

Implementasi Metode *Boosting* Untuk Prediksi Jenis Tanaman Berdasarkan Kondisi Tanah

Rafif Huda Aditya^{1*}, Ghufron¹

¹Universitas Islam Sultan Agung, Semarang, Indonesia

*E-mail: rafifhudaaditya@gmail.com

ABSTRAK

Pertanian merupakan sektor strategis dalam mendukung ketahanan pangan dan pembangunan berkelanjutan, khususnya terkait SDG 2 (Zero Hunger) dan SDG 9 (Industry, Innovation, and Infrastructure). Tantangan global seperti degradasi tanah dan perubahan iklim menuntut adanya inovasi berbasis teknologi. Penelitian ini bertujuan mengembangkan sistem prediksi jenis tanaman berdasarkan parameter tanah (N, P, K, suhu, kelembaban, pH) dengan membandingkan algoritma Gradient Boosting Machine (GBM) dan Light Gradient Boosting Machine (LightGBM). Dataset diperoleh dari Kaggle dan diproses melalui normalisasi serta label encoding, kemudian dilakukan hyperparameter tuning dengan RandomizedSearch. Hasil menunjukkan GBM memiliki akurasi 96,14% dan LightGBM 96,82%. Temuan penelitian menunjukkan bahwa metode boosting efektif digunakan pada sistem rekomendasi tanaman berbasis kondisi tanah, sekaligus mendukung penerapan pertanian presisi yang berkelanjutan.

Kata Kunci : Pertanian Presisi, *Machine Learning*, *Gradient Boosting*, LightGBM, Prediksi Tanaman

ABSTRACT

Agriculture is a strategic sector in supporting food security and sustainable development, particularly in achieving SDG 2 (Zero Hunger) and SDG 9 (Industry, Innovation, and Infrastructure). Global challenges such as soil degradation and climate change require technological innovations. This study aims to develop a crop prediction system based on soil parameters (N, P, K, temperature, humidity, pH) by comparing Gradient Boosting Machine (GBM) and Light Gradient Boosting Machine (LightGBM). The dataset was obtained from Kaggle, processed using normalization and label encoding, and optimized through RandomizedSearch for hyperparameter tuning. Results show that GBM achieved 96.14% accuracy, while LightGBM achieved 96.82%. The system was implemented into a web-based interface using Streamlit. The findings highlight that boosting methods are effective for crop recommendation systems based on soil conditions while supporting the adoption of sustainable precision agriculture.

Keywords : Precision Agriculture, *Machine Learning*, *Gradient Boosting*, LightGBM, Crop Prediction

PENDAHULUAN DAN TINJAUAN PUSTAKA

Pendahuluan

Sektor pertanian memegang peranan fundamental dalam upaya mencapai beberapa Tujuan Pembangunan Berkelanjutan (SDGs), terutama SDG 2: Tanpa Kelaparan (*Zero Hunger*), dengan menopang ketahanan pangan dan stabilitas ekonomi dunia, sebagaimana ditegaskan dalam berbagai studi yang menyatakan bahwa pertanian merupakan fondasi bagi keberlangsungan hidup manusia sekaligus pilar utama pembangunan ekonomi dan ketahanan pangan global[1]. Seiring dengan proyeksi pertumbuhan populasi global, tuntutan untuk meningkatkan produktivitas agrikultur secara berkelanjutan menjadi semakin mendesak. Namun, sektor ini menghadapi tantangan signifikan berskala global, seperti degradasi tanah, cuaca ekstrem, dan penurunan kesuburan lahan, yang berisiko menurunkan produktivitas pertanian hingga 14% pada tahun 2050 jika tidak ada upaya adaptasi[2]. Kondisi ini diperparah oleh fakta bahwa 33% tanah dunia mengalami degradasi sedang hingga berat[3], mengancam ketahanan pangan dan keberlanjutan ekosistem.

Untuk mengatasi kompleksitas ini, inovasi yang sejalan dengan SDG 9: Industri, Inovasi, dan Infrastruktur menjadi krusial, di mana pendekatan pertanian presisi (*precision agriculture*) menjadi kunci untuk optimalisasi hasil panen yang efisien dan berkelanjutan. Selain itu, pemanfaatan data terbuka dan perangkat *open-source* memungkinkan praktik pertanian presisi berbasis bukti[4].

