



Sistem Prediksi Harga Tandan Buah Segar (TBS) Kelapa Sawit Menggunakan Data Analytics Berbasis Web Scraping Informasi Pasar

Rasendriya Fadla Ayyadana | 04221058

Program Studi Teknik Elektro — Institut Teknologi Kalimantan

Dosen Pembimbing 1

Himawan Wicaksono, S.ST., M.T.

Dosen Penguji 1

Muhammad Ridho Dewanto, S.T., M.T

Dosen Pembimbing 2

Amalia Rizqi Utami, S.T., M.T.

Dosen Penguji 2

Muhammad Agung Nursyeha, S.T., M.T.





Latar Belakang

1

Harga TBS sulit diprediksi sehingga petani bergantung pada tengkulak

2

Fluktuasi dipengaruhi oleh harga CPO global dan sentimen berita

3

Sistem prediksi berbasis data numerik & non-numerik



Rumusan Masalah

- 1 **Bagaimana mengintegrasikan data numerik (harga historis) dan non-numerik (berita/sentimen)?**
- 2 **Bagaimana menerapkan ARIMAX dan IndoBERT untuk prediksi yang akurat?**
- 3 **Sejauh mana korelasi antara sentimen berita dan historis harga terhadap perubahan harga TBS kelapa sawit?**



Tujuan

1

Integrasi data numerik & non-numerik melalui Big Data Analytics

2

Implementasi model ARIMAX & IndoBERT untuk analisis pola dan sentimen

3

Menganalisis korelasi antara historis harga dan sentimen berita terhadap perubahan harga TBS.



Dua Sumber Data, Dua Model Komparatif



Data Numerik

Harga historis TBS dari portal SIDIK HTBS (sidikhtbs.id), dikumpulkan via web scraping



Data Non-Numerik

Artikel berita dari portal infosawit.com, dikumpulkan via web scraping



Cleaning &
Preprocessing



Pelabelan
Sentimen



Fine-tuning
IndoBERT



Agregasi
Mingguan



Gabung dengan
Harga TBS



ARIMA vs
ARIMAX

Evaluasi akhir menggunakan tiga metrik objektif: MAE, RMSE, dan MAPE



Lima Ahli Memberi Penilaian Sentimen Berita

Setiap ahli menilai sampel berita sawit menggunakan skala Likert 5-poin yang dimodifikasi untuk mengukur potensi dampak berita terhadap harga TBS.



Panel Lima Ahli

-  Bapak Indrawan, S.Pd., M.Si., Pembina Galeri Investasi Digital ITK
-  Ibu Isti Kamila, S.Pd., M.Si., Dosen Aktuaria
-  Bapak Aldila Bandaro, Deputi Kepala Perwakilan BEI Kalimantan Timur
-  Ibu Aryani, S.Pd, Pemilik usaha loding buah sawit
-  Bapak Junaedi, Pemilik usaha bibit pohon sawit



Web Scraping Otomatis dari Dua Sumber Daring



Harga TBS — SIDIK HTBS

- Selenium WebDriver mengakses sidikhtbs.id, mengunduh file Excel otomatis (Jan 2020–sekarang)
- Data dikelompokkan ke 3 kategori umur tanaman: 3–9, 10–20, 21–25 tahun
- Setiap baris dipetakan ke tanggal Jum'at mingguan menggunakan logika week offset
- Data valid mulai 14 Januari 2022, disimpan sebagai data_avg_tbs.csv



Berita — Infosawit.com

- Sumber awal iqplus.info dialihkan karena update tidak konsisten & artikel terbatas
- REST API WordPress endpoint /wp-json/wp/v2/posts dengan kata kunci "sawit"
- Incremental scraping — hanya ambil artikel baru berdasarkan kecocokan URL, tanpa duplikasi
- Total terkumpul ±8.995 artikel, periode 2022 hingga saat penelitian

BeautifulSoup + Selenium

Library scraping

sidikhtbs.id & infosawit.com

Sumber data

Jan 2022 – Jun 2026

Rentang waktu



Hubungi Team Marketing IQ Plus:
021-5155450

Username Password

REGISTER LOG IN

Search  

Jun 21, 2026 - 02:53 PM

BISNIS-27 427.142 -0.498 (%) CO

HOME

PRODUCT

NEWS

MARKET SUMMARY

TRADING VIEW

ABOUT US



SEARCH FOR KEYWORD "sawit"

