Jurnal Komputer dan Informatika
Vol 18 No 1, June 2025: page

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Explainable AI for LSTM-based Stock Price Forecasting: A SHAP-powered Interpretation

Teddy Agustinus[1]
[1] Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
E-mail: [1]teddy.825230152@stu.untar.ac.id

Abstract

Forecasting stock prices with deep learning models like Long Short-Term Memory (LSTM) provides impressive predictive accuracy; however, the 'black box' nature of these models presents a significant hurdle to transparency and trust. This study creates and analyzes stacked LSTM models to predict the daily closing values of four prominent technology stocks—Meta, Tesla, Microsoft, and Oracle—utilizing a 60-day look-back window on historical data spanning from 2015 to 2025. The models exhibit strong predictive performance, consistently achieving accuracy rates exceeding 95% for all stocks, assessed through RMSE and MAPE. To clarify the decision-making process of the model, the SHAP framework was utilized for interpretation. The SHAP analysis revealed a consistent and surprising primary conclusion: for each of the four stocks, the most distant historical data point (t-59) was overwhelmingly the most significant factor in predicting the following day's price. Conversely, the influence of more recent prices was found to be relatively minimal. This research confirms the high predictive strength of LSTMs but, more importantly, emphasizes the crucial role of Explainable AI (XAI) in uncovering unexpected model reasoning. The results challenge the traditional belief in recency bias within financial markets and underscore the importance of validating a model's internal logic beyond its apparent accuracy.

Keywords: *Stock Price Forecasting, Long Short-Term Memory (LSTM), Explainable AI (XAI), SHAP, Time Series Analysis.*

Abstrak

Peramalan harga saham menggunakan model deep learning seperti *Long Short-Term Memory* (LSTM) menawarkan akurasi prediksi yang tinggi, namun sifatnya sebagai 'kotak hitam' menjadi tantangan signifikan terhadap transparansi dan kepercayaan. Penelitian ini mengembangkan dan menginterpretasikan model *stacked* LSTM untuk meramalkan harga penutupan harian dari empat saham teknologi utama — Meta, Tesla, Microsoft, dan Oracle —menggunakan periode look-back 60 hari pada data historis dari tahun 2015 hingga 2025. Model menunjukkan kemampuan prediksi yang tangguh, secara konsisten mencapai tingkat akurasi di atas 95% untuk semua saham, yang diukur dengan metrik RMSE dan MAPE. Untuk membuka proses pengambilan keputusan model, kerangka kerja SHAP (*Shapley Additive exPlanations*) diterapkan untuk interpretasi. Analisis SHAP menghasilkan temuan utama yang konsisten dan berlawanan dengan intuisi: untuk keempat saham, titik data historis terjauh (t-59) secara luar biasa menjadi fitur paling berpengaruh dalam menentukan prediksi harga hari berikutnya, sementara dampak dari harga yang lebih baru ditemukan dapat diabaikan. Studi ini memvalidasi kekuatan prediksi LSTM, namun yang lebih penting, menggarisbawahi peran krusial dari *Explainable AI* (XAI) dalam mengungkap heuristik model yang tidak terduga. Temuan ini menantang asumsi umum mengenai bias rekam jejak terbaru (*recency bias*) di pasar keuangan dan menyoroti perlunya memvalidasi logika internal sebuah model di luar akurasi permukaannya.

Kata kunci*: Prediksi Harga Saham, Long Short-Term Memory (LSTM), Explainable AI (XAI), SHAP, Analisis Deret Waktu.*

1. Introduction

Forecasting stock market prices remains a fundamental challenge in the financial domain, owing to the market's intrinsically random, non-linear, and ever-changing nature. For decades, investors and analysts have pursued reliable methods for projecting market movements to maximize returns and mitigate risks. The sheer complexity of financial time series, influenced by a multitude of economic, political, and social factors, renders accurate prediction a formidable task. The emergence of machine learning introduced a paradigm shift in addressing this problem. Machine learning and deep learning algorithms have demonstrated a superior ability to outperform traditional statistical models by capturing complex, non-linear patterns within historical price data. Consequently, this has spurred the research community to explore higher-order information in time series and conduct comparative analyses of various algorithms to identify the most effective models for predicting market trends.

In the area of deep learning, LSTM networks have garnered significant attention. These networks represent a specific type of RNN designed to identify patterns within lengthy sequences of data. Their inherent ability to handle long-term dependencies makes them a preferred choice for time series forecasting, with stock price prediction being a prime example of their application. Their efficacy has been consistently validated across various financial instruments, including the Indonesian Syariah stock index, individual company stocks like Tesla, and even volatile cryptocurrencies like Bitcoin [5]–[8]. The successful application of LSTM and its variants, such as the Gated Recurrent Unit (GRU), in other time-series domains like hydrological forecasting further underscores their robustness [9].

To further enhance predictive accuracy, recent research has focused on augmenting LSTM models with external, unstructured data sources. A prominent trend has been the integration of sentiment analysis from social media platforms, particularly Twitter, and news outlets. The rationale is that public sentiment often serves as a leading indicator of market behavior. Numerous studies have successfully combined historical price data with sentiment scores to create more powerful and context-aware forecasting models, demonstrating improved performance under conditions of high uncertainty [10]–[14].

However, the increasing complexity and performance of models like LSTMs have an Achilles' heel: their lack of transparency. The 'black box' nature of these models hinders a clear understanding of their decision-making process, making the rationale for their predictions difficult to trace. This opacity is a significant barrier to adoption in high-stakes environments, such as finance, where understanding the "why" is as crucial as the "what." This very problem has catalyzed the growth of Explainable AI (XAI), a discipline focused on developing techniques to provide insight into the internal workings of otherwise opaque models. The integration of XAI has been identified as a critical step in extending the capabilities and trustworthiness of machine learning in financial forecasting [15].

