Fresh Fruit Bunch (FFB) Price Prediction System for Palm Oil Using Data Analytics Based on Market Information Web Scraping

1st Rasendriya Fadla Ayyadana
Department of Electrical Engineering
Institut Teknologi Kalimantan
Balikpapan, Indonesia
04221058@student.itk.ac.id

2nd Himawan Wicaksono
Department of Electrical Engineering
Institut Teknologi Kalimantan
Balikpapan, Indonesia
himawan@lecturer.itk.ac.id

3rd Amalia Rizqi Utami
Department of Electrical Engineering
Institut Teknologi Kalimantan
Balikpapan, Indonesia
amalia.risqi@lecturer.itk.ac.id

Abstract*—The unpredictable fluctuation of Fresh Fruit Bunch (FFB) palm oil prices poses a major challenge for farmers due to their dependence on volatile global Crude Palm Oil (CPO) price dynamics and market news sentiment. This study develops an FFB price prediction system based on Data Analytics by integrating numerical data (historical prices) and non-numerical data (news articles) through a hybrid approach combining ARIMAX and IndoBERT. Data were collected automatically from sidikhtbs.id and infosawit.com through web scraping, producing 187 weekly FFB price observations across three plantation-age categories (3–9, 10–20, and 21–25 years) and more than 8,929 news articles. Sentiment labeling was performed using a rule-based approach and the Groq API (LLaMA 3.3 70B); Groq labels were treated as LLM-assisted pseudo-labels, validated on a limited scale through expert assessment, and used as training data to fine-tune IndoBERT. Weekly sentiment scores from IndoBERT inference were integrated as exogenous variables in the ARIMAX model, with ARIMA as the baseline. Primary evaluation using a 38-week rolling forecast showed that ARIMAX consistently achieved lower MAPE than ARIMA across all age categories (3–9 years: 1.76%, 10–20 years: 1.69%, 21–25 years: 1.73%), with marginal yet consistent accuracy gains, while both models outperformed a naive-forecast baseline. The research output is an interactive web dashboard intended to support farmers' decision-making in harvest planning and bargaining power.*

Keywords*: ARIMAX, Data Analytics, IndoBERT, FFB Price Prediction, Web Scraping.*

I. INTRODUCTION

The Fresh Fruit Bunch (FFB) price received by palm oil farmers is not determined through a free-market mechanism, but rather through an administrative pricing formula regulated by the Provincial FFB Price-Setting Team under Indonesia's Ministry of Agriculture Regulation No. 01/PERMENTAN/KB.120/1/2018. The formula combines the Crude Palm Oil (CPO) price, the Palm Kernel (PK) price, and a proportion coefficient known as the "K-index," which represents the share of CPO and PK sale value passed on to farmers after deducting processing and marketing costs at the mill level [1], [2]. Because this formula depends on globally fluctuating CPO/PK prices and on a K-index that varies by region and period, the FFB price actually received by farmers reflects an interaction between global market fundamentals and regional administrative policy rather than a pure supply-demand mechanism [2].

FFB prices are vital to the welfare of millions of smallholder farmers, yet global commodity price fluctuations frequently create economic uncertainty at the farm level [3]. This condition is worsened by farmers' limited access to transparent, real-time market price information, forcing many to rely on unilateral pricing set by intermediaries or middlemen [4], [5]. Such dependence can be detrimental because the price received often fails to reflect the true dynamics of the highly volatile CPO market [6]. CPO prices, in turn, are strongly influenced by external factors such as export–import policy, demand from major consumer countries such as India and China, and sentiment from macroeconomic news [7]. Consequently, an accessible, data-driven FFB price prediction system capable of producing accurate estimates is urgently needed to support more strategic harvest and distribution planning for smallholders.

Prior efforts to forecast palm oil commodity prices have largely relied on simple statistical methods or standalone machine-learning approaches. A Moving Average algorithm has been used to forecast FFB prices with 93.91% accuracy, demonstrating the effectiveness of simple statistical methods in capturing short-term trend patterns [8]. Meanwhile, a Backpropagation Artificial Neural Network (ANN) applied to three-month-ahead CPO price prediction achieved low MSE values (0.0227 training, 0.0176 testing), demonstrating the model's ability to capture non-linear patterns in historical data [7]. However, a key limitation of both approaches is their inability to simultaneously analyze complex technical-historical patterns while integrating contextual information from market news: Moving Average is unresponsive to sudden changes and ignores external sentiment, while ANN tends to behave as a black box and depends heavily on feature-engineering quality. This gap indicates that a single-modality approach cannot yet produce comprehensive predictions that integrate large-scale numerical and non-numerical data to address the complex dynamics of the palm oil market.