- 17/06/26 - 12:58 EKOM: MENTAN SEBUT 130 PERUSAHAAN SAWIT BELUM NAIKKAN HARGA TBS
- 10/06/26 - 10:20 COMD: HARGA TBS SAWIT RIAU NAIK RP3.674,93 PER KG
- 08/06/26 - 14:44 COMD: MENTAN: PRESIDEN MINTA HARGA TBS SAWIT NAIK 10 PERSEN
- 08/06/26 - 11:53 EKOM: MENTAN : 300 PERUSAHAAN DIPERIKSA AKIBAT TAK NAIKKAN TBS SAWIT
- 04/06/26 - 15:11 COMD: EKSPOR SAWIT MALAYSIA AKAN HADAPI TEKANAN AKIBAT ATURAN BARU INDONESIA
- 04/06/26 - 11:13 GZCO: GENJOT PRODUKTIVITAS KELAPA SAWIT, GZCO SIAPKAN BELANJA MODAL RP161 MILIAR
- 03/06/26 - 15:04 COMD: HARGA TBS SAWIT RIAU TURUN JADI RP3.271,91 PER KG
- 03/06/26 - 08:40 SSMS: SAWIT SUMBERMAS (SSMS) ALOKASIKAN RP136,17 MILIAR UNTUK JASA ANALISIS LABORATORIUM
- 02/06/26 - 11:00 COMD: KETUA DPD RI: HARGA SAWIT AKAN KEMBALI NORMAL BERTAHAP
- 29/05/26 - 07:06 EKOM: ASOSIASI PETANI SAWIT DUKUNG DSI MINTA HARGA TBS SEGERA DIPULIHKAN
- 26/05/26 - 16:26 EKOM: WAMENTAN PASTIKAN PETANI DAN PELAKU USAHA SAWIT TERLINDUNGI
- 26/05/26 - 16:10 EKOM: MENKEU : MUSIM MAS DAN WILMAR DIANTARA PERUSAHAAN SAWIT YANG DISELIDIKI
- 26/05/26 - 15:21 EKOM: GAPKI: INOVASI DAN TEKNOLOGI JADI KEBUTUHAN STRATEGIS INDUSTRI SAWIT
- 20/05/26 - 12:00 EKOM: PRABOWO TERBITKAN TATA KELOLA EKSPOR, BUMN EKSPORTIR TUNGGAL SAWIT

LAST QUOTE

Code: COMPOSIT SUBMIT

COMPOSITE

6,177
Last

High
0

Change
4.799

%Change
0.078

Close
6,172

Open
—

Low
0

TVol
29,316,
903,462

Last Update:
2026-06-19 16:15:00

MARKET SUMMARY

IDX INDEX

REGIONAL INDEX

Name	Last	Chg	% Chg
BISNIS-27	427.00	-2.00	-0.50 ↓
COMPOSITE	6,163.00	-9.00	-0.14 ↓
DBX	2,904.00	2.00	0.05 ↑
ESGQKEHATI	102.00	0.00	-0.18 ↓
ESGSKEHATI	103.00	0.00	-0.16 ↓
I-GRADE	154.00	-1.00	-0.42 ↓





Dua Pendekatan Pelabelan Sentimen secara Bertahap

Bullish

Berpotensi menaikkan harga

Bearish

Berpotensi menurunkan harga

Neutral

Tidak berdampak langsung

Irrelevant

Tidak berkaitan dengan harga

A. Rule-Based Labeling

- Mencocokkan teks berita dengan pola regex per kategori (kata kunci kenaikan/penurunan harga, permintaan, pasokan)
- Booster score untuk angka harga konkret yang disertai "naik"/"turun"
- Tiap label disertai tingkat keyakinan: high, medium, low
- Keterbatasan: tidak menangkap diksi implisit atau tidak baku

B. API Groq Labeling

- Model Llama 3.3 70B Versatile diposisikan sebagai trader CPO & analisis fundamental komoditas
- Mengevaluasi konteks kalimat secara kontekstual, bukan sekadar kata kunci
- Aturan prioritas: hasil lelang KPBN jadi acuan utama; jika withdraw, rujuk harga CPO Bursa Malaysia
- Menghasilkan ground truth final untuk fine-tuning IndoBERT



Dari Probabilitas Model ke Variabel Eksogen ARIMAX

Skor Sentimen Mentah

Skor = $P(\text{bullish}) - P(\text{bearish})$, bernilai pada rentang $[-1, +1]$. Berita irrelevant diberi skor 0.

Amplifikasi Non-Linear

$\text{Samp} = \text{sign}(s) \times |s|^{2.5} \times 100$ — eksponen 2,5 dipilih karena menghasilkan korelasi Pearson tertinggi terhadap pergerakan harga aktual dibandingkan eksponen 2 dan 3.