Kondisi degradasi tanah dan perubahan iklim tersebut secara langsung memengaruhi ketersediaan unsur hara utama seperti nitrogen (N), fosfor (P), dan kalium (K), serta parameter lingkungan seperti suhu dan kelembaban tanah. Variabel-variabel ini sangat menentukan kesesuaian suatu lahan untuk jenis tanaman tertentu, sehingga analisis berbasis data terhadap parameter tersebut menjadi langkah strategis dalam mengoptimalkan pemilihan tanaman dan meningkatkan produktivitas pertanian.

Dalam ekosistem pertanian presisi, *machine learning* muncul sebagai teknologi transformatif yang mampu memberikan solusi berbasis data. Berbagai studi menunjukkan bahwa *machine learning* efektif dalam menganalisis variabel lingkungan yang kompleks seperti komposisi nutrisi tanah (N, P, K), suhu, kelembaban, dan tingkat pH untuk memprediksi jenis tanaman yang paling cocok untuk suatu lokasi[5]. Dengan mendeteksi pola yang sulit dikenali secara manual, teknologi ini memungkinkan petani dan praktisi agrikultur untuk membuat keputusan yang lebih akurat, mengurangi risiko kegagalan panen, dan meningkatkan efisiensi penggunaan lahan[6],[7].

Di antara beragam teknik *machine learning*, metode ensemble boosting seperti *Gradient Boosting Machine* (GBM) dan *Light Gradient Boosting Machine* (LightGBM) telah menunjukkan performa yang superior untuk berbagai tugas klasifikasi. Sebuah studi relevan yang [8] menyoroti adanya *trade-off* antara kedua model ini, di mana satu model mungkin unggul dalam kecepatan sementara yang lain menawarkan stabilitas

akurasi. Temuan ini membuka celah penelitian (*research gap*) yang penting: perlunya analisis perbandingan komprehensif untuk menentukan model mana yang memberikan keseimbangan terbaik antara akurasi prediksi dan efisiensi komputasi saat dihadapkan pada dataset agrikultur yang beragam dan mencakup berbagai kondisi iklim. Lebih jauh, eksplorasi mengenai dampak *hyperparameter tuning* terhadap performa masing-masing model dalam konteks global ini masih perlu didalami.

Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis perbandingan kinerja dan performa antara model GBM dan LightGBM untuk prediksi jenis tanaman menggunakan dataset yang representatif secara global. Penelitian akan berfokus pada evaluasi pengaruh hyperparameter tuning terhadap akurasi dan efisiensi komputasi. Hasilnya diharapkan dapat memberikan wawasan berharga bagi komunitas ilmiah dan praktisi precision agriculture. Dengan demikian, penelitian ini tidak hanya berkontribusi secara teknis, tetapi juga secara langsung mendukung pencapaian SDG 2 dengan mendorong produktivitas pertanian, serta sejalan dengan semangat inovasi pada SDG 9 untuk membangun solusi teknologi yang andal bagi tantangan global.

