Web-Based Decision Support Assistant for Sectoral Stock Technical Analysis Using Walk-Forward Analysis

Agung Prasetyo[1], Agus Herwanto[2], Vitri Tundjungsari[3], Diah Aryani[4]
[1,2,3,4] Universitas Esa Unggul, Jakarta, Indonesia
Email: agung11prs@student.esaunggul.ac.id

Abstract

This study developed a web-based decision support assistant for sectoral stock technical analysis using Walk-Forward Analysis and Large Language Model-based explanations. The system was designed to identify the best-performing technical indicator for each sector based on historical out-of-sample evaluation and to present the results in an interpretable form. The study evaluated Moving Average Crossover, Moving Average Convergence Divergence, and Relative Strength Index using daily historical data from 40 stocks representing the Financial, Technology, Industrial, and Energy sectors on the Indonesia Stock Exchange. The evaluation applied a fixed-length rolling scheme consisting of a six-month in-sample period, a three-month out-of-sample period, and a three-month window shift. Signal performance was assessed using Average Forward Return at T+1, T+3, T+5, and T+10 trading days, with Directional Accuracy as the primary metric. The results showed that Relative Strength Index achieved the highest observed out-of-sample performance in the financial sector with a Directional Accuracy of 51.52%, while Moving Average Crossover achieved the highest performance in the Technology, Industrial, and Energy sectors with 50.00%, 57.14%, and 71.43%, respectively. The system presented sector-based indicator results, BUY, SELL, or HOLD signals, performance metrics, price visualizations, explanatory narratives, and PDF reports. All 20 predefined Black Box Testing scenarios were completed successfully, User Acceptance Testing achieved a success rate of 99.83%, and the System Usability Scale produced a score of 78.63, which was categorized as Good, Grade B+, and Acceptable. The findings indicate that technical indicator performance varies across sectors and evaluation periods, supporting the use of chronological out-of-sample validation in sector-specific technical analysis.

Keywords: *Decision Support System; Stock Technical Analysis; Fixed-Length Rolling Walk-Forward Analysis; Large Language Model; System Usability Scale.*

A. INTRODUCTION

The growth of digital capital markets has expanded public access to stock price data, charts, technical indicators, and trading services. However, access to information does not necessarily ensure that users can interpret it correctly or select an appropriate technical indicator for a particular stock or market condition. Historical Open, High, Low, Close, and Volume data remain an important basis for technical analysis because they can be transformed into indicators and trading signals through predefined computational rules (Hasan et al., 2024; Htun et al., 2023).

Moving Average Crossover, Moving Average Convergence Divergence, and Relative Strength Index are widely used indicators that represent different analytical characteristics. Moving Average Crossover emphasizes trend direction, Moving Average Convergence Divergence combines trend and momentum information, while Relative Strength Index identifies momentum conditions associated with overbought and oversold levels. Their performance, however, may vary across assets, periods, and market conditions (Hasan et al., 2024; Lento & Gradojevic, 2022).

A preliminary needs survey involving 205 university students with experience in stock investment or trading and technical analysis indicated a clear demand for contextual indicator evaluation. The results showed that 46.34% of respondents experienced difficulties selecting technical indicators, 92.18% believed that indicator effectiveness may differ across sectors, and 85.84% supported the development of a system that presents indicator evaluation results according to sector characteristics. The survey also identified Moving Average Crossover, Moving Average Convergence Divergence, and Relative Strength Index as frequently used indicators, while the Financial, Technology, Industrial, and Energy sectors were selected as the primary evaluation domains.

Stock market data are chronological, so indicator evaluation should preserve temporal order and prevent future observations from influencing earlier selection stages. A single static split may also produce results that depend heavily on one particular market period. This study therefore applies Fixed-Length Rolling Walk-Forward Analysis. In each window, a six-month in-sample period is used to compare indicators, followed by a three-month out-of-sample period for validation. The window is then shifted forward by three months and the procedure is repeated. This structure separates selection from validation and reduces dependence on a single data partition (Cerqueira et al., 2021; Deep et al., 2025).

Sector classification is incorporated because stock sectors may exhibit different return dynamics and risk transmission patterns. Fitriyah & Mukminatin (2025) found that Indonesian sectoral stock indices differed in their roles as transmitters and receivers of return-related risk, while Sahoo and Kumar (2024) reported differences in volatility spillovers across sectors. These findings support the inclusion of sectoral context in technical indicator evaluation.

The evaluation results are implemented in a web-based decision support assistant. The system presents the indicator with the highest observed historical out-of-sample performance for the relevant sector, together with BUY, SELL, or HOLD signals, evaluation metrics, price visualizations, explanations, and downloadable reports. All analytical calculations, including indicator computation, signal generation, Walk-Forward Analysis, and performance measurement, are conducted by deterministic modules.

A Large Language Model is used only as an explanation layer. Although such models can improve the accessibility of financial information, they may also generate inconsistent or unsupported statements (Kang & Liu, 2023; Lee et al., 2025; Li et al., 2023). Therefore, the model is not used to calculate indicators, generate signals, select the best-performing indicator, predict stock prices, or provide final investment recommendations. A deterministic fallback mechanism is used when the model service is unavailable.

This study aims to develop a web-based decision support assistant for sectoral stock technical analysis, evaluate Moving Average Crossover, Moving Average

Convergence Divergence, and Relative Strength Index using Fixed-Length Rolling Walk-Forward Analysis, and assess the system through Black Box Testing, User Acceptance Testing, and the System Usability Scale. Its contribution lies in integrating sector-specific indicator evaluation, rolling in-sample selection, chronological out-of-sample validation, multi-horizon signal assessment, and controlled natural-language explanation within a single web-based framework.

B. LITERATURE REVIEW

Hasan et al. (2024) showed that Moving Average, Moving Average Convergence Divergence, and Relative Strength Index can extract useful information from historical stock prices, although their performance may differ across assets and evaluation periods. Similarly, Htun et al. (2023) emphasized the importance of OHLCV (Open, High, Low, Close, and Volume) data and technical features for stock market analysis. Although their work primarily focused on price prediction and feature selection, it confirmed that historical price and volume data provide a reliable foundation for technical indicator computation.

Lento and Gradojevic (2022) found that the performance of technical trading rules varies across assets and market regimes. Their findings indicate that technical indicators should not be assumed to produce consistent results under all market conditions. This supports the need for repeated chronological evaluation rather than relying on a single historical period.

Sector-specific studies provide additional evidence supporting this perspective. Fitriyah & Mukminatin (2025) found that Indonesian sectoral stock indices differed in their roles as transmitters and receivers of return-related risk.Sahoo and Kumar (2024) similarly reported that volatility spillovers vary across sectors in emerging and developed markets. Al-Nefaie & Aldhyani (2022) applied Multilayer Perceptron and Long Short-Term Memory models to stock price forecasting across several sectors of the Saudi Stock Exchange. Although their study focused on forecasting rather than technical indicator evaluation, it demonstrates the feasibility of analyzing stock data across distinct sectoral contexts.