Agregasi Mingguan

- Berita irrelevant & skor sangat kecil dieliminasi
- Weighted mean berbobot tingkat keyakinan sentimen
- Periode week ending Friday, selaras publikasi harga TBS

Penggabungan Dataset

- Left join dengan harga TBS sebagai dataset utama
- Sentimen kosong diisi 0 (numerik) / neutral (kategorikal)
- Seluruh observasi harga TBS tetap dipertahankan

4.1 PENGAMBILAN DATA



187 Observasi Harga & 8.928 Artikel Berita Terkumpul

187

observasi mingguan
harga TBS

14 Jan 2022 – 12 Jun 2026

8.995

artikel berita
infosawit.com

1 Jan 2022 – 21 Jun 2026

3

kategori umur
tanaman sawit

3–9, 10–20, 21–25 tahun

Statistik Deskriptif Harga TBS (Rp/kg)

Statistik	Umur 3–9 Th	Umur 10–20 Th	Umur 21–25 Th
Rata-rata	2.528,71	2.899,38	2.744,74
Std. Deviasi	456,25	518,73	500,52
Minimum	1.374,95	1.578,47	1.489,53
Maksimum	3.515,54	4.041,54	3.826,58

Kategori umur 10–20 tahun konsisten lebih tinggi (fase produktif puncak); kategori 3–9 tahun terendah (fase awal produksi).



Konsensus Tinggi pada Sinyal Eksplisit, Bervariasi pada Berita Implikatif

Lima ahli (R1–R5) menilai 10 pernyataan berita CPO/TBS menggunakan skala –2 (sangat bearish) hingga +2 (sangat bullish).

No	Pernyataan Singkat	R1	R2	R3	R4	R5	Rata-rata
1	Kemenhut-Gakkum	0	0	-1	2	1	0,4
2	turun Rp525/kg	-1	-1	-2	-1	-2	-1,4
3	B50 Mandatori	2	2	2	1	2	1,8
4	Buku Diman Ponidjan	0	0	0	0	0	0
5	India impor minyak	-2	-2	-2	2	-2	-1,2
6	Implementasi B50	2	1	2	2	1	1,6
7	Apical Medanlicious	0	0	0	0	0	0
8	turun Rp90/kg	-2	-1	1	2	0	0
9	CPO Rp14.400/kg	2	2	2	2	1	1,8
10	CPO Rp16.275/kg	-1	-1	0	1	0	-0,2

R1 = Bapak Junaedi
R2 = Ibu Aryani, S.Pd
R3 = Bapak Aldila Bandaro
R4 = Ibu Isti Kamila, S.Pd., M.Si.
R5 = Bapak Indrawan, S.Pd., M.Si.

Konsensus Tertinggi

B50 Mandatori & CPO Rp14.400/kg (skor 1,8) — bullish konsisten di seluruh responden

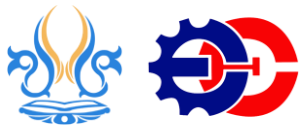
Konsensus Bearish

turun Rp525/kg (skor –1,4) — dinilai bearish secara seragam

Variasi Tertinggi

India impor minyak & turun Rp90/kg — rentang –2 hingga +2, mencerminkan ambiguitas konteks

Skema pelabelan cukup mampu menangkap sentimen dominan pada berita bersinyal harga eksplisit, namun perlu konteks tambahan untuk berita kebijakan atau implikatif.