Tinjauan Pustaka

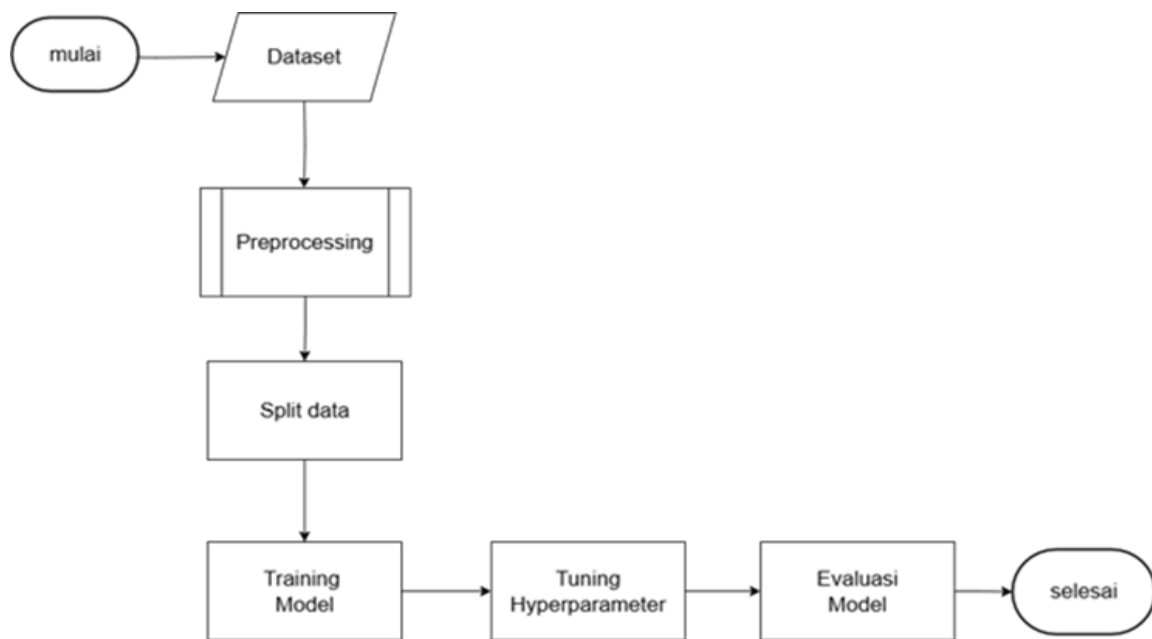
Berbagai penelitian sebelumnya menunjukkan bahwa metode boosting, seperti Gradient Boosting Machine (GBM) dan Light Gradient Boosting Machine (LGBM), mampu memberikan performa tinggi dalam tugas klasifikasi. Beberapa studi membandingkan kedua model ini dan menemukan bahwa GBM cenderung lebih stabil dalam prediksi, sedangkan LGBM sering kali unggul dalam efisiensi komputasi meskipun dengan tingkat akurasi yang sebanding[8], [9].

Penelitian lain menunjukkan bahwa metode boosting dapat mencapai akurasi tinggi pada berbagai domain, seperti prediksi harga sewa rumah, klasifikasi kelulusan mahasiswa, hingga deteksi penyakit, dengan capaian akurasi antara 71% hingga 100% pada dataset tertentu[10]–[12]. Hasil-hasil ini menegaskan bahwa baik GBM maupun LGBM memiliki potensi besar untuk diterapkan pada berbagai kasus klasifikasi dengan performa yang kompetitif.

Secara keseluruhan, studi-studi tersebut menegaskan bahwa baik GBM maupun LGBM memiliki keunggulan masing-masing: GBM unggul dalam stabilitas prediksi, sedangkan LGBM menawarkan efisiensi komputasi dan akurasi tinggi pada dataset besar. Hal ini menjadi landasan penting bagi penelitian ini untuk mengeksplorasi performa kedua model dalam konteks klasifikasi jenis tanaman berbasis kondisi tanah.

METODE/EKSPERIMEN

Alur penelitian ini sebagaimana dapat dilihat pada Gambar 1



Gambar 1. Alur Penelitian

1. Pengumpulan Dataset

Langkah ini merupakan proses pengumpulan data yang akan digunakan sebagai bahan utama dalam pelatihan dan pengujian model *machine learning*. Penelitian ini menggunakan dataset publik yang diperoleh dari platform Kaggle. Dataset ini berisi parameter tanah dan lingkungan yang terdiri atas nitrogen (N), fosfor (P), kalium (K), suhu, kelembaban, dan pH, dengan label target berupa jenis tanaman. Rincian struktur data ditunjukkan pada Tabel 1.

Table 1. Tipe Data Penelitian

No	Nama Kolom	Tipe Data	Keterangan
1	N	Int64	Rasio kandungan Nitrogen dalam tanah
2	P	Int64	Kandungan fosfor (Phosphorus) dalam tanah.
3	K	Int64	Kandungan Kalium dalam tanah
4	Temperature	Float64	Suhu dalam tanah dengan satuan derajat Celcius
5	Humidity	Float64	Kelembaban dalam tanah dengan satuan persen (%)
6	pH	Float64	Tingkat keasaman tanah
7	label	String	nama tanaman sesuai parameter tanah

Selain itu, dataset terdiri dari 22 jenis tanaman, masing-masing memiliki 100 data sampel, sebagaimana ditunjukkan pada Gambar 2.