From a methodological perspective, Deep et al. (2025) demonstrated that Walk-Forward Validation can preserve chronological order and separate model or strategy selection from subsequent out-of-sample testing. This approach is relevant to financial time-series evaluation because it prevents future observations from influencing earlier selection stages.

From the perspective of decision support systems, Cinelli et al. (2022) emphasized that such systems are designed to assist users in evaluating alternatives systematically rather than replacing human decision-making. In financial applications, Large Language Models may improve the accessibility of complex information through natural-language explanations. However, their outputs may also contain unsupported or inconsistent statements, particularly in high-risk domains such as finance (Kang & Liu, 2023; Lee et al., 2025; Li et al., 2023).

Previous studies reveal several consistent observations. Classical technical indicators such as Moving Average Crossover, Moving Average Convergence Divergence, and Relative Strength Index remain relevant for stock market analysis; however, their effectiveness cannot be assumed to be uniform across different stocks or industry sectors. Historical OHLCV data continue to serve as the primary foundation for technical analysis, while sector characteristics introduce substantial variation in indicator performance. Because stock market data are inherently sequential, technical indicator evaluation should preserve the chronological order of observations to avoid conclusions that depend excessively on a single data partition (Cerqueira et al., 2021; Deep et al., 2025).

Another challenge concerns the communication of analytical results. While many systems successfully calculate technical indicators, relatively few provide explanations that are accessible to non-expert users. In this context, Large Language Models offer a practical solution, provided that their role is carefully constrained. Rather than participating in indicator selection, signal generation, or core computational processes, the LLM should function solely as an explanation layer that translates deterministic analytical outputs into natural language (Kang & Liu, 2023; Lee et al., 2025; Li et al., 2023).

Despite extensive research on technical trading strategies and stock price prediction, few studies have integrated sector-specific technical indicator evaluation, rolling in-sample selection, chronological out-of-sample validation, multi-horizon Average Forward Return assessment, web-based decision support, and controlled natural-language explanations generated from deterministic analytical outputs within a single framework. This research addresses that gap by combining these components into a single analytical system.

Technical analysis uses historical price and trading volume data to interpret market trends, momentum, and potential reversals. Rather than predicting future prices with certainty, it applies predefined rules to transform historical market data into technical signals. In this study, technical analysis is used to generate BUY, SELL, and HOLD signals through Moving Average Crossover, Moving Average Convergence Divergence, and Relative Strength Index (Hasan et al., 2024; Lento & Gradojevic, 2022).

The three indicators represent complementary analytical characteristics. Moving Average Crossover is a trend-following indicator that identifies changes in trend direction through the crossing of short-term and long-term moving averages. Moving Average Convergence Divergence combines trend and momentum information through the relationship between exponential moving averages and a signal line. Relative Strength Index measures momentum and identifies conditions associated with overbought and oversold price movements. The detailed parameters and signal-generation rules are described in the Method section.

Sectoral context is included because stock sectors may exhibit different return dynamics and volatility spillover patterns. Fitriyah & Mukminatin (2025) showed that Indonesian sectoral stock indices differed in their roles as transmitters and receivers of return-related risk, while Sahoo and Kumar (2024) reported that volatility spillovers vary across sectors. These findings support the evaluation of technical indicators within separate sectoral contexts rather than assuming uniform performance across all stocks.

A Decision Support System is designed to assist users in evaluating alternatives systematically without replacing human judgment. In the proposed system, the Decision Support System organizes sector-based indicator evaluation results, technical signals, performance metrics, price visualizations, explanatory information, and downloadable reports. The Large Language Model is used only to generate natural-language explanations from deterministic analytical outputs. All core calculations remain within deterministic modules to maintain transparency, consistency, and reproducibility (Cinelli et al., 2022; Kang & Liu, 2023).

Historical stock market data are commonly represented through Open, High, Low, Close, and Volume variables, which provide the basis for technical indicator computation and signal generation. Technical indicators are evaluated using Fixed-Length Rolling Walk-Forward Analysis. This methodology preserves chronological order by repeatedly partitioning the dataset into fixed-duration in-sample and out-of-sample periods. The in-sample period is used for indicator comparison, while the subsequent out-of-sample period validates the selected indicator on later observations. This structure reduces dependence on a single static data partition (Cerqueira et al., 2021; Deep et al., 2025).

Signal performance can be evaluated using Average Forward Return and Directional Accuracy. Average Forward Return summarizes subsequent returns across predefined trading-day horizons, while Directional Accuracy measures the proportion of active signals whose direction is consistent with subsequent price movement. Hit Rate, Total Active Signals, and Correct Signals provide supporting context regarding consistency and sample size.

System quality is assessed through Black Box Testing, User Acceptance Testing, and the System Usability Scale. Black Box Testing evaluates whether system functions produce the expected outputs without examining the internal program structure. User Acceptance Testing measures the successful completion of predefined user scenarios, while the System Usability Scale measures users perceived usability after interacting with the system. Together, these methods assess functional correctness, user task completion, and perceived usability (Gustinov et al., 2023; Aliyah et al., 2025; Grier et al., 2013; Hyzy et al., 2022).

C. METHOD

1. Research Design

This study employed a Research and Development approach with a prototyping model to develop a web-based decision support assistant for sectoral stock technical analysis. This approach was selected because the study not only evaluated technical indicators but also produced a functional software system for end users. The development process was conducted iteratively through needs identification, system design, prototype development, evaluation, refinement, and final product completion. User needs were identified through a preliminary survey, while prototype evaluation informed interface and functional improvements before the final system was tested.

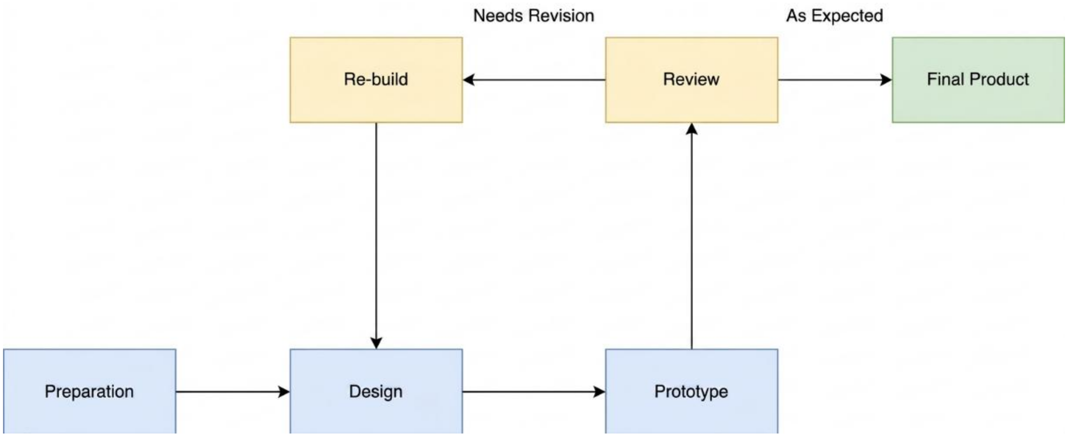


Figure 1. Research Method Flowchart (R&D–Prototyping)

The development process comprised six stages: needs identification, system design, prototype development, evaluation, refinement, and final product completion.

