

Optimasi Prediksi Harga Sawit Menggunakan Teknik Stacking Algoritma Machine Learning dan Deep Learning dengan SMOTE

Abdul Karim*, Budianto Bangun, Sugeng Prayetno, Mohammad Afrendi*

Fakultas Sains dan Teknologi, Program Studi Teknologi Informasi, Universitas Labuhanbatu, Rantauprapat, Indonesia

Email: ^{1,*}abdkarim6@gmail.com@email.com, ²budiantobangun44@email.com, ³sprayetno835@gmail.com,

⁴afrendi220924@gmail.com

Email Penulis Korespondensi: abdkarim6@gmail.com@email.com

Submitted: 30/04/2025; Accepted: 25/06/2025; Published: 25/06/2025

Abstrak—Prediksi harga kelapa sawit memiliki peran strategis dalam pengambilan keputusan di sektor agribisnis, khususnya dalam menghadapi volatilitas pasar dan ketidakseimbangan distribusi data historis. Penelitian ini bertujuan untuk mengoptimalkan akurasi prediksi harga sawit dengan menerapkan pendekatan stacking yang menggabungkan algoritma machine learning dan deep learning, serta mengintegrasikan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk menangani permasalahan ketidakseimbangan kelas pada data. Tiga model utama yang digunakan dalam studi ini adalah Random Forest, Long Short-Term Memory (LSTM), dan model dengan integrasi SMOTE. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score, serta didukung oleh analisis confusion matrix. Hasil penelitian menunjukkan bahwa model dengan integrasi SMOTE memiliki performa terbaik secara keseluruhan dengan nilai accuracy sebesar 0,5447, precision 0,5512, recall 0,5447, dan F1-score 0,5462. Model ini juga menunjukkan distribusi klasifikasi yang lebih seimbang dibandingkan dua model lainnya. Temuan ini menegaskan bahwa penerapan teknik oversampling seperti SMOTE, apabila dikombinasikan dengan algoritma yang tepat, dapat meningkatkan kinerja prediktif secara signifikan pada data yang tidak seimbang. Penelitian ini memberikan kontribusi dalam pengembangan model prediksi harga komoditas berbasis data historis dan membuka ruang bagi eksplorasi metode hybrid yang lebih adaptif di masa depan.

Kata Kunci: Harga Sawit; LSTM; Random Fores; SMOTE; Stacking

Abstract—The prediction of palm oil prices plays a strategic role in decision-making within the agribusiness sector, particularly in addressing market volatility and imbalanced historical data distribution. This study aims to optimize the accuracy of palm oil price prediction by applying a stacking approach that combines machine learning and deep learning algorithms, while integrating the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance issues. Three main models were employed in this study: Random Forest, Long Short-Term Memory (LSTM), and a model enhanced with SMOTE. The evaluation was conducted using accuracy, precision, recall, and F1-score metrics, supported by confusion matrix analysis. The results indicate that the model integrated with SMOTE outperforms the others, achieving an accuracy of 0.5447, precision of 0.5512, recall of 0.5447, and F1-score of 0.5462. This model also demonstrates a more balanced classification performance compared to the LSTM and Random Forest models. These findings confirm that the application of oversampling techniques such as SMOTE, when combined with appropriate algorithms, can significantly enhance predictive performance in imbalanced datasets. The study contributes to the development of predictive models for commodity prices based on historical data and opens opportunities for further exploration of more adaptive hybrid methods in future research.

Keywords: Palm Oil Price; LSTM; Random Forest; SMOTE; Stacking

1. PENDAHULUAN

Kelapa sawit merupakan salah satu komoditas unggulan Indonesia yang berperan strategis dalam perekonomian nasional maupun global, baik sebagai penyumbang devisa negara melalui ekspor maupun sebagai penyedia lapangan kerja bagi jutaan tenaga kerja di sektor agribisnis. Namun demikian, harga kelapa sawit di pasar global cenderung sangat fluktuatif, dipengaruhi oleh berbagai faktor seperti permintaan internasional, kebijakan perdagangan, kondisi iklim, hingga spekulasi pasar. Ketidakstabilan ini menjadi tantangan besar bagi para pemangku kepentingan, khususnya dalam pengambilan keputusan strategis terkait produksi, distribusi, dan investasi. Oleh karena itu, ketersediaan model prediksi harga kelapa sawit yang akurat dan adaptif menjadi kebutuhan mendesak untuk memitigasi risiko ekonomi yang ditimbulkan oleh volatilitas tersebut [1]

Dalam beberapa tahun terakhir, penerapan Machine Learning (ML) dan Deep Learning (DL) telah berkembang pesat sebagai pendekatan untuk prediksi harga komoditas. Namun, tantangan utama yang masih sering dihadapi adalah ketidakseimbangan distribusi data historis, di mana kelas-kelas tertentu (misalnya tren harga turun) jauh lebih sedikit dibandingkan kelas lainnya (tren harga naik). [2][3]. Ketidakseimbangan ini menyebabkan model prediktif cenderung bias terhadap kelas mayoritas, sehingga akurasi dalam memprediksi peristiwa yang jarang terjadi menjadi sangat rendah. Selain itu, sebagian besar penelitian sebelumnya masih mengandalkan pendekatan algoritma tunggal, seperti Random Forest atau LSTM secara terpisah, yang memiliki keterbatasan dalam menangkap kompleksitas non-linier dan temporal data. [4], [5]