To address these limitations, this research proposes an FFB price prediction system built on Data Analytics with web-scraped market information, integrating ARIMAX and IndoBERT. The problem statements addressed are: (1) how to integrate numerical (historical price) and non-numerical (news/market sentiment) data into a comprehensive analytical basis for price prediction; (2) how to design, implement, and test ARIMAX and IndoBERT models for FFB price prediction, including validation of sentiment labels through expert assessment; and (3) to what extent the news-sentiment variable contributes to improved ARIMAX accuracy relative to an ARIMA model without a sentiment variable. Correspondingly, the objectives of this study are to integrate numerical and non-numerical data through a Data Analytics approach; to design, implement, and test the ARIMAX and IndoBERT models, including validating Groq API sentiment pseudo-labels through expert assessment; and to measure the contribution of the news-sentiment variable to prediction accuracy using MAE and MAPE. The scope of the research is limited to the FFB reference price published by the SIDIK HTBS portal (sidikhtbs.id) for a specific region over the last five years, combined with palm-oil news scraped from infosawit.com; results are therefore region- and period-specific and intended for academic purposes rather than industrial decision-making.

II. RELATED WORK

A. FFB Price Formation Mechanism

The FFB purchase price paid to farmers is calculated using an official formula defined in Ministry of Agriculture Regulation No. 01/PERMENTAN/KB.120/1/2018, combining the weighted CPO and PK realization prices, their respective extraction rates, and the K-index [1]. A study of this mechanism in Muaro Jambi Regency found that among the cost components tested through regression analysis, only indirect operational cost significantly affected the magnitude of the K-index ($R^2 = 41.9\%$), and that the real price paid to farmers can differ substantially between mills depending on the partnership scheme applied [2]. This confirms that news sentiment, as modeled as an exogenous variable in this study, is one of several factors influencing FFB price rather than a single deterministic factor.

B. Time-Series Forecasting for Palm Oil Commodities

ARIMAX extends the ARIMA model by incorporating exogenous variables (X) to improve time-series forecasting accuracy, allowing the model to account for external factors such as economic indicators or market sentiment in addition to historical patterns [9]. Studies applying ARIMAX to crude-oil and palm-oil price forecasting have shown that incorporating relevant exogenous variables produces lower error metrics than ARIMA alone [9], while ARIMA itself has been shown to reliably project FFB price trends when the model satisfies stationarity, information-criterion, and white-noise requirements [10].

C. Transformer-Based Sentiment Analysis

The Transformer architecture, introduced by Vaswani et al. [11], employs a self-attention mechanism that enables a model to weigh the influence of every token in an input sequence in parallel, which is effective for capturing long-term dependencies in commodity and financial time series. BERT [12] builds on this architecture through bidirectional context modeling trained via Masked Language Modeling and Next Sentence Prediction. IndoBERT [13] is a monolingual pre-trained language model for Indonesian, trained on 220 million words drawn from Indonesian Wikipedia, major news outlets, and a general web corpus, and is well-suited for classifying sentiment in Indonesian-language commodity news.

D. LLM-Assisted Pseudo-Labeling

Manually labeling large-scale sentiment datasets is costly and time-consuming, particularly for domain-specific classification tasks. Weak-supervision frameworks such as Snorkel allow training data to be constructed from imperfect but informative label sources, such as heuristic rules or existing models, rather than relying solely on full manual annotation [14]. The emergence of large language models (LLMs) has enabled LLM-assisted annotation, in which an LLM produces pseudo-labels subsequently used to train a smaller, more computationally efficient model; this approach has been applied in financial sentiment analysis, for example in the construction of the News With GPT Instruction (NWGI) dataset used to fine-tune lightweight language models [15]. In this study, labels produced by the Groq API are likewise treated as LLM-assisted pseudo-labels rather than absolute ground truth, consistent with this weak-supervision framework.

E. Summary of Prior Studies

Table I summarizes prior studies relevant to FFB/CPO price prediction and sentiment-based commodity analysis that informed the design of this research.