2. Research Data and Sample

The study used daily Open, High, Low, Close, and Volume data from 40 stocks listed on the Indonesia Stock Exchange. The sample consisted of four sectors, namely Financial, Technology, Industrial, and Energy, with ten stocks representing each sector. The indicators and sectors were selected based on a preliminary needs survey involving 205 respondents who had experience in stock investment or trading and had used technical analysis.

Historical market data were retrieved from Yahoo Finance through the Python yfinance library. The warm-up period began on July 1, 2024, to provide sufficient observations for the initial calculation of technical indicators. The main Fixed-Length Rolling Walk-Forward Analysis period extended from October 21, 2024, to April 20, 2026. The dataset was checked for the completeness of Open, High, Low, Close, and Volume variables, chronological consistency, and sufficient observations for indicator calculation. Active BUY or SELL signals without complete T+1, T+3, T+5, and T+10 trading-day observations were excluded from performance evaluation.

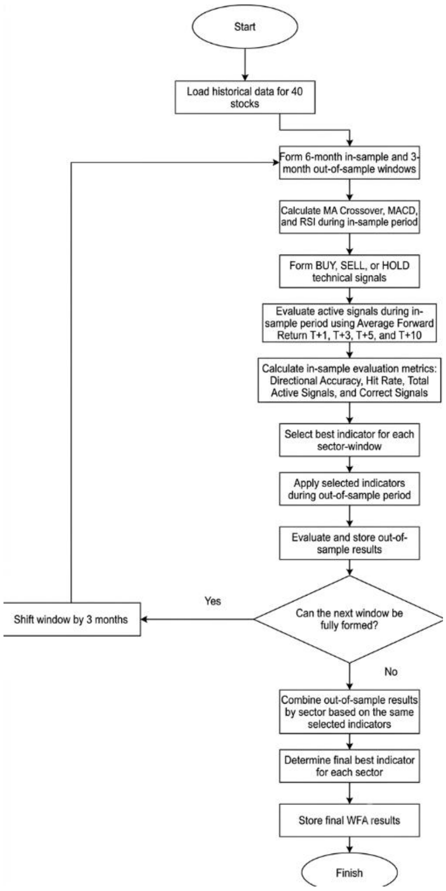


Figure 2. Flow Fixed-Length Rolling Walk-Forward Analysis

3. System Development

The system was developed using a web-based client-server architecture, with Flask serving as the backend framework. Users can submit a stock ticker, company name, or alias. The system validates the input against the stock mapping, identifies the corresponding sector, retrieves the stored sector-level evaluation result, calculates the latest technical indicator values and signals, and presents the results through an analysis dashboard and downloadable PDF report.

Fixed-Length Rolling Walk-Forward Analysis was conducted offline, and the resulting sector-level evaluation records were stored for use by the web application. This design avoids rerunning the complete historical evaluation for every user request while allowing the latest technical signals to be calculated from the most recently available market data.

The Large Language Model was restricted to generating natural-language explanations from structured deterministic outputs. Indicator calculation, signal generation, Walk-Forward Analysis, metric calculation, and best-performing indicator selection were performed by deterministic modules. If the model service was unavailable, the system used a template-based deterministic fallback to maintain explanation availability without modifying the analytical results. The system used the OpenAI 5.4 mini model through the application programming interface.

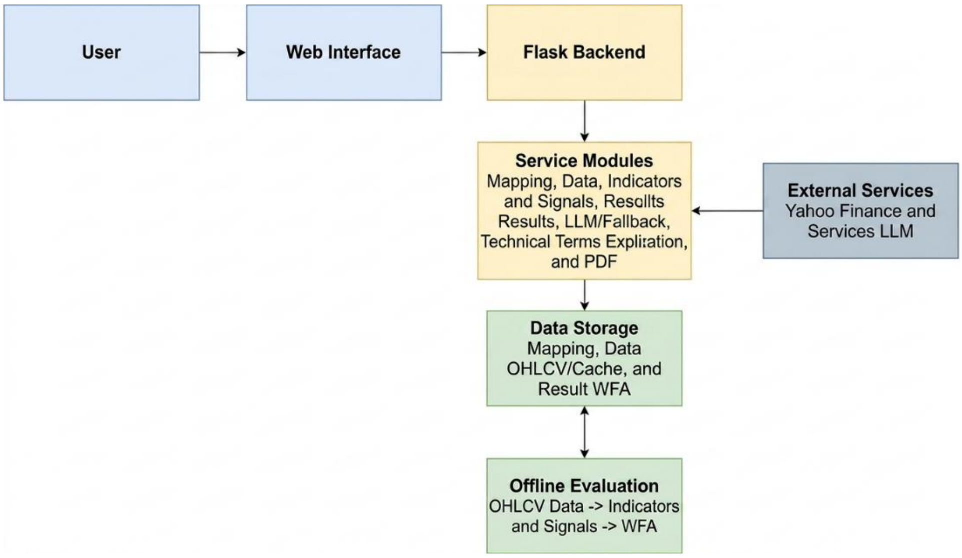


Figure 3. Proposed System Architecture

The diagram simply displays:

a. Users
b. Web Application (Flask)
c. Stock Mapping Module
d. Historical Data Module
e. Technical Indicator Module
f. Fixed-Length Rolling Walk-Forward Analysis Module
g. Evaluation Results Database
h. LLM/Fallback Module
i. Analysis Results Dashboard

4. Technical Indicator and Signal Rules

Three technical indicators were evaluated using fixed parameters without parameter optimization: Moving Average Crossover with 10-period and 50-period Simple Moving Averages, Moving Average Convergence Divergence with parameters of 12, 26, and 9, and Relative Strength Index using Wilder's 14-period formulation with thresholds of 30 and 70.

A Moving Average Crossover BUY signal was generated when the 10-period Simple Moving Average crossed above the 50-period Simple Moving Average, while a SELL signal was generated when it crossed below. For Moving Average Convergence Divergence, BUY and SELL signals were generated from crossovers between the Moving Average Convergence Divergence line and the signal line. The histogram was calculated but was not used as an additional signal filter. For Relative Strength Index, a BUY signal was generated when the previous value was below 30 and the current value reached or exceeded 30. A SELL signal was generated when the previous value was above 70 and the current value reached or fell below 70. When no active signal condition was satisfied, the output was classified as HOLD.

All indicator calculations and signal-generation rules were implemented through deterministic modules. No parameter optimization was conducted during the rolling evaluation.