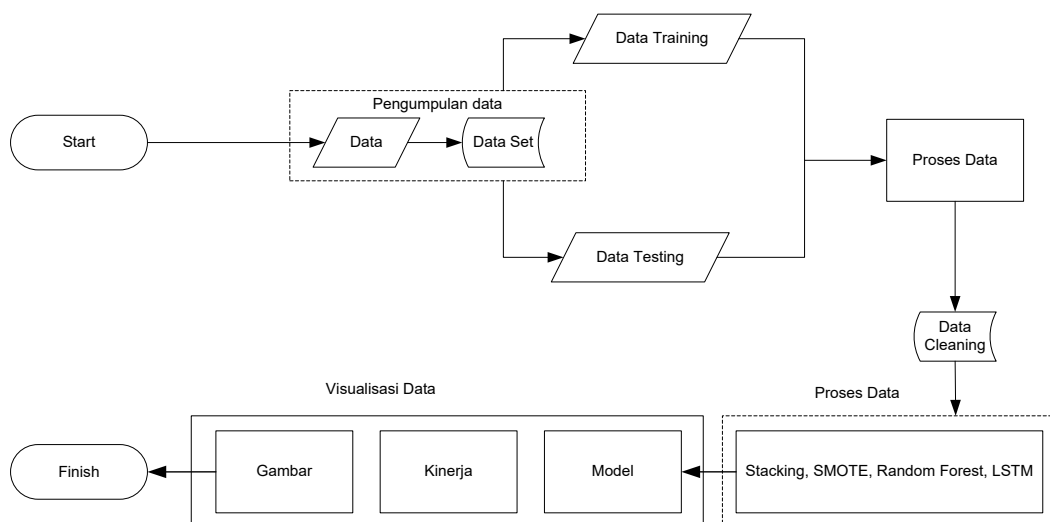
Meskipun penelitian terkait prediksi harga komoditas berbasis Machine Learning (ML) dan Deep Learning (DL) sudah cukup berkembang, masih terdapat beberapa kelemahan yang perlu diperbaiki. Salah satu masalah utama adalah ketidakseimbangan data (imbalanced data), di mana model cenderung mengabaikan pola penting pada data minoritas yang dapat memengaruhi hasil prediksi. Selain itu, sebagian besar penelitian sebelumnya menggunakan pendekatan berbasis algoritma tunggal, yang sering kali kurang optimal dalam menangkap kompleksitas data. Teknik Stacking, yang menggabungkan kekuatan beberapa algoritma prediktif melalui pendekatan ensemble, telah terbukti

efektif meningkatkan akurasi di berbagai aplikasi, seperti prediksi pasar saham dan komoditas lainnya[2][3]. Namun, penerapan teknik Stacking dalam konteks prediksi harga kelapa sawit masih sangat terbatas. Selain itu, metode Synthetic Minority Oversampling Technique (SMOTE) [4], [5] untuk menangani ketidakseimbangan data juga belum dimanfaatkan secara maksimal, meskipun telah menunjukkan hasil yang baik pada berbagai domain penelitian lainnya

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan integrasi dua pendekatan: (1) teknik ensemble berbasis stacking, yang menggabungkan keunggulan beberapa model prediktif dalam satu kerangka kerja, serta (2) teknik oversampling SMOTE (Synthetic Minority Over-sampling Technique) untuk menyeimbangkan distribusi kelas dalam data pelatihan. Meskipun keduanya telah terbukti efektif dalam studi terpisah, penerapan keduanya secara simultan dalam konteks prediksi harga kelapa sawit masih sangat terbatas dalam literatur. Dengan mengombinasikan Random Forest, LSTM, dan SMOTE dalam kerangka stacking, penelitian ini bertujuan untuk meningkatkan akurasi prediksi secara signifikan, sekaligus memberikan kontribusi praktis bagi pengambilan keputusan berbasis data di industri kelapa sawit.

2. METODOLOGI PENELITIAN

Berikut cara yang dilakukan untuk mencapai tujuan yang telah ditetapkan secara details dan berdasarkan kegiatan selama satu tahun yang dibuat dalam diagram alir pada Gambar 2.



Gambar 2. Diagram alir Penelitian

Keterangan:

- a. Start
Proses dimulai dengan langkah awal pengolahan data untuk keperluan analisis atau pemodelan.
- b. Pengumpulan Data
Data yang diperlukan dikumpulkan dari berbagai sumber. Data yang terkumpul disimpan dalam bentuk dataset, yang siap untuk diproses lebih lanjut.
- c. Data Training dan Data Testing
Dataset yang terkumpul dibagi menjadi dua bagian:
 1. Data Training: Digunakan untuk melatih model.
 2. Data Testing: Digunakan untuk menguji performa model.
- d. Proses Data
Sebelum model dibuat, data perlu melalui tahap pemrosesan:
Data Cleaning: Langkah ini melibatkan pembersihan data dari nilai yang hilang, duplikasi, atau ketidakkonsistenan agar data lebih siap untuk analisis.
- e. Pemodelan dengan Algoritma
Data yang sudah bersih kemudian diproses menggunakan algoritma Stacking, SMOTE, Long Short-Term Memory (LSTM), Random Forest
- f. Visualisasi Data
Setelah model selesai dibuat, hasilnya divisualisasikan untuk mempermudah interpretasi:
Gambar: Menampilkan hasil dalam bentuk visual, seperti grafik atau diagram.
Kinerja: Mengukur performa model berdasarkan metrik evaluasi seperti akurasi, presisi, atau recall.
Model: Visualisasi struktur model, seperti pohon keputusan atau aturan klasifikasi.
- g. Finish
Proses selesai setelah hasil analisis atau pemodelan data telah divisualisasikan dan siap untuk disimpulkan atau digunakan lebih lanjut.

Proses ini menggambarkan alur kerja yang terstruktur mulai dari pengumpulan data hingga evaluasi model. Dengan adanya visualisasi, hasil akhir menjadi lebih mudah dipahami dan dapat digunakan untuk pengambilan keputusan.

Validasi hasil penelitian dilakukan dengan membandingkan model prediksi harga kelapa sawit yang dikembangkan menggunakan teknik Stacking Machine Learning & Deep Learning dengan SMOTE terhadap beberapa baseline model. Tujuan validasi ini adalah untuk memastikan bahwa pendekatan yang diusulkan memberikan peningkatan akurasi dan keandalan prediksi dibandingkan dengan metode konvensional. Penelitian tidak hanya mengukur akurasi model, tetapi juga memastikan bahwa model yang dikembangkan memberikan peningkatan signifikan dibandingkan metode baseline. Teknik Stacking dan SMOTE diharapkan mampu memberikan hasil yang lebih akurat, stabil, dan siap diimplementasikan dalam industri kelapa sawit.

2.1 Machine Learning

Machine learning [6] adalah cabang dari kecerdasan buatan [7] yang memungkinkan sistem komputer untuk belajar dari data dan membuat keputusan tanpa instruksi eksplisit. Ini mencakup berbagai teknik untuk mengenali pola dan memecahkan masalah di berbagai bidang. [8], [9], [10], [11], [12], [13].

Machine learning (ML) adalah cabang dari kecerdasan buatan (AI) yang memungkinkan komputer untuk belajar dari data dan membuat keputusan atau prediksi tanpa diprogram secara eksplisit. Dalam machine learning, algoritma digunakan untuk menganalisis data, mengenali pola, dan kemudian membuat keputusan atau prediksi berdasarkan data tersebut. Tujuan utama dari machine learning adalah untuk membuat sistem yang dapat belajar dan berkembang seiring waktu dengan memberikan data dan pengalaman baru, tanpa memerlukan pemrograman manual untuk setiap skenario.

2.2 Stacking Algoritma Machine Learning

Stacking (atau Stacked Generalization) [14], [15], [16] adalah teknik ensemble dalam machine learning yang menggabungkan beberapa model prediktif untuk meningkatkan akurasi dan kinerja prediksi. Teknik ini melibatkan pelatihan beberapa model dasar (base models) dan kemudian menggunakan model lain (sering disebut sebagai meta-model atau blender) untuk menggabungkan prediksi dari model-model dasar tersebut. [17], [18], [19], [20].

2.3 Deep Learning

Deep learning [21], [22] adalah subbidang dari machine learning yang berfokus pada penggunaan jaringan saraf tiruan dengan banyak lapisan (deep neural networks) untuk menganalisis dan memproses data. Konsep dasar dari deep learning berasal dari cara kerja otak manusia, di mana neuron-neuron saling terhubung dan berkomunikasi untuk memproses informasi. Dalam konteks komputer, jaringan saraf tiruan terdiri dari lapisan input, lapisan tersembunyi, dan lapisan output. Setiap lapisan terdiri dari sejumlah neuron yang saling terhubung, dan setiap koneksi memiliki bobot yang dapat disesuaikan selama proses pelatihan. [23], [24], [25].

2.4 SMOTE

SMOTE, atau Synthetic Minority Over-sampling Technique [25], [26], adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam dataset, yang sering kali menjadi tantangan dalam analisis data dan machine learning. Ketidakseimbangan kelas terjadi ketika jumlah contoh dari satu kelas jauh lebih sedikit dibandingkan dengan kelas lainnya, yang dapat menyebabkan model machine learning cenderung bias terhadap kelas mayoritas. Dalam konteks ini, SMOTE berfungsi untuk meningkatkan jumlah contoh dari kelas minoritas dengan cara yang cerdas dan efektif [27].

Proses SMOTE dimulai dengan mengidentifikasi contoh dari kelas minoritas dalam dataset. Setelah contoh tersebut diidentifikasi, SMOTE menciptakan contoh sintetis baru dengan cara mengambil dua langkah. Pertama, untuk setiap contoh dari kelas minoritas, SMOTE mencari k tetangga terdekat (nearest neighbors) dalam kelas yang sama. Kemudian, SMOTE menghasilkan contoh baru dengan cara menginterpolasi antara contoh yang ada dan salah satu tetangga terdekat yang dipilih secara acak. Proses ini melibatkan pengambilan nilai acak antara fitur dari contoh yang ada dan tetangga terdekat, sehingga menghasilkan contoh baru yang berada di antara keduanya dalam ruang fitur [25], [28], [29].

