

OPTIMALISASI MODEL ENSEMBLE PADA FRAMEWORK DATA INTELLIGENCE UNTUK MEMPREDIKSI PRODUKTIVITAS KELAPA SAWIT BERBASIS POWER BI

Anggun Agnesia Silaen, Agustia Hananto, Elvina Novalia
Sistem Informasi, Universitas Buana Perjuangan Karawang
Jalan HS. Ronggo Waluyo, Telukjambe Timur, Puseurjaya, Karawang
si22.anggunsilaen@mhs.ubpkarawang.ac.id

ABSTRAK

Produktivitas kelapa sawit pada kebun rakyat di Indonesia masih rendah akibat pengelolaan agronomi yang belum berbasis data. Permasalahan utama terletak pada pengambilan keputusan yang masih mengandalkan pengalaman sehingga kurang optimal dalam meningkatkan hasil panen. Penelitian ini bertujuan untuk mengembangkan framework data intelligence yang mampu memprediksi produktivitas kelapa sawit secara akurat serta mengidentifikasi faktor-faktor yang berpengaruh. Metode yang digunakan adalah pendekatan kuantitatif dengan machine learning berbasis model ensemble Random Forest, meliputi tahap pengumpulan data, pra-pemrosesan, pembangunan model, evaluasi, tuning, serta analisis feature importance, yang kemudian divisualisasikan menggunakan dashboard Power BI. Data yang digunakan berasal dari tiga kebun periode 2020–2024. Hasil penelitian menunjukkan bahwa tuning model meningkatkan kinerja dengan nilai R^2 sebesar 0,7065 serta penurunan MAE dan RMSE, yang menandakan peningkatan akurasi dan stabilitas prediksi. Analisis feature importance menunjukkan bahwa kondisi tanah dan luas lahan merupakan faktor paling dominan dalam menentukan produktivitas. Framework yang dihasilkan mampu memberikan prediksi yang andal serta insight agronomi yang mendukung pengambilan keputusan berbasis data.

Kata kunci: Data Intelligence, Machine Learning, Produktivitas Kelapa Sawit, Feature Importance, Power BI

1. PENDAHULUAN

Kelapa sawit merupakan komoditas strategis di Indonesia yang menjadi sumber utama pendapatan bagi petani. Namun, produktivitas kebun rakyat masih sering berada di bawah potensi optimal akibat praktik agronomi yang belum konsisten, seperti pemupukan, interval panen, pengendalian gulma, dan pengelolaan tanaman. Kondisi ini menyebabkan variasi hasil panen yang cukup besar meskipun berada pada lingkungan yang relatif serupa. Selain itu, pengambilan keputusan yang masih berbasis intuisi membuat petani kurang mampu merespons perubahan kondisi lingkungan, seperti kualitas tanah, pemupukan secara efektif dan terukur.

Perkembangan teknologi digital, khususnya machine learning (ML), menawarkan solusi melalui kemampuan analisis data agronomi untuk memprediksi produktivitas dan mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil panen [1] [2]. ML mampu mengolah data secara kompleks dan menemukan pola yang tidak terlihat melalui pendekatan tradisional [1]. Namun, kompleksitas hasil analisis ML serta keterbatasan literasi digital petani menjadi tantangan dalam implementasinya. Selain itu, penelitian yang berfokus pada petani kecil masih relatif terbatas, sehingga diperlukan pendekatan yang lebih aplikatif, sederhana, dan mudah dipahami agar teknologi dapat dimanfaatkan secara optimal.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan sebuah framework data intelligence yang mampu mengolah dan menganalisis

data agronomi kelapa sawit menggunakan metode model ensemble dan analisis feature importance untuk mengidentifikasi faktor utama yang memengaruhi produktivitas [3] [4]. Selain itu, penelitian ini juga bertujuan menyajikan hasil analisis dalam bentuk dashboard interaktif berbasis Power BI yang informatif dan mudah digunakan oleh petani kecil dalam memantau kondisi kebun secara berkelanjutan [5].

Sebagai rencana penyelesaian masalah, penelitian ini mengintegrasikan machine learning dan visualisasi data dalam satu sistem yang sederhana dan informatif. Model ensemble digunakan untuk meningkatkan akurasi prediksi [3] [6], sedangkan feature importance memberikan pemahaman mengenai kontribusi setiap variabel terhadap hasil panen [7]. Hasil analisis tersebut kemudian divisualisasikan dalam dashboard interaktif agar lebih mudah dipahami oleh pengguna non-teknis [5]. Dengan pendekatan ini, diharapkan petani dapat beralih dari pengambilan keputusan berbasis intuisi menuju pendekatan berbasis data, sehingga produktivitas kelapa sawit dapat ditingkatkan secara lebih efektif dan berkelanjutan.

