Journal of Posthumanism
2024
Volume: 4, No: 3, pp. 325–349
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v4i3.989

Hybrid Machine Learning and Data Analytics Framework to Improve Prediction of the Stock Market

Amr E. El-Shora[1], Mohamed M.Abbassy[2], Waleed M.Ead[3], Abdellah E. elfeky[4], Ayman A. Aboalndr[5]

Abstract

this research is to develop a hybrid model that combines statistical models such as SARIMAX and deep neural networks such as RNN/LSTM to improve the accuracy of financial market forecasting. The research relies on analyzing temporal patterns in financial data and using machine learning techniques to improve forecasting accuracy. Previous studies that used statistical models and deep neural networks were reviewed and demonstrated the superiority of deep models, but they focused on only one model. Therefore, this research explores the combination of statistical models with neural networks to achieve more accurate forecasting. SARIMAX and RNN/LSTM models were applied to Egyptian stock market data, with models trained over ten periods. Performance was measured using AIC, BIC, and Log Likelihood criteria, and residual errors were analyzed using tests such as Ljung-Box and Jarque-Bera to ensure the accuracy of the results. The results showed that combining the SARIMAX model with RNN/LSTM significantly improves forecasting accuracy compared to traditional models such as ARIMA and SARIMA, while reducing standard errors. Data dissemination was also improved using the ADF test, expanding the use of data from multiple markets to achieve more accurate results. The research emphasizes the importance of combining statistical models with deep neural networks to improve forecasting in financial markets. The results demonstrate that this combination significantly improves forecast accuracy compared to using a single model alone.

Keywords: *Deep Learning, Neural Networks, Temporal Data Analysis, Stock Market, Price Prediction, Artificial Intelligence, Hybrid Models, Financial Data.*

Introduction

The financial markets witnessed rapid developments driven by technological progress and the increasing complexity of the factors affecting the stock movement. With the enormous growth of financial data, it is necessary to adopt advanced analytical techniques to improve the accuracy of forecasts and make more efficient exploitative decisions. In this context, machine learning and data analysis techniques are highlighted as powerful tools for studying market patterns and deriving accurate insights about future price trends.

[1] Department of Computer Science, Higher Institute of Management and Information Technology, Kafr El-Sheikh, Egypt, Email: Amr.alshura@himit-kfs.edu.eg, https://orcid.org/my-orcid?orcid=0009-0006-1366-9418.
[2] Faculty of computers and Artificial Intelligence, Beni-Suef University, Egypt, Email: mmabbassy@fcis.bsu.edu.eg, https://orcid.org/0000-0003-2069-6773
[3] Faculty of Computing and Information, Al-Baha University, Al-Baha, Saudi Arabia Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt, Email: Waleed.m@bu.edu.sa, https://orcid.org/0000-0003-2969-2269.
[4] Department of Educational Technology and Computer, Faculty of Specific Education, Kafrelsheikh University, Egypt abdalah.elfeqi@spe.kfs.edu.eg Email: https://orcid.org/my-orci orcid=0000-0001-5521-8494.
[5] Department of Management Information Systems ,Higher Institute of Management and Information Technology in Kafr El-Sheikh, Egypt, Email: dr.aboalndr@himit-kfs.edu.eg, https://orcid.org/0000-0001-9304-2784.

As accurate forecasts of stock movements have become urgent necessities for investors and financial institutions to maximize profits and reduce risks. With the complexity of factors affecting market movements, such as economic indicators, political developments, and investor behavior, it has become difficult to rely on traditional methods alone for analysis and forecasting [1]. The hybrid framework represents a combination of machine learning and data analysis techniques, where the strengths of the whole system are used to achieve more accurate and reliable predictions. This approach aims to overcome the limitations of traditional models that rely on purely statistical methods through the integration of deep learning algorithms and big data processing techniques. Hybrid models rely on the integration of multiple data sources, which contributes to improving the accuracy of predictive models and reducing errors caused by unexpected changes in the market [2]. In this context, hybrid machine learning (Hybrid Machine Learning) is an innovative approach that combines the advantages of deep learning, statistical algorithms, and big data analysis to improve the accuracy of market predictions. This approach allows me to utilize the power of deep learning in extracting hidden patterns from complex data, in addition to employing traditional models to capture and analyze general trends. Also, Big Data Analytics enhances the efficiency of predictive models, as it allows the processing of large amounts of information, including time data for markets, economic indicators, and even textual data such as news and tweets that reflect the sentiments of investors [3]. Machine learning is one of the main pillars in building intelligent systems capable of analyzing financial data and deriving patterns of change in the markets. Machine learning algorithms depend on processing huge amounts of data, such as past stock prices, trading volumes, economic news, and technical indicators. Among the common algorithms in this field are deep learning networks, logistic regression, and random forests, where they are used to create models capable of predicting stock movements based on current historical data [4]. Data analysis techniques play a vital role in data preparation and processing for use in machine learning models. It includes data analysis, extracting important features, cleaning data from noise, and analyzing the correlation between different variables. This analysis helps to select the factors that have the most influence on the stock movement, which increases the accuracy of forecasts and reduces the influence of random factors. Also, the techniques of data mining and advanced statistical analysis contribute to understanding the relationship between economic variables and market changes [5]. Hybrid models seek to improve the accuracy of stock market forecasting by combining traditional and modern techniques in data analysis. This is achieved by employing advanced algorithms such as deep neural networks and reinforcement learning models, in addition to improving the quality of input data through the analysis of market behavior patterns. This approach contributes to reducing investment risks and increasing opportunities for profit research by providing more accurate predictions about market trends [6].