Dokumentasi Bersama Responden

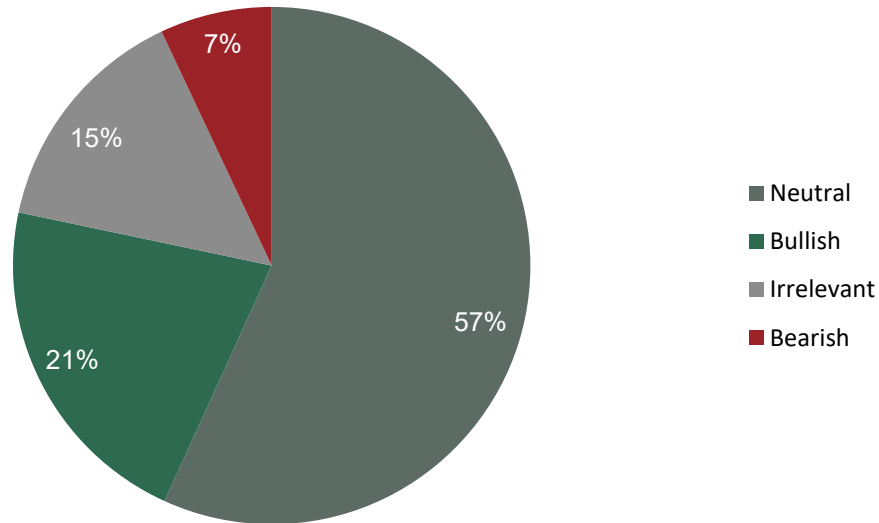




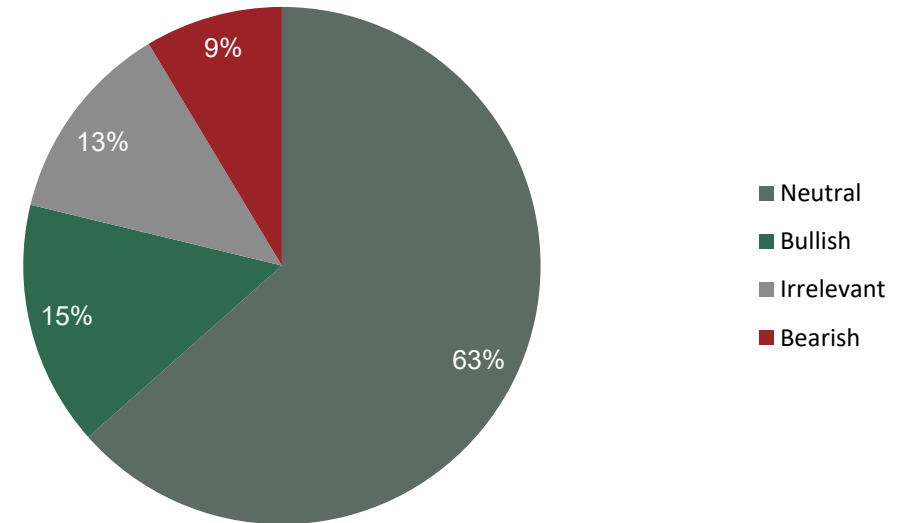
API Groq Menghasilkan Distribusi Label Lebih Realistis

Diterapkan pada 6.979 artikel yang sama, dibandingkan berdasarkan proporsi label dan tingkat keyakinan (confidence)

Rule-Based Labeling



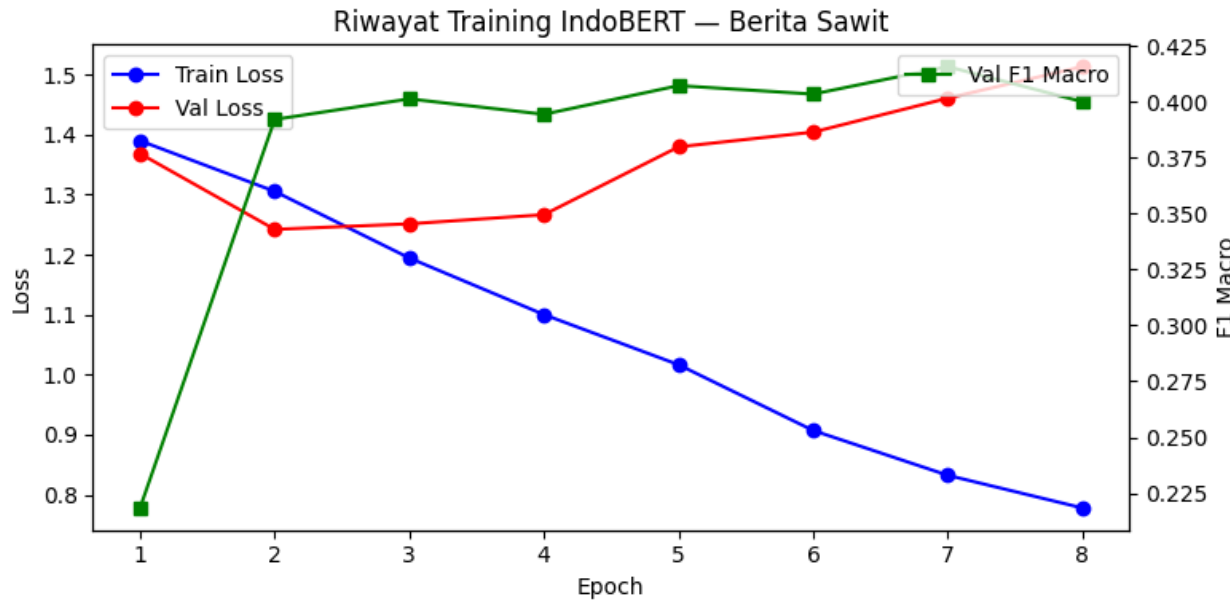
API Groq (Llama 3.3 70B)



Confidence Low pada rule-based mencapai 72,1% ,pelabelan sangat bergantung pencocokan kata kunci literal. Pada Groq, 0% Low confidence, didominasi Medium (61,4%) dan High (38,5%) model memahami konteks kalimat secara utuh dan dipakai sebagai ground truth final.



Early Stopping pada Epoch ke-7, F1 Macro 0,42



Model Terbaik: Epoch 7

Early stopping (patience 3) menyimpan model terbaik dengan F1 macro 0,4156, sebelum val loss terus naik akibat overfitting.