2. TINJAUAN PUSTAKA

2.1. Framework Data intelligence untuk Optimalisasi Produktivitas

Framework data intelligence dalam agrikultur merupakan pendekatan terintegrasi yang mencakup pengumpulan, pengolahan, analisis, dan visualisasi data untuk mendukung pengambilan keputusan

berbasis data [8]. Dalam konteks kelapa sawit, pendekatan ini memungkinkan data agronomi diolah menjadi informasi yang meningkatkan efisiensi produksi [9]. Prosesnya meliputi *preprocessing*, pemodelan *machine learning* berbasis *ensemble*, serta interpretasi menggunakan *feature importance* untuk mengidentifikasi faktor utama produktivitas [3] [5]. Hasil analisis kemudian disajikan dalam bentuk *dashboard* interaktif guna mempermudah pengambilan keputusan [7].

2.2. Machine Learning dalam Agrikultur

Machine learning (ML) merupakan teknologi utama dalam smart agriculture karena mampu menganalisis data agronomi yang kompleks dan non-linear [1] [2]. Dalam perkebunan kelapa sawit, ML digunakan untuk memodelkan hubungan variabel seperti umur tanaman, kondisi tanah, dan pemupukan guna meningkatkan akurasi prediksi produktivitas [9]. Algoritma berbasis *ensemble* seperti Random Forest dan boosting banyak digunakan karena mampu menangani data beragam dan interaksi kompleks antarvariabel [3] [4]. Selain itu, teknik interpretabilitas seperti SHAP dan *feature importance* membantu mengidentifikasi faktor utama yang memengaruhi hasil produksi [5] [6]. Dengan demikian, ML tidak hanya berfungsi sebagai alat prediksi, tetapi juga mendukung pengambilan keputusan berbasis data.

2.3. Model ensemble dalam Model Prediktif Argonomi

Model *ensemble* merupakan pendekatan *machine learning* yang efektif untuk memodelkan hubungan non-linear dan kompleks pada data agrikultur [1]. Metode boosting membangun model secara bertahap, di mana setiap model memperbaiki kesalahan sebelumnya sehingga meningkatkan akurasi prediksi [3] [4]. Pendekatan ini sesuai untuk analisis produktivitas kelapa sawit karena mampu menangani interaksi variabel agronomi seperti umur tanaman, pemupukan, dan kondisi tanah [9] [5]. Perkembangan algoritma seperti XGBoost, LightGBM, dan CatBoost semakin meningkatkan kinerja model *ensemble* dalam mengolah data multisumber secara efisien [4]. Selain itu, integrasi data agronomi dengan indeks vegetasi dan kelembapan tanah terbukti meningkatkan akurasi prediksi produktivitas [9].

2.4. Feature importance dalam Analitik Agronomi

Feature importance merupakan teknik interpretabilitas dalam *machine learning* yang digunakan untuk mengidentifikasi kontribusi setiap variabel terhadap hasil prediksi model [5] [6]. Dalam analitik agronomi, teknik ini penting karena produktivitas dipengaruhi oleh banyak faktor yang saling berinteraksi, sehingga hasil model perlu dapat dijelaskan untuk mendukung pengambilan keputusan [1]. Metode explainable artificial intelligence seperti SHAP membantu meningkatkan transparansi model

serta mengidentifikasi variabel agronomi yang paling berpengaruh [6]. Dalam penelitian kelapa sawit, pendekatan ini sering dikombinasikan dengan model *ensemble* untuk menganalisis faktor yang memengaruhi hasil panen secara lebih jelas [9] [5].

2.5. Power BI sebagai Platform Business Intelligence

Power BI merupakan platform Business Intelligence yang memungkinkan integrasi, analisis, dan visualisasi data dari berbagai sumber sehingga informasi lebih mudah dipahami [7] [8]. Dalam agrikultur, Power BI digunakan untuk menyajikan data agronomi dan hasil analisis *machine learning* dalam bentuk *dashboard* interaktif guna mendukung pengambilan keputusan [7]. *Dashboard* ini memungkinkan pengguna memantau tren produktivitas, mengevaluasi faktor agronomi, serta mengidentifikasi pola produksi secara lebih intuitif [8]. Dengan demikian, integrasi Power BI dengan analitik data mempermudah interpretasi hasil dan mendukung pengelolaan perkebunan secara lebih efektif.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis analitik data dan *machine learning* untuk mengidentifikasi faktor-faktor yang memengaruhi produktivitas kelapa sawit serta membangun model prediksi berbasis data. Proses penelitian dilakukan secara sistematis melalui beberapa tahapan yang saling terintegrasi, mulai dari pengumpulan data hingga visualisasi hasil dalam *dashboard* interaktif.