Keunggulan utama dari SMOTE adalah bahwa ia tidak hanya menambah jumlah contoh dari kelas minoritas, tetapi juga menciptakan variasi dalam data, yang membantu model machine learning belajar lebih baik tentang pola yang ada dalam kelas minoritas. Dengan cara ini, SMOTE dapat meningkatkan kinerja model dalam memprediksi kelas minoritas, yang sering kali diabaikan dalam dataset yang tidak seimbang. Selain itu, SMOTE dapat digunakan dalam berbagai aplikasi, termasuk deteksi penipuan, diagnosis medis, dan analisis sentimen, di mana kelas minoritas sering kali sangat penting tetapi sulit untuk diidentifikasi.

Namun, meskipun SMOTE memiliki banyak manfaat, ada beberapa tantangan yang perlu diperhatikan. Salah satunya adalah risiko overfitting, di mana model dapat belajar terlalu banyak dari contoh sintetis yang dihasilkan, sehingga mengurangi kemampuannya untuk generalisasi pada data baru. Selain itu, SMOTE dapat memperkenalkan noise jika tidak diterapkan dengan hati-hati, terutama jika data asli memiliki outlier. Oleh karena itu, penting untuk melakukan evaluasi yang cermat dan mempertimbangkan teknik lain, seperti pengurangan dimensi atau kombinasi

dengan metode lain, untuk mencapai hasil yang optimal. Secara keseluruhan, SMOTE merupakan alat yang sangat berguna dalam menangani ketidakseimbangan kelas dan meningkatkan kinerja model machine learning dalam berbagai konteks

3. HASIL DAN PEMBAHASAN

Data yang dikumpulkan dari data historis harga kelapa sawit serta data cuaca dari BMKG dalam jangka waktu lima tahun dengan jumlah data 1226 data. Data yang digunakan dalam penelitian ini dapat di lihat dari Tabel 1.

Tabel 1 Data Awal Penelitian

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Curah Hujan	Perubahan%	Hasil
02/01/2020	3.125,00	3.060,00	3.125,00	3.046,00	0,07K	0	2,76%	Naik
03/01/2020	3.110,00	3.129,00	3.129,00	3.100,00	0,06K	0	-0,48%	Turun
06/01/2020	3.036,00	3.060,00	3.060,00	3.060,00	0,02K	0	-2,38%	Turun
07/01/2020	3.038,00	3.061,00	3.090,00	3.040,00	0,07K	0	0,07%	Naik
08/01/2020	3.038,00	3.043,00	3.044,00	3.043,00	0,02K	0	0,00%	Naik
....
31/12/2024	4.861,00	4.912,00	4.920,00	4.800,00	1,31K	0,1	-1,30%	Turun

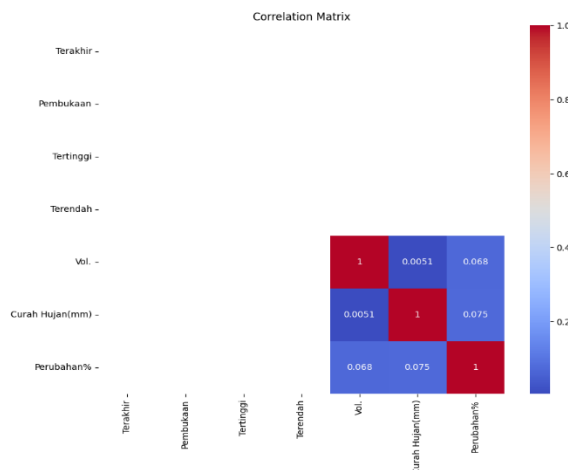
3.1. Eklorasi data

Tabel 1 merupakan tabel yang telah bersih dan sesuai format sehingga data bisa dilakukan dengan cara di eklorasi data untuk memahami struktur dataset, termasuk bentuk data, tipe variabel, keberadaan data yang hilang, statistik deskriptif, keunikan nilai, serta distribusi dari fitur numerik. Selain itu, juga akan menganalisis korelasi antar variabel dan mengidentifikasi pola atau tren berdasarkan aspek temporal

Tabel 2. Hasil Eklorasi data

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Curah Hujan(mm)	Perubahan%	Hasil
02/01/2020	NaN	NaN	NaN	NaN	0.07	0.0	2.76	Naik
03/01/2020	NaN	NaN	NaN	NaN	0.06	0.0	0.48	Turun
06/01/2020	NaN	NaN	NaN	NaN	0.02	0.0	2.38	Turun
07/01/2020	NaN	NaN	NaN	NaN	0.07	0.0	0.07	Naik
08/01/2020	NaN	NaN	NaN	NaN	0.02	0.0	0.00	Naik