TABLE I. SUMMARY OF PRIOR STUDIES

Study	Focus	Key Result
Sayarizki et al. (2024)	IndoBERT for sentiment analysis of presidential-candidate tweets	Fine-tuned IndoBERT reached 80% accuracy; performance dropped for the negative class due to data imbalance.
Amri (2023)	ARIMAX for world crude-oil price forecasting	ARIMAX(0,1,2) with exchange-rate and production exogenous variables outperformed ARIMA(2,1,1), MAPE 8.88% vs 14.45%.
Sujarwo (2025)	ARIMA for FFB price trend projection in Jambi	ARIMA selected as best model; satisfied stationarity, AIC/BIC and white-noise criteria.
Halmi & Putri (2024)	Backpropagation ANN for CPO price prediction	MSE 0.0227 (training), 0.0176 (testing) for 3-month-ahead prediction.
Siregar & Santoso (2025)	Moving Average for FFB price forecasting	RMSE 137.19, MAPE 6.09% (93.91% accuracy).

III. METHODOLOGY

A. Research Overview

This study builds an FFB price prediction system by integrating a Data Analytics approach with web-scraped market information. Numerical data (historical FFB prices) were scraped from the SIDIK HTBS portal, while non-numerical data (news articles) were scraped from infosawit.com. Both datasets underwent cleaning and preprocessing before sentiment labeling was performed to generate pseudo-labels for fine-tuning IndoBERT. Labeling proceeded in two stages: a rule-based method using regular-expression patterns, followed by an LLM-based method using the Groq API, which produced more contextual labels used for fine-tuning. The fine-tuned IndoBERT model produced weekly aggregated sentiment scores, which were merged with weekly FFB prices to build a single integrated dataset. Two models were then trained comparatively: ARIMA as a baseline relying solely on historical price patterns, and ARIMAX as the enhanced model that enriches ARIMA with the exogenous sentiment variable. Prior to modeling, stationarity was tested via the Augmented Dickey-Fuller (ADF) test and optimal order (p, d, q) was identified via ACF/PACF analysis and Auto-ARIMA. The resulting forecasts were evaluated using MAE and MAPE to quantify the extent to which sentiment integration improves prediction accuracy.

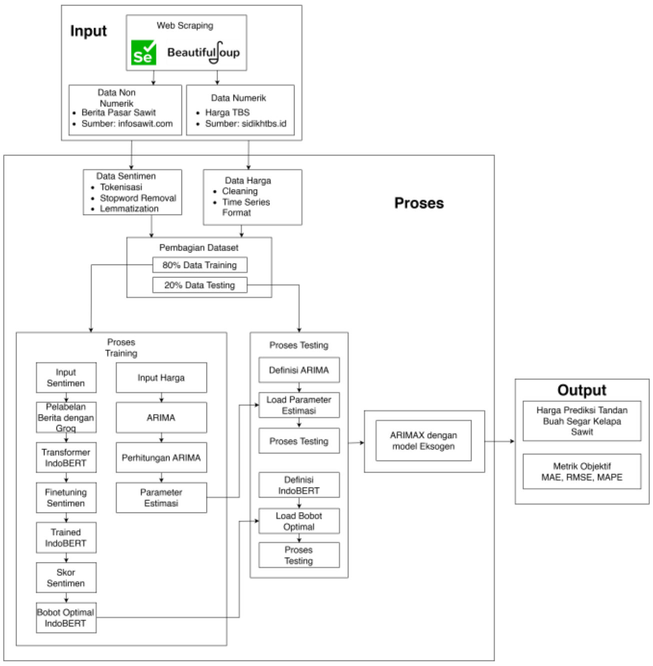

Fig. 1. Block diagram of the system's data, processing, and output pipeline.

B. Expert Validation of Sentiment Labels

A small-scale sanity check on sentiment-label consistency was carried out prior to IndoBERT fine-tuning, involving five experts with backgrounds in economics, commodity markets, and data analysis. Ten news statements were purposively selected to represent explicit bullish, explicit bearish, and ambiguous/implicit sentiment signals. Each expert scored the statements on a modified 5-point Likert scale ranging from −2 (strong potential to depress price) to +2 (strong potential to raise price). This validation is treated as a qualitative sanity check on labeling-direction consistency rather than a statistical estimate of label accuracy across the full 6,979-article population.