This research addresses this critical gap by leveraging SHAP (SHapley Additive exPlanations), a state-of-the-art XAI framework, to interpret the predictions of an LSTM-based stock forecasting model. By applying SHAP, this study aims not only to predict future stock prices but also to provide a clear, quantifiable explanation of which features, such as past prices or specific time lags, contribute most significantly to each forecast. This approach seeks to bridge the gap between predictive power and interpretability, fostering greater trust and insight into AI-driven financial decision-making.

2. Related Work

Bacco et al. in [1] investigate the prediction of stock price movements for 26 major banks in

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

North American and European markets throughout 2022, a period marked by significant economic and geopolitical uncertainty. They combine Long Short-Term Memory (LSTM) networks with FinBERT to model financial time series and perform sentiment analysis on approximately 14,000 tweets from key economic figures. Their experiments examine the effects of market region, bank-level versus aggregated-bank data, varying historical data spans, and the inclusion of public sentiment on model accuracy. The results reveal that models trained on more recent data and augmented with Twitter-derived sentiment achieve higher F1-scores, particularly for large North American banks, which also exhibit stronger inter-bank correlations. In contrast, the more fragmented European market benefits most from a global aggregation approach. These findings underscore the need to tailor predictive strategies to specific market dynamics and leverage non-traditional data sources, such as social media sentiment.

Wen et al. [2] propose a novel framework for predicting up or down trends in financial time series by first extracting frequent motifs to denoise and compact raw price data, then reconstructing the series by stitching these motifs in chronological order. They feed the reconstructed sequences into a one-dimensional The model utilizes a Convolutional Neural Network (CNN) engineered with two convolutional layers and fully connected layers, a design capable of learning sophisticated spatial relationships Experiments on S&P 500 and individual stocks (e.g., GOOGL, IBM, BA) show their method outperforms ARIMA, Wavelet+ARIMA, HMM, LSTM, and SFM baselines by 4 - 7 % in accuracy, with notable gains in recall and precision. They demonstrate that motif-based reconstruction filters out low-frequency noise while preserving trend information, enabling more efficient training than recurrent models. Overall, this work highlights the advantage of leveraging high-order structural features and CNNs for more accurate and computationally efficient stock trend forecasting.

Comparative studies have been frequently conducted to evaluate a diverse range of predictive models. For instance, research has compared the efficacy of Several machine learning techniques such as Decision Tree, Random Forest, and Support Vector Classification with deep learning approaches like RNN and LSTM for predicting up/down trends in four sectors of the Tehran Stock Exchange, utilizing ten technical indicators over a ten-year period. They evaluate each model on both continuous input data and binary-transformed data that encodes indicator-specific up or down signals. Their experiment assessed via F1-score, accuracy, and ROC-AUC shows that all models improve substantially when fed binary data, with deep learning approaches (RNN and LSTM) achieving the highest F1-scores (up to ~90%). Machine learning models also perform better under binary preprocessing, narrowing the gap with deep learners. Overall, this study demonstrates that preprocessing financial features into trend-deterministic binary form and leveraging recurrent architectures yield more accurate and efficient stock trend forecasts.

Khan et al. in [4] develop a blended model for stock market forecasting that combines the strengths of various machine learning classifiers. with sentiment features extracted from social media (Twitter) and financial news headlines. They preprocess data by performing feature selection and spam-tweet reduction before training ten classifiers (e.g., SVM, Random Forest, AdaBoost) and deep learning models to forecast ten-day market trends. Their experiments on multiple markets reveal peak accuracies of 80.53% using social media sentiment and 75.16% using news sentiment, while an ensemble of Random Forest classifiers achieves the highest overall accuracy of 83.22%. They also identify which stocks are hardest to predict (e.g., New York, Red Hat) and which are most influenced by social media (e.g., NYSE, IBM) versus financial news (e.g., LSE, Microsoft). Their results demonstrate the value of combining external sentiment data with traditional price features and ensembled learning for more robust market forecasts.

Thormann et al. in [5] address the challenge of predicting stock price trends by utilizing neural networks, with a specific focus on sentiment analysis of Twitter data and Long Short-Term Memory (LSTM) networks. The study highlights Twitter's potential for such analysis due to its vast user base and real-time information flow, making it valuable for forecasting financial trends. It introduces financial feature engineering and LSTM architecture for forecasting, presenting a method to collect and process tweets for sentiment analysis and combine them with financial technical indicators. The research demonstrates that a model combining financial and Twitter features can outperform a baseline LSTM model (using only lagged close prices) in forecasting Apple's stock prices 30 and 60 minutes ahead, with the code for replication provided.

Peivandizadeh et al. [6] propose a two-stage framework that enhances short-term stock price forecasting by fusing social media sentiment with historical market data. First, the authors address the challenge of imbalanced sentiment classes by introducing an Off-policy Proximal Policy Optimization (PPO) algorithm whose reward function is tailored to prioritize the correct classification of minority sentiments. Next, they deploy a Transductive Long Short-Term Memory (TLSTM) network that dynamically weights training examples based on their temporal closeness to the prediction point, thereby better capturing recent market dynamics. Their dataset comprises over 12,000 daily news articles and detailed price and volume records from 50 stocks and 10 ETFs on the Indian NSE, spanning 2015–2020. The text has been preprocessed via BERT embeddings and three dilated convolutional layers for feature extraction. Empirical evaluation shows that the integrated model achieves an RMSE of 2.147 for stock forecasting and an F-measure of 89% for sentiment analysis, outperforming ablated variants and confirming the value of both Off-policy PPO and TLSTM components.