Dataset yang digunakan terdiri dari beberapa variabel, antara lain "Tanggal", "Terakhir", "Pembukaan", "Tertinggi", "Terendah", "Volume", "Curah Hujan (mm)", "Perubahan%", dan "Hasil". Hasil inspeksi awal menunjukkan bahwa sebagian besar nilai pada kolom harga pasar, seperti "Terakhir", "Pembukaan", "Tertinggi", dan "Terendah", memiliki banyak nilai yang hilang (missing values). Sementara itu, variabel "Volume", "Curah Hujan (mm)", dan "Perubahan%" memiliki data yang lengkap dan siap untuk dianalisis lebih lanjut. Distribusi nilai numerik pada ketiga variabel tersebut menunjukkan variasi yang rendah, khususnya pada curah hujan yang cenderung konstan di angka nol. Selain itu, dilakukan analisis korelasi antar variabel numerik untuk melihat keterkaitan linear. Hasil visualisasi matriks korelasi menunjukkan bahwa hubungan antar variabel sangat lemah, dengan nilai korelasi tertinggi hanya sebesar 0,075. Temuan ini mengindikasikan bahwa tidak terdapat hubungan linear yang kuat di antara variabel-variabel numerik yang tersedia. Oleh karena itu, dibutuhkan pendekatan analitik lanjutan, seperti pemodelan non-linear atau analisis berbasis waktu, untuk mengidentifikasi pola tersembunyi yang mungkin tidak terdeteksi melalui korelasi linier semata.



Gambar 2. Korelasi Matrik

Gambar 2 menunjukkan matriks korelasi yang divisualisasikan dalam bentuk heatmap untuk beberapa variabel numerik, yakni Volume (Vol.), Curah Hujan (mm), dan Perubahan Persentase (%). Korelasi ini mengukur hubungan linear antar variabel, dengan nilai antara -1 hingga 1. Dari visualisasi tersebut, tampak bahwa korelasi antara Volume dan Curah Hujan sangat lemah, hanya sebesar 0,0051, begitu pula antara Volume dan Perubahan% yang hanya sebesar 0,068. Sementara itu, korelasi antara Curah Hujan dan Perubahan% juga lemah dengan nilai 0,075. Hal ini menunjukkan bahwa tidak terdapat hubungan linear yang signifikan antara ketiga variabel tersebut. Sementara beberapa variabel lain seperti "Terakhir", "Pembukaan", "Tertinggi", dan "Terendah" tidak muncul dalam matriks, kemungkinan disebabkan oleh data yang tidak memenuhi syarat perhitungan korelasi, seperti keberadaan nilai non-numerik, nilai konstan, atau missing value. Berdasarkan temuan ini, dapat disimpulkan bahwa eksplorasi lebih lanjut dibutuhkan, misalnya melalui analisis non-linear atau pemodelan temporal, untuk mengetahui apakah terdapat pola atau hubungan lain yang lebih kompleks antar variabel. Setelah data di cening selanjutnya dilakukan tranning dan testing dengan data sebagai berikut:

X_train shape: (980, 7)

X_test shape: (246, 7)

y_train shape: (980,)

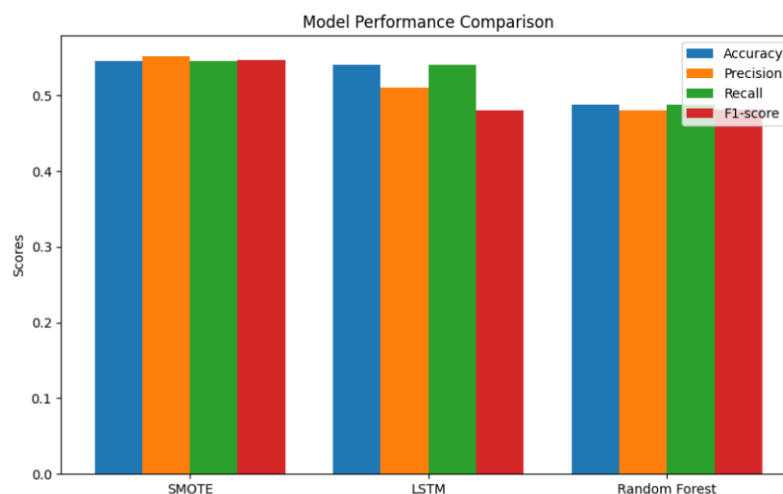
y_test shape: (246,)

Setelah data tranning dan testing di temukan selanjutnya dilakukan evaluasi model dalam hal ini menggunakan model SMOTE, LSTM, dan Random Forest yang telah dilatih pada data uji (X_test, y_test) dengan hasil pada Tabel 3.

Tabel 3. Hasil Pengujian Model

	SMOTE	LSTM	Random Forest
Accuracy	0.5447154471544715	0.540650406504065	0.4878048780487805
Precision	0.5511930729063389	0.5102869208303733	0.4795222398776802
Recall	0.5447154471544715	0.540650406504065	0.4878048780487805
F1-score	0.546194820038386	0.48049373854726274	0.48164944332953064
Confusion Matrix:	[[74 63] [49 60]]	[[115 22] [91 18]]	[[82 55] [71 38]]