C. Data Collection

Historical FFB price data were collected automatically using Selenium WebDriver to access the SIDIK HTBS portal and download the available Excel export, covering January 2020 to the current period; valid weekly records begin on 3 January 2022 and are grouped into three plantation-age categories (3–9, 10–20, and 21–25 years). News data were collected from infosawit.com through its WordPress REST API endpoint (/wp-json/wp/v2/posts) using the keyword "sawit" (palm oil), applying incremental scraping that stops once a previously stored article URL is encountered, yielding roughly 8,929 articles spanning 2022 to the research period. An earlier candidate news source, iqplus.info, was evaluated but discarded due to inconsistent update frequency and insufficient article volume.

D. Data Preprocessing

Price columns were cleaned through numeric conversion to remove invalid values, averaged per age category, and mapped to a weekly Monday-based date using a week-offset scheme derived from the monthly period labels (I–IV). News articles with empty content were removed; each text was cleaned to strip "Baca Juga" (Read Also) cross-reference phrases, location headers, trailing artefacts, and redundant whitespace; articles shorter than 200 clean characters or exact duplicates were discarded, producing a cleaned dataset of date text pairs ready for labeling.

E. Sentiment Labeling

News articles were labeled into four categories bullish, bearish, neutral, and irrelevant using two sequential approaches. The rule-based approach matched article text against curated regular-expression keyword patterns for each category (e.g., confirmed KPBN/Bursa Malaysia price increases for bullish, confirmed decreases for bearish, routine regulatory reports for neutral, and unrelated news such as crime, CSR, or seminars for irrelevant), supplemented by a booster mechanism for texts containing explicit price figures. The Groq-API approach used the Llama 3.3 70B Versatile model, prompted to act as a CPO trader and fundamental commodity analyst, applying the same four-category scheme with an explicit priority rule that treats confirmed KPBN auction outcomes as the primary sentiment signal. Each labeling method also produced a confidence level (high/medium/low) based on inter-category score separation.

F. IndoBERT Fine-Tuning

The pre-trained IndoBERT model was fine-tuned on the Groq-labeled dataset of 6,979 articles (neutral: 4,425; bullish: 1,068; irrelevant: 879; bearish: 600), split 70:15:15 into 4,885 training, 1,047 validation, and 1,047 test samples. Class weighting was applied to the cross-entropy loss based on inverse class-frequency proportion, assigning weights of 1.631 (bullish), 2.905 (bearish), 0.394 (neutral), and 1.984 (irrelevant) to counter class imbalance.

G. Sentiment Score Aggregation

Model output probabilities were converted into a raw sentiment score using P(bullish) − P(bearish), producing a value in the range [−1, +1]; irrelevant articles were assigned a score of 0. Because this range is small relative to the FFB price scale, a non-linear amplification function using an exponent of 2.5 was applied, chosen as a midpoint between the quadratic (2) and cubic (3) functions based on ablation experiments maximizing the Pearson correlation between sentiment score and actual price movement. Daily article-level scores were then filtered (removing irrelevant and near-zero-sentiment articles), adaptively selected per week to prioritize higher-strength signals during active weeks, and aggregated into a weekly score using a confidence-weighted mean for the week ending Friday, aligned with the FFB price publication schedule.

H. Dataset Integration and Modeling

The weekly sentiment scores were merged with the weekly FFB price series using a left join on the date field, with the price data as the primary series so that all price observations are retained even for weeks without qualifying sentiment-bearing news (missing sentiment was imputed as 0). Stationarity of the merged series was assessed via the ADF test; where required, first-order differencing (d = 1) was applied. Optimal ARIMA order (p, d, q) was selected via a stepwise Auto-ARIMA search minimizing the Akaike Information Criterion (AIC). The ARIMAX model was then built using the SARIMAX implementation from the statsmodels library, using the same order as the ARIMA baseline, with the exogenous feature set determined via AIC-based feature selection.

I. Evaluation Metrics

Model performance was evaluated using rolling forecast (walk-forward validation), in which the model is retrained on all available historical data up to a given point and used to predict one step ahead, repeated across the test window. Of 187 weekly observations, 149 weeks (80%) were used for training and 38 weeks (20%) for testing. Two complementary error metrics were used: Mean Absolute Error (MAE), which reports error in the same unit as price (Rp/kg) and is less sensitive to outliers, and Mean Absolute Percentage Error (MAPE), which reports error as a percentage of actual price and enables scale-independent comparison. As additional validation, forecasts were compared against a naive baseline $\hat{y}(t+1) = y(t)$ [16], [17], and residual white-noise behavior was assessed using the Ljung-Box test [18].