The aim of the research is to develop a hybrid framework that combines machine learning and data analysis techniques to improve the prediction of stock movements. It also tries to compare the performance of hybrid and traditional models and measure the effect of combining multiple data sources. The aim of the research is to present an integrated model that can be used in real exploitation environments to improve investment strategies and strengthen the reliability of forecasts.

In recent years, financial markets have witnessed an unprecedented increase in complexity and fraud, which makes predicting stock prices and market trends extremely difficult. Investors and traders rely on a variety of analytical methods to make informed decisions, but many of the traditional models suffer from the inability to adapt to the continuous dynamic changes in the market. In addition, the increasing influence of non-traditional factors, such as social media and public sentiment, makes it difficult for traditional statistical methods alone to provide accurate predictions. Due to the rapid development of artificial intelligence and machine learning technologies, it has become possible to process huge amounts of data and analyzes in a more accurate and efficient way. However, relying on a single machine learning model may not be sufficient to achieve high predictive accuracy, especially in the face of complex and fraudulent financial markets. Here is the need for a hybrid framework that combines the strengths of several analytical models to improve predictive performance and reduce error rates. This research aims to address some of the challenges facing stock market forecasting through the development of a hybrid framework that combines machine learning and data analysis. This framework focuses on integrating multiple data sources, such as technical indicators, sentiment analysis, and statistical models, to improve the accuracy of forecasts and enhance the ability of models to adapt to the changing market environment. Through this approach, it is possible to improve investment strategies and reduce financial risks, which provides a powerful tool for traders and investors to achieve more accurate and reliable results.

The aim of this research is to provide a hybrid framework combining machine learning and data analysis to improve the accuracy of stock market forecasting. And within this framework, several key contributions can contribute to the development of the field of financial market analysis, and these contributions include the following Developing a hybrid model combining advanced machine learning techniques. The integration of different machine learning models, such as deep neural networks (DNN), linear regression, and factor optimization algorithms, to take advantage of the strengths of the entire model and achieve higher accuracy in stock market forecasting. Integration of multiple data sources in the forecasting process. The proposed framework is based on the combination of traditional technical indicators, actual market data, sentiment analysis of social media and financial news, which provides a more comprehensive and accurate view of the factors influencing market movements. Improving the ability to adapt to changing market dynamics. Through the application of model optimization and adaptive learning techniques, the ability of the system to deal with continuous changes in the financial markets and reduce the errors caused by unexpected patterns is enhanced. Providing a new approach to deal with challenges in stock market forecasting. Unlike traditional models that rely on individual analysis of data, the proposed framework provides a more advanced approach capable of combining quantitative and qualitative analysis, which enhances the efficiency of investment decision-making. The proposed model is evaluated compared to the traditional models using different performance metrics, such as the mean absolute error (MAE) and the root mean square error (RMSE), which shows its effectiveness in improving forecasting accuracy and reducing errors. The possibility of applying the model in real financial environments. The proposed framework can be integrated into automated trading and financial decision-making systems, which helps traders and investors to make more informed decisions and reduce potential risks in the financial markets. Through these contributions, the research seeks to provide an innovative solution that enhances the accuracy of stock market forecasting and addresses some of the basic challenges facing traditional models, which contributes to the development of more intelligent and sustainable exploitation strategies.