Indicator performance was evaluated using Average Forward Return, Directional Accuracy, Hit Rate, Total Active Signals, and Correct Signals. Average Forward Return was calculated as the mean of returns at T+1, T+3, T+5, and T+10 trading days after signal formation and was used to determine whether each active BUY or SELL signal was directionally correct. Directional Accuracy served as the primary metric for selecting the best-performing indicator, while Hit Rate, Total Active Signals, and Correct Signals provided supporting information regarding performance consistency and the number of evaluated signals.

5. Performance Evaluation Metrics

a. Directional Accuracy

$$\boldsymbol{Directional\ Accuracy = \frac{CS}{TAS} \times 100\%}$$

Information:

DA = Directional Accuracy;

CS = Correct Signals or the number of active signals assessed as correct;

TAS = Total Active Signals or the total number of all active BUY and SELL signals being evaluated.

b. Hit Rate

$$\boldsymbol{HR_i = \frac{CS_i}{TAS_i} \times 100\%}$$

Information:

HR_i = signal success rate on the i-th stock-window unit;

CS_i = Number of correct Signals on the i-th active stock-window unit;

TAS_i = Total number of Active Signals on the i-th active stock-window unit.

After the success rate of each window is obtained, the Hit Rate value is calculated using the formula:

$$\boldsymbol{HR = \frac{\sum_{i=1}^{n} HR_i}{n}}$$

Information:

HR = Hit Rate;

HR_i = success rate in window i;

n = the number of active stock-window units that have at least one BUY or SELL signal.

c. Average Forward Return

$$\boldsymbol{AFR_t = \frac{R_{T+1} + R_{T+3} + R_{T+5} + R_{T+10}}{4}}$$

Information:

AFR_t = Average Forward Return for a signal formed at time t;

R_{T+1} = return on one trading day after the signal;

R_{T+3} = return on three trading days after the signal;

R_{T+5} = return on five trading days after the signal;
R_{T+10} = return on ten trading days after the signal.

6. System Testing

System quality was evaluated through three testing phases, involving 40 respondents based on the characteristics of the target users. Functional testing was conducted using Black Box Testing to ensure each feature performed as designed. Black Box Testing was conducted by the developer using 20 predefined functional scenarios. User Acceptance Testing involved respondents who completed 15 operational scenarios, resulting in 600 task responses. The same respondents subsequently completed the 10-item System Usability Scale questionnaire using a five-point response scale. User Acceptance Testing (UAT) was then conducted to measure user success in completing key usage scenarios. The system's usability was evaluated using the System Usability Scale (SUS), Individual System Usability Scale scores were calculated using the standard odd- and even-item conversion procedure and multiplied by 2.5, after which the mean score across respondents was reported.

The combination of these three methods was used to assess the system's functionality, user acceptance, and usability, Together, these methods assessed functional correctness, successful user task completion, and perceived usability of the developed system.

D. RESULT AND DISCUSSION

1. Results of Technical Indicator Evaluation in Each Sector

Technical indicator evaluation was conducted using Fixed-Length Rolling Walk-Forward Analysis with a six-month in-sample period, a three-month out-of-sample period, and a three-month window shift. In each sector and window, Moving Average Crossover, Moving Average Convergence Divergence, and Relative Strength Index were compared using in-sample Directional Accuracy. The indicator with the highest in-sample value was then evaluated during the subsequent out-of-sample period. After all four windows were completed, the out-of-sample results were grouped according to the indicator selected in the corresponding in-sample period and aggregated at the sector level.

Average Forward Return was used to classify each active BUY or SELL signal as directionally correct or incorrect. Directional Accuracy served as the primary performance metric, while Hit Rate, Total Active Signals, and Correct Signals provided supporting context regarding consistency and the number of evaluated signals.

Table 1. Results of Technical Indicator Evaluation in Each Sector

Sector	Best-Performing Indicator	Directional Accuracy	Hit Rate	Total Active Signals	Correct Signals
Financial	RSI	51.52%	57.89%	33	17
Technology	MA Crossover	50.00%	56.67%	14	7
Industrial	MA Crossover	57.14%	62.50%	35	20
Energy	MA Crossover	71.43%	76.67%	21	15

Source: data proceed

The evaluation results presented in Table 1 indicate that the Relative Strength Index (RSI) achieved the best performance in the financial sector, whereas the Moving Average Crossover (MA Crossover) outperformed the other indicators in the Technology, Industrial, and Energy sectors. These findings suggest that the effectiveness of technical indicators depends on sector-specific price dynamics, indicating that no single indicator consistently performs best across all sectors.

In the Financial sector, Relative Strength Index produced 17 correct signals from 33 active signals, resulting in a Directional Accuracy of 51.52% and a Hit Rate of 57.89%. Although its Directional Accuracy was only slightly above 50%, it achieved the highest aggregated out-of-sample value among the indicators that were selected and validated in this sector. The result should therefore be interpreted as the highest observed historical performance within the selected sample, period, parameters, and evaluation rules.

In the Technology sector, Moving Average Crossover produced 7 correct signals from 14 active signals, resulting in a Directional Accuracy of 50.00% and a Hit Rate of 56.67%. This result indicates that the number of correct and incorrect signals was equal. Therefore, although Moving Average Crossover achieved the highest aggregated out-of-sample Directional Accuracy among the validated indicators in this sector, the result does not indicate strong directional superiority.

In the Industrial sector, Moving Average Crossover produced 20 correct signals from 35 active signals, resulting in a Directional Accuracy of 57.14% and a Hit Rate of 62.50%. The larger number of active signals provides a broader empirical basis than the Technology-sector result, although the findings remain limited to the selected stocks and evaluation period.

The highest observed sector-level result occurred in the Energy sector, where Moving Average Crossover produced 15 correct signals from 21 active signals. This yielded a Directional Accuracy of 71.43% and a Hit Rate of 76.67%. Among the four sectors, this result represented the largest proportion of directionally correct signals, although it remains a historical result derived from a limited number of active signals and does not guarantee similar performance in later periods.

Sector-level results differed substantially. Moving Average Crossover achieved Directional Accuracy values of 50.00% in Technology, 57.14% in Industrial, and 71.43% in Energy, while Relative Strength Index achieved the highest observed value in

Financial at 51.52%. These differences show that the highest-performing indicator and its observed performance varied across the evaluated sectors.

The interpretation of these results should consider the number of active signals underlying each percentage. A high Directional Accuracy based on a small number of signals provides a more limited empirical basis than a similar value derived from a larger number of observations. For this reason, Directional Accuracy was interpreted together with Hit Rate, Total Active Signals, and Correct Signals.

The evaluation identified the indicator with the highest observed aggregated out-of-sample Directional Accuracy for each sector under the predefined research configuration. Rather than applying a single technical indicator uniformly across all stocks, the proposed decision support system selects the indicator with the strongest historical performance according to the sector classification of the analyzed company. These results represent historical comparative performance and should not be interpreted as universal indicator superiority or as final investment recommendations.