Jumlah data untuk semua kelas:

label	
rice	100
maize	100
chickpea	100
kidneybeans	100
pigeonpeas	100
mothbeans	100
mungbean	100
blackgram	100
lentil	100
pomegranate	100
banana	100
mango	100
grapes	100
watermelon	100
muskmelon	100
apple	100
orange	100
papaya	100
coconut	100
cotton	100
jute	100
coffee	100

Gambar 2. Jumlah Data per Label

Dengan demikian, total dataset terdiri dari 2200 baris data yang seimbang pada setiap label tanaman, sehingga sangat sesuai untuk digunakan dalam eksperimen klasifikasi berbasis *machine learning*.

Unsur hara tersebut merupakan faktor penting dalam pertumbuhan tanaman. Nitrogen berkontribusi pada pertumbuhan vegetatif, fosfor mendukung perkembangan akar, sedangkan kalium memengaruhi regulasi air dan kualitas buah. Faktor eksternal seperti suhu, kelembaban, dan pH tanah turut menentukan kesesuaian lahan pertanian[7].

Pra-preprosesan

2. Penghapusan fitur

Pada tahap ini dilakukan penghapusan fitur yang dianggap tidak relevan terhadap tujuan penelitian. Fitur rainfall dihapus dari dataset karena dinilai tidak memberikan kontribusi signifikan dalam proses prediksi jenis tanaman. Langkah ini bertujuan untuk menyederhanakan model, mengurangi kompleksitas, serta menyesuaikan input sistem prediksi agar lebih fokus pada variabel yang berpengaruh langsung.

Dataset yang digunakan selanjutnya hanya memuat parameter yang relevan, yaitu nitrogen (N), fosfor (P), kalium (K), suhu, kelembaban, pH, dan label tanaman. Setelah tahap ini, dilakukan eksplorasi awal untuk memahami distribusi data, termasuk pemeriksaan representasi label tanaman guna memastikan bahwa dataset memiliki proporsi kelas yang seimbang.

3. Label Encoding

Pada tahap pra-pemrosesan data, dilakukan encoding terhadap variabel target (label tanaman) agar dapat diproses oleh algoritma machine learning. Kolom label yang semula berbentuk teks, seperti *rice*, *maize*, dan lainnya, dikonversi menjadi nilai numerik menggunakan *LabelEncoder* dari pustaka *Scikit-Learn*. Proses ini juga menyimpan hasil encoding beserta label asli untuk keperluan interpretasi hasil.

Proses encoding dilakukan dengan memanggil kelas *LabelEncoder*, kemudian menggunakan fungsi *fit_transform()* untuk mengubah nilai string pada kolom label menjadi format numerik. Hasil encoding tersebut disimpan ke dalam variabel *y*, yang selanjutnya digunakan sebagai target (label) dalam proses pelatihan model.

	label_asli	label_encoded
0	rice	20
1	rice	20
2	rice	20
3	rice	20
4	rice	20
...
2195	coffee	5
2196	coffee	5
2197	coffee	5
2198	coffee	5
2199	coffee	5

[2200 rows x 2 columns]

Gambar 3. Hasil Encoding

Gambar 3 menunjukkan hasil encoding. Pada hasil tersebut, kolom *label_asli* berisi nama tanaman, misalnya *rice* dan *coffee*, sementara kolom *label_encoded* berisi representasi numerik dari label tersebut, misalnya *rice* menjadi 20 dan *coffee* menjadi 5. Setiap label unik dalam dataset memperoleh kode angka yang bersifat diskret dan tidak memiliki urutan semantik, namun sangat diperlukan agar algoritma machine learning dapat memproses data dengan baik.

4. Split Data

Sebelum proses pelatihan model machine learning, dataset dibagi menjadi dua bagian utama, yaitu data training dan data testing. Pada penelitian ini, pembagian dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian, menggunakan fungsi *train_test_split* dari pustaka *Scikit-Learn*.

Pembagian data dilakukan dengan teknik stratifikasi berdasarkan label agar distribusi kelas tetap proporsional pada kedua subset data. Teknik ini penting untuk menghindari bias distribusi kelas yang dapat memengaruhi akurasi model serta menjaga representasi data tetap seimbang.