Otabek and Choi in [7]. This study investigates how classifying Bitcoin-related tweets by four key attributes poster followers, comments, likes, and retweets—can improve price prediction while reducing computational costs. The authors collect over 5 million tweets from 2014 to 2018, preprocess them (noise removal and sentiment scoring via VADER), and split the data into four attribute-based subsets. They formulate Bitcoin price forecasting as a reinforcement learning task, defining states as actual prices, actions as percentage price change predictions, and rewards through three custom functions within a Q-learning framework. Experiments identify tweets from users with the most followers as the most predictive subset. Using only that subset, their Q-learning model achieves predictions that are 12.5% more accurate than a classic approach using all tweets, while consuming 88.8% less CPU time and 80% less overall runtime. The findings demonstrate that targeted tweet attribute filtering combined with Q-learning enhances both forecast accuracy and resource efficiency.

Anandita and Wahyuningsih conducted a study in [8] that examined the predictability of the Indonesian Sharia Stock Index (ISSI) using a Long Short-Term Memory (LSTM) model, which is a type of recurrent neural network recognized for its efficacy in analyzing time series data. The research leveraged historical ISSI data from 2017 to 2022 for the training phase. To improve model convergence and performance, the entire dataset underwent a preprocessing step using a Min-Max Scaler for normalization. The architecture of the LSTM included an input layer, one LSTM layer, and a dense layer, with hyperparameters adjusted through experimentation to optimize accuracy. When evaluated on the test set, the model achieved an exceptionally low Mean Absolute Percentage Error (MAPE) of 0.018%, reflecting a high level of forecasting accuracy. This result implies that LSTM serves as a reliable approach for predicting Sharia-compliant indices and can act as an efficient decision-support tool for investors.

Arfan & Lussiana in [9]. This study compares the performance of Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) algorithms in forecasting Indonesian stock prices, focusing on PT Unilever Tbk, PT Kimia Farma, and PT Gudang Garam over the period from 2016 to 2019. The research process involved collecting historical stock and currency data, then preprocessing it through merging, imputation, and Min–Max normalization, followed by splitting the data into training and test sets and designing the LSTM model. The LSTM was optimized by varying the number of hidden layers, memory cells, input sample size, and training epochs, while the SVR was evaluated across different input windows (1, 30, and 60 days). Test results reveal that LSTM consistently achieved much lower Mean Squared Error (MSE) than SVR, approximately 0.0013 versus 0.0025 for 1-day windows, with the gap widening at 60-day windows. Although LSTM training time (around 8 minutes) was significantly longer than SVR's (about 10 seconds), LSTM demonstrated superior capability in capturing long-term dependencies and yielding more accurate forecasts. The best LSTM configuration, with one hidden layer, 150 memory cells, 10-day samples, and 20 epochs, produced the lowest MSE across all three stocks (UNVR: 0.0016, KAEF: 0.0013, GGRM: 0.0017).

Hanafiah et al. in [10] conducted a study that utilized a Recurrent Neural Network employing a Long Short-Term Memory (LSTM) methodology to predict stock prices for three Indonesian companies, UNVR.JK, KAEF.JK, and GGRM.JK, during the period from 2016 to 2019. Historical price data were sourced from Yahoo Finance, and currency exchange rates were collected from OFX, then merged, filled for missing entries, and normalized using Min–Max scaling prior to being divided into training and testing datasets. The architecture of the LSTM was meticulously optimized by adjusting the number of hidden layers (ranging from 1 to 5), memory cells (between 50 and 150),

 ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

sample window size (from 10 to 50), and training epochs (set between 10 and 50), utilizing the RMSProp optimizer until achieving convergence. The ideal configuration consisted of one hidden layer, 150 memory cells, a batch size of 10, and 20 epochs, which resulted in the lowest Mean Squared Error (MSE) of 0.0016 for UNVR, 0.0013 for KAEF, and 0.0017 for GGRM.

Febriansyah et al. conducted a study in [12] that explored the application of a Long Short-Term Memory (LSTM) network for predicting daily Bitcoin prices, leveraging the model's capabilities in managing time series data. The research employed a historical dataset covering the period from December 2020 to April 2024, which was preprocessed to address missing values and normalize features using a Min-Max scaler. To identify the best configuration, the LSTM model was trained through three distinct experiments (for 1, 10, and 20 epochs) utilizing the Nadam optimizer. The evaluation of performance demonstrated a strong in-sample fit, resulting in an RMSE of 17,318.40 and a MAPE of 3.24% on the training dataset. However, a significant drop in accuracy was noted on the test dataset, which produced an RMSE of 27,921.84 and a MAPE of 5.36%. This difference indicates a possible overfitting of the training data. The authors concluded that although LSTM effectively captures price dynamics, future investigations should aim to improve generalization through enhanced regularization methods and the inclusion of external macroeconomic factors.

Çelik, T. B., and colleagues in [13] have successfully utilized machine learning techniques to predict the movement of stock market prices. Nevertheless, two significant limitations are present: the opaque nature of ML, which complicates the explanation of prediction reasoning, and the difficulty in attaining notable improvements in accuracy that warrant the added computational effort. To tackle this issue, this research suggests an explainable Artificial Intelligence (XAI) methodology to evaluate the trustworthiness of predictions, allowing decision-makers to prioritize more reliable forecasts while steering clear of lower-quality ones. The research introduces a novel two-step stacking ensemble approach that combines machine learning (ML), empirical mode decomposition (EMD), and explainable artificial intelligence (XAI) utilizing local interpretable model-agnostic explanations (LIME). The experimental findings indicate that this proposed model achieved a peak accuracy of 0.9913 with trusted predictions on the KOSPI dataset.