2. Walk-Forward Evaluation Across Windows

Fixed-Length Rolling Walk-Forward Analysis produced four complete evaluation windows for each sector. In every window, the indicator with the highest in-sample Directional Accuracy was selected and subsequently evaluated during the following out-of-sample period. The results showed that the indicator selected during the in-sample period did not always retain the same level of performance during out-of-sample validation.

In the Financial sector, Relative Strength Index was selected in the first three windows, while Moving Average Crossover was selected in the fourth window. Relative Strength Index increased from an in-sample Directional Accuracy of 43.75% to an out-of-sample value of 66.67% in the first window. Its performance remained at 50.00% in the second window but declined from 54.17% to 30.00% in the third window. In the fourth window, Moving Average Crossover decreased from 47.37% in-sample to 40.00% out-of-sample.

In the Technology sector, Relative Strength Index was selected in the first two windows, while Moving Average Crossover was selected in the third and fourth windows. Out-of-sample Directional Accuracy was lower than the corresponding in-sample value in the first three windows. However, Moving Average Crossover increased from 45.45% in-sample to 80.00% out-of-sample in the fourth window. This value was based on only five active signals and should therefore be interpreted cautiously.

In the Industrial sector, Relative Strength Index was selected in the first window, while Moving Average Crossover was selected in the remaining three windows. Moving Average Crossover increased from 61.90% in-sample to 87.50% out-of-sample in the second window. In contrast, its Directional Accuracy declined from 72.73% to 50.00% in the third window and from 68.75% to 47.37% in the fourth window. The result demonstrates that a high in-sample value did not necessarily persist during the subsequent validation period.

In the Energy sector, Moving Average Crossover was selected in the first and fourth windows, while Relative Strength Index was selected in the second and third windows. Moving Average Crossover increased from 51.72% to 66.67% in the first window and from 64.29% to 83.33% in the fourth window. Relative Strength Index increased slightly from 60.53% to 63.64% in the second window but declined substantially from 68.42% to 35.71% in the third window.

Across the four sectors, Moving Average Crossover and Relative Strength Index were each selected in eight of the sixteen sector-window combinations, while Moving Average Convergence Divergence was not selected in any window because it did not obtain the highest in-sample Directional Accuracy. The equal selection frequency of Moving Average Crossover and Relative Strength Index indicates that indicator preference varied across sectors and periods rather than being dominated by one indicator during the in-sample stage.

The differences between in-sample and out-of-sample results demonstrate the importance of chronological validation. Indicators that performed strongly during the selection period occasionally experienced substantial declines during the subsequent validation period, while other indicators improved. Therefore, the final sector-level indicator was not determined from in-sample performance or selection frequency alone. Instead, it was based on aggregated out-of-sample Directional Accuracy across the windows in which the corresponding indicator had been selected.

The results do not establish that one indicator is universally superior. They show that historical indicator performance depended on the evaluated sector, window, selected stocks, parameters, signal rules, and observation period. Fixed-Length Rolling Walk-Forward Analysis therefore provided a structured method for separating indicator selection from later validation and for observing changes in performance across chronological periods.

Table 2. In-Sample Selection and Out-of-Sample Validation by Sector and Window

Sector	Window	Selected Indicator	In-Sample DA	Out-of-Sample DA	Active Signals	Correct Signals
Financial	1	RSI	43.75%	66.67%	15	10
Financial	2	RSI	50.00%	50.00%	8	4
Financial	3	RSI	54.17%	30.00%	10	3
Financial	4	MA Crossover	47.37%	40.00%	20	8
Technology	1	RSI	63.89%	50.00%	6	3
Technology	2	RSI	53.85%	30.00%	20	6
Technology	3	MA Crossover	58.82%	33.33%	9	3
Technology	4	MA Crossover	45.45%	80.00%	5	4

Industrial	1	RSI	56.67%	50.00%	10	5
Industrial	2	MA Crossover	61.90%	87.50%	8	7
Industrial	3	MA Crossover	72.73%	50.00%	8	4
Industrial	4	MA Crossover	68.75%	47.37%	19	9
Energy	1	MA Crossover	51.72%	66.67%	15	10
Energy	2	RSI	60.53%	63.64%	11	7
Energy	3	RSI	68.42%	35.71%	14	5
Energy	4	MA Crossover	64.29%	83.33%	6	5

3. Implementation of Decision Support Assistant

The research results are implemented in the form of a web-based decision support assistant that helps users conduct technical stock analysis based on the best-performing indicator in each sector. The system utilizes the results of the Fixed-Length Rolling Walk-Forward Analysis (WFA) evaluation as the basis for decision-making, using stored sector-level results derived from historical out-of-sample validation. With this approach, the system does not select indicators randomly but instead uses those with the best historical performance for the stock sector being analyzed.

Users begin the analysis process by entering a stock code on the main page. The system then identifies the stock sector, retrieves the best-performing indicator stored in the WFA results database, and generates technical signals along with supporting information displayed in an interactive dashboard.