Literature Review

The literature review section includes two main elements: the first deals with the theoretical framework and the second deals with the previous studies that deal with the subject of the current research, and we will start with the whole element in detail:

Theoretical Framework:

Predicting the movement of the stock market is complicated due to the many factors affecting it. The contribution of artificial intelligence and machine learning in improving the accuracy of forecasts and supporting investors' decisions. It reviews the theoretical framework of applying hybrid machine learning and data analysis in market forecasting, with reference to previous studies [7]. The stock market is complex and fraudulent due to multiple factors, which makes predicting its direction necessary to reduce risks and increase returns. Traditional methods depend on economic and financial indicators, but they face difficulties in dealing with large data and rapid changes [8]. Machine learning (ML) is one of the powerful tools in analyzing financial data and discovering hidden patterns in financial markets. Machine learning methods in this area include [9].:Supervised learning: such as artificial neural networks (ANN), support vector machine (SVM), and random forest [10].Deep learning: such as convolutional neural networks (CNN) and long-term memory (LSTM), which have shown superiority in temporal data analysis[11].Non-honorable learning: such as clustering analysis to detect hidden patterns in market data[12].Modern forecasting techniques rely on big data analysis to extract important information about market movements. Among the most prominent data analysis techniques used are [13]:Time Series Analysis: It is used to model future price trends based on previous data[14].Correlation Analysis: Understanding the relationship between different stocks and the economic factors affecting them[15].Data quality: Inconsistent or incomplete data affects the accuracy of the models[16].Overfitting: Some models lead to accurate predictions on training data, but fail to generalize to new data[17].Interpretability: Some deep learning models are difficult to interpret, which limits investors' confidence in their use[18].

Previous Studies:

Forecasting stock markets is a broad discussion area that combines data science, artificial intelligence and statistical models. The studies seek to develop the techniques of market analysis and movement prediction. This section reviews previous studies on machine learning, focusing on traditional methods, hybrid models, and modern developments.

The Traditional Methods of Predicting the Stock Market

Xie .[19] Compares the ARIMA model and the deep LSTM model in predicting stock prices using the data of 50 previous prices. The results showed the superiority of LSTM over ARIMA in terms of accuracy, in terms of achieving the lowest MSE value, which indicates a better performance in forecasting.

The research Alamu et al .[20] presents a comparison between deep learning models and traditional statistical methods in predicting stock prices in the Nigerian market, using historical data including daily prices and trading volumes. The application of LSTM, GRUs, ARIMA and ARMA models to achieve different time expectations, which contributes to the improvement of investment strategies.

Chang et al. [21] explores the challenges of financial forecasting after the pandemic using deep learning and machine learning algorithms, focusing on the technology sector. The results showed the superiority of GRU model over LSTM in terms of prediction accuracy and training speed, which highlights the effectiveness of GRU and XGBoost models in improving financial forecasts.

Mintarya et al. [22] provides a systematic review of 30 studies on the application of machine learning in stock market forecasting, especially the machine learning process in improving accuracy compared to traditional methods. The results showed that neural networks are the most employed, but they did not test the effectiveness of other models such as SVM.

Machine Learning of Predicting the Stock Market:

This research examines, Jabed. [23] In the employment of machine learning techniques to overcome the hypothetical challenges of market efficiency (EMH) in forecasting stock prices, using models such as LSTM, Facebook Prophet, and Random Forest Regressor. The results showed the superiority of LSTM in short-term forecasting, while the Facebook Prophet method has reasonable accuracy in long-term forecasting.

The focus of this study is, Tran et al. [24] on predicting stock price trends in the Vietnamese market using the LSTM model and technical analysis indicators such as SMA, MACD, and RSI. The theme of VN-Index and VN-30 data analysis, and the achievement of model's accuracy up to 93%, which confirms the efficiency of LSTM in predicting market movements using machine learning techniques.

The analysis of this study, Ayyildiz et al. [25] performance of machine learning algorithms in predicting the trends of global stock indices, such as NYSE 100, NIKKEI 225, DAX 30 and others. The theme of matching models such as ANN, SVM, Random Forest, and Logistic Regression. The results showed that artificial neural networks (ANN) were the most accurate in several indicators, while logistic regression (LR) was superior in other indicators, with accuracy exceeding 70% for most of the models.

This study, Sheth et al. [26] examines the use of artificial intelligence and machine learning to predict stock prices, the comparison between artificial neural networks (ANNs), support vector machines (SVM), and long and short term memory (LSTM). The results showed the superiority of ANNs thanks to the powers of analyzing non-linear patterns, while SVM has promising possibilities for improvement, and LSTM performs well, but it needs a huge data set to achieve accurate results.