3.1. Objek Penelitian

Objek penelitian berupa data agronomi dan produktivitas kelapa sawit dari tiga kebun rakyat di wilayah Hulu Kalimantan Timur dengan periode 2020–2024. Data mencakup variabel agronomi seperti luas lahan, umur tanaman, jenis bibit, kondisi tanah, dan status pemupukan, serta variabel produktivitas berupa total hasil panen (kg)

3.2. Tahapan Penelitian

Penelitian ini menggunakan pendekatan analitik data dan *machine learning* untuk membangun *framework data intelligence* yang mampu mengidentifikasi faktor dominan yang mempengaruhi produktivitas kelapa sawit.

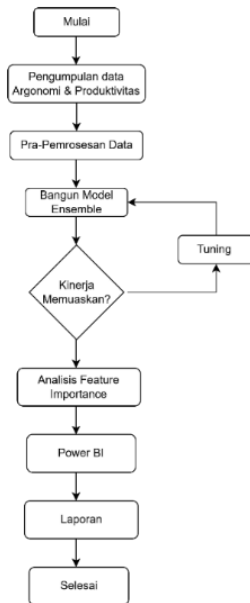
a. Pengumpulan Data

Tahap awal dilakukan dengan mengumpulkan data agronomi dan produktivitas dari tiga kebun. Data yang diperoleh meliputi kondisi tanaman, riwayat pemupukan, serta hasil panen tandan buah segar (TBS). Data ini menjadi dasar utama dalam proses analisis dan pemodelan.

b. Pra-Pemrosesan Data

Data yang telah dikumpulkan kemudian dibersihkan dan disiapkan melalui proses

preprocessing. Tahapan ini meliputi pembersihan data, penanganan missing value, perbaikan format (terutama kolom tanggal), serta transformasi variabel kategorikal menjadi numerik. Tujuan tahap ini adalah memastikan kualitas data agar model dapat bekerja secara optimal.



Gambar 1. Flowchart Tahapan Penelitian.

c. Pembangunan Model Ensemble

Model prediksi dibangun menggunakan metode *ensemble* dengan algoritma Random Forest. Variabel agronomi digunakan sebagai fitur input (X), sedangkan total hasil panen sebagai target (y). Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%) untuk melatih dan menguji performa model.

d. Evaluasi Model dan Pengambilan Keputusan

Tahap evaluasi dilakukan untuk menilai kinerja model menggunakan R^2 Score, MAE, dan RMSE pada data uji. Hasil evaluasi disajikan dalam tabel dan divisualisasikan menggunakan *scatter plot* untuk membandingkan nilai prediksi dan nilai aktual. Berdasarkan hasil tersebut, ditentukan apakah model sudah memuaskan atau perlu dilakukan *tuning hyperparameter*.

Tabel 1 Keputusan Kinerja Model

Metrik	Kriteria "Memuaskan"	Sangat Baik
R^2	> 0.70	> 0.85
MAE	$< 15\%$ dari rata-rata hasil panen	$< 10\%$
RMSE	Tidak jauh lebih besar dari MAE	Mendekati MAE

e. Tahap Tuning Model

Jika hasil evaluasi belum optimal, dilakukan proses *tuning hyperparameter* menggunakan metode seperti GridSearchCV untuk meningkatkan performa model. Model kemudian dievaluasi kembali untuk memastikan peningkatan akurasi.

f. Analisis Feature Importance

Setelah model optimal diperoleh, dilakukan analisis *feature importance* untuk mengidentifikasi variabel yang paling berpengaruh terhadap produktivitas kelapa sawit. Hasil ini memberikan wawasan agronomi yang dapat digunakan sebagai dasar pengambilan keputusan.

g. Integrasi ke Power BI

Hasil prediksi dan analisis *feature importance* kemudian divisualisasikan dalam *dashboard* interaktif menggunakan Power BI. *Dashboard* ini menampilkan tren produktivitas, perbandingan antar kebun, serta faktor dominan secara visual sehingga mudah dipahami oleh pengguna. learning secara jelas dan informatif.

4. HASIL DAN PEMBAHASAN

Menyajikan hasil penelitian dan pembahasan yang diperoleh dari seluruh tahapan analisis dalam *framework data intelligence* untuk memprediksi produktivitas kelapa sawit. Proses penelitian dilakukan secara sistematis, dimulai dari pengumpulan data agronomi dan produktivitas, dilanjutkan dengan pra-pemrosesan data, pembangunan model *machine learning*, evaluasi dan *tuning* model, analisis *feature importance*, hingga integrasi hasil ke dalam *dashboard* Power BI. Melalui tahapan tersebut, penelitian ini tidak hanya menghasilkan model prediksi, tetapi juga menghasilkan informasi yang dapat digunakan untuk mendukung pengambilan keputusan agronomi berbasis data.