Test Set: Akurasi 45%, F1 Macro 0,36

- Kelas neutral terbaik (F1 0,61) karena dominan pada dataset
- Bullish & bearish kelas paling relevan justru rendah (F1 0,25 & 0,26) akibat sampel minoritas terbatas dan bahasa implisit

Dataset 6.979 artikel berlabel Groq (70:15:15) dengan class weighting untuk mengatasi imbalance (neutral 4.425, bullish 1.068, irrelevant 879, bearish 600 artikel).

Meski performa classifier moderat, sinyal arah sentimen tetap bermanfaat sebagai variabel skoring dalam pipeline agregasi mingguan — bukan sebagai classifier final yang berdiri sendiri.



Groq Lebih Mampu Membedakan Berita Relevan dan Irrelevant

Tanggal	Ringkasan Berita	Rule-Based	API Groq
21 Mei 2026	Harga CPO KPBN withdraw, turun 5,77%	Bearish -98,74	Bearish -82,48
2 Apr 2026	Program B50 perkuat pemanfaatan sawit domestik	Bullish +96,68	Bullish +53,96
1 Apr 2026	Insentif Rp737 juta untuk pekebun via kredit RSPO	Bullish +98,07	Irrelevant 0,00
31 Mar 2026	Sawit dikaitkan banjir-longsor di Sumatera	Bearish -40,32	Neutral +0,84
24 Apr 2022	DSNG raih dua penghargaan emisi korporasi	Bullish +95,40	Irrelevant 0,00

Rule-based melabeli berita 1, 5 sebagai Bullish hanya karena kata-kata terkait kenaikan/penghargaan terdeteksi secara leksikal, meski konteks aktualnya tidak relevan terhadap harga.

Groq berhasil mengoreksi kasus ini menjadi Irrelevant/Neutral, menunjukkan pemahaman konteks yang lebih baik — ini menjadi alasan utama Groq dipilih sebagai ground truth final.



Differencing d=1 Menstasionerkan Seluruh Deret Harga

H0: data memiliki akar unit (tidak stasioner) • H0 ditolak jika p-value < 0,05

Data Level (Sebelum Differencing)

Kategori Umur	ADF Statistic	p-value	Status
3–9 Tahun	-1,3194	0,6202	Tidak Stasioner
10–20 Tahun	-1,3929	0,5857	Tidak Stasioner
21–25 Tahun	-1,3227	0,6187	Tidak Stasioner

Setelah Differencing (d=1)

Kategori Umur	ADF Statistic	p-value	Status
3–9 Tahun	-6,1381	0,0000	Stasioner
10–20 Tahun	-6,1985	0,0000	Stasioner
21–25 Tahun	-6,1721	0,0000	Stasioner

Variabel Eksogen

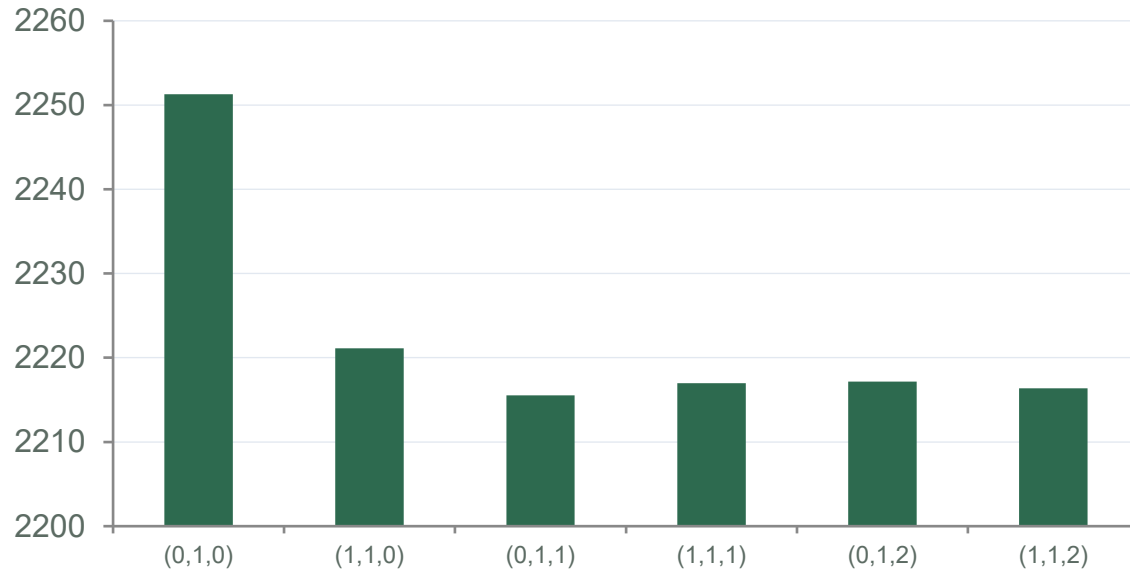
Variabel	ADF Statistic	p-value	Status
Weekly Sentiment	-10,8831	0,0000	Stasioner
Weekly News Count	-1,5408	0,5133	Tidak Stasioner

Pada ARIMAX, variabel eksogen tidak disyaratkan stasioner mutlak — ketidakstasioneran variabel endogen sudah ditangani melalui differencing. Parameter d=1 ditetapkan untuk seluruh pemodelan berikutnya.