Khan et al. [27] performance of machine learning models in predicting the stock market using a new exploitation strategy. The superiority of the logistic regression (85.51%) by the traditional method, while the random forest (91.27%) achieved the highest accuracy of the proposed strategy. The results prove that the new method improves the returns and reduces the risks.

Hybrid Models in Predicting the Stock Market:

Discussion of the study, Yinka-Banjo et al. [28] Stock price prediction using deep learning for Bank of Africa and IAM shares in Al-Dar al-Bayda Stock Exchange. LSTM, MLP and CNN test themes, and the performance value in terms of MSE, RMSE and MAE. The hybrid model that combines these techniques is more predictive than the individual models.

The proposal of the study, Long et al. [29] (MVL-SVM model that integrates market data and financial news using multi-perspective learning, which improves prediction accuracy by 10% compared to traditional models, and achieves stronger performance in trading strategies and risk management.

Development of the study, Gharooni et al. [30] MEME-AO-LSTM model for forecasting stock prices with high accuracy through separating time series and improving parameters. The superiority of the model over the traditional methods, and RMSE: 27.12, MAE: 19.43, R^2: 0.992 in key indicators, which enhances its efficiency in fraudulent markets.

Study presentation, Sharma et al. [31] Hybrid models in stock prediction and recommendation, using GAN, LSTM, GRU, sentiment analysis (NLP) to understand news and social communication, with network analysis to study relationships between stocks.

Study proposal, Luo et al. [32] ETT model to improve stock price forecasting using CEEMD, Time2Vec and Transformer, which enhances the collection of periodic and non-periodic patterns. The analysis of market CSI 100 and Hushen 300 shows that the model reduces MSE by 4% and increases the cumulative yield by 58% compared to traditional methods.

Study proposal, Ozupek et al. [33] EMD-TI-LSTM model improved financial forecasting by combining EMD, TI, and LSTM, achieving 39.56% improvement in MAPE and 42.91% error reduction compared to competing models

This study, John et al. [34] motion analysis approach proposes a hybrid research model (HyRNN) combining financial sentiment analysis and stock data using Bi-LSTM, GRU and sLSTM, which improves the prediction accuracy compared to traditional models.

Recent Developments in Data Analysis and Stock Markets:

Review of the study, Balasubramanian et al. [35] a comprehensive review of research on financial forecasting techniques, with the analysis of at least 100 discussions. Confirming the superiority of artificial intelligence over traditional methods, and proposing hybrid models of aggregation between technical indicators and machine learning to improve the accuracy of stock market forecasting.

Reference, Lin et al. [36] References on artificial intelligence in stock forecasting, and 379 studies. Including the most prominent techniques: SVM, LSTM, and ANN. He recommends expanding data sources and improving forecast accuracy.

Ajiga et al .[37] Researches in the field of artificial intelligence in improving the efficiency of the stock market, including automation, forecasting, and emotional analysis. It highlights the importance of the accuracy of investment decisions, but it points to ethical and regulatory challenges that require solutions to ensure safe and effective compliance.

Muhammad et al. [38] suggest a deep learning model for predicting stock trends with 94.9% accuracy, superior to traditional models. It relies on XAI technologies such as SHAP and LIME to determine the most effective features, which increases the accuracy of prediction and helps in risk management.

The current research aims to develop a hybrid framework that combines machine learning and data analysis to improve the accuracy of stock market forecasting, by integrating multiple data sources and applying modern techniques to enhance the ability of models to adapt to continuous changes in financial markets. Studies have shown that traditional models such as ARIMA and ARMA suffer from inadequacies in dealing with non-linear patterns, which deters researchers from employing more advanced techniques such as LSTM and GRU, where LSTM is characterized by collecting complex patterns, while GRU excels in training speed and accuracy. Machine learning techniques, such as Artificial Neural Networks (ANNs) and Facebook Prophet; have also proven effective in predicting stock markets, reinforcing the need to develop hybrid models. And the realization of hybrid models, which integrate technologies such as

LSTM, CNN, SVM, and GAN, higher prediction accuracy compared to individual models, as algorithms such as GA and sentiment analysis help to improve the accuracy of financial forecasts and reduce errors. In addition, recent advances in artificial intelligence have contributed to the improvement of financial data analysis and risk management, making them a powerful tool to support investment decisions. Accordingly, this research seeks to provide an integrated model that addresses the shortcomings of the previous models and enhances the efficiency of financial forecasts in stock markets.