4.1. Pengumpulan Data Agronomi dan Produktivitas

Tahap awal penelitian dilakukan dengan mengumpulkan data agronomi dan produktivitas kelapa sawit yang digunakan sebagai dasar analisis. Dataset disusun dari aktivitas panen pada tiga kebun rakyat dengan periode pengamatan tahun 2020–2024. Data yang digunakan mencakup variabel agronomi seperti luas lahan, umur tanaman, jenis bibit, kondisi tanah, dan status pemupukan, serta variabel produktivitas berupa total hasil panen, harga, dan pendapatan.

Secara keseluruhan, dataset terdiri dari 360 data yang disimpan dalam format spreadsheet (.xlsx) dan disusun secara terstruktur untuk mendukung proses analisis *machine learning*.

No	Tanggal Panen	Hasil Panen	Luas Lahan	Umur Tanaman (Tahun)	Jenis Bibit	Kondisi Tanah	Status Pemupukan	Total Hasil (kg)	Harga (Rp/kg)	Total Pendapatan (Rp)
1	20/01/2020	Kebun A	10	14	Tenaga	Sukur	Tidak Disuplai	5125	2020.26	Rp5.654.562.50
2	20/01/2020	Kebun A	10	14	Tenaga	Sukur	Tidak Disuplai	5670	2048.68	Rp11.613.853.84
3	10/02/2020	Kebun B	5	11	Daf	Setengah	Tidak Disuplai	2070	2099.86	Rp4.352.924.42
4	20/02/2020	Kebun B	5	11	Daf	Setengah	Disuplai	4876	2044.23	Rp10.045.575.25
5	10/03/2020	Kebun C	5	9	Tenaga	Kuning Sukur	Disuplai	2880	2055.52	Rp5.921.897.60
6	20/03/2020	Kebun C	5	9	Tenaga	Kuning Sukur	Disuplai	4062	2058.68	Rp8.358.955.16
7	10/04/2020	Kebun A	10	14.08	Tenaga	Sukur	Tidak Disuplai	12350	2075.94	Rp25.549.007.00
8	20/04/2020	Kebun A	10	14.08	Tenaga	Sukur	Tidak Disuplai	10893	2022.32	Rp22.032.504.62
9	10/05/2020	Kebun B	5	11.08	Daf	Setengah	Disuplai	6462	2057.42	Rp13.304.678.42
10	20/05/2020	Kebun B	5	11.08	Daf	Setengah	Tidak Disuplai	4554	2060.34	Rp9.394.546.58
11	10/06/2020	Kebun C	5	9.08	Tenaga	Kuning Sukur	Tidak Disuplai	2887	2093.84	Rp6.054.364.78
12	20/06/2020	Kebun C	5	9.08	Tenaga	Kuning Sukur	Disuplai	5066	2094.34	Rp10.610.077.98
13	10/07/2020	Kebun A	10	14.17	Tenaga	Sukur	Tidak Disuplai	10770	2055.08	Rp22.155.377.40
14	20/07/2020	Kebun A	10	14.17	Tenaga	Sukur	Tidak Disuplai	7771	2026.22	Rp15.839.849.62
15	10/08/2020	Kebun B	5	11.17	Daf	Setengah	Tidak Disuplai	2887	2062.77	Rp5.959.077.06
16	20/08/2020	Kebun B	5	11.17	Daf	Setengah	Disuplai	4930	2062.71	Rp10.182.798.30
17	10/09/2020	Kebun C	5	9.17	Tenaga	Kuning Sukur	Tidak Disuplai	4042	2064.66	Rp8.352.488.06
18	20/09/2020	Kebun C	5	9.17	Tenaga	Kuning Sukur	Tidak Disuplai	4000	2068.26	Rp8.281.864.48
19	10/10/2020	Kebun A	10	14.25	Tenaga	Sukur	Tidak Disuplai	8779	2066.34	Rp18.152.225.06
20	20/10/2020	Kebun A	10	14.25	Tenaga	Sukur	Disuplai	8430	2063.62	Rp17.404.773.46
21	10/11/2020	Kebun B	5	11.25	Daf	Setengah	Disuplai	4840	2047.07	Rp9.914.576.15
22	20/11/2020	Kebun B	5	11.25	Daf	Setengah	Disuplai	5528	2060.39	Rp11.495.552.42
23	10/12/2020	Kebun C	5	9.25	Tenaga	Kuning Sukur	Disuplai	8020	2088.10	Rp16.740.222.80
24	20/12/2020	Kebun C	5	9.25	Tenaga	Kuning Sukur	Disuplai	7943	2099.35	Rp16.669.462.35
25	10/01/2021	Kebun A	10	14.33	Tenaga	Sukur	Tidak Disuplai	7702	2081.52	Rp16.251.467.04

Gambar 2. Dataset Penelitian

Dataset ini selanjutnya digunakan sebagai input dalam proses analisis untuk mengidentifikasi faktor-faktor yang memengaruhi produktivitas kelapa sawit serta membangun model prediksi berbasis data.