TABLE II. DESCRIPTIVE STATISTICS OF FFB PRICE DATA (RP/KG)

Statistic	3–9 yrs	10–20 yrs	21–25 yrs
Mean	2,528.71	2,899.38	2,744.74
Std. Dev.	456.25	518.73	500.52
Minimum	1,374.95	1,578.47	1,489.53
Median	2,559.81	2,941.02	2,779.82
Maximum	3,515.54	4,041.54	3,826.58

IV. RESULTS AND DISCUSSION

A. Dataset Description

Data collection from sidikhtbs.id produced 187 weekly FFB price observations spanning 14 January 2022 to 12 June 2026 across three plantation-age categories. As summarized in Table II, the 10–20-year category consistently commanded the highest average price (Rp2,899.38/kg), consistent with peak agronomic productivity at that plantation age, while the 3–9-year category had the lowest average price (Rp2,528.71/kg) as young palms are still in their early production phase. News scraping from infosawit.com yielded 8,928 cleaned articles spanning 1 January 2022 to 7 June 2026, averaging over 160 articles per month.

B. Expert Validation Results

Across the ten purposively sampled statements, the highest-consensus items were an explicit B50 biodiesel-mandate announcement and a confirmed KPBN price increase, both averaging +1.8 (strongly bullish) across the five experts, while a confirmed Rp525/kg price decline averaged −1.4 (bearish). Two statements involving small price movements (below Rp100/kg) and one involving import-volume news showed substantial inter-rater variation (ranging from −2 to +2), reflecting genuine ambiguity in how experts interpret borderline commodity news.

C. Rule-Based vs. Groq API Labeling

Rule-based labeling of 6,979 articles produced a neutral-dominated distribution (56.8% neutral, 21.5% bullish, 14.7% irrelevant, 7.0% bearish), with the majority of labels (72.1%) falling into the low-confidence tier, reflecting the method's dependence on literal keyword matches and its difficulty capturing implicit or non-standard phrasing, particularly for bearish sentiment (Fig. 2).

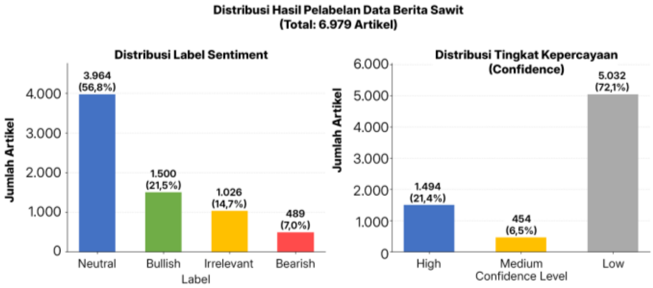

Fig. 2. Distribution of rule-based sentiment labels.

Re-labeling the same dataset using the Groq API produced a markedly different and more balanced distribution (63.4% neutral, 15.3% bullish, 12.6% irrelevant, 8.6% bearish), with no low-confidence labels at all 61.4% medium and 38.5% high confidence (Fig. 3). The bearish share increased from 489 to 600 articles while bullish decreased from 1,500 to 1,068, indicating that the LLM-based approach better captures implicit negative sentiment that literal keyword matching misses.

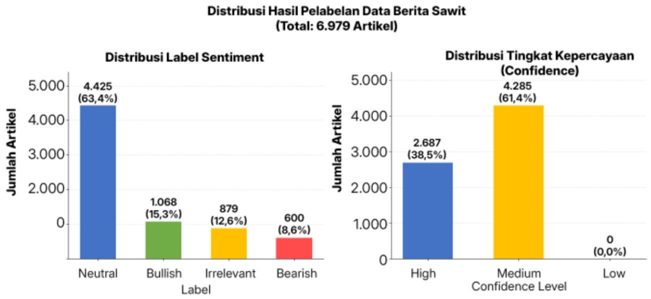

Fig. 3. Distribution of Groq-API (LLaMA 3.3 70B) sentiment labels.

Comparing the categorical Groq-API labels against the expert-derived sentiment direction on the same ten statements, the system achieved agreement on 7 of 10 samples (70%), exceeding the pre-defined minimum accuracy target of 60% and improving substantially over an earlier prompt iteration (40%). Remaining disagreements involved forward-looking policy expectations (where experts anticipated future impact while the system labeled based on explicit textual facts) and marginal price changes below Rp100/kg, both of which represent genuinely ambiguous interpretation cases rather than systematic model error. Given this limited-scale validation, Groq-API labels are treated as LLM-assisted pseudo-labels rather than absolute ground truth, consistent with the weak-supervision framework discussed in Section II-D.