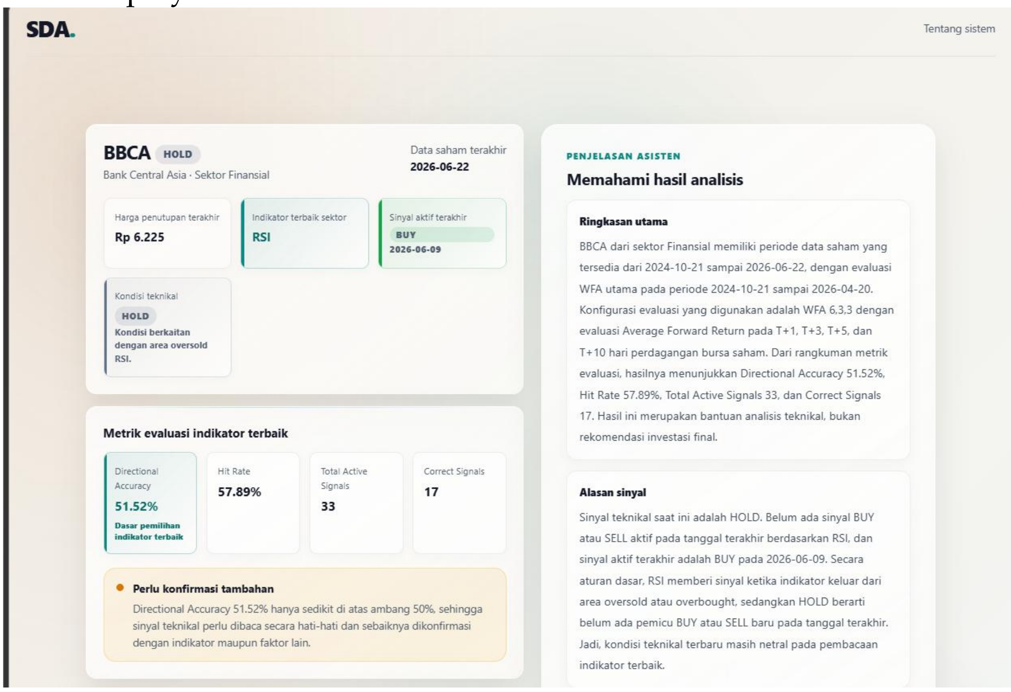
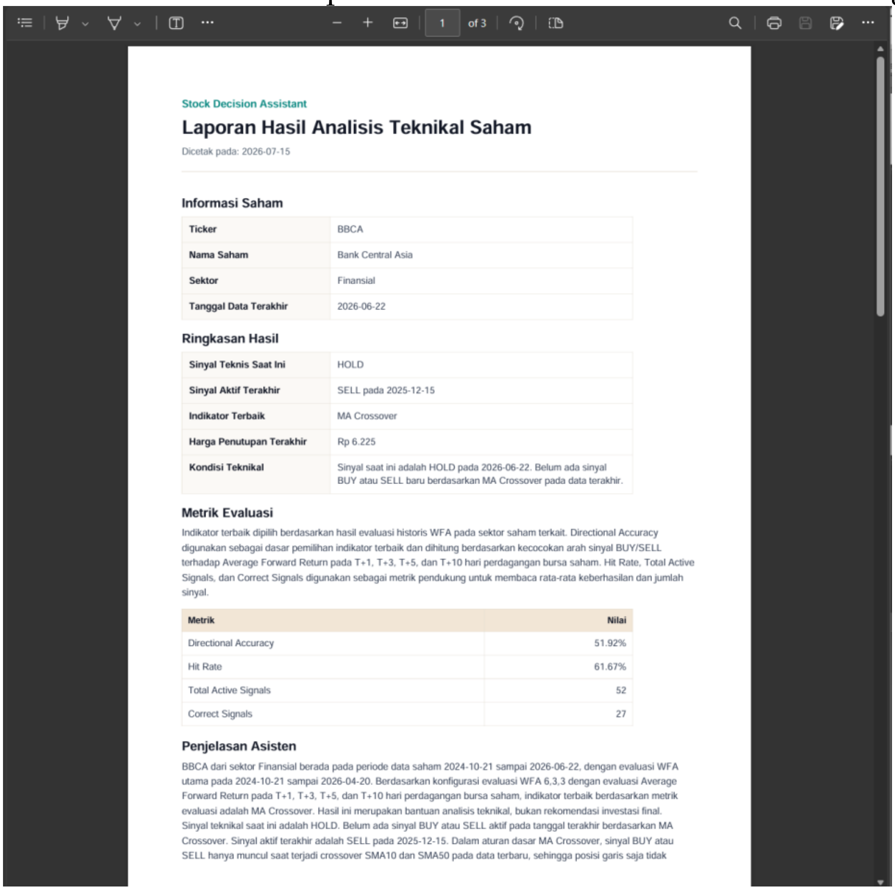


Figure 4. Decision Support Assistant Dashboard

The dashboard is designed as the primary user interface for rapid stock analysis. In addition to accepting stock code input, this page also displays concise information about the selected issuer, allowing analysis to be performed through a single interface without requiring additional configuration.

After the analysis is complete, the system displays the evaluation results, including the technical indicators used, technical signals (Buy, Sell, or Hold), a stock price movement chart, and an explanation of the analysis results generated by the Large Language Model (LLM). The LLM serves as an explanation layer that translates technical analysis results into a narrative that is easier for users to understand, without affecting the indicator calculation process or deterministic decision-making.


1 of 3

Stock Decision Assistant

Laporan Hasil Analisis Teknikal Saham

Dicetak pada: 2026-07-15

Informasi Saham

Ticker	BBCA
Nama Saham	Bank Central Asia
Sektor	Finansial
Tanggal Data Terakhir	2026-06-22

Ringkasan Hasil

Sinyal Teknis Saat Ini	HOLD
Sinyal Aktif Terakhir	SELL pada 2025-12-15
Indikator Terbaik	MA Crossover
Harga Penutupan Terakhir	Rp 6.225
Kondisi Teknikal	Sinyal saat ini adalah HOLD pada 2026-06-22. Belum ada sinyal BUY atau SELL baru berdasarkan MA Crossover pada data terakhir.

Metrik Evaluasi

Indikator terbaik dipilih berdasarkan hasil evaluasi historis WFA pada sektor saham terkait. Directional Accuracy digunakan sebagai dasar pemilihan indikator terbaik dan dihitung berdasarkan kecocokan arah sinyal BUY/SELL terhadap Average Forward Return pada T+1, T+3, T+5, dan T+10 hari perdagangan bursa saham. Hit Rate, Total Active Signals, dan Correct Signals digunakan sebagai metrik pendukung untuk membaca rata-rata keberhasilan dan jumlah sinyal.

Metrik	Nilai
Directional Accuracy	51.92%
Hit Rate	61.67%
Total Active Signals	52
Correct Signals	27

Penjelasan Asisten

BBCA dari sektor Finansial berada pada periode data saham 2024-10-21 sampai 2026-06-22, dengan evaluasi WFA utama pada 2024-10-21 sampai 2026-04-20. Berdasarkan konfigurasi evaluasi WFA 6,3,3 dengan evaluasi Average Forward Return pada T+1, T+3, T+5, dan T+10 hari perdagangan bursa saham, indikator terbaik berdasarkan metrik evaluasi adalah MA Crossover. Hasil ini merupakan bantuan analisis teknikal, bukan rekomendasi investasi final. Sinyal teknikal saat ini adalah HOLD. Belum ada sinyal BUY atau SELL aktif pada tanggal terakhir berdasarkan MA Crossover. Sinyal aktif terakhir adalah SELL pada 2025-12-15. Dalam aturan dasar MA Crossover, sinyal BUY atau SELL hanya muncul saat terjadi crossover SMA10 dan SMA50 pada data terbaru, sehingga posisi garis saja tidak

Figure 5. Stock Analysis Results

Figure 5 presents an example of the generated PDF report. The report preserved the principal information shown on the dashboard and provided a portable record of the analysis. Its contents functioned as supporting technical information and were not intended as price predictions, automated trading instructions, or final investment recommendations.

4. System Testing

System testing was conducted to evaluate the functionality, user acceptance, and usability of the developed decision support assistant. Testing consisted of Black Box Testing, User Acceptance Testing (UAT), and System Usability Scale (SUS). The combination of these three methods was used to ensure that the system not only functioned as required but was also easy to use and acceptable to users.

a. Black Box Testing

Black Box Testing was conducted by the developer using 20 predefined functional scenarios. The scenarios covered page access, stock input validation, unavailable-stock handling, retrieval of stored sector-level Walk-Forward Analysis results, technical signal generation, dashboard presentation, explanatory output and deterministic fallback, technical-term explanations, PDF report generation, report consistency, and repeated stock analysis.

All 20 scenarios produced outputs that matched the predefined expected results. This result indicates that the tested functions operated according to the specified functional requirements during the testing process.