4.2. Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam pemodelan *machine learning*. Proses ini meliputi pembersihan data, perbaikan format, serta pemeriksaan kelengkapan data agar dataset siap diolah secara optimal.

Pada penelitian ini, fokus utama pra-pemrosesan adalah normalisasi format pada kolom *Tanggal Panen*. Ditemukan adanya variasi format penulisan tanggal serta data dalam bentuk angka (serial Excel), sehingga dilakukan proses konversi dan standarisasi ke dalam format *datetime* yang konsisten.

Tanggal Panen	Nama Kebun	Luas (Ha)	Umur Tanaman (Tahun)	Jenis Bibit	Kondisi Tanah	Status Pemupukan	Total Hasil (kg)	Harga (Rp/kg)	Total Pendapatan (Rp)	
9	2025-01-05	Kebun A	10	14.0	Terena	Sulur	Total Dipupuk	5185	2596.55	14905059.85
1	2025-01-20	Kebun A	10	14.0	Terena	Sulur	Total Dipupuk	6673	3546.68	17203915.64
2	2025-01-10	Kebun B	5	11.0	DuP	Sedang	Total Dipupuk	2979	3095.98	9222924.42
3	2025-01-25	Kebun B	5	11.0	DuP	Sedang	Dipupuk	4675	2940.23	13745675.25
4	2025-01-15	Kebun C	5	9.0	Terena	Kurang Sulur	Dipupuk	2888	2866.52	8273897.68

Gambar 3. Perbaikan Format Kolom Tanggal

Hasil perbaikan memperlihatkan data tanggal yang telah seragam dan siap digunakan dalam analisis berbasis waktu.

Selanjutnya, dilakukan pemeriksaan *missing value* pada seluruh kolom dataset untuk memastikan tidak terdapat data yang hilang.

	Kolom	Jumlah Missing
0	Tanggal Panen	0
1	Nama Kebun	0
2	Luas (Ha)	0
3	Umur Tanaman (Tahun)	0
4	Jenis Bibit	0
5	Kondisi Tanah	0
6	Status Pemupukan	0
7	Total Hasil (kg)	0
8	Harga (Rp/kg)	0
9	Total Pendapatan (Rp)	0

Gambar 4. Pemeriksaan Missing Value

Hasil menunjukkan bahwa seluruh variabel memiliki nilai nol, yang berarti dataset tidak memiliki *missing value* dan tidak memerlukan proses penanganan tambahan.

Dengan demikian, data yang telah melalui tahap pra-pemrosesan memiliki kualitas yang baik dan siap digunakan pada tahap pembangunan model *machine learning*.

4.3. Pembangunan Model Ensemble

Setelah data melalui tahap pra-pemrosesan, langkah selanjutnya adalah pembangunan model *machine learning* menggunakan pendekatan *ensemble*.

Metode ini dipilih karena mampu menangani hubungan non-linear dan interaksi kompleks antar variabel agronomi sehingga menghasilkan prediksi yang lebih stabil dan akurat. Proses pembangunan model meliputi penentuan variabel, pembagian data, dan pelatihan model.

4.4. Penentuan Variabel Input dan Target

Pada tahap ini dilakukan pemisahan variabel input (X) dan variabel target (y). Variabel input terdiri dari luas lahan, umur tanaman, jenis bibit, kondisi tanah, status pemupukan, dan bulan, sedangkan variabel target adalah total hasil panen (kg).