D. IndoBERT Fine-Tuning Performance

Training loss decreased steadily from 1.39 to 0.78 over 8 epochs, while validation loss increased after epoch 2 (from 1.24 to 1.51); early stopping (patience = 3) retained the best checkpoint at epoch 7 with a macro-F1 of 0.4156. On the held-out test set (n = 1,309), the model achieved 45% accuracy and a macro-F1 of 0.36 (Table III). The neutral class, which dominates the dataset (63.4%), achieved the best performance (F1 = 0.61, precision = 0.75), while the bullish and bearish classes most relevant to the study's purpose showed markedly lower precision (0.19 and 0.21, respectively), indicating a tendency to over-predict non-neutral sentiment. Confusion-matrix analysis showed this error was concentrated in bidirectional confusion between neutral and bullish/bearish classes rather than confusion between bullish and bearish themselves (only 38 and 22 misclassifications, respectively, out of 200 and 113 true instances).

E. Justification for Using Probability Scores

Although article-level classification accuracy is modest, IndoBERT is not used in this pipeline as a standalone final classifier but as a scoring component: the continuous probability difference P(bullish) − P(bearish) is used directly as the raw sentiment score rather than the discrete predicted label, making the pipeline inherently more tolerant of per-article classification uncertainty. Weak individual-article signals are further stabilized through confidence-weighted weekly aggregation, which acts as a natural noise filter so that random per-article misclassifications do not directly propagate into the weekly exogenous sentiment score used by ARIMAX.

F. Stationarity Testing

ADF testing on the level series for all three FFB price categories failed to reject the null hypothesis of a unit root (p-values 0.585–0.620, all above 0.05), confirming non-stationarity. After first-order differencing (d = 1), ADF statistics for all three categories fell well below the 5% critical value (p = 0.0000 in all cases), confirming stationarity. Testing on candidate exogenous variables showed weekly sentiment was already stationary at the level (ADF = −10.88, p = 0.0000), while weekly news count was not (p = 0.513); both were nonetheless retained as candidate exogenous inputs given ARIMAX's tolerance for non-stationary exogenous series when the endogenous series is properly differenced.

G. ARIMA and ARIMAX Modeling

Stepwise Auto-ARIMA search based on AIC selected ARIMA(0,1,1) as the optimal order for all three age categories, indicating that recent forecast-error patterns (MA component) rather than deeper autoregressive history explain most of the price dynamics; the MA(1) coefficient was statistically significant ($p < 0.001$) and consistent across categories (≈0.51). Using the same (0,1,1) order for a fair comparison, ARIMAX was constructed with two AIC-selected exogenous features: bull_l2 (bullish-news proportion, lag 2 weeks) and bear_l1 (bearish-news proportion, lag 1 week), reflecting the causal assumption that news sentiment affects price with a short delay. Table IV summarizes the 38-week rolling-forecast performance of both models.

TABLE III. INDOBERT TEST-SET PERFORMANCE (N = 1,309)

Class	Precision	Recall	F1-score	Support
Bullish	0.19	0.36	0.25	200
Bearish	0.21	0.35	0.26	113
Neutral	0.75	0.51	0.61	830
Irrelevant	0.32	0.34	0.33	166
Macro avg.	0.37	0.39	0.36	1,309

TABLE IV. 38-WEEK ROLLING-FORECAST PERFORMANCE: NAIVE, ARIMA, AND ARIMAX

Category	Model	MAE (Rp)	MAPE (%)
3–9 yrs	Naive	57.28	1.878
3–9 yrs	ARIMA	55.18	1.798
3–9 yrs	ARIMAX	54.19	1.763
10–20 yrs	Naive	63.18	1.827
10–20 yrs	ARIMA	60.02	1.723
10–20 yrs	ARIMAX	59.04	1.693
21–25 yrs	Naive	61.05	1.846
21–25 yrs	ARIMA	58.44	1.756
21–25 yrs	ARIMAX	57.50	1.725

As shown in Table IV, ARIMAX consistently achieved lower MAPE than ARIMA across all three plantation-age categories (1.76% vs. 1.80% for 3–9 years; 1.69% vs. 1.72% for 10–20 years; 1.73% vs. 1.76% for 21–25 years), and both models consistently outperformed the naive baseline, confirming that the modeling approach adds genuine value over a no-model heuristic. The margin between ARIMA and ARIMAX, while small, was directionally consistent across all categories, indicating that the bull_l2 and bear_l1 sentiment features contribute additional, relevant information beyond what is captured by historical price patterns alone.