Waqas et al. [14] have highlighted the rapid advancements in Artificial Intelligence (AI), particularly its significant impact on the water sector, especially in hydrological predictions. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models have become central to this progress. While RNNs serve as foundational models, they face challenges such as vanishing gradients, which limit their capacity to model long-term dependencies effectively. To address these issues, variants like LSTM and Gated Recurrent Units (GRU) were introduced, offering improved efficacy in handling nonlinear and time-variant hydrological data. Moreover, incorporating attention mechanisms and hybrid models that merge different deep learning approaches boosts prediction accuracy by effectively capturing both temporal and spatial dependencies. Nevertheless, the deployment of these complex models presents difficulties due to the need for extensive datasets and considerable computational power. Consequently, future research ought to focus on creating interpretable architectures, improving data quality, and utilizing transfer learning to enhance model generalization.

Cho et al. [15] developed a predictive model for water levels, aimed at reducing the economic effects and loss of life caused by flooding, which is an increasing global issue. This model employs sophisticated deep learning methodologies, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) models, which are utilized. For its predictions, the model relies on various meteorological data inputs, including upstream and downstream water levels, temperature, humidity, and precipitation data gathered from the Automated Synoptic Observing System (ASOS). Furthermore, the model has demonstrated its effectiveness in accurately predicting historical peak water levels with minimal error. In the future, it is anticipated that this model will be incorporated into a flood risk management system to aid in early evacuation planning.

3. Materials and Methods

Figure 1 displays the flowchart outlining the research process, which commences with Data Acquisition and continues with Data Preprocessing, Model Development, Model Training, and XAI utilizing SHAP. This study constructs and interprets Long Short-Term Memory (LSTM) models to predict stock prices for major technology companies, including Meta, Microsoft, Oracle, and Tesla, utilizing Explainable AI (SHAP) to identify and clarify the elements influencing each forecast.

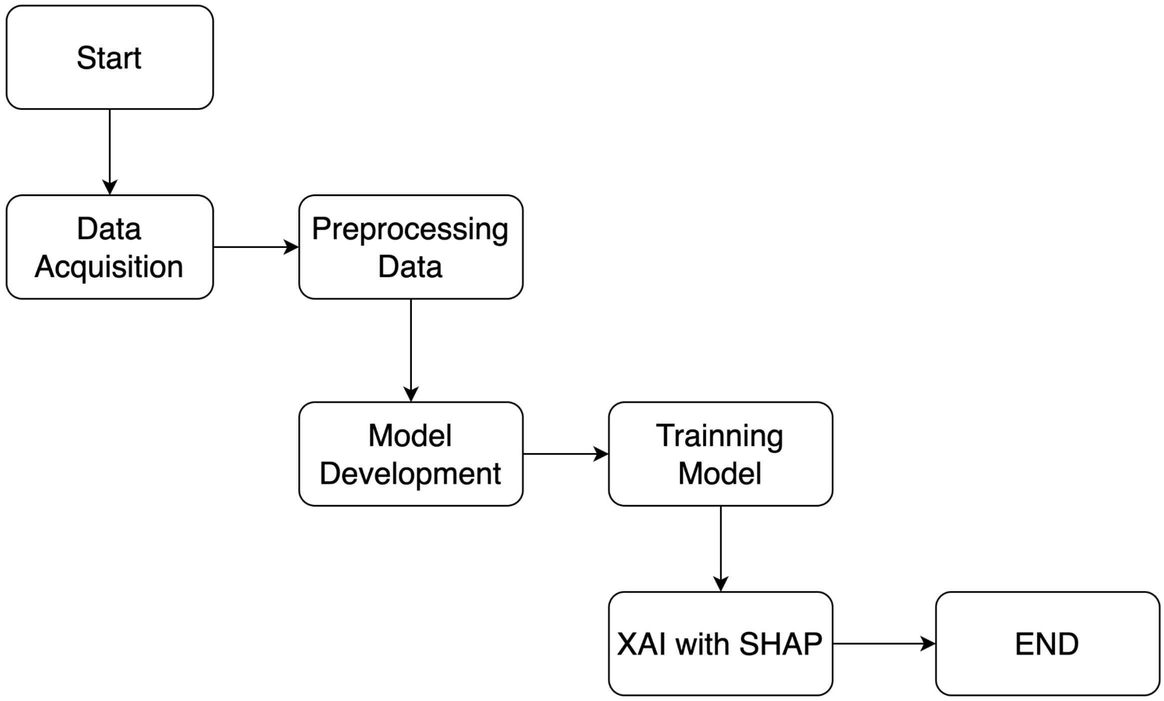

Figure 1. Research Procedure

3.1 Data Acquisition

The dataset for this research comprises historical daily stock price data for four leading American technology companies: Meta Platforms, Inc. (META), Tesla, Inc. (TESLA), Oracle Corporation (ORACLE), and Microsoft Corporation (MSFT). These companies were selected to represent influential and often volatile entities within the technology sector, providing a robust test for the forecasting model. The data was obtained from investing.com, a publicly accessible financial data portal, and spans a comprehensive 10-year period from January 2015 to June 2025. This extensive timeframe was chosen to ensure the model is trained on a wide variety of market conditions, including periods of growth, stability, and high volatility. For each company, the dataset contains several key features, including the daily opening price ('Open'), highest price ('High'), lowest price ('Low'), closing price ('Close'), trading volume ('Vol.'), and the daily percentage change ('Change %'). While multiple features were available, this study focuses specifically on forecasting the daily 'Close' price, as it is the most standard indicator of a stock's value. A summary of the datasets is detailed in Table 1.