=== (X) Input Data ===									
Luas Lahan (Ha)	Umur Tanaman (Tahun)	Jenis Bibit	Kondisi Tanah	Status Pemupukan	Total Hasil (kg)	Harga (Rp/kg)	Total Pendapatan (Rp)		
10	14.0	Terena	Sulur	Total Dipupuk	5185	2596.55	14905059.85	87	
10	14.0	Terena	Sulur	Total Dipupuk	6673	3546.68	17203915.64	085	
5	11.0	DuP	Sedang	Total Dipupuk	2979	3095.98	9222924.42	285	
5	11.0	DuP	Sedang	Dipupuk	4675	2940.23	13745675.25	835	
5	9.0	Terena	Kurang Sulur	Dipupuk	2888	2866.52	8273897.68	881	
5	11.0	DuP	Sedang	Total Dipupuk	2979	3095.98	9222924.42	28	
5	11.0	DuP	Sedang	Total Dipupuk	2979	3095.98	9222924.42	14	
5	11.0	DuP	Sedang	Total Dipupuk	2979	3095.98	9222924.42	25	

Gambar 5. variabel input (x)

Data input (X) yang memiliki variasi kondisi agronomi pada setiap kebun. Variasi ini penting agar model dapat mempelajari pola hubungan antar variabel.

=== Contoh Data Target (y) ===	
Total Hasil (kg)	
345	4299
311	1553
156	12697
56	6178
235	8039
26	2037
355	9306
241	14535

Gambar 6. Variabel input (y)

Selanjutnya, data target (y) yang menunjukkan variasi hasil panen dengan rentang yang cukup besar. Hal ini menunjukkan adanya perbedaan produktivitas antar kebun yang dipengaruhi oleh kondisi agronomi.

4.5. Pembagian Data Latih dan Data Uji

Dataset kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Dari total 360 data, diperoleh 288 data latih dan 72 data uji.

Jumlah data latih: 288
Jumlah data uji: 72

Gambar 7. Pembagian Data uji dan Data Latih

Pembagian ini bertujuan agar model dapat dilatih menggunakan sebagian besar data, serta diuji menggunakan data yang belum pernah dilihat

sebelumnya. Dengan demikian, kemampuan model dalam melakukan generalisasi dapat dievaluasi secara lebih objektif.

4.6. Pelatihan Model *Ensemble*

Tahap selanjutnya adalah pelatihan model menggunakan algoritma Random Forest Regressor. Model ini bekerja dengan menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi yang lebih stabil.

```
*** Model berhasil dilatih
```

Gambar 8. Pelatihan Model *Ensemble*

Model berhasil dilatih menggunakan data latih dan siap digunakan untuk tahap evaluasi. Dalam penelitian ini, model menggunakan 100 pohon keputusan untuk mempelajari hubungan antara variabel agronomi dan hasil panen.

Model yang telah dilatih ini diharapkan mampu mengenali pola dalam data dan menghasilkan prediksi yang akurat pada data uji.

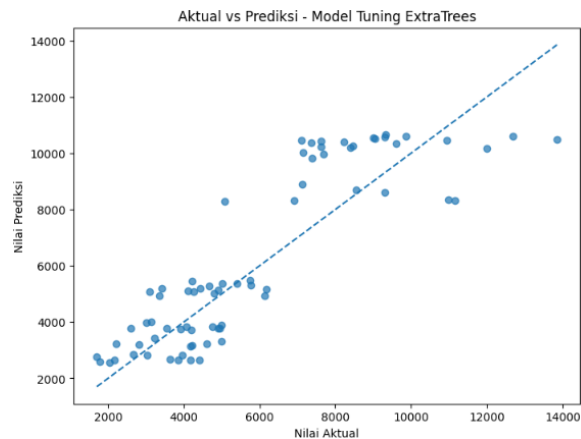
4.7. Evaluasi Model

Tahap evaluasi dilakukan untuk mengukur kinerja model Random Forest dalam memprediksi hasil panen kelapa sawit dengan membandingkan nilai prediksi dan nilai aktual pada data uji. Evaluasi menggunakan metrik R^2 , Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE), serta didukung dengan visualisasi scatter plot untuk melihat kesesuaian antara nilai aktual dan prediksi.

```
=== HASIL EVALUASI MODEL AWAL ===
R2 : 0.6078
MAE : 1456.23 kg
RMSE : 1770.67 kg
Rata-rata hasil panen : 5856.08 kg
MAE (%) terhadap rata-rata hasil panen : 24.87%
Selisih RMSE - MAE : 314.44 kg
```

Gambar 9. Evaluasi Model

Evaluasi model nilai R^2 sebesar 0,6078 yang menunjukkan bahwa model mampu menjelaskan sekitar 60,78% variasi data. Nilai MAE sebesar 1456,23 kg dibandingkan dengan rata-rata hasil panen sebesar 5856,08 kg menghasilkan tingkat kesalahan sebesar 24,87%, yang menunjukkan bahwa akurasi model masih belum optimal. Selain itu, nilai RMSE sebesar 1770,67 kg lebih besar dibandingkan MAE dengan selisih 314,44 kg yang menunjukkan adanya beberapa prediksi dengan error yang cukup tinggi.