H. Residual Diagnostics

Ljung-Box testing on model residuals at lag 10 produced p-values well above the 0.05 significance threshold for all six model, category combinations (ranging 0.75–0.91), indicating that the null hypothesis of no residual autocorrelation cannot be rejected. Both ARIMA and ARIMAX residuals are therefore consistent with white noise, confirming that both models adequately capture the temporal structure of the data without leaving a systematic unexplained pattern.

I. Case Study: April–June 2026 Forecast Comparison

A focused 11-observation-point case study spanning April–June 2026 compared ARIMA and ARIMAX forecasts against actual prices for all three age categories (Fig. 4). Both models tracked the general price trend, though ARIMAX showed a slight edge in the 10–20-year and 21–25-year categories (MAPE 3.09% vs. 3.11% and 3.09% vs. 3.10%, respectively), while the 3–9-year category showed identical performance (3.13% for both). During a sharp, unanticipated price spike on 27 April 2026, neither model fully captured the sudden change expected given the inherent limitation of linear time-series models in handling abrupt shifts. During the sharpest price decline of the test window (29 May 2026), ARIMAX slightly underperformed ARIMA because the lagged sentiment signal was not synchronized with the sudden actual movement. Conversely, during the price rebound in early-to-mid June 2026, ARIMAX outperformed ARIMA in the 10–20-year and 21–25-year categories, suggesting that lagged sentiment features help the model anticipate recovery signals not visible from price history alone. Overall, ARIMA remains a reliable and efficient choice in sideways (range-bound) market conditions, whereas ARIMAX offers a marginal but consistent advantage during trending periods driven by identifiable external triggers, such as biodiesel-mandate policy news or Malaysian CPO stock data, that precede their effect on local FFB prices by one to two weeks.

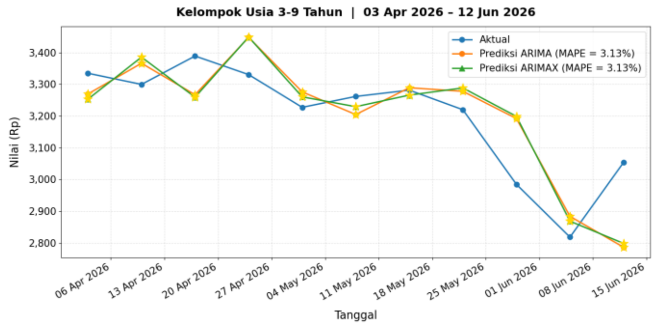

Fig. 4. Actual vs. ARIMA vs. ARIMAX forecasts, April–June 2026 (3–9-year category).

J. Web Dashboard Interface

The research output includes an interactive web dashboard (HTML/CSS/JavaScript, currently running locally as a prototype) that presents next-week price predictions from both ARIMA and ARIMAX side-by-side, an interactive historical-and-forecast price chart with train/test split markers, and a filterable feed of recent news articles labeled by sentiment (Fig. 5). The dashboard is intended to help farmers and industry stakeholders interpret both the quantitative forecast and the qualitative market context behind it.

Fig. 5. Web dashboard: next-week FFB price prediction summary.

V. CONCLUSION

This study successfully integrated numerical and non-numerical palm-oil market data via BeautifulSoup- and Selenium-based web scraping, producing 187 weekly FFB price observations across three plantation-age categories and 8,929 news articles from sidikhtbs.id and infosawit.com. IndoBERT, fine-tuned on Groq-API pseudo-labels, produced a more proportional and contextually informed sentiment score than the rule-based alternative, and when integrated as an exogenous variable in ARIMAX(0,1,1), yielded 38-week rolling-forecast MAPE values of 1.76%, 1.69%, and 1.73% for the three age categories. ARIMAX, using the exogenous features bull_l2 and bear_l1, consistently produced lower MAPE than ARIMA, demonstrating a measurable, if marginal, effect of news sentiment on FFB price fluctuation, while ARIMA remained competitive during sideways market conditions where historical price momentum alone was sufficient to explain price movement.

Future work could incorporate additional exogenous variables such as Bursa Malaysia CPO futures prices, the Rupiah–USD exchange rate, or national production indicators; improve sentiment-label quality through domain-expert annotation or LLM ensembles; explore hybrid ARIMAX, deep-learning architectures such as LSTM or Temporal Fusion Transformer for high-volatility regimes; extend the local dashboard prototype into an online web application with automated alerting; and validate the approach using data from other palm-oil-producing provinces to test generalizability.