Table 1. Dataset Information

Dataset name	Description	File format
META	Daily stock prices for Meta Platforms Inc. (META), including open, high, low, close, and trading volume.	csv
MICROSOFT	Daily stock prices for Microsoft Corporation (MICROSOFT), including open, high, low, close, and trading volume.	csv
ORACLE	Daily stock prices for Oracle Corporation (ORACLE), including open, high, low, close, and trading volume.	csv
TESLA	Daily stock prices for Tesla Inc. (TESLA) including open, high, low, close, and trading volume.	csv

3.2 Preprocessing data

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

During the data preprocessing stage, the raw stock price data for each company (Meta, Microsoft, Oracle, and Tesla) was obtained from the respective CSV files. The 'Date' column was initially changed into a datetime object, and then the DataFrame's index was designated to facilitate time-series operations. Essential elements for forecasting include 'Open', 'High', 'Low', 'Close', and 'Volume'. Were extracted next. All numerical features were then normalized using MinMaxScaler to bring their values within the range of 0 to 1, which is essential for the stable and efficient training of neural networks, such as Long Short-Term Memory (LSTM) networks. After normalization, the data was organized into sequential inputs suitable for the LSTM model, where each input sequence consists of a specified sequence length (e.g., 60) previous timesteps to forecast the value at the next timestep. Ultimately, the arranged sequential data was split into training, validation, and testing sets utilizing a time-series split strategy, with 80% designated for training and the remaining 20% allocated for testing. This method guaranteed that there was no information transfer from future time points into the past.

3.3 Model Development

The predictive model utilized in this study was created using a stacked Long Short-Term Memory (LSTM) network, a deep learning framework skilled at understanding temporal dependencies within time series data. The model was built with the Keras Sequential API, where layers are incorporated in a linear arrangement.

3.3.1 Network Architecture

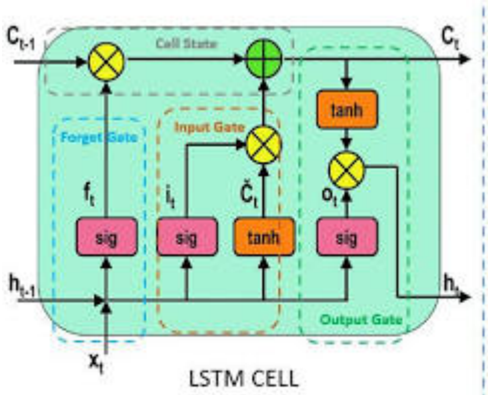

Figure 2. LSTM Architecture

The LSTM network is structured with the following components:

- The first layer is an input LSTM layer that contains 50 LSTM units and is designed with an input shape to handle sequences of 60 time steps.
- Next, a Dropout layer is included, set at a rate of 0.2, which serves as a regularization method. This technique reduces overfitting by randomly deactivating 20% of neuron activations during training.
- Following this is a hidden LSTM layer that also has 50 units[11], which processes the output from the preceding layer. This layer is configured to produce only the result from the final time step, effectively condensing the sequence information into one vector.
- The last layer is a Dense layer featuring a single neuron that outputs a continuous numerical value, representing the predicted stock price for the next day.

3.3.2 Model Compilation

Prior to commencing training, the model is set up with essential elements that direct the learning process.

- Optimizer: The Adam optimizer is selected due to its computational efficiency and its capability to adjust the learning rate, which makes it especially effective for training deep neural networks.
- Loss Function: The Mean Squared Error (MSE) acts as the loss function, quantifying

the disparity between the model's predictions and the actual target values. The objective of the model is to minimize this MSE value during the training process.

3.4 Model Training and Evaluation

Following the model's architectural design, this section outlines the crucial phases of model training and performance evaluation. It begins by detailing the training procedure, in which the model learns patterns from historical data through an iterative process. Subsequently, the methodology for quantitatively assessing the trained model's predictive accuracy is described, focusing on the standard regression metrics used to measure its effectiveness and generalization capability on the test set.

3.4.1 Training Process

The model is trained utilizing the fit method on the prepared training dataset (X_train, y_train). Training occurs over 50 epochs, with each epoch representing a complete pass through the entire training dataset. The data is processed in batches of 32, meaning the model's internal parameters are updated after every 32 samples. A validation set (validation_data) is utilized at the conclusion of each epoch to assess the model's performance on previously unseen data, which is essential for tracking its ability to generalize and for identifying overfitting.

3.4.2 Performance Evaluation Metrics

After completing the training phase, a thorough evaluation of the model's predictive performance is carried out to assess its capacity to generalize to novel, unseen data. This numerical evaluation is performed solely on the test set, which was reserved during the training phase, employing two standard regression metrics to deliver a detailed assessment of its accuracy.

A. Root Mean Squared Error (RMSE)

RMSE measures the square root of the average of the squared differences between the predicted and actual values. It is a widely used metric that heavily penalizes larger errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Where:

- n Is the total number of samples
- y_i Is the actual stock price
- $\hat{y}_i$ Is the predicted price

The result is in the same unit as the predicted value (e.g., USD), providing an interpretable measure of the model's average error magnitude.

B. Mean Absolute Percentage Error (MAPE)

MAPE computes the mean absolute percentage error. This provides an intuitive measure of relative error, which helps compare performance across stocks with different price scales.

$$MAPE = \sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right| \times 100\backslash\%$$

Where:

- n Is the total number of samples
- y_i Is the actual stock price
- $\hat{y}_i$ Is the predicted price

3.4.3 Explainable AI (SHAP) Method

To achieve interpretability in models, this study employs the SHAP (Shapley Additive exPlanations) framework. SHAP leverages principles from game theory to clarify the predictions of any machine learning model by fairly assessing the impact of each feature on a given outcome. At the heart of SHAP is an additive feature attribution model, where an

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

explanation function, g, approximates the original complex model by expressing a single prediction as the sum of a base value along with the contributions from each input feature. The subsequent formula outlines this explanation model:

$$g(z') = \phi_0 + \sum_{i=1}^{m} \phi_i \ z'_i$$

Where:

- $g(z')$ Is the explanation model suitable for a simplified input?
- ϕ_0 It represents the base value, which is the average prediction across the entire dataset.
- m Is the number of input features (in this case, 60 time steps).
- ϕ_i Is the SHAP value for feature i, representing the specific contribution of that feature to push the prediction away from the base value.
- z'_i It is a binary variable representing the presence (1) or absence (0) of feature i.