Table 2. hasil pembagian data

1	Jumlah data training	1760 data
2	Data Testing	440 data

Proses split data ini memiliki peran penting dalam menghindari overfitting, serta memungkinkan pengukuran performa model secara objektif. Dengan pembagian yang acak namun tetap terstruktur, model dapat dilatih menggunakan data training dan dievaluasi pada data testing untuk menilai sejauh mana kemampuan generalisasinya terhadap data baru yang belum pernah dikenali sebelumnya.

5. Hyperparameter Tuning

Untuk meningkatkan performa model dalam prediksi jenis tanaman berbasis data tanah, dilakukan proses *hyperparameter tuning* pada algoritma Gradient Boosting Machine (GBM) dan Light Gradient Boosting Machine (LGBM). *Hyperparameter* merupakan parameter yang mengatur mekanisme internal algoritma dan tidak dapat dipelajari langsung dari data, sehingga pemilihan nilai yang optimal sangat penting untuk menghasilkan model yang akurat, stabil, dan mampu melakukan generalisasi terhadap data baru.

Dalam penelitian ini digunakan teknik *RandomizedSearchCV*, yaitu metode pencarian *hyperparameter* secara acak dari ruang parameter yang telah ditentukan. Pendekatan ini dinilai lebih efisien secara komputasi dibandingkan *GridSearchCV*, terutama pada ruang pencarian yang luas. Meskipun tidak selalu menemukan kombinasi absolut terbaik, metode ini mampu menghasilkan konfigurasi yang kompetitif dalam waktu relatif singkat.

Adapun parameter utama yang dioptimalkan meliputi:

- *n_estimators*, jumlah pohon dalam *ensemble*;
- *max_depth*, kedalaman maksimum pohon;
- *learning_rate*, laju pembaruan bobot pada setiap iterasi *boosting*;
- *subsample*, proporsi data pelatihan yang digunakan untuk membangun setiap pohon.

Proses tuning dilakukan pada data latih dengan teknik *cross-validation* guna memastikan performa model konsisten pada berbagai subset data. Kombinasi parameter terbaik kemudian dipilih berdasarkan nilai akurasi tertinggi dari proses validasi, dan digunakan untuk melatih ulang model akhir (*final model*). Model inilah yang kemudian dievaluasi pada data uji untuk menilai kemampuan klasifikasi terhadap jenis tanaman berdasarkan kondisi tanah.

6. Gradient Boosting Machine(GBM)

Gradient Boosting Machine merupakan salah satu metode *ensemble learning* yang membangun model prediksi melalui kombinasi sejumlah pohon keputusan (*decision trees*) secara berurutan. Setiap pohon baru dibentuk untuk memperbaiki kesalahan dari pohon sebelumnya dengan meminimalkan fungsi kerugian (*loss function*) menggunakan algoritma gradien.

7. LightGBM

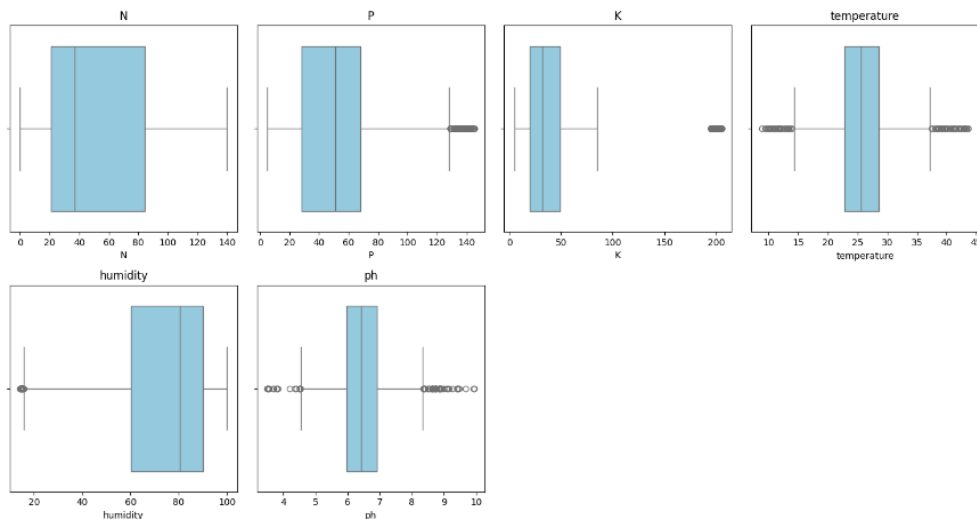
LightGBM merupakan pengembangan dari GBM yang menggunakan pendekatan *leaf-wise growth* dengan *depth limitation*. Pendekatan ini memungkinkan proses pembelajaran lebih cepat dan efisien dalam penggunaan memori dibandingkan metode *level-wise growth* yang digunakan pada GBM.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen yang diperoleh dari penerapan dua algoritma *ensemble boosting*, yaitu *Gradient Boosting Machine* (GBM) dan *Light Gradient Boosting Machine* (LightGBM), pada dataset kondisi tanah. Analisis dilakukan dengan tujuan mengevaluasi performa kedua model dalam memprediksi jenis tanaman berdasarkan parameter tanah berupa nitrogen (N), fosfor (P), kalium (K), suhu, kelembaban, dan pH.