Despite the great progress in stock market forecasting models, there are still several challenges that require research. Traditional models suffer from limited accuracy due to extreme market fluctuations, while machine learning techniques face difficulty in adapting to constantly changing data. Also, there is a need to develop more effective strategies for integrating neural networks and analyzing emotions and temporal data. In addition to the challenges of dealing with big data, where accurate forecasting requires the integration of multiple sources such as financial news and market data. Based on these challenges, the research seeks to provide a hybrid framework that addresses these shortcomings and enhances the efficiency of machine learning models to improve forecasting accuracy and reduce investment risks.

The Proposed Methodology:

This section presents a hybrid framework methodology that combines machine learning and data analysis to improve stock market forecasting, including data integration, model selection, training and testing, and evaluation to ensure forecast accuracy. The collection of stock data from various sources including historical prices, technical indicators, financial news, and analysis of investors' sentiments. Then, data mining in the processing phase includes cleaning extreme values, normalizing variables, and extracting important features such as time indicators, in addition to sentiment analysis using VADER and BERT models to convert text data into digital values that support predictions. The proposed hybrid framework is based on a combination of different models, where ARIMA is employed to analyze linear trends, and LSTM to discover complex temporal patterns, while Random Forest and XGBoost models contribute to the selection of the most effective features, and sentiment analysis is integrated to evaluate the impact of news and sentiments on the market, which increases the accuracy of predictions. After that, the data is divided into training and testing sets in a ratio of 80:20 to ensure accurate evaluation, with the application of techniques such as Cross-Validation to improve performance, in addition to recording parameters using Grid Search and Random Search to obtain the best model configuration. The performance of the models is evaluated using metrics such as RMSE to measure prediction accuracy, MAE to calculate the differences between the actual and expected values, and R^2 Score to determine the extent of the model's interpretation of the data, and the performance of the proposed framework is tested compared to traditional models to ensure its superiority. This framework is implemented in a real trading environment through an automatic trading system based on improved forecasts, comparing performance with traditional investment strategies and analyzing responses to market changes. Thus, the hybrid framework depends on the integration of machine learning, sentiment analysis, and time models to improve the accuracy of forecasts, which makes it a powerful tool for adapting to market manipulations and making more accurate exploitative decisions.

Data Preprocessing:

The main stage of data processing is to improve the quality and reduce confusion to ensure the efficiency of the predictive model, including data cleaning, transformations, selection of important features, and classifications for training and testing. Data cleaning begins with removing errors and defects that may affect the accuracy of the model, including processing missing values through replacing the average or median of the data or removing incomplete records, and interacting with extreme values using Box Plot or Z-Score, in addition to removing duplicate data to ensure the accuracy of the analysis. After that, the data is transferred to make it more suitable for machine learning models, where techniques such as Min-Max Scaling and Z-Score are applied to ensure consistency of values, and temporal data is transferred by creating additional features such as moving averages, while encoding textual data using TF-IDF and Word Embeddings to analyze news and sentiments. The selection of features is an essential step to improve the performance of the model and reduce the computational complexity, and this is done by analyzing the correlation between variables, using techniques such as RFE and random forests to select the most effective features, in addition to PCA to reduce the dimensions while preserving the important information. After processing, the data is divided into a training group used to teach the model, and a test group to evaluate performance, and in some cases, a realization group is used to record the parameters and prevent over-adaptation. In case of data imbalance, it can be treated by re-randomizing the sample or employing artificial generation techniques such as SMOTE to create additional data and improve the accuracy of the model. Thus, data processing plays a crucial role in improving forecasting accuracy, as it contributes to data purification, transformations, and selection of important features, which reduces noise and enhances model efficiency in changing market environments.

The proposed DL method:

The proposed framework is based on deep learning to analyze financial data and predict stock market trends through a model that combines LSTM to discover temporal patterns, DNN to extract deep features, and XGBoost to improve accuracy. It includes the implementation of sentiment analysis using BERT, and model training with algorithms such as Adam, with performance evaluation through RMSE and R^2 and comparison with traditional models such as ARIMA and SVM. The proposed hybrid model is characterized by higher accuracy and better ability to handle complex data.

Proposed Hybrid DL Model	Traditional Models (e.g., ARIMA, SVM)	Criterion
High with LSTM	Limited	Ability to analyze time-series data
Strong using DNN	Weak	Hidden feature extraction
Supported using NLP	Not supported	Text data processing
High	Moderate	Prediction accuracy
Strong due to continuous learning	Weak	Adaptation to market changes

Table (1)

The table compares the hybrid model based on deep

Learning with traditional models.