ARIMA(0,1,1) Terpilih sebagai Model Optimal

Perbandingan AIC Kandidat Model



AIC terkecil → ARIMA(0,1,1) optimal untuk ketiga kategori umur (konsisten antar kelompok)

Interpretasi ARIMA(0,1,1)

$p=0$ (tanpa AR), $d=1$ (satu kali differencing), $q=1$ (satu komponen MA). Koefisien MA lag-1 signifikan ($\approx 0,50-0,52$, $p < 0,001$) pada ketiga kategori — kesalahan periode sebelumnya berpengaruh kuat.

Evaluasi Rolling Forecast — ARIMA(0,1,1)

Kategori	MAE (Rp)	RMSE (Rp)	MAPE
3–9 Tahun	54,65	79,12	1,78%
10–20 Tahun	59,39	85,58	1,71%
21–25 Tahun	57,96	83,49	1,74%

80% data latih (149 minggu), 20% data uji (38 minggu) — seluruh kategori MAPE < 2% (kategori sangat akurat)



Dua Variabel Eksogen Berbasis Lag Sentimen

bull_l2

Proporsi berita bullish dua minggu sebelumnya (lag 2)

bear_l1

Proporsi berita bearish satu minggu sebelumnya (lag 1)

Dipilih melalui analisis AIC — variabel ber-lag terbukti lebih signifikan karena sentimen butuh waktu untuk tercermin pada harga. Hanya dua variabel dipilih untuk menjaga parsimoni & menghindari overfitting.

Evaluasi Rolling Forecast — ARIMAX(0,1,1)

Kategori	MAE (Rp)	RMSE (Rp)	MAPE
3–9 Tahun	54,19	77,98	1,76%
10–20 Tahun	59,04	85,04	1,69%
21–25 Tahun	57,50	82,52	1,73%