1. Analisa *Boxplot*

Pada analisis distribusi dan outlier menyajikan visualisasi *boxplot* untuk setiap variabel numerik pada dataset, *boxplot* digunakan untuk menggambarkan distribusi data, rentang nilai, serta mendeteksi keberadaan outlier pada masing-masing fitur. Pada Gambar 3 merupakan membantu dalam memahami pola sebaran data dan potensi nilai ekstrem yang dapat mempengaruhi proses pemodelan.



Gambar 3. Visualisasi *Boxplot*

Dari visualisasi Gambar 3, terlihat bahwa beberapa fitur memiliki nilai-nilai ekstrem (outlier), terutama pada fitur P dan K, yang menunjukkan sebaran data ke arah atas secara signifikan. Fitur temperature, humidity, dan ph juga mengandung sejumlah outlier, meskipun dalam jumlah yang relatif kecil. Sementara itu, fitur N memiliki distribusi yang cukup luas, namun tidak menunjukkan outlier yang mencolok.

2. Normalisasi

Normalisasi ini bertujuan untuk memastikan bahwa setiap fitur memiliki skala yang seragam, sehingga tidak ada satu fitur pun yang mendominasi proses pembelajaran akibat perbedaan skala antar fitur. *RobustScaler* bekerja dengan median dan interkuartil range (IQR), sehingga lebih tahan terhadap nilai-nilai ekstrem atau outlier. Ini membuatnya sangat sesuai digunakan dalam dataset yang mengandung variabel dengan sebaran tidak simetris atau memiliki nilai ekstrim. Hal ini didukung pula oleh Singh dkk, yang menunjukkan bahwa strategi estimasi *robust* dapat menjaga akurasi model meskipun data mengandung nilai ekstrem[13].

3. Akurasi sebelum *hyperparameter tuning*

Proses pelatihan dimulai dengan membangun model dasar dari masing-masing algoritma, yaitu *Gradient Boosting Machine* (GBM) dan *LightGBM* (LGBM). Model dilatih menggunakan parameter default sebagai baseline untuk mengetahui sejauh

mana masing-masing algoritma mampu mempelajari pola data. Hasil dari pelatihan awal ini akan menjadi dasar dalam melakukan *hyperparameter tuning*.

Table 3. akurasi sebelum hyperparameter tuning

Model	akurasi	precision	recall	F1-Score
GBM	96.36%	0.96	0.96	0.96
LGBM	96.36%	0.96	0.96	0.96

Tabel 3 menunjukkan hasil evaluasi performa awal dari kedua model. Model GBM menghasilkan akurasi sebesar 96,36%, dengan nilai precision, recall, dan F1-score rata-rata sebesar 0,96. Hasil ini menunjukkan bahwa model GBM sudah cukup andal dalam melakukan klasifikasi multikelas berdasarkan kondisi tanah.