Gambar 10. Visualisasi Scatter Plot

Hal ini mengindikasikan adanya beberapa prediksi dengan error yang cukup tinggi, sehingga model belum sepenuhnya stabil. Berdasarkan hasil tersebut, dapat disimpulkan bahwa performa model masih perlu ditingkatkan melalui proses *tuning* pada tahap selanjutnya.

Visualisasi hubungan antara nilai aktual dan prediksi ditunjukkan melalui scatter plot. Titik-titik yang menyebar di sekitar garis diagonal menunjukkan bahwa model telah mampu mengikuti pola data, namun masih terdapat penyimpangan pada beberapa titik yang cukup jauh dari garis ideal. Hal ini menandakan bahwa model belum sepenuhnya stabil dalam melakukan prediksi.

4.8. *Tuning* Model

Berdasarkan hasil evaluasi sebelumnya, model masih memiliki tingkat kesalahan yang cukup tinggi sehingga dilakukan proses *tuning* untuk meningkatkan performa. *Tuning* dilakukan menggunakan metode GridSearchCV guna memperoleh kombinasi parameter terbaik pada model Random Forest.

```
=== HASIL EVALUASI SETELAH TUNING ===
R2 : 0.7065
MAE : 1266.76 kg
RMSE : 1531.79 kg
MAE (%) terhadap rata-rata hasil panen : 21.63%
Selisih RMSE - MAE : 265.04 kg
```

Gambar 11. Evaluasi Setelah *Tuning*

Evaluasi setelah *tuning* nilai R^2 sebesar 0,7065, MAE sebesar 1266,76 kg, dan RMSE sebesar 1531,79 kg. Nilai MAE (%) terhadap rata-rata hasil panen juga menurun menjadi 21,63%, yang menunjukkan adanya penurunan tingkat kesalahan prediksi.

Perbandingan kinerja model sebelum dan sesudah *tuning* pada Gambar 4.10 menunjukkan adanya peningkatan performa yang signifikan. Nilai R^2 meningkat dari 0,6078 menjadi 0,7065, sedangkan MAE dan RMSE mengalami penurunan masing-masing dari 1456,23 kg menjadi 1266,76 kg dan dari 1770,67 kg menjadi 1531,79 kg. Selain itu, selisih antara RMSE dan MAE juga menurun, yang mengindikasikan bahwa model menjadi lebih stabil.

4.9. Analisis Feature Importance

Setelah model optimal diperoleh, dilakukan analisis *feature importance* untuk mengidentifikasi variabel yang paling berpengaruh terhadap prediksi produktivitas kelapa sawit.

	Fitur	Importance
11	Kondisi Tanah_Subur	0.226294
0	Luas (Ha)	0.193628
4	Nama Kebun_Kebun A	0.159949
13	Status Pemupukan_Tidak Dipupuk	0.074997
6	Nama Kebun_Kebun C	0.069123
12	Status Pemupukan_Dipupuk	0.067013
9	Kondisi Tanah_Kurang Subur	0.055937
1	Umur Tanaman (Tahun)	0.036998
5	Nama Kebun_Kebun B	0.030217
10	Kondisi Tanah_Sedang	0.028030
8	Jenis Bibit_Tenera	0.022804
7	Jenis Bibit_DxP	0.018917
2	Harga (Rp/kg)	0.009352
3	Bulan	0.006742

Gambar 12. Analisis Feature Importance

Berdasarkan hasil tersebut, variabel yang paling dominan adalah kondisi tanah subur (0,226), diikuti oleh luas lahan (0,193) dan perbedaan kebun (0,159). Hal ini menunjukkan bahwa faktor kesuburan tanah dan skala lahan memiliki peran utama dalam menentukan hasil panen.

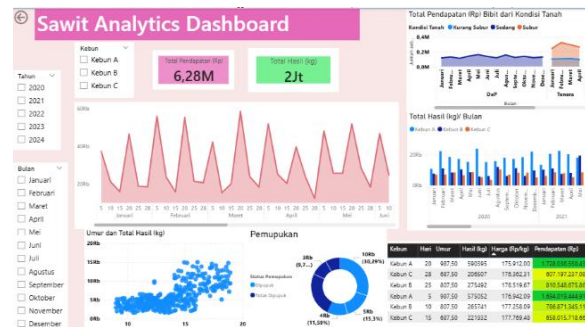
Variabel lain seperti status pemupukan, kondisi tanah kurang subur, dan umur tanaman juga memberikan kontribusi yang cukup signifikan terhadap model. Sementara itu, variabel seperti jenis bibit, harga, dan bulan memiliki pengaruh yang relatif lebih rendah.