In the context of this research, SHAP values are calculated for each of the 60 historical time steps to determine which days most significantly influence the stock price prediction. The aggregate of these values is then presented as a feature importance plot in Chapter 4, providing a clear and quantifiable measure of each feature's impact on the model's output.

4. Results and Discussion

This chapter details and evaluates the performance of the trained LSTM model. It starts with a numerical assessment of the model's predictions for four distinct stocks, then moves on to a visual qualitative analysis. Next, the dynamics of the training process are explored, and the findings are discussed regarding their implications, limitations, and possible directions for future research.

4.1. Performance Evaluation

Table 2 presents a comparative analysis of the effectiveness of the prediction model for four different stocks: Meta (META), Tesla (TESLA), Microsoft (MICROSOFT), and Oracle (ORACLE). The performance of the model was assessed using three main metrics: Root Mean Squared Error, Mean Absolute Percentage Error, and Accuracy, which is calculated as (100% - MAPE). This table aims to investigate and compare the model's performance across each individual stock dataset, each of which displays distinct price characteristics and varying levels of volatility.

Table 2. Training result comparison

Stock Evaluation Results	RMSE	MAPE (%)	Accuracy (%)
META Price	15.6301	2,32	97,68
TESLA Price	13.2610	3,90	96,10
MICROSOFT Price	21.6992	4,87	95,13
ORACLE Price	4.5925	2,27	97,73

From the table, it can be seen that the model shows its best performance when applied to the ORACLE Price data. This is evidenced by the lowest RMSE value of 4.5925 and the lowest MAPE of 2.27%, resulting in a 97.73% accuracy rate. Conversely, the lowest performance was recorded for the MICROSOFT Price stock, which had the highest RMSE (21.6992) and MAPE (4.87%) values, with an accuracy of 95.13%.

Interestingly, the META Price stock ranks second-best in terms of accuracy (97.68%), despite having a higher RMSE value (15.6301) compared to the TESLA Price (13.2610). This phenomenon indicates that although there were some predictions for META with significant absolute value differences, the majority of its predictions were, in percentage terms, very close to the actual price.

4.2. Visual Analysis of Prediction Performance

While the quantitative metrics in Table 2 provide a concise summary of model performance, a visual analysis offers deeper qualitative insights into the model's behavior. This section examines two types of plots generated for each stock: a retrospective plot comparing predicted versus actual prices on the test set, and a prospective plot forecasting the price for the next 30 days.

4.2.1. Retrospective Analysis: Prediction vs. Actual Prices

The first form of visualization, presented for Meta Price in Figure 3, directly overlays the model's predictions (red, dashed line) onto the actual historical prices from the test set (blue, solid line). For META, the plot reveals an exceptionally high degree of accuracy. The prediction line tracks the actual price with remarkable fidelity, adeptly capturing not only the strong upward trend observed from early 2023 but also the more minor, high-frequency fluctuations within that trend. The minimal divergence between the two lines provides an explicit visual confirmation of the low MAPE (2.32%) and high accuracy (97.68%) reported in Table 2. This tight fit indicates the model's robustness in learning the specific patterns of META's stock.

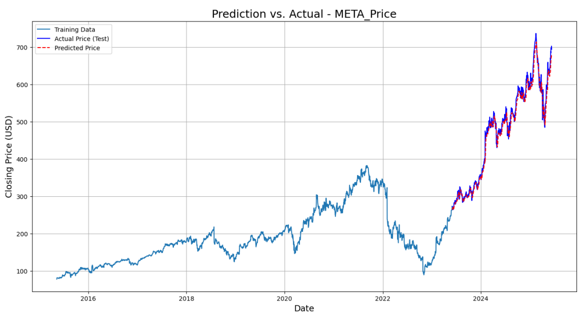

Figure 3 Retrospective Analysis of Meta stock price

Figure 4 Retrospective Analysis of Tesla stock price

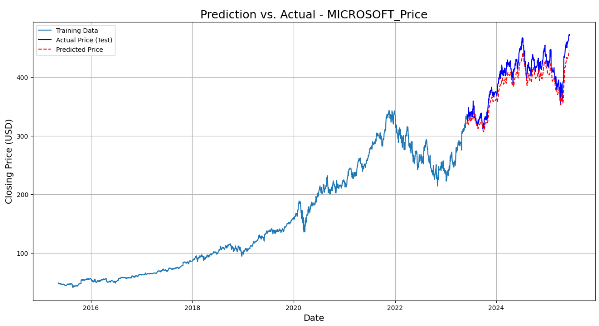

Figure 5 Retrospective Analysis of Microsoft stock price

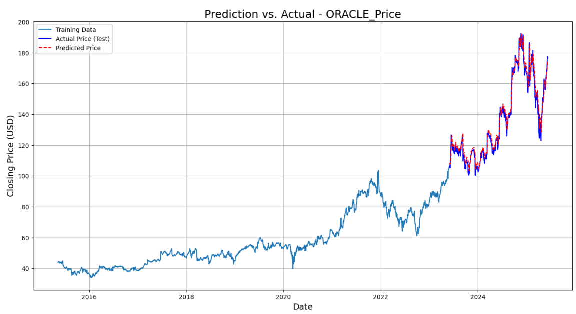

Figure 6 Retrospective Analysis of Oracle stock price

A similar analysis was performed for the other stocks. The plot for Oracle in Figure 6 showed a similarly tight fit, corresponding to its low error metrics. In contrast, the plot for Microsoft in Figure 5 visually demonstrates a greater degree of prediction lag during periods of high volatility, aligning with its higher reported RMSE and MAPE values. The plot for Tesla in Figure 4 presents a balanced

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

case, where the model successfully captures the overall trend but shows minor deviations during sharp price swings.