REFERENCES

[1] Kementerian Pertanian Republik Indonesia, Peraturan Menteri Pertanian Nomor 01/PERMENTAN/KB.120/1/2018 tentang Pedoman Penetapan Harga Pembelian Tandan Buah Segar Kelapa Sawit Produksi Pekebun, Kementerian Pertanian RI, 2018.

[2] A. Rahman, A. Malik, and R. Siata, "Identifikasi faktor penentu besaran indeks 'K' dan kebijakan penetapan harga TBS kelapa sawit pola kemitraan di Kabupaten Muaro Jambi," Journal of Agribusiness and Local Wisdom, vol. 1, no. 1, pp. 87–99, 2018.

[3] K. Kanchymalay, U. R. Hashim, R. Krishnan, R. R. Ikram, and R. R. Kuhan, "Deep learning-based prediction model for crude palm oil prices using news sentiment analysis with sliding window," Journal of Artificial Intelligence and Technology, pp. 40–47, 2025.

[4] I. Amal, Tarno, and Suparti, "Crude palm oil price prediction using multilayer perceptron and long short-term memory," J. Math. Comput. Sci., vol. 11, no. 6, pp. 8034–8045, 2021.

[5] H. Basheer and A. Khamis, "Modelling and forecasting of monthly crude palm oil price of Malaysia using hybrid wavelet-modified GMDH model," ASM Sc. J., vol. 17, 2022.

[6] C. W. Goh, J. Chai, A. Rahman, and W. E. Ong, "Crude palm oil price prediction using simulated annealing-based support vector regression (SA-SVR)," Asian Academy of Management Journal of Accounting and Finance, vol. 20, no. 1, pp. 305–333, 2024.

[7] M. R. Halmi and R. A. Putri, "Implementation of artificial neural network in predicting CPO prices using backpropagation," Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, vol. 9, no. 2, pp. 161–165, 2024.

[8] Z. U. Siregar and H. Santoso, "Algoritma moving average untuk peramalan harga TBS kelapa sawit," Sistemasi: Jurnal Sistem Informasi, vol. 14, pp. 455–469, 2025.

[9] I. Amri, A. Wulandari, K. Abidah, A. Irawan, and M. Haris, "Pemodelan ARIMAX untuk meramalkan harga minyak mentah dunia," SQUARE: Journal of Mathematics and Mathematics Education, vol. 5, no. 1, pp. 47–58, 2023.

[10] R. M. Sujarwo, "Penerapan model ARIMA untuk memproyeksi tren harga TBS sawit di Provinsi Jambi," Scientific Journal of Reflection: Economic, Accounting, Management and Business, vol. 8, no. 1, pp. 251–261, 2025.

[11] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint, 2018.

[13] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," in Proc. 28th Int. Conf. Computational Linguistics (COLING 2020), pp. 757–770, 2020.

[14] A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré, "Snorkel: Rapid training data creation with weak supervision," Proceedings of the VLDB Endowment, vol. 11, no. 3, pp. 269–282, 2017.

[15] H. Yang, X. Y. Liu, and C. D. Wang, "FinGPT: Democratizing internet-scale data for financial large language models," NeurIPS Workshop on Instruction Tuning and Instruction Following, 2023.

[16] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed. OTexts, 2018.

[17] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Mind the naive forecast! A rigorous evaluation of forecasting models for time series with low predictability," Applied Intelligence, 2026, Advance online publication.

[18] I. U. Moffat and E. A. Akpan, "White noise analysis: A measure of time series model adequacy," Applied Mathematics, vol. 10, no. 11, pp. 989–1003, 2019.

[19] A. Abodayeh, R. Hejazi, W. Najjar, L. Shihadeh, and R. Latif, "Web scraping for data analytics: A BeautifulSoup implementation," in 2023

Sixth Int. Conf. Women in Data Science at Prince Sultan University (WiDS PSU), Riyadh, 2023.

[20] Y. Santur, M. Ulaş, and M. Karabatak, "A web scraping-based approach for fundamental analysis platform in financial assets," Journal of New Results in Science, vol. 11, no. 3, pp. 222–232, 2022.

[21] H. Khoshvaght, R. R. Permala, A. Razmjou, and M. Khiadani, "A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction," Journal of Environmental Chemical Engineering, p. 119675, 2025.