4. LightGBM

Setelah *tuning* selesai, diperoleh kombinasi hyperparameter terbaik sebagaimana ditampilkan pada Tabel 4.6, yaitu:

Table 4. Best Parameter LGBM

<i>colsample_bytree</i>	<i>subsample</i>	<i>num_leaves</i>	<i>n_estimators</i>	<i>max_depth</i>	<i>learning_rate</i>
0.79	0.6	55	180	5	0.01

Kombinasi pada tabel 4 menunjukkan bahwa model memperoleh performa optimal dengan jumlah pohon yang besar, kedalaman pohon sedang, dan laju pembelajaran yang kecil sehingga pembaruan bobot berjalan secara bertahap. Nilai *subsample* yang rendah juga memberikan efek regularisasi dengan meningkatkan variasi antar pohon, sehingga membantu mengurangi risiko *overfitting*.

Secara keseluruhan, *hyperparameter tuning* yang dilakukan pada model LightGBM terbukti efektif dalam meningkatkan kemampuan klasifikasi sistem rekomendasi tanaman berdasarkan kondisi tanah. Kombinasi parameter optimal ini digunakan dalam pelatihan ulang model untuk memperoleh performa terbaik pada tahap evaluasi akhir.

5. Gradient Boosting Machine(GBM)

Untuk model *Gradient Boosting Machine* (GBM), proses *tuning* menghasilkan kombinasi *hyperparameter* terbaik yaitu :

Table 5. best parameter GBM

<i>subsample</i>	<i>n_estimators</i>	<i>min_samples_split</i>	<i>max_depth</i>	<i>learning_rate</i>
0.68	489	3	6	0.0359

Setelah *tuning* selesai, diperoleh kombinasi hyperparameter terbaik seperti yang ditampilkan pada tabel 5, yaitu: *subsample*=0.68, *n_estimators*=489, *max_depth*=7, dan *learning_rate*=0.0359. Kombinasi ini menunjukkan bahwa GBM bekerja lebih optimal ketika hanya menggunakan sebagian data pelatihan pada setiap iterasi dan

membangun pohon dengan kedalaman sedang untuk menangkap pola yang kompleks. Nilai *learning rate* yang sedang juga memberikan keseimbangan antara ketepatan dan kestabilan dalam pembelajaran model.

Dengan konfigurasi ini, model GBM berhasil mencapai performa yang baik dalam melakukan klasifikasi jenis tanaman berdasarkan parameter tanah. Hasil *tuning* ini kemudian digunakan untuk pelatihan ulang model akhir dan dievaluasi pada data uji guna menilai kemampuan generalisasi dari model yang telah dioptimalkan.

6. Akurasi sesudah *hyperparameter tuning*

Hasil pengujian menunjukkan bahwa kedua algoritma *boosting* mampu memberikan performa prediksi yang tinggi pada dataset agrikultur. Model GBM menghasilkan akurasi sebesar 96,14%, sedangkan LightGBM sedikit lebih unggul dengan akurasi 96,82%. Meskipun perbedaan akurasi relatif kecil, perbedaan signifikan terlihat pada waktu komputasi. Proses pelatihan GBM membutuhkan waktu sekitar 3378,41 detik, sedangkan LightGBM hanya memerlukan 281,66 detik.

Hal ini mengindikasikan bahwa LightGBM lebih efisien secara komputasi tanpa mengorbankan akurasi prediksi. Sebaliknya, GBM cenderung lebih stabil pada variasi dataset tertentu, meskipun memerlukan waktu yang lebih lama untuk dilatih. *Trade-off* ini sesuai dengan temuan penelitian sebelumnya yang menyoroti adanya keunggulan relatif masing-masing algoritma.

Analisis *confusion matrix* memperlihatkan bahwa kedua model mampu mengklasifikasikan jenis tanaman dengan tingkat kesalahan yang rendah. Implementasi berbasis Streamlit memungkinkan visualisasi hasil secara interaktif, sehingga pengguna dapat dengan mudah memahami prediksi dan performa model. Selain itu, penelitian ini menegaskan relevansi *machine learning* dalam mendukung pertanian presisi, di mana pengambilan keputusan berbasis data dapat mengurangi risiko gagal panen dan meningkatkan efisiensi penggunaan lahan.

PENUTUP

Kesimpulan

Penelitian ini menunjukkan bahwa *Gradient Boosting Machine* (GBM) dan *Light Gradient Boosting Machine* (LightGBM) sama-sama efektif digunakan dalam prediksi jenis tanaman berdasarkan parameter tanah. LightGBM terbukti lebih unggul dalam efisiensi komputasi dengan akurasi yang hampir setara dengan GBM. Implementasi sistem berbasis Streamlit juga berhasil menyediakan antarmuka interaktif yang mendukung penggunaan praktis.