4.2.2. Prospective Analysis: 30-Day Forecast

Beyond evaluating performance on historical data, the trained model was also used to generate a 30-day price forecast for each of the four stocks, as illustrated in Figures 7, 8, 9, and 10. These forward-looking plots extend from the last known data point for each stock to project the anticipated price trend for the subsequent month.

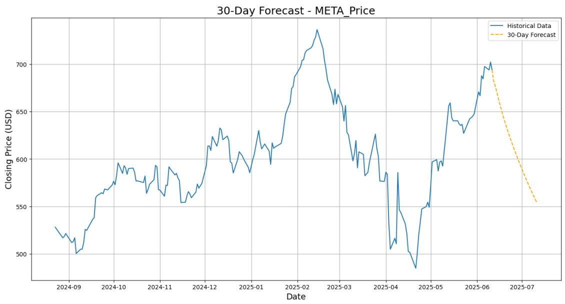

Figure 7 30-day forecast of Meta stock price

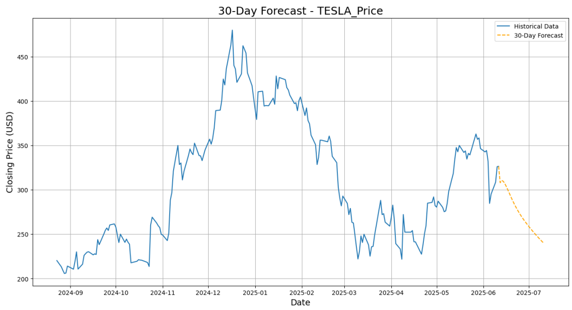

Figure 8 30-day forecast of Tesla stock price

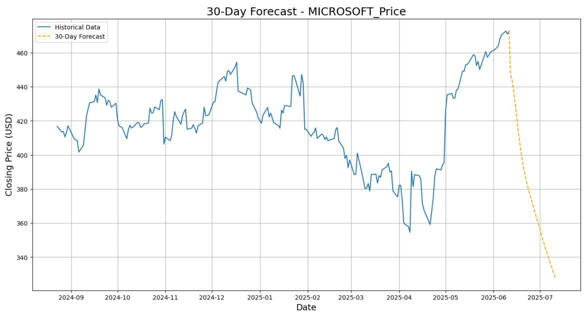

Figure 9 30-day forecast of company stock price

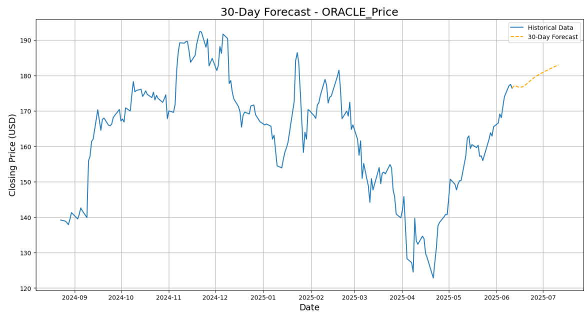

Figure 10 30-day forecast of Oracle stock price

The resulting forecasts reveal divergent potential trends, reflective of each stock's unique historical patterns learned by the model. The projection for Meta Price, for instance, indicates a potential downward correction following its recent recovery period. In contrast, the forecast for Oracle in Figure 10, the stock on which the model performed best, suggests a trend of continued stability with a slight upward trajectory. Meanwhile, the model anticipates a period of sideways consolidation for Tesla, as shown in Figure 8, indicating no strong directional momentum. For Microsoft, as shown in Figure 9, the model projects a continuation of a gentle downward trend, reflecting the challenges observed in the retrospective analysis.

4.2.3. Summary of Visual Findings

In conclusion, the dual approach of visual analysis, encompassing both retrospective and prospective perspectives, provides a comprehensive qualitative understanding of the model's capabilities. The prediction-versus-actual plots confirm the quantitative findings in **Table 2** by illustrating *how* and *where* the model achieves its accuracy on historical data. Meanwhile, the 30-day forecasts demonstrate a practical application of the model, offering a data-driven glimpse into potential future trends based on past performance.

4.3 Explainable AI (XAI) Analysis with SHAP

Figure 11 SHAP Value

In order to go beyond just performance metrics and examine the internal logic of the model, an Explainable AI (XAI) analysis was performed using SHAP (Shapley Additive exPlanations). The main goal of this analysis was to determine which of the 60 previous time steps (features ranging from t-59 to t-0) had the greatest impact on the model's predictions for the following day. The results regarding feature importance for each of the four stocks are illustrated in Figure 11.

The SHAP analysis revealed a striking and consistent pattern across all four models. As shown in the summary plots for META, TESLA, MICROSOFT, and ORACLE, the most distant time step, t-59 (corresponding to the price 60 days prior), was consistently identified as the single most dominant feature influencing the predictions. The mean absolute SHAP value for t-59 was significantly higher than for all other time steps, whose contributions were comparatively negligible and do not appear on the summary bar plots.

This counterintuitive finding suggests that the LSTM models, in their effort to capture long-term dependencies, may have assigned a disproportionate weight to the starting point of the input sequence. This could imply that the model learned a simplified heuristic, such as anchoring its prediction based on the momentum or trend initiated 60 days ago, rather than continuously re-evaluating based on more recent price action. This challenges the common assumption in financial markets that the most recent prices (recency bias) are always the most critical predictors. Instead, the model has identified a significant, long-range signal originating from the 60-day historical mark, which could correspond to a monthly or quarterly market cycle pivot point captured within the data.