Table 3. Black Box Testing Results

Testing Result	Number of Scenarios	Percentage
Passed	20	100%
Failed	0	0%
Total	**20**	**100%**

Source: Authors' testing results

Based on the test results in Table 3, all the main functions of the system run in accordance with the functional requirements that have been determined. The system is capable of receiving stock input, identifying issuer sectors, taking the best-performing indicator based on WFA evaluation results, producing technical signals and analysis results, displaying LLM-based explanations, and producing analysis reports in PDF format, no failure was observed within the 20 predefined testing scenarios.

b. User Acceptance Testing (UAT)

User Acceptance Testing is conducted to determine the level of user acceptance of the system based on several assessment aspects, including ease of use, clarity of information, ease of operation, and the benefits of the system in assisting the stock analysis process.

Table 4. User Acceptance Testing Result

Parameter	Result
Number of respondents	40
Number of test scenarios	15
Total responses	600
Successful responses	599
Failed responses	1
UAT success rate	99.83%

Based on Table 4, the system achieved a 99.83% success rate, with 599 successful responses out of a total of 600 test responses. Only one unsuccessful response occurred, specifically in the scenario of the main summary matching between the dashboard and the PDF report. This condition did not impact the analysis process or the system's core functionality, as all other features could be implemented according to user requirements.

The very high UAT score indicates that the system met functional requirements from a user perspective. All key processes, from stock search and sector identification, analysis result presentation, information visualization, LLM-based explanations, and PDF report generation, were successfully utilized by the majority of respondents. These results demonstrate that the developed decision support assistant is capable of supporting the technical stock analysis process effectively and is easy for users to operate.

c. System Usability Scale (SUS)

The System Usability Scale (SUS) is used to evaluate the usability level of a decision support assistant based on user perceptions after trying all of the system's main functions. The test involved the same 40 respondents as in User Acceptance Testing (UAT), so the assessment is based on direct experience using the system. The SUS instrument consists of 10 statements with a five-point Likert scale, then calculated using standard procedures to produce a score ranging from 0 to 100.

Table 5. System Usability Scale Evaluation Results

Parameter	Result
Number of respondents	40
Average SUS score	78.63
Benchmark SUS	> 68
Grade	B+
Acceptability	Acceptable
Adjective Rating	Good

Based on Table 5, the system obtained a SUS score of 78.63, which is above the benchmark value of 68. These results indicate that the system's usability level is in the good category, obtaining a Grade B+, and is included in the Acceptability Range: Acceptable. The developed interface and interaction mechanisms have met a good level of ease of use according to user perception.

The SUS scores obtained indicate that most respondents were able to use the system without experiencing significant difficulties. The stock analysis process, presentation of technical indicator evaluation results, information visualization, LLM-based explanations, and PDF report generation were easily understood and operated. However, these results also indicate that there is still room for improvement in the user experience, particularly in terms of interface simplification and information presentation to make the system more user-friendly.

It should be noted that the System Usability Scale measures user perceptions of the system's ease of use, not the performance of technical indicators, the accuracy of Fixed-Length Rolling Walk-Forward Analysis results, or the quality of investment recommendations generated. The SUS score is used as an indicator of interface quality and user experience, while Technical-indicator performance was evaluated through the WFA results, Black Box Testing and User Acceptance Testing assessed system functionality and user task completion, as discussed in the previous subsection.

Overall, Black Box Testing showed that all 20 predefined functional scenarios passed, User Acceptance Testing produced a success rate of 99.83%, and the System Usability Scale yielded a mean score of 78.63. These results indicate that the developed system operated according to the tested functional requirements, supported successful completion of nearly all predefined user tasks, and achieved a positive perceived-usability evaluation. These findings concern system functionality and usability and should not be interpreted as evidence of future indicator performance or investment effectiveness

The findings show that Fixed-Length Rolling Walk-Forward Analysis identified the technical indicator with the highest observed aggregated out-of-sample Directional Accuracy for each evaluated sector. Relative Strength Index achieved the highest result in the financial sector, whereas Moving Average Crossover achieved the highest results in the Technology, Industrial, and Energy sectors. These results indicate that observed technical-indicator performance differed across sectors and evaluation periods, rather than remaining uniform across all stocks. This finding is consistent with previous studies reporting that technical-rule performance may vary across assets, market conditions, and evaluation periods (Hasan et al., 2024; Lento & Gradojevic, 2022).

The adoption of Fixed-Length Rolling Walk-Forward Analysis provided a chronological evaluation structure that separated in-sample indicator selection from subsequent out-of-sample validation. Unlike a single static split, the rolling configuration allowed indicator performance to be observed across multiple sequential periods and reduced dependence on one particular data partition. The final sector-level results were based on aggregated out-of-sample performance rather than in-sample performance alone (Cerqueira et al., 2021; Deep et al., 2025).

The evaluation results were subsequently implemented in a web-based decision support assistant that integrates sector identification, sector-specific indicator selection, technical signal generation, data visualization, and Large Language Model (LLM)-based explanations. In the proposed system, the LLM is not responsible for calculating technical indicators or generating technical signals. Instead, it functions exclusively as an explanation layer that translates deterministic computational outputs into natural-language descriptions. This design preserves analytical consistency while minimizing the potential errors associated with relying on generative models for core computational tasks.

System evaluation showed that all 20 predefined Black Box Testing scenarios passed, User Acceptance Testing achieved a success rate of 99.83%, and the System Usability Scale yielded a mean score of 78.63, corresponding to the Good, Grade B+, and Acceptable categories under the interpretation framework used in this study. These results indicate that the tested functions operated according to the specified requirements, nearly all predefined user tasks were completed successfully, and respondents generally perceived the system as usable. However, these results concern functionality, task completion, and perceived usability and do not demonstrate future technical-indicator performance or investment effectiveness.

This study contributes by integrating sector-specific technical-indicator evaluation using Fixed-Length Rolling Walk-Forward Analysis into a web-based decision support system supported by Large Language Model-generated explanations. The proposed framework enables users to access sector-level indicator results derived from historical out-of-sample validation together with concise and interpretable analytical explanations. Nevertheless, the study is limited to four sectors, three technical indicators, fixed parameters, four Walk-Forward Analysis windows, and the selected observation period. Future research may incorporate additional sectors, a broader range of indicators, alternative Walk-Forward configurations, and longer evaluation periods to examine the consistency and broader applicability of the proposed framework.