Secara keseluruhan, hasil penelitian ini memberikan kontribusi dalam pengembangan sistem rekomendasi tanaman berbasis *machine learning*, yang tidak hanya bermanfaat bagi praktisi pertanian, tetapi juga mendukung pencapaian SDG 2 dan SDG 9.

Saran

Untuk pengembangan lebih lanjut, disarankan menambahkan parameter lingkungan seperti curah hujan, jenis tanah, dan lokasi geografis, serta memperluas cakupan data agar model lebih adaptif terhadap kondisi nyata. Selain itu, integrasi dengan sensor IoT atau platform mobile/web dapat meningkatkan kemanfaatan sistem secara real-time dalam mendukung pertanian presisi.

DAFTAR PUSTAKA

- M. S. Allahyari dan A. Poursaeed, "Sustainable Agriculture: Implication for SDG2 (Zero Hunger)," vol. 2, no. May 2019, 2021, doi: 10.1007/978-3-319-69626-3.
- A. A. Farah, M. A. Mohamed, O. S. H. Musse, dan B. A. Nor, "The multifaceted impact of climate change on agricultural productivity: a systematic literature review of SCOPUS-indexed studies (2015–2024)," *Discov. Sustain.*, vol. 6, no. 1, 2025, doi: 10.1007/s43621-025-01229-2.
- P. Smith *dkk.*, "Status of the World ' s Soils," hal. 73–104.
- J. H. Jeppesen, R. H. Jacobsen, R. Nyholm Jørgensen, dan T. S. Toftegaard, "Towards Data-Driven Precision Agriculture using Open Data and Open Source Software," *Int. Conf. Agric. Eng. 2016*, hal. 1–6, 2022, [Daring]. Tersedia pada: <https://arxiv.org/abs/2204.05582>
<https://arxiv.org/pdf/2204.05582>
- D. Singh Mohan, V. Dhote, P. Mishra, P. Singh, dan A. Srivastav, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING IoT Framework for Precision Agriculture: Machine Learning Crop Prediction," *Orig. Res. Pap. Int. J. Intell. Syst. Appl. Eng. IJISAE*, vol. 2023, no. 5s, hal. 300–313, 2023, [Daring]. Tersedia pada: www.ijisae.org
- V. Meshram, K. Patil, V. Meshram, D. Hanchate, dan S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artif. Intell. Life Sci.*, vol. 1, no. October, 2021, doi: 10.1016/j.aillsci.2021.100010.
- A. C. Muhammad *dkk.*, "Dasar-dasar Pembelajaran Mesin," hal. 131, 2023.
- I. Wardhana, Musi Ariawijaya, Vandri Ahmad Isnaini, dan Rahmi Putri Wirman, "Gradient Boosting Machine, Random Forest dan Light GBM untuk Klasifikasi Kacang Kering," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, hal. 92–99, 2022, doi: 10.29207/resti.v6i1.3682.
- R. Dahlia dan C. I. Agustyaningrum, "Perbandingan Gradient Boosting dan Light Gradient Boosting Dalam Melakukan Klasifikasi Rumah Sewa," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 6, hal. 1016–1020, 2022, doi: 10.32672/jnkti.v5i6.5460.
- S. S. Dahlia Rizka, Fitriana Lady Agustin, "ANALISIS ALGORITMA GRADIENT BOOSTING DALAM PENGARUH," vol. 8, hal. 36–44, 2025.
- V. Atlantic, E. Sulistianingsih, dan H. Perdana, "Gradient Boosting Machine Pada Klasifikasi Kelulusan Mahasiswa," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 2, hal. 165–174, 2024.
- L. M. Cendani dan A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *J. Masy. Inform.*, vol. 13, no. 1, hal. 33–44, 2022, doi: 10.14710/jmasif.13.1.42912.
- G. Singh, D. Bhattacharyya, dan A. Bandyopadhyay, "Robust estimation strategy for handling outliers," *Commun. Stat. - Theory Methods*, vol. 53, hal. 5311–5330, 2023, doi: 10.1080/03610926.2023.2218567.