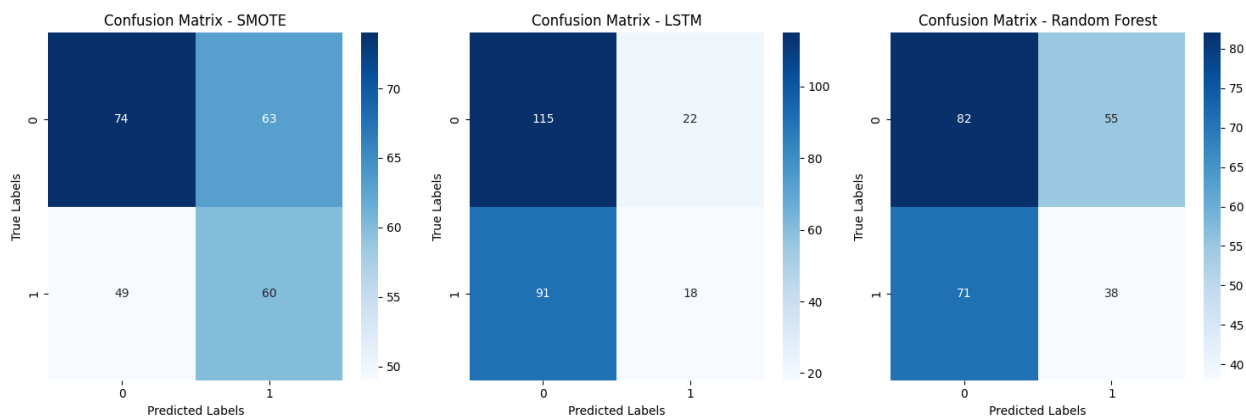
Setelah evaluasi model selanjutnya visualisasi model untuk melihat performa model mana yang lebih baik untuk perbandingan model dan tren data utama.



Gambar 3. Performa Perbandingan Model

Berdasarkan hasil visualisasi perbandingan performa model, dapat disimpulkan bahwa pendekatan yang mengintegrasikan teknik SMOTE menunjukkan kinerja paling unggul dibandingkan model lainnya, yaitu LSTM dan Random Forest, dalam konteks prediksi harga sawit. Hal ini ditunjukkan oleh skor metrik evaluasi yang relatif lebih tinggi dan seimbang, khususnya pada nilai akurasi, presisi, recall, dan F1-score, yang seluruhnya berada di atas angka 0.54. Penerapan SMOTE tampaknya berhasil meningkatkan kemampuan model dalam mengenali distribusi kelas minoritas, sehingga meningkatkan keseimbangan dalam prediksi.

Sementara itu, model LSTM menunjukkan akurasi dan recall yang cukup kompetitif, namun mengalami penurunan pada presisi dan F1-score, yang mengindikasikan adanya ketidakseimbangan antara prediksi positif dan aktual. Di sisi lain, Random Forest menempati posisi terbawah dalam keseluruhan metrik performa, dengan skor akurasi, presisi, recall, dan F1-score yang berada di kisaran 0.48 hingga 0.49. Temuan ini memperkuat argumen bahwa penggunaan teknik penyeimbangan data seperti SMOTE dapat memberikan kontribusi signifikan terhadap peningkatan akurasi model, terutama dalam skenario klasifikasi dengan distribusi kelas yang tidak seimbang.



Gambar 4. Confusion Matrix

Dalam penelitian ini, dilakukan evaluasi terhadap tiga pendekatan pemodelan dalam rangka prediksi harga kelapa sawit, yaitu model berbasis SMOTE (Synthetic Minority Oversampling Technique), LSTM (Long Short-Term Memory), dan Random Forest. Evaluasi dilakukan dengan mengacu pada empat metrik utama yaitu accuracy, precision, recall, dan F1-score, serta didukung oleh analisis matriks kebingungan (confusion matrix) guna memahami pola kesalahan klasifikasi.

Berdasarkan visualisasi performa model (Gambar 3), diketahui bahwa model dengan integrasi SMOTE secara konsisten menunjukkan nilai tertinggi pada seluruh metrik evaluasi. Nilai accuracy, precision, recall, dan F1-score seluruhnya berada di kisaran 0.54 hingga 0.56, menandakan performa yang stabil dan seimbang dalam mengklasifikasikan kedua kelas. Model LSTM menunjukkan accuracy dan recall yang cukup tinggi, namun mengalami penurunan yang signifikan pada precision dan F1-score, yang menunjukkan adanya bias terhadap kelas mayoritas. Sementara itu, model Random Forest memiliki performa yang paling rendah pada seluruh metrik, dengan skor mendekati 0.48, sehingga kurang optimal dalam konteks data yang tidak seimbang.

Analisis lebih lanjut melalui matriks kebingungan (Gambar 4) mengonfirmasi temuan tersebut. Model SMOTE berhasil mengklasifikasikan 74 sampel kelas 0 dan 60 sampel kelas 1 dengan benar, meskipun masih terdapat kesalahan klasifikasi berupa 63 false positive dan 49 false negative. Hal ini menunjukkan bahwa SMOTE mampu meningkatkan sensitivitas model terhadap kelas minoritas. Di sisi lain, LSTM menunjukkan akurasi tinggi untuk kelas mayoritas dengan 115 true negative, namun mengalami kegagalan dalam mengidentifikasi kelas minoritas, ditunjukkan oleh 91 false negative, yang menurunkan nilai recall dan F1-score. Adapun model Random Forest menghasilkan 82 true negative dan 38 true positive, namun kesalahan klasifikasinya masih cukup tinggi, dengan 55 false positive dan 71 false negative, menunjukkan ketidakseimbangan dalam generalisasi antar kelas.