E. CONCLUSION

This study developed a web-based decision support assistant for sector-specific stock technical analysis by integrating Fixed-Length Rolling Walk-Forward Analysis (WFA) and a Large Language Model (LLM). The results show that the best-performing technical indicator varied across sectors, with the Relative Strength Index (RSI) achieving the highest observed out-of-sample performance in the financial sector and the Moving Average Crossover (MA Crossover) achieving the highest performance in the Technology, Industrial, and Energy sectors. The proposed system integrates sector identification, indicator selection, technical signal generation, data visualization, and LLM-based explanations into a unified workflow while maintaining deterministic analytical computations. System evaluation showed that all predefined Black Box Testing scenarios passed, User Acceptance Testing achieved a success rate of 99.83%, and the System Usability Scale (SUS) produced a score of 78.63. These findings indicate that the proposed system can function as a practical decision support tool for sector-specific stock technical analysis within the scope of the study. Future research may include additional sectors, technical indicators, and longer evaluation periods.

REFERENCES

1. Aliyah, A., Hartono, N., & Muin, A. A. (2025). Penggunaan User Acceptance Testing (UAT) pada pengujian sistem informasi pengelolaan keuangan dan inventaris barang. *Switch: Jurnal Sains dan Teknologi Informasi*, *3*(2), 42–58.
2. Al-Nefaie, A. H., & Aldhyani, T. H. H. (2022). Predicting close price in emerging Saudi stock exchange: Time series models. *Electronics*, *11*(21), 3443.
3. Cerqueira, V., Torgo, L., & Soares, C. (2021). Model selection for time series forecasting: Empirical analysis of different estimators. *ArXiv Preprint ArXiv:2104.00584*.
4. Cinelli, M., Kadziński, M., Miebs, G., Gonzalez, M., & Słowiński, R. (2022). Recommending multiple criteria decision analysis methods with a new taxonomy-

based decision support system. *European Journal of Operational Research*, *302*(2), 633–651.

5. Deep, G., Deep, A., & Lamptey, W. (2025). Interpretable Hypothesis-Driven Trading: A Rigorous Walk-Forward Validation Framework for Market Microstructure Signals. *ArXiv Preprint ArXiv:2512.12924*.
6. Fitriyah, F., & Mukminatin, Z. N. (2025). Return connectedness of sectoral stock indices on the Indonesian stock exchange during Russian-Ukraine conflict. *Journal of Accounting and Investment*, *26*(1), 155–172.
7. Grier, R. A., Bangor, A., Kortum, P., & Peres, S. C. (2013). The system usability scale: Beyond standard usability testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 187–191.
8. Gustinov, M. D., Azani, N. W., Al Ghani, R., Auliani, S. N., Maharani, S., Hamzah, M. L., & Rizki, M. (2023). Analysis of web-based e-commerce testing using black box and white box methods. *International Journal of Information System and Innovation Management (IJISIM)*, *1*(1), 20–31.
9. Hasan, S., Nurhasanah, S., & Santoso, W. P. (2024). Analisis teknikal menggunakan moving average (MA), moving average convergence-divergence (MACD), dan relative strength index (RSI) untuk mengoptimalkan dalam pengambilan keputusan investasi pada saham sektor manufaktur index LQ45 BEI tahun 2022-2023. *El-Mal: Jurnal Kajian Ekonomi & Bisnis Islam*, *5*(4), 3318–3334.
10. Hermansah, L., Murhadi, M., & Saputro, W. T. (2025). User Acceptance Testing to Assess User Receptiveness Toward a Soft Skills Training Information System. *Sistemasi: Jurnal Sistem Informasi*, *14*(5), 2097–2112.
11. Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation*, *9*(1), 26.
12. Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Leigh, S., & Hunt, S. (2022). System usability scale benchmarking for digital health apps: meta-analysis. *JMIR MHealth and UHealth*, *10*(8), e37290.
13. Kang, H., & Liu, X.-Y. (2023). Deficiency of large language models in finance: An empirical examination of hallucination. *ArXiv Preprint ArXiv:2311.15548*.
14. Kasmar, A. F., Hadi, R., & Asri, E. (2025). Black-Box Testing pada Sistem Evaluasi Capaian Pembelajaran Berdasarkan Outcome Based Education. *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, *6*(2), 203–212.
15. Kostikova, A., Wang, Z., Bajri, D., Pütz, O., Paaßen, B., & Eger, S. (2026). Lllms: A data-driven survey of evolving research on limitations of large language models. *ACM Computing Surveys*, *58*(11), 1–33.
16. Lee, J., Stevens, N., & Han, S. C. (2025). Large language models in finance (finllms). *Neural Computing and Applications*, *37*(30), 24853–24867.
17. Lento, C., & Gradojevic, N. (2022). The profitability of technical analysis during the COVID-19 market meltdown. *Journal of Risk and Financial Management*, *15*(5), 192.

18. Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large language models in finance: A survey. *Proceedings of the Fourth ACM International Conference on AI in Finance*, 374–382.
19. Lin, H.-Y., ChiangLin, C.-Y., & Tseng, H.-W. (2024). Optimal Parameter Selection and Indicator Design for Technical Analysis Strategies by Computer Software: An Empirical Analysis of the Taiwan Futures Market. *Engineering Proceedings*, *74*(1), 56.
20. Mazumder, S., Neogy, S., & Das, S. (2023). Simple moving average (SMA) crossover strategy with buy sell indicator. *Advancement in Management and Technology (AMT)*, *3*(4), 26–40.
21. Mejasson, P., Petridis, M., Knight, B., Soper, A., & Norman, P. (2001). Intelligent design assistant (IDA): a case base reasoning system for material and design. *Materials & Design*, *22*(3), 163–170.
22. Mirza, F. K., Pekcan, Ö., Hekimoğlu, M., & Baykaş, T. (2025). Stock price forecasting through symbolic dynamics and state transition graphs with a convolutional recurrent neural network architecture. *Neural Computing and Applications*, *37*(20), 15855–15890.
23. Nemtinov, V., Zazulya, A., Kapustin, V., & Nemtinova, Y. (2019). Analysis of decision-making options in complex technical system design. *Journal of Physics: Conference Series*, *1278*(1), 012018.
24. Otto, P., Fassò, A., & Maranzano, P. (2024). A review of regularised estimation methods and cross-validation in spatiotemporal statistics. *Statistic Surveys*, *18*, 299–340.
25. Rajpal, S. S., Mahadeva, R., Goyal, A. K., & Sarda, V. (2025). Improving forecasting accuracy of stock market indices utilizing attention-based LSTM networks with a novel asymmetric loss function. *AI*, *6*(10), 268.
26. Sahoo, S., & Kumar, S. (2024). Volatility spillover among the sectors of emerging and developed markets: a hedging perspective. *Cogent Economics & Finance*, *12*(1), 2316048.
27. Vanesha, N. A., Rizky, R., & Purwanto, A. (2024). Comparison Between Usability and User Acceptance Testing on Educational Game Assessment. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *13*(2), 210–215.
28. Yasunobu, C., Kosaka, M., Yokomura, K., & Honda, K. (1992). A decision support system building tool with fuzzy logic and its application to chart technical analysis. In *Adaptive Intelligent Systems* (pp. 19–32). Elsevier.
29. Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., & Dong, Z. (2023). A survey of large language models. *ArXiv Preprint ArXiv:2303.18223*, *1*(2), 1–124.