pengujian menggunakan data training yang berjumlah 1.200 data, model menunjukkan performa yang sangat tinggi dengan nilai akurasi mencapai 100%. Hal ini ditunjukkan oleh confusion matrix, di mana seluruh data pada masing-masing kelas berhasil diprediksi dengan benar. Nilai precision, recall, dan F1-score pada ketiga kelas, yaitu rendah, sedang, dan tinggi, juga mencapai nilai 1,00. Hal ini menunjukkan bahwa model mampu mempelajari pola data training secara sempurna.

Selanjutnya, dilakukan pengujian menggunakan data testing yang berjumlah 300 data. Pada tahap pengujian ini, nilai akurasi mengalami penurunan. Untuk penjelasan lebih lanjut, gambar dan rumus perhitungannya dapat dilihat pada bagian di bawah ini.



Gambar 4 Confusion Matrix Testing CatBoost

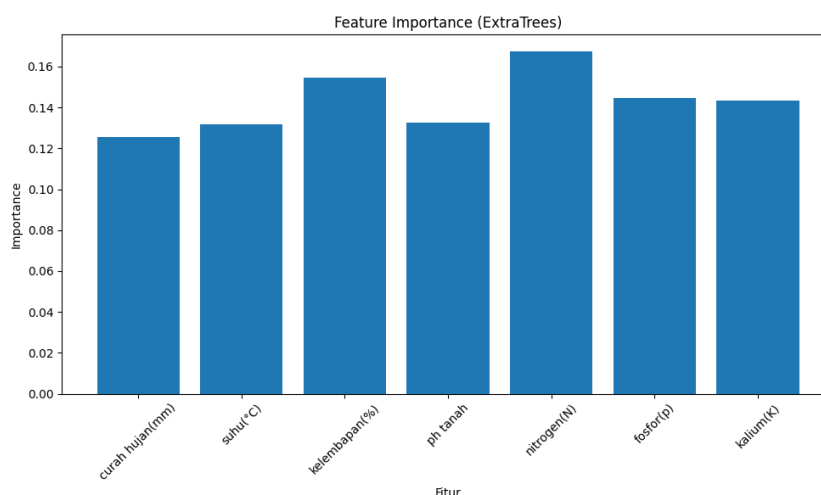
Feature Importance

Feature importance merupakan metode yang digunakan untuk mengukur seberapa besar kontribusi setiap fitur terhadap hasil prediksi model machine learning. Metode ini banyak digunakan dalam Explainable Artificial Intelligence (XAI) untuk memahami bagaimana model mengambil keputusan. Secara umum, feature importance mengukur pengaruh suatu variabel terhadap perubahan nilai prediksi model. Semakin besar nilai importance suatu fitur, maka semakin besar pula kontribusi fitur tersebut dalam menentukan hasil prediksi model.

Menurut Fang et al. (2020), pada model berbasis tree ensemble seperti ExtraTrees dan CatBoost, nilai importance dihitung berdasarkan seberapa besar fitur tersebut digunakan dalam proses pemisahan node pada pohon keputusan serta seberapa besar pengurangan impurity yang dihasilkan.

Analisis Feature Importance Pada Model Extra Trees Classifier

Pada model ExtraTrees, fitur yang paling berpengaruh adalah kandungan nitrogen dalam tanah. Nitrogen merupakan unsur hara utama yang berperan dalam pertumbuhan vegetatif tanaman padi, terutama dalam pembentukan daun dan batang tanaman. Oleh karena itu, tingginya nilai importance pada nitrogen menunjukkan bahwa ketersediaan unsur hara dalam tanah sangat berpengaruh terhadap hasil panen.



Gambar 5 *Feature Importance Extra Trees Classifier*

Algoritma ExtraTrees bekerja dengan membangun banyak pohon keputusan secara acak (randomized trees) kemudian menggabungkan hasil prediksi dari seluruh pohon tersebut. Metode ini menghitung feature importance berdasarkan penurunan impurity yang dihasilkan oleh setiap fitur saat digunakan dalam proses pemisahan data pada node pohon keputusan.

Secara umum dapat disimpulkan bahwa faktor yang paling berpengaruh terhadap hasil panen dalam penelitian ini adalah curah hujan, pH tanah, nitrogen, suhu, dan kelembapan. Hal ini menunjukkan bahwa faktor iklim dan kondisi tanah merupakan variabel utama yang menentukan keberhasilan produksi padi.

Berdasarkan hasil analisis *SHAP Beeswarm Plot*, dapat disimpulkan bahwa kedua model berhasil mengidentifikasi faktor-faktor penting yang mempengaruhi tingkat hasil panen padi, terutama unsur hara tanah seperti nitrogen, fosfor, dan kalium, serta kondisi lingkungan seperti kelembapan dan suhu.