Secara keseluruhan, pendekatan dengan teknik SMOTE terbukti paling efektif dalam menangani permasalahan ketidakseimbangan kelas yang umum terjadi dalam data historis harga sawit. Meskipun belum mencapai performa yang sangat tinggi, kombinasi SMOTE dengan model pembelajaran mesin lainnya memiliki potensi untuk dioptimalkan lebih lanjut, baik melalui tuning parameter, pemilihan fitur, maupun pendekatan ensemble seperti stacking atau boosting.

4. KESIMPULAN

Penelitian ini telah mengevaluasi kinerja tiga pendekatan pemodelan dalam prediksi harga minyak kelapa sawit, yaitu algoritma SMOTE, LSTM, dan Random Forest, dengan menggunakan data historis yang memiliki ketidakseimbangan kelas. Berdasarkan hasil evaluasi kuantitatif, model yang dilatih dengan teknik SMOTE menunjukkan kinerja yang paling optimal di antara ketiganya. Hal ini ditunjukkan dengan nilai metrik evaluasi yang lebih tinggi dan seimbang, yaitu akurasi sebesar 0.5447, presisi sebesar 0.5512, recall sebesar 0.5447, dan F1-score sebesar 0.5462. Model SMOTE juga menunjukkan distribusi klasifikasi yang lebih adil untuk kedua kelas, seperti yang tercermin dalam matriks kebingungan yang mencatat 60 Positif Benar dan 74 Negatif Benar, dengan jumlah kesalahan klasifikasi yang relatif moderat. Sebaliknya, model LSTM, meskipun memiliki akurasi keseluruhan yang mendekati SMOTE, menunjukkan kelemahan yang signifikan dalam mengenali kelas minoritas, yang ditandai dengan nilai False Negative yang tinggi dan nilai True Positive yang rendah. Model Random Forest menunjukkan kinerja keseluruhan terendah di semua metrik yang diuji, dan kurang mampu menangani distribusi kelas yang tidak seimbang. Dengan demikian, dapat disimpulkan bahwa pendekatan yang menggunakan teknik oversampling seperti SMOTE dapat berkontribusi secara signifikan dalam meningkatkan kinerja model klasifikasi pada data yang tidak seimbang, terutama dalam konteks prediksi harga kelapa sawit. Temuan ini memperkuat pentingnya strategi penyeimbangan data sebagai tahap penting dalam pemodelan prediktif, dan membuka peluang untuk pengembangan metode hibrida yang lebih adaptif melalui kombinasi algoritme pembelajaran mesin dan teknik peningkatan kualitas data.