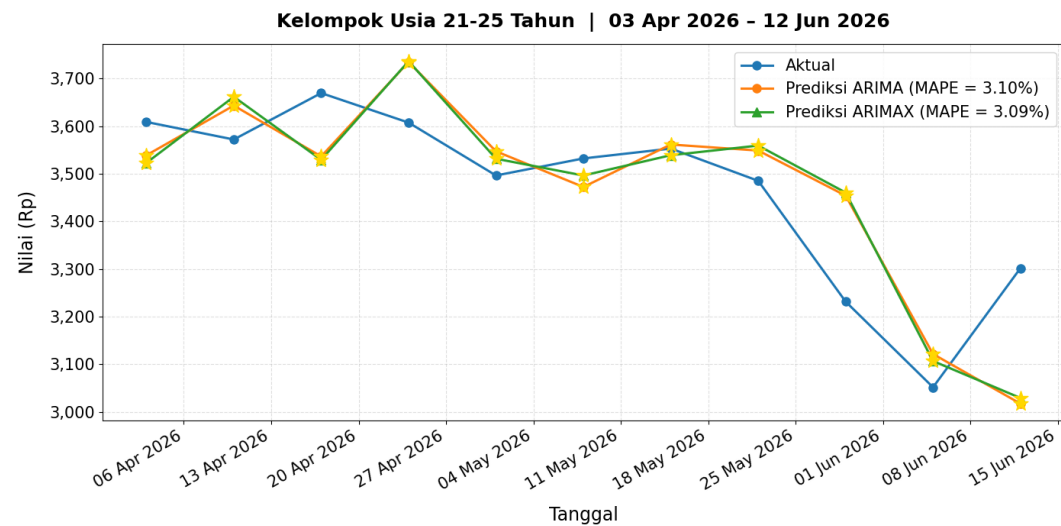
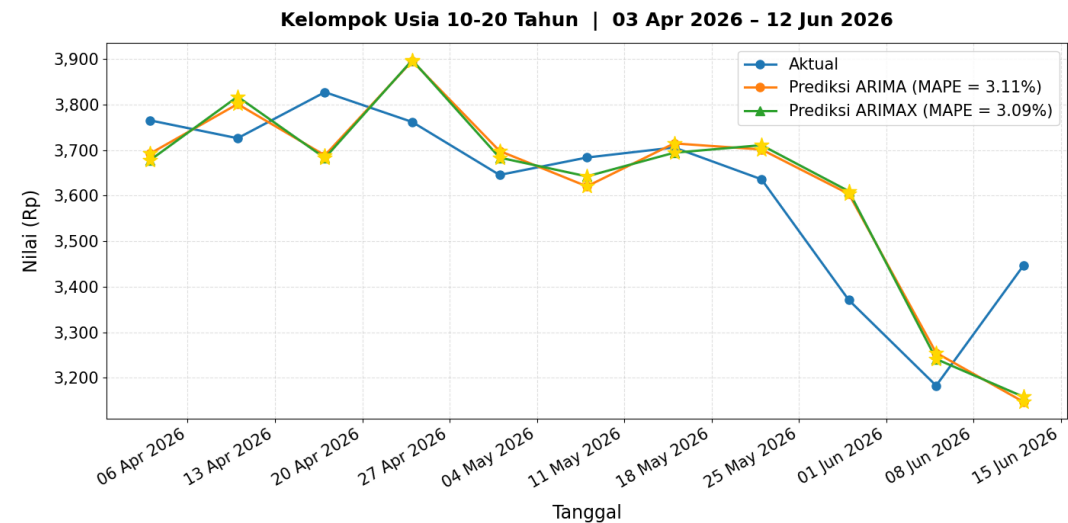
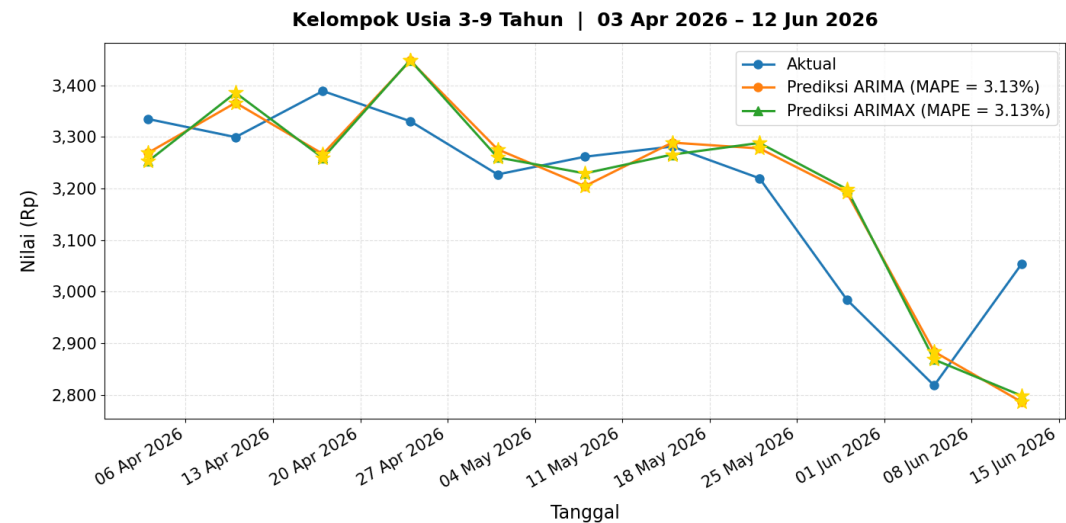
MAPE 1,69%–1,76%

Konsisten lebih rendah dari ARIMA baseline di seluruh kategori, meski selisihnya kecil (0,03–0,04 poin persentase).

Mengindikasikan kontribusi sentimen berita yang nyata namun terbatas.



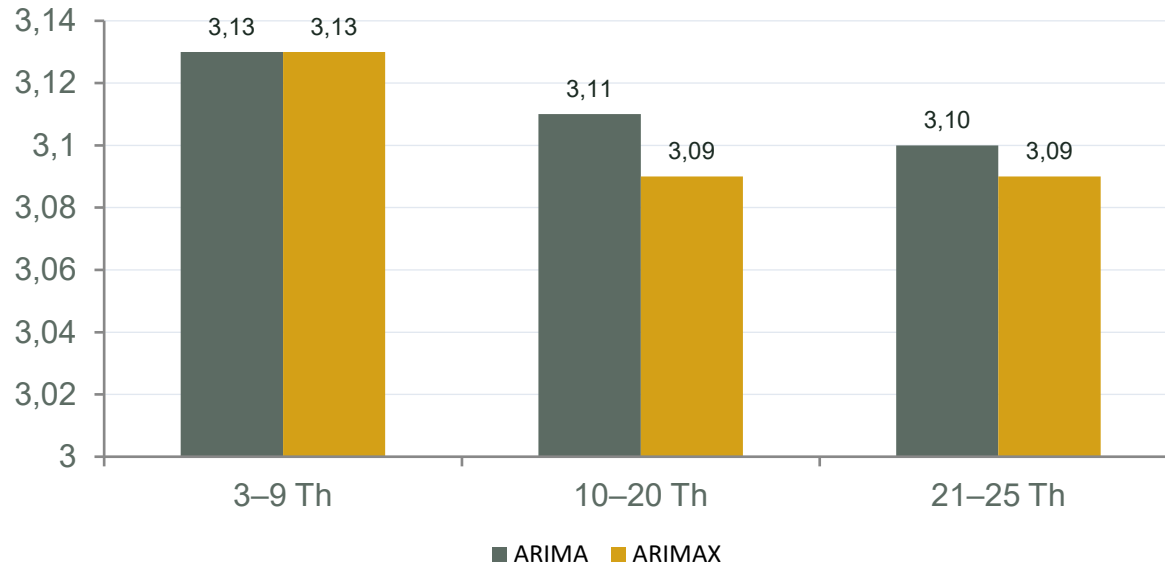
Perbandingan Hasil Prediksi Arima & Arimax