Namun demikian, model *CatBoost* menunjukkan interpretasi yang lebih stabil dan lebih jelas dalam menggambarkan kontribusi fitur terhadap prediksi model, sedangkan model ExtraTrees memberikan distribusi kontribusi fitur yang lebih bervariasi karena sifat randomisasi pada algoritma ensemble-nya.

Tabel 1 Karakteristik Interpretasi

Aspek	<i>CatBoost</i>	<i>ExtraTrees</i>
Konsistensi interpretasi	Tinggi	Sedang
Stabilitas <i>SHAP value</i>	Lebih Stabil	Lebih Menyebar
Kemampuan menangkap interaksi fitur	Sangat baik	baik
Kompleksitas model	Tinggi	Sedang

Dari perbandingan tersebut dapat disimpulkan bahwa *CatBoost* memiliki kemampuan interpretasi yang lebih stabil, sedangkan *ExtraTrees* memberikan

interpretasi yang lebih eksploratif karena variasi pohon keputusan yang lebih tinggi. Dengan demikian, dalam konteks interpretasi model menggunakan Explainable AI (XAI), CatBoost cenderung lebih mudah diinterpretasikan dibandingkan ExtraTrees, meskipun kedua model tetap memberikan informasi yang konsisten mengenai faktor-faktor yang memengaruhi produktivitas panen padi.

Lingkungan Pengujian

Lingkungan pengujian memberikan gambaran mengenai spesifikasi hardware dan software yang digunakan dalam proses pengujian sistem, baik untuk pengujian validasi maupun kualitas. Spesifikasi tersebut diperoleh melalui proses observasi berdasarkan aspek sistem. Berikut merupakan ringkasan spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam proses pengujian.

1. Hardware

- 12th Gen Intel Core i5-1235U (1.30 GHz)
- RAM 8.00 GB (7.68 GB usable)
- Sistem operasi 64-bit, prosesor berbasis x64

2. Software

- *Laragon*
- *PHP*
- *Mysql*
- *Web Browser(google Chrome)*

Perangkat lunak yang digunakan untuk membangun prototipe aplikasi prediksi tingkat keberhasilan panen menggunakan perangkat lunak open source, yaitu PHP, Laragon, dan MySQL sebagai basis data (database). Ditinjau dari segi keandalan, perangkat lunak tersebut telah teruji dan banyak digunakan dalam pengembangan sistem tanpa terkendala masalah lisensi karena bersifat open source.

KESIMPULAN

Berdasarkan pengukuran kinerja melalui komparasi dua algoritma yang telah dilakukan terhadap jumlah data, dapat disimpulkan bahwa algoritma CatBoost memiliki kemampuan yang lebih baik dalam menentukan tingkat keberhasilan panen padi dengan indikator rendah, sedang, dan tinggi. Kedua algoritma sebenarnya dapat digunakan untuk memprediksi tingkat keberhasilan panen padi, namun setelah dilakukan pengujian akurasi, algoritma dengan tingkat akurasi tertinggi dipilih sebagai model terbaik. Adapun hasil penelitian dari berbagai percobaan yang telah dilakukan dapat disimpulkan sebagai berikut:

1. Algoritma CatBoost memperoleh tingkat akurasi tertinggi, yaitu sebesar 91,33%, sedangkan Extra Trees Classifier memperoleh akurasi sebesar 77,33%, sehingga terdapat selisih sebesar 14,00%. Hasil evaluasi model menunjukkan bahwa CatBoost memiliki performa yang lebih unggul dibandingkan Extra Trees dalam melakukan klasifikasi hasil panen padi. Berdasarkan analisis ROC Curve, model CatBoost memperoleh nilai AUC sebesar 0,99 pada kelas rendah,

0,98 pada kelas sedang, dan 0,99 pada kelas tinggi, yang termasuk dalam kategori excellent classification performance. Sementara itu, model Extra Trees memperoleh nilai AUC sebesar 0,96 pada kelas rendah, 0,81 pada kelas sedang, dan 0,93 pada kelas tinggi.

2. Aturan (rule) yang dihasilkan oleh algoritma CatBoost diterapkan pada prototype sistem prediksi keberhasilan panen padi dan menunjukkan hasil akurasi pengujian yang stabil. Berdasarkan tingkat akurasi yang diperoleh, prototype tersebut dinilai telah mampu menerapkan metode prediksi dengan baik dalam menentukan tingkat keberhasilan panen padi.
3. Implementasi metode Explainable AI menggunakan SHAP (SHapley Additive Explanations) mampu memberikan interpretasi terhadap model machine learning dengan menunjukkan kontribusi masing-masing variabel terhadap hasil prediksi. Hasil analisis SHAP menunjukkan bahwa variabel curah hujan, suhu, kelembapan, dan nitrogen merupakan faktor yang paling berpengaruh dalam menentukan hasil prediksi panen padi.