REFERENCES

- [1] FAO, “The State of World Fisheries and Aquaculture 2022.” [Online]. Available: <https://openknowledge.fao.org/items/11a4abd8-4e09-4bef-9c12-900fb4605a02>
- [2] P. Triya, N. Suarna, and N. Dienwati Nuris, “Penerapan Machine Learning Dalam Melakukan Prediksi Harga Saham Pt. Bank Mandiri (Persero) Tbk Dengan Algoritma Linear Regression,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 1207–1214, 2024, doi: 10.36040/jati.v8i1.8958.
- [3] K. Puteri and A. Silvanie, “Machine Learning untuk Model Prediksi Harga Sembako,” *Jurnal Nasional Informatika*, vol. 1, no. 2, pp. 82–94, 2020.
- [4] L. Qadrini, H. Hikmah, and M. Megasari, “Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun 2017,” *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 386–391, 2022, doi: 10.47065/josyc.v3i4.2154.
- [5] M. Sulistiyono, Y. Pristyanto, S. Adi, and G. Gumelar, “Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi,” *Sistemasi*, vol. 10, no. 2, p. 445, 2021, doi: 10.32520/stmsi.v10i2.1303.
- [6] F. Kausar, M. Awadalla, M. Mesbah, and T. AlBadi, “Automated Machine Learning based Elderly Fall Detection Classification,” *Procedia Comput Sci*, vol. 203, pp. 16–23, Jan. 2022, doi: 10.1016/J.PROCS.2022.07.005.
- [7] M. Bahri, B. Veloso, A. Bifet, and J. Gama, “AutoML for Stream k-Nearest Neighbors Classification,” *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, pp. 597–602, Dec. 2020, doi: 10.1109/BIGDATA50022.2020.9378396.
- [8] Z. Sadeqi-Arani and A. Kadkhodaie, “A bibliometric analysis of the application of machine learning methods in the petroleum industry,” *Results in Engineering*, vol. 20, p. 101518, Dec. 2023, doi: 10.1016/J.RINENG.2023.101518.
- [9] F. Kausar, M. Awadalla, M. Mesbah, and T. AlBadi, “Automated Machine Learning based Elderly Fall Detection Classification,” *Procedia Computer Science*, vol. 203, pp. 16–23, Jan. 2022, doi: 10.1016/J.PROCS.2022.07.005.
- [10] N. Mohamudally, “Paving the Way Towards Collective Intelligence at the IoT Edge,” *Procedia Computer Science*, vol. 203, pp. 8–15, 2022, doi: 10.1016/j.procs.2022.07.004.
- [11] S. Wang, T. K. Nguyen, and T. Bhatt, “Trip-Related Fall Risk Prediction Based on Gait Pattern in Healthy Older Adults: A Machine-Learning Approach,” *Sensors*, vol. 23, no. 12, Jun. 2023, doi: 10.3390/S23125536.
- [12] D. Razum, G. Seketa, J. Vugrin, and I. Lackovic, “Optimal threshold selection for threshold-based fall detection algorithms with multiple features,” *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, pp. 1513–1516, Jun. 2018, doi: 10.23919/MIPRO.2018.8400272.
- [13] M. Wever, A. Tornede, F. Mohr, and E. Hullermeier, “AutoML for Multi-Label Classification: Overview and Empirical Evaluation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 9, pp. 3037–3054, Sep. 2021, doi: 10.1109/TPAMI.2021.3051276.
- [14] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [15] T. H. Pham and P. Le-Hong, “End-to-End Recurrent Neural Network Models for Vietnamese Named Entity Recognition: Word-Level Vs. Character-Level,” *Communications in Computer and Information Science*, vol. 781, pp. 219–232, 2018, doi: 10.1007/978-981-10-8438-6_18.
- [16] N. Peng and M. Dredze, “Improving named entity recognition for Chinese social media with word segmentation representation learning,” *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Short Papers*, pp. 149–155, 2016, doi: 10.18653/v1/p16-2025.
- [17] Q. Tran, A. MacKinlay, and A. J. Yepes, “Named Entity Recognition with stack residual LSTM and trainable bias decoding,” *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol 1, Jun. 2017.
- [18] J. P. C. Chiu and E. Nichols, “Named Entity Recognition with Bidirectional LSTM-CNNs,” *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 357–370, Dec. 2016, doi: 10.1162/tacl_a_00104.
- [19] W. Gunawan, D. Suhartono, F. Purnomo, and A. Ongko, “Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs,” *Procedia Computer Science*, vol. 135, pp. 425–432, Jan. 2018, doi: 10.1016/J.PROCS.2018.08.193.
- [20] L. Wang, “Automatic Test Paper Generation Technology for Mandarin Based on Hilbert Huang Algorithm,” *Procedia Computer Science*, vol. 228, pp. 808–816, Jan. 2023, doi: 10.1016/J.PROCS.2023.11.099.
- [21] J. Zheng, H. Zhou, X. Liu, Z. Yang, and Z. Ge, “Local deep learning of principal component regression model for spectroscopic calibration of time-varying spectra data,” *Measurement*, vol. 247, p. 116855, Apr. 2025, doi: 10.1016/J.MEASUREMENT.2025.116855.
- [22] B. Liu *et al.*, “Characterization of lacustrine shale oil reservoirs based on a hybrid deep learning model: A data-driven approach to predict lithofacies, vitrinite reflectance, and TOC,” *Marine and Petroleum Geology*, p. 107309, Jan. 2025, doi: 10.1016/J.MARPETGEO.2025.107309.
- [23] Y. Huo *et al.*, “A survey of deep learning-based microscopic cell image understanding,” *Displays*, vol. 87, p. 102968, Apr. 2025, doi: 10.1016/J.DISPLA.2025.102968.
- [24] S. Liang *et al.*, “Quantitative determination of acid value in palm oil during thermal oxidation using Raman spectroscopy combined with deep learning models,” *Food Chemistry*, p. 143107, Jan. 2025, doi: 10.1016/J.FOODCHEM.2025.143107.
- [25] S. B. Rabbani, I. V. Medri, and M. D. Samad, “Deep clustering of tabular data by weighted Gaussian distribution learning,” *Neurocomputing*, vol. 623, p. 129359, Mar. 2025, doi: 10.1016/J.NEUCOM.2025.129359.
- [26] A. Ali *et al.*, “An intelligent computing methodology for two-phase flow performance assessment of electrical submersible pump using artificial neural network and synthetic minority over-sampling technique,” *Measurement*, vol. 244, p. 116512, Feb. 2025, doi: 10.1016/J.MEASUREMENT.2024.116512.



- [27] S. A. Alex, J. Jesu Vedha Nayahi, and S. Kaddoura, “Deep convolutional neural networks with genetic algorithm-based synthetic minority over-sampling technique for improved imbalanced data classification,” *Applied Soft Computing*, vol. 156, p. 111491, May 2024, doi: 10.1016/J.ASOC.2024.111491.
- [28] R. A. Prasajo *et al.*, “Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique,” *Electric Power Systems Research*, vol. 220, p. 109361, Jul. 2023, doi: 10.1016/J.EPSR.2023.109361.
- [29] N.-T. Ho *et al.*, “Machine Learning Approach With Random Forest And Synthetic Minority Over-Sampling Technique May Be Optimal In Non-Invasive Euploidy Detection: A Preliminary Study,” *Fertility and Sterility*, vol. 120, no. 4, p. e107, Oct. 2023, doi: 10.1016/J.FERTNSTERT.2023.08.352.