Ultimately, the SHAP analysis provides a crucial, albeit unexpected, insight into the model's "black box" behavior. It demonstrates that the model's high accuracy, as reported in Table 2, may be heavily reliant on this single, long-term historical data point rather than a complex combination of recent price movements. This highlights the importance of XAI in uncovering the underlying mechanisms of complex models and validating whether their decision-making process aligns with financial theory.

5. Conclusion

This research successfully developed and evaluated Long Short-Term Memory (LSTM) models for forecasting the stock prices of four major technology companies: Meta, Tesla, Microsoft, and Oracle. The study pursued a dual objective: first, to build models capable of achieving high predictive accuracy, and second, to interpret the models' decision-making processes through the application of Explainable AI (XAI) [13], explicitly using the SHAP framework. The empirical results demonstrated that the LSTM models were robust, consistently achieving prediction accuracies above 95% across all tested stocks, with the highest accuracy of 97.73% recorded for ORACLE Price.

The primary contribution of this study, however, lies not in the predictive performance itself but in the insights generated by the SHAP analysis. Contrary to the common assumption that the most recent prices are the most influential predictors in time-series forecasting, the study revealed a consistent and unexpected pattern: the price from 60 days prior (t-59) was identified as the single most dominant feature driving the models' predictions for all four stocks. This suggests that the LSTM, in its effort to capture long-term dependencies, may have learned a simplified heuristic anchored on this distant data point rather than a more complex combination of recent price fluctuations. This discovery underscores the critical importance of XAI; without it, the high accuracy of the models would have masked this counterintuitive internal logic, highlighting a potential divergence between a model's performance and its alignment with financial theory.

The primary limitation of this study is its reliance on univariate data (historical prices only),

 ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

which neglects external factors such as trading volume, market sentiment from news, and macroeconomic indicators. Therefore, future research should focus on developing multivariate models that integrate these diverse data sources to create more context-aware predictions. Furthermore, applying SHAP analysis to these more complex models would be a valuable next step to understand how the model weighs historical prices against external factors. Investigating different look-back periods could also determine if the t-59 dominance is a consistent phenomenon or an artifact of the chosen window size. Ultimately, this research validates the predictive power of LSTMs while strongly advocating for the integration of XAI to ensure that complex financial models are not only accurate but also transparent and trustworthy.

6. ACKNOWLEDGMENT

I would like to express my sincere gratitude to all individuals who have supported me throughout the completion of this research. Although this study was conducted independently, I am truly thankful for the encouragement from my lecturers at Universitas Tarumanagara, as well as the respondents who generously shared their time and insights. I also extend my appreciation to my family and close friends for their unwavering moral support, which motivated me to persevere and complete this research successfully.

7. Reference

[1] Bacco L, Petrosino L, Arganese D, Vollero L, Papi M, Merone M. Investigating stock prediction using LSTM networks and sentiment analysis of tweets under high uncertainty: A case study of North American and European banks. IEEE Access. 2024 Aug 26.

[2] Wen M, Li P, Zhang L, Chen Y. Stock market trend prediction using high-order information of time series. IEEE Access. 2019 Feb 26;7:28299-308.

[3] Nabipour M, Nayyeri P, Jabani H, Mosavi A. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. IEEE Access. 2020 Aug 12;8:150199-212.

[4] Khan W, Ghazanfar MA, Azam MA, Karami A, Alyoubi KH, Alfakeeh AS. Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing. 2022 Jul 1:1-24.

[5] Thormann ML, Farchmin J, Weisser C, Kruse RM, Safken B, Silbersdorff A. Stock price predictions with LSTM neural networks and Twitter sentiment. Stat Optim Inf Comput 9 (2): 268–287 [Internet]. 2021

[6] Peivandizadeh A, Hatami S, Nakhjavani A, Khoshsima L, Qazani MR, Haleem M, Alizadehsani R. Stock market prediction with transductive long short-term memory and social media sentiment analysis. IEEE Access. 2024 May 10.

[7] Otabek S, Choi J. Twitter attribute classification with Q-learning on bitcoin price prediction. IEEE Access. 2022 Sep;10:96136-48.

[8] Anandita A, Wahyuningsih T. Prediksi Indeks Saham Syariah Menggunakan Model Long Short-Term Memory (LSTM). Jurnal Ilmu Manajemen (JIMMU). 2024 Mar;9(1):60-9.

[9] Arfan A, Lussiana ET. Perbandingan algoritma long short-term memory dengan SVR pada prediksi harga saham di Indonesia.

[10] Hanafiah A, Arta Y, Nasution HO, Lestari YD. Penerapan Metode Recurrent Neural Network dengan Pendekatan Long Short-Term Memory (LSTM) Untuk Prediksi Harga Saham. Bulletin of Computer Science Research. 2023 Dec 31;4(1):27-33.

[11] Saputra R, Alamsyah AG, Tjoanda M, Nick K, Cornelius A. Analisis Prediksi Saham Tesla Menggunakan Algoritma Long Short Term Memory (LSTM). J Comput Sci Inf Technol. 2024 Dec;2(1):81-90.

[12] Sujjada A, Sembiring F. Prediksi Harga Bitcoin Menggunakan Algoritma Long ShortTerm Memory. JURNAL INOVTEK POLBENG. 2024;9:450-9.

[13] Çelik TB, İcan Ö, Bulut E. Extending machine learning prediction capabilities by explainable AI in financial time series prediction. Applied Soft Computing. 2023 Jan 1;132:109876.

[14] Waqas M, Humphries UW. A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX. 2024 Sep 12:102946.

[15] Cho M, Kim C, Jung K, Jung H. Water level prediction model applying a long short-term memory (LSTM)–gated recurrent unit (GRU) method for flood prediction. Water. 2022 Jul;14(14):2221.