Berdasarkan hasil penelitian ini, algoritma CatBoost diharapkan dapat digunakan sebagai metode yang efektif dalam memprediksi tingkat keberhasilan panen padi secara cepat dan akurat, sehingga mampu mendukung peningkatan kinerja dan pengambilan keputusan bagi para petani.

DAFTAR PUSTAKA

- Azhari, M., & Nasution, P. (2025). Application of data mining to determine the performance of family planning field officers (PLKB) using the C4.5 algorithm.
- Azhari, M., & Rahman, M. (2022). Analisis perbandingan algoritma WP dan TOPSIS dalam menentukan kandidat peserta lomba kompetensi siswa. *IT (Informatic Technique) Journal*, 10(1), 42–55. <https://doi.org/10.22303/it.10.1.2022.42-55>
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan akurasi, recall, dan presisi klasifikasi pada algoritma C4.5, Random Forest, SVM, dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640–650. <https://doi.org/10.30865/mib.v5i2.2937>
- Badshah, A., Alkazemi, B. Y., Din, F., Zamli, K. Z., & Haris, M. (2024). Crop classification and yield prediction using robust machine learning models for agricultural sustainability. *IEEE Access*, 12, 162799–162813. <https://doi.org/10.1109/ACCESS.2024.3486653>
- Das, A., & Rad, P. (2020). *Opportunities and challenges in explainable artificial intelligence (XAI): A survey*. arXiv. <http://arxiv.org/abs/2006.11371>

- Dinh, T., Wong, H., Lisik, D., Koren, M., Tran, D., Yu, P. S., & Torres-Sospedra, J. (2025). Data clustering: A fundamental method in data science and management. *Data Science and Management*. https://doi.org/10.1016/j.dsm.2025.08.001
- Diyanti, Martanto, & Bahtiar, A. (2023). Prediksi hasil panen padi tahun 2023 menggunakan metode regresi linear di Kabupaten Indramayu. *Jurnal Informatika Terpadu*, 9(1), 18–23.
- Elbeltagi, A., Srivastava, A., Cao, X., Bilali, A. E., Raza, A., Khadke, L., & Salem, A. (2025). An interpretable machine learning approach based on SHAP, Sobol, and LIME values for precise estimation of daily soybean crop coefficients. *Scientific Reports*, 15(1), 1–20. https://doi.org/10.1038/s41598-025-20386-y
- Fang, F., Ventre, C., Li, L., Kanthan, L., Wu, F., & Basios, M. (2020). *Better model selection with a new definition of feature importance*. arXiv. http://arxiv.org/abs/2009.07708
- Filippi, P., Han, S. Y., & Bishop, T. F. A. (2025). *Addressing the common issues in published studies*.
- Foody, G. M. (2023). Challenges in the real-world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLoS ONE*, 18(10), 1–27. https://doi.org/10.1371/journal.pone.0291908
- Gopi, S. R., & Karthikeyan, M. (2023). Effectiveness of crop recommendation and yield prediction using hybrid moth flame optimization with machine learning. *Engineering, Technology and Applied Science Research*, 13(4), 11360–11365. https://doi.org/10.48084/etasr.6092
- Himayanta, K. L., & Wardhani, D. F. (2025). Prediksi hasil panen padi dengan machine learning. *Jurnal Kelitbangan*, 13(1), 1–14.
- Islam, M. M., Alharthi, M., Alkadi, R. S., Islam, R., & Masum, A. K. M. (2024). Crop yield prediction through machine learning: A path towards sustainable agriculture and climate resilience in Saudi Arabia. *AIMS Agriculture and Food*, 9(4), 980–1003. https://doi.org/10.3934/agrfood.2024053
- Jabed, M. A., & Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10(24). https://doi.org/10.1016/j.heliyon.2024.e40836
- Jhajharia, K., Sharma, N. V., & Mathur, P. (2025). A machine learning model for crop yield prediction using remote sensing data. 6, 577–590.