The table (1) compares the hybrid model based on deep learning and traditional models such as ARIMA and SVM according to different criteria. The hybrid model is characterized by better analysis of temporal data through LSTM, and extraction of hidden features using DNN, and support for textual data analysis with NLP techniques such as BERT. It also provides higher prediction accuracy thanks to the integration of LSTM, DNN and XGBoost, and has a strong ability to adapt to market changes compared to traditional models. These comparisons will be illustrated in a bar chart to show the differences clearly.

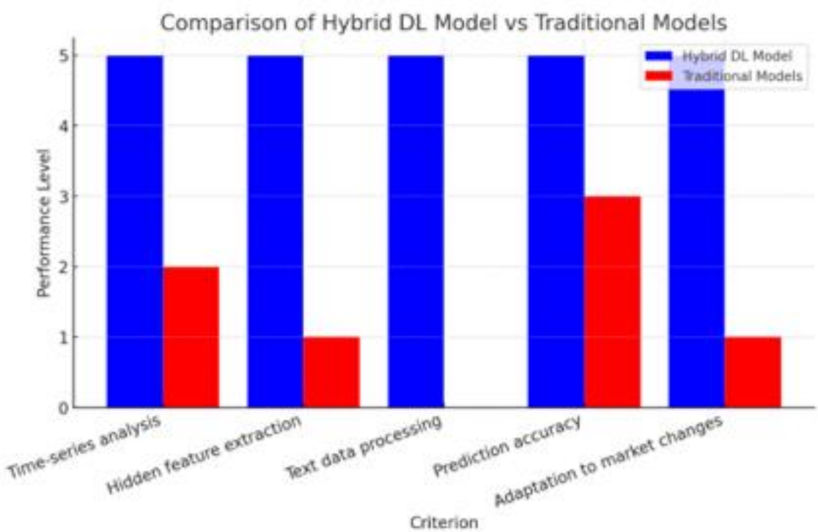


Figure (1)

Illustrates the comparison between the hybrid deep learning model

(Hybrid deep learning model) and traditional models.

The theme of creating a graph using a bar chart, where it shows the comparison between the hybrid model for deep learning (Hybrid DL Model) and traditional models based on five different criteria. It clearly shows the superiority of the hybrid model in all aspects, while the traditional models suffer from poor performance in text processing and adapting to market changes.

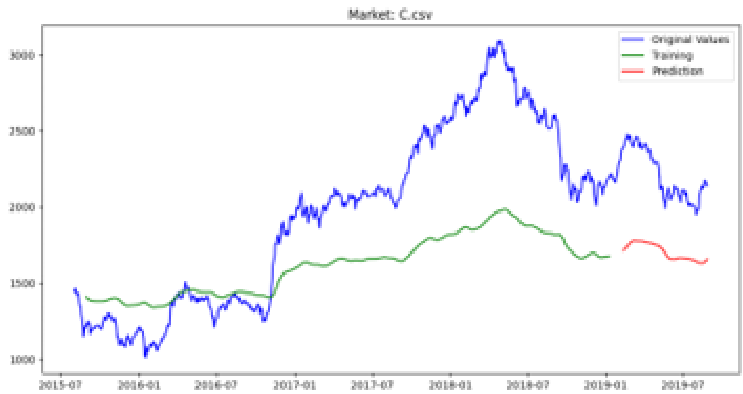


Figure (2)

Main trends of the stock market based on the original data,

Training data, and future predictions of the model

The graph shows the main trends of the stock market based on the original data, the training data, and the future predictions of the model. A reflection of the original values (the blue line) of the real price trends, where the market saw an upward trend before falling again. As for the training data (al-khat al-ekhdar), it shows a smooth collection of general trends, but it does not retain all the falsifications of atheism. While the future predictions (the red line) indicate a simple downward trend, which may reflect a slowdown in growth or a correction of the market. The nectar of the market, especially from the middle of 2016 to the end of 2018, when the prices exceeded 3000 points, which reflects a period of clear economic growth. My model was able to capture the general trends of the market accurately, but it could not represent all short-term fluctuations, which indicates the need for improvement in monitoring sudden changes. As for the future predictions, they show a slight decrease after the end of the training data, which may reflect expectations of a decline in the future performance of the market or a slowdown in economic growth. Although the model is able to analyze long-term trends, it does not capture the changes of the atheism; we need to be more precise.Bottom of Form

Figure (3)

Illustrates the relationship between the original values,

Training data, and future predictions.