Secara keseluruhan, hasil ini menunjukkan bahwa faktor agronomi utama, khususnya kondisi tanah dan luas lahan, menjadi penentu dominan produktivitas kelapa sawit. Temuan ini juga memperkuat bahwa model tidak hanya mampu melakukan prediksi, tetapi juga memberikan insight yang relevan untuk mendukung pengambilan keputusan berbasis data.

4.10. Integrasi Dashboard Power BI

Hasil pemodelan *machine learning* dan analisis *feature importance* diimplementasikan dalam dashboard interaktif menggunakan Power BI untuk mendukung interpretasi dan pengambilan keputusan.

Dashboard "Sawit Analytics" menampilkan ringkasan kinerja utama berupa total pendapatan sebesar 6,28M dan total hasil produksi sebesar 2 juta kg.



Gambar 13. Dashboard Power BI

Visualisasi yang disajikan meliputi tren hasil panen, perbandingan produktivitas antar kebun, serta pendapatan berdasarkan kondisi tanah, yang menunjukkan bahwa tanah subur menghasilkan performa lebih tinggi. Selain itu, hubungan antara umur tanaman dan hasil panen, serta distribusi pemupukan juga divisualisasikan untuk memberikan insight agronomi yang lebih jelas.

Fitur filter seperti tahun, bulan, dan kebun memungkinkan eksplorasi data secara dinamis. Secara keseluruhan, dashboard ini mampu merepresentasikan hasil analisis model dan mempermudah pemahaman informasi sehingga dapat digunakan sebagai alat pendukung keputusan dalam meningkatkan produktivitas kelapa sawit.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun *framework data intelligence* berbasis *machine learning* (Random Forest) untuk memprediksi produktivitas kelapa sawit, dengan kondisi tanah dan luas lahan sebagai faktor dominan. Hasil *tuning* meningkatkan performa model (R^2 dari 0,6078 menjadi 0,7065 serta penurunan MAE dan RMSE), menunjukkan model lebih akurat dan stabil. Hasil analisis juga diimplementasikan dalam dashboard Power BI yang interaktif untuk mendukung pengambilan keputusan berbasis data. Ke depan, penelitian dapat dikembangkan dengan penambahan variabel dan data yang lebih luas, serta pengujian metode lain, sementara dashboard dapat dimanfaatkan sebagai alat bantu pengelolaan kebun secara optimal.

DAFTAR PUSTAKA

- [1] R. K dan B. V, "Optimizing Crop Yield Prediction through Multiple Models: An Ensemble Stacking Approach," International Journal of Data Informatics and Intelligent Computing, vol. 3, no. 2, p. 52–58, 2024.
- [2] S. Z. Yumni dan Widowati, "Implementasi Microsoft Power BI Dalam Memantau Kehadiran dan Transportasi Pegawai," Jurnal Sains dan Edukasi Sains, vol. 4, no. 1, pp. 1-8, 2021.
- [3] M. A. Javed dan M. A. A. Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," Heliyon, vol. 10, p. e40836, 2024.

-
- [4] D. Sharma, "Artificial Intelligence in Agriculture," *Just Agriculture*, vol. 2, no. 4, pp. 1-4, 2021.
- [5] M. H. Zainudin, M. A. M. Razman, M. F. Abdul Aziz dan N. Nawi, "Comparison of accuracy performance between GRU, LSTM, and BiLSTM for stock price prediction," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, pp. 146-152, 2021.
- [6] E. J. Jamshidi, Y. Yusup, C. W. Hooy, M. A. Kamaruddin, H. M. Hassan, S. A. Muhammad, H. Z. M. Shafri, K. H. Then, M. S. Norizan dan C. C. Tan, "Predicting oil palm yield using a comprehensive agronomy dataset and 17 machine learning and deep learning models," *Ecological Informatics*, vol. 81, p. 102595, 2024.
- [7] K. V. Sharma, "Data Mining and Business Analytics," CVR College of Engineering, Hyderabad, 2024.
- [8] G. N. Elwirehardja, T. Suparyanto, Miftakhurrokhmat dan B. Pardamean, "Determining Variables Associated with Annual Oil Palm Yield: An Explainable Gradient Boosting Approach," *Procedia Computer Science*, vol. 227, p. 262–271, 2023.
- [9] P. Mahesh dan R. Soundrapandiyan, "Yield prediction for crops by gradient-based algorithms," *PLOS ONE*, vol. 19, no. 8, p. 1–20, 2024.
- [10] B. Rozemberczki, L. Watson, P. Bayer, H.-T. Yang, O. Kiss, S. Nilsson dan R. Sarkar, "The Shapley Value in Machine Learning," *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2022)*, vol. Leave blank, pp. 5572-5579, 2022.