- <https://doi.org/10.47857/irjms.2025.v06i02.03182>
- Manurung, D., Zealtiel, B., & Lubis, A. H. (2025). Prediksi produksi tanaman padi di Indonesia dengan menggunakan algoritma Random Forest Regressor. *Journal of Computing and Informatics Research*, 4(3), 345–352. <https://doi.org/10.47065/comforch.v4i3.2125>
- Masahid, M., Dawud, M. Y., & Abryandoko, E. W. (2025). Pengaruh fluktuasi unsur iklim tahunan terhadap produksi padi: Studi empiris berdasarkan data historis. *Agrikultura*, 36(2), 216–227. <https://doi.org/10.24198/agrikultura.v36i2.63758>
- Mugemangango, C., Nzabanita, J., Muhoza, D. N., & Cahill, N. D. (2024). Comparative analysis of machine learning models for predicting rice yield: Insights from agricultural inputs and practices in Rwanda. *Research on World Agricultural Economy*, 5(4), 350–366. <https://doi.org/10.36956/rwae.v5i4.1247>
- Nareindra Bayutama Wibisono, S. S. (2025). Crop yield prediction using Random Forest algorithm and XGBoost machine learning model. *International Journal of Research and Innovation in Social Science*, 8, 1175–1189.
- Nikhil, U. V., Pandiyan, A. M., Raja, S. P., & Stamenkovic, Z. (2024). Machine learning-based crop yield prediction in South India: Performance analysis of various models. *Computers*, 13(6). <https://doi.org/10.3390/computers13060137>
- Nizami, T., Mustaqim, M. A., & Ariannor, W. (2025). Analisis kinerja model machine learning dalam prediksi gagal panen gabah. *Progresif: Jurnal Ilmiah Komputer*, 21(1), 184–193. <https://doi.org/10.35889/progresif.v21i1.2501>
- Nuraini, D., Violina, D., Anamisa, D. R., Khotimah, B. K., Jauhari, A., & Mufarroha, F. A. (2025). Prediksi hasil panen padi dengan metode multiple linear regression dan particle swarm optimization untuk meningkatkan produksi padi di Madura. *JUSIFOR: Jurnal Sistem Informasi dan Informatika*, 4(1), 1–8. <https://doi.org/10.70609/jusifor.v4i1.5857>
- Petropoulos, T., Beinos, L., Berruto, R., Miserendino, G., Marinoudi, V., Busato, P., Zisis, C., & Bochtis, D. (2025). *Interpretable machine learning for legume yield prediction using satellite remote sensing data*.

- Purwaningrum, Y., Asbur, Y., Atmaja Nasution, S., & Nuh, M. (2025). Pengaruh perubahan iklim terhadap produktivitas tanaman padi gogo (**Oryza sativa** L.). *AGRILAND Jurnal Ilmu Pertanian*, 13(1), 19–23.
- Quille-Mamani, J., Ramos-Fernández, L., Huanuqueño-Murillo, J., Quispe-Tito, D., Cruz-Villacorta, L., Pino-Vargas, E., Flores del Pino, L., Heros-Aguilar, E., & Ángel Ruiz, L. (2025). Rice yield prediction using spectral and textural indices derived from UAV imagery and machine learning models in Lambayeque, Peru. *Remote Sensing*, 17(4). https://doi.org/10.3390/rs17040632
- Razavi, M. A., Nejadhashemi, A. P., Majidi, B., Razavi, H. S., Kpodo, J., Eiswaran, R., Ciampitti, I., & Prasad, P. V. V. (2024). Enhancing crop yield prediction in Senegal using advanced machine learning techniques and synthetic data. *Artificial Intelligence in Agriculture*, 14, 99–114. https://doi.org/10.1016/j.aiia.2024.11.005
- Saha, S., Kucher, O. D., Utkina, A. O., & Rebouh, N. Y. (2025). Precision agriculture for improving crop yield predictions: A literature review. https://doi.org/10.3389/fagro.2025.1566201
- Shawon, S. M., Eima, F. B., Mahi, A. K., Niha, F. L., & Zubair, H. T. (2025). Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology*, 10, 100718. https://doi.org/10.1016/j.atech.2024.100718
- Sweet, L., Müller, C., Anand, M., & Zscheischler, J. (2023). Cross-validation strategy impacts the performance and interpretation of machine learning models. *Artificial Intelligence for the Earth Systems*, 2(4), 1–14. https://doi.org/10.1175/aies-d-23-0026.1
- Tasneem, K. T., Shahzad, M. U., Rashid, J., Othman, K. M., Zafar, T., & Faheem, M. (2025). Predicting rice yield and impact of climate change on rice production using machine learning models. *Theoretical and Applied Climatology*, 156(12). https://doi.org/10.1007/s00704-025-05912-2
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. https://doi.org/10.1016/j.compag.2020.105709
- Yenkikar, A., Mishra, V. P., Bali, M., & Ara, T. (2025). An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction. *MethodsX*, 15, 103442.

<https://doi.org/10.1016/j.mex.2025.103442>

- Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review. *Drones*, 8(10). <https://doi.org/10.3390/drones8100559>
- Yunis, R., Sudarto, & Adiputra Pardosi, I. (2024). Enhancing rice production prediction: A comparative machine learning analysis of climate variables. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 13(1), 91–104. <https://doi.org/10.23887/janapati.v13i1.71527>