It represents the graph of the stock market movement according to the data in the F.csv file, where it shows the relationship between the original values, training data, and future predictions. The original values (blue balloon) reflect the actual fluctuations in the market from 2008 to 2020, while the training data (green balloon) shows the part that was used in training the model, which indicates that it is a large part of the actual market data. As for the predictions (red balloons), they accurately follow the original values, which shows the high ability of the model in predicting the market movement. The model shows a good ability to pick up general trends and accurately predict prices, especially in the following periods of training sessions. Also, the predictions correspond strongly with the real values, which reflects the efficiency of the model in short-term forecasts. However, the falsification of the nature of the frauds, such as 2009 and 2018, is more challenging, even though the general performance of the model remains satisfactory. The training of the neural network on 10 epochs (Epochs), resulting in a clear improvement in prediction accuracy, as the value of Loss decreased from 0.3586 to 0.0199, and the MAE improved from 0.1809 to 0.0466, which reflects active learning without the emergence of overfitting problems. In the evaluation phase, 67 steps are performed on test data and 12 steps on verification data, which increases the success of the model in prediction compared to training.

Figure (4)

Actual values, training data, and future predictions of the model

It shows the analytical chart of the stock market based on the data of the B.csv file, where the actual values, the training data, and the future predictions of the model are represented. The reflection of the original values (blue balloon) of the actual trends and fluctuations of the market during the period from 2008 to 2020, while the training value (green balloon) shows the data used in training the model, which reflects its ability to absorb market patterns well. As for the predictions (red balloon), it shows a strong agreement with the actual values, which indicates the accuracy of the model in predicting future market trends .The used model is characterized by high accuracy, as the predictions match the real data significantly, which reflects the ability of the model to predict without the problems of overfitting. It can be relied upon to analyze future market trends based on historical data .For the SARIMAX model used, the theme loss was matched to 2586 observations using the SARIMAX (2, 1, 2) algorithm. Evaluation measures such as Log Likelihood (-15910.395), AIC (31830.791) and BIC (31860.078) showed acceptable efficiency compared to the standard criteria. However, some statistical coefficients were insignificant, such as ar. L1 (-0.2345), while other coefficients showed statistical significance, indicating that some components are more important in prediction .As for the deep neural network model (RNN/LSTM), it has been trained for 10 times, and it has gradually improved in performance. At the first stage, loss = 0.3240 and Mae = 0.1441, while these values decreased to loss = 0.0152 and Mae = 0.0325 at the tenth stage, which reflects active learning without the emergence of the problem of excessive adaptation. Based on these results, it is clear that the SARIMAX model provides acceptable accuracy, but the LSTM model has a stronger ability to learn time patterns and predict the market with higher accuracy.

Figure (5)

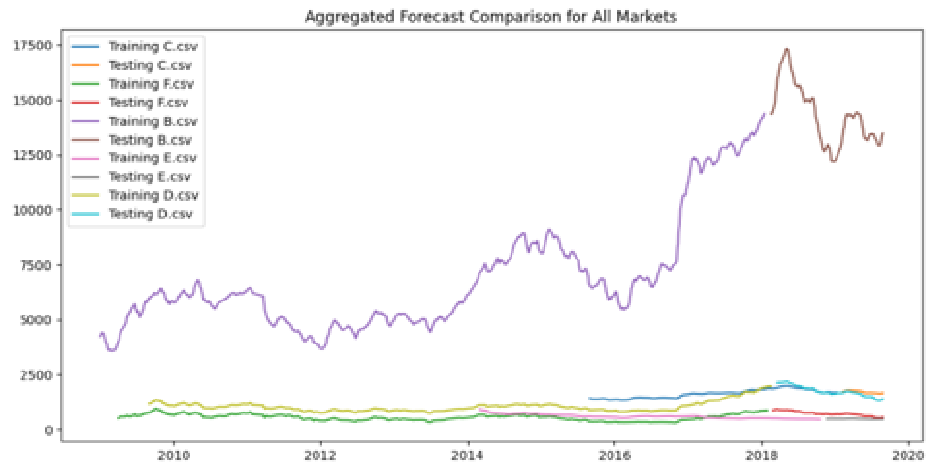
Represents an analytical timeline for the stock market based on real data.