Keunggulan ARIMAX Konsisten walaupun selisih tipis

Perbandingan MAPE (%) per Kategori Umur



Model Terbaik per Metrik & Kategori

Kategori	MAE	RMSE	MAPE
3–9 Tahun	ARIMA	ARIMA	ARIMAX
10–20 Tahun	ARIMAX	ARIMAX	ARIMAX
21–25 Tahun	ARIMAX	ARIMA	ARIMAX

Seluruh MAPE < 10% → kategori highly accurate (kriteria Lewis). ARIMAX unggul di seluruh kategori pada MAPE, tapi selisih MAE/RMSE hanya berkisar Rp0,08–0,47.

Selisih performa yang kecil sebagian dijelaskan oleh sifat lag variabel eksogen: sentimen berita lag 1–2 minggu kurang responsif terhadap perubahan harga yang mendadak dan tajam, seperti yang terjadi pada Mei 2026.



ARIMA Unggul saat Sideways, ARIMAX Unggul saat Trending



Pasar Sideways

- ARIMA lebih stabil dan kompetitif
- Contoh: kategori 3–9 tahun, ARIMA MAE Rp98,46 vs ARIMAX Rp98,64
- Pergerakan harga ditentukan inersia historis — komponen moving average ARIMA sudah memadai
- Sentimen berlag justru bisa jadi noise jika tidak selaras dengan harga yang datar



Pasar Trending

- ARIMAX menunjukkan keunggulan
- Periode rebound 5 & 12 Juni 2026 — ARIMAX lebih baik mengestimasi arah pemulihan (umur 10–20 & 21–25 tahun)
- Variabel bull_l2 dan bear_l1 menangkap konteks arah pasar yang tidak terlihat ARIMA
- Trigger eksternal (biodiesel, stok CPO Malaysia, permintaan India/Tiongkok) muncul beberapa hari–dua minggu sebelum berdampak ke harga lokal

Kesimpulan: ARIMA andal dan efisien pada pasar stabil; ARIMAX memberi keunggulan marginal namun konsisten pada pasar dinamis dengan sinyal sentimen yang mendahului pergerakan harga.



KESIMPULAN

Sentimen Berita Berkontribusi Nyata, namun Terbatas



Integrasi data berhasil

Web scraping menghasilkan 187 observasi mingguan harga TBS (3 kategori umur) dan 8.889–8.929 artikel berita dari sidikhtbs.id dan infosawit.com.



IndoBERT + ARIMAX terintegrasi

Model IndoBERT yang di-fine-tune dengan label API Groq menghasilkan skor sentimen lebih proporsional, diintegrasikan sebagai eksogen ARIMAX(0,1,1) dengan MAPE 1,69%–1,76% (38 minggu rolling forecast).



ARIMAX konsisten lebih baik, walau selisih tipis

Variabel bull_l2 dan bear_l1 terbukti berpengaruh terukur terhadap fluktuasi harga TBS — namun pada kondisi sideways, ARIMA tetap kompetitif karena pola historis harga masih dominan.



Lima Arah Pengembangan ke Depan

- 1 Tambahan variabel eksogen seperti harga CPO futures Bursa Malaysia, kurs Rupiah/USD, atau indikator produksi nasional
- 2 Peningkatan kualitas pelabelan sentimen melalui domain expert sawit atau ensemble beberapa LLM sebagai anotator
- 3 Eksplorasi model hibrida ARIMAX dengan deep learning (LSTM atau Temporal Fusion Transformer) untuk pola non-linear
- 4 Pengembangan dashboard lokal menjadi aplikasi web daring dengan notifikasi otomatis saat prediksi berubah signifikan
- 5 Pengujian pada data provinsi penghasil sawit lain (Riau, Sumatera Utara, Kalimantan Timur) untuk uji generalisabilitas

Terima Kasih

Sistem Prediksi Harga TBS Kelapa Sawit — Pertanyaan & Diskusi

Rasendriya Fadla Ayyadana | NIM 04221058 | Institut Teknologi Kalimantan