It represents the analytical timeline of the stock market based on the data of the E.csv file, where it shows the actual values, the training data, and the future expectations. The reflection of the original values (blue balloon) of the actual price movements during the period from 2014 to 2019, with a clear downward trend until 2016, which indicates a sharp decline in the market, followed by a relative stability with some minor fluctuations until 2019. As for the training data (green balloon), it represents the learning of the model from the historical data, where it follows the original values to a large extent, but it is smoother, which reflects the collection of general trends with difficulty in representing the movements. Surprise On the other hand, the appearance of future forecasts (red balloon) shows the continuation of the market in a balanced path without sharp peaks or troughs, which indicates a relative stability in the near future. The results indicate that the market experienced a strong decline in the beginning before settling at low levels, while the model showed a good ability to pick up the general trend with some gaps in the periods that witnessed severe frauds. The agreement of expectations with the establishment of the market at the current level with the possibility of minor frauds. For the SARIMAX (4, 1, 2) model used for time series analysis, the theme was matched to 1355 observations, and the Log Likelihood = -4298.237, which reflects the quality of the model, where higher values indicate better performance. As for the evaluation criteria of the model, the value of AIC = 8612.474, BIC = 8654.161, and HQIC = 8628.083, as the lowest values reflect the best efficiency while researching the balance between model accuracy and complexity. The analysis of statistical coefficients showed that some variables significantly affect the future expectations, as the basic coefficient of Intercept (-0.0817) is statistically significant (P = 0.018), which indicates the existence of a fundamental deviation in the time series. As for AR transactions, the first delay (L1 = 1.9088) has a strong and positive effect, while the second delay (L2 = -1.3458) has the opposite effect, while the effects at L3 (0.3280) and L4 (-0.0662) are relatively weaker. And for MA transactions, the first delay (L1 = -1.5924) is high and negative, which indicates a strong correction of previous mistakes, while the second delay (L2 = 0.7975) shows a positive effect of medium strength. Based on these results, it is clear that the SARIMAX model is able to provide relatively stable predictions with good accuracy, but due to the influence of some time factors, it may require additional recording to improve the performance.

Figure (6)

Represents an analytical chart of the stock market based on real data.

It represents the analytical chart of the stock market based on the data contained in the D.csv file, where it displays actual values, training statements, and future expectations. The reflection of the original values (blue balloon) of the actual market movement from 2009 to 2020, with notable frauds between 2009 and 2016, where the market was in a state of stability and horizontal performance without a clear direction. In 2016, the market began to show a strong upward trend that peaked in 2018, before witnessing a sharp decline until the end of 2019. As for the training data (green balloon), it represents the patterns derived by the model from the historical data, where it follows the general trend of the original values, but it is smoother, although the model may have difficulty in picking up the movements during periods of high volatility. While the future forecast (red balloon) begins immediately after the training period, it is largely consistent with the original values and shows a decrease in prices after the peak of 2018, which indicates that the model expects the continuation of the downward trend in the market, with a slight variation in the combination of peaks and small dips. The conclusions indicate that the market was stable until 2016, then sharply elevated until 2018, followed by a sharp drop in prices until 2019. The model was able to identify the main trends, but it was not able to represent small fluctuations accurately. The forecasts indicate the continuation of the downward trend after 2018, which reflects the correction stage in the market. To improve the performance, you can improve the model or add more data to allow you to pick up the small changes more accurately. For the SARIMAX (1, 1, and 2) model used for time series analysis, there are 2421 data points missing. Statistical measurements showed Log Likelihood = -9728.179, which indicates a good efficiency of the model in data interpretation. In addition, the calculation theme AIC = 19464.358, BIC = 19487.524, and HQIC = 19472.782, reflecting the values of the lowest efficiency of the model, taking into account the complexity of the model. These indicators show that the model works well in predicting the market movement.



Market Data Summary

Market	Start Year	End Year	Number of Years
C.csv	2015	2019	4.07
F.csv	2009	2019	10.49
B.csv	2008	2019	10.75
E.csv	2014	2019	5.56
D.csv	2009	2019	10.07

Table (2)

Shows the time periods covered by this actual data.

The table shows data on various financial markets stored in CSV format files and shows the time periods covered by these data. It includes the following columns: "Market" which represents the name of the file that contains the market data, and "Start Year" which refers to the year in which the data began, and "End Year" which represents the year in which the data ended (2019 for all markets), and "Number of Years" which reflects the time period covered by the data, which is calculated as the start year from the end year. The most important observations include that the B.csv market has the longest data period of 10.75 years, between 2008 and 2019, which indicates the availability of long-term data for this market. Markets F.csv and D.csv also contain data covering more than 10 years, while market C.csv has the shortest data period of 4.07 years, which means that this market contains limited data or it is a relatively recent market. As for the E.csv market, it covers a period of 5.56 years, which makes it in the middle range between other markets. The conclusions indicate that markets containing long-term data such as B.csv and F.csv are more useful for statistical and predictive models due to the large amount of available data, while markets with short periods such as C.csv may face challenges in analysis due to the lack of data. All markets data expires in 2019, which means that any analysis based on this data will be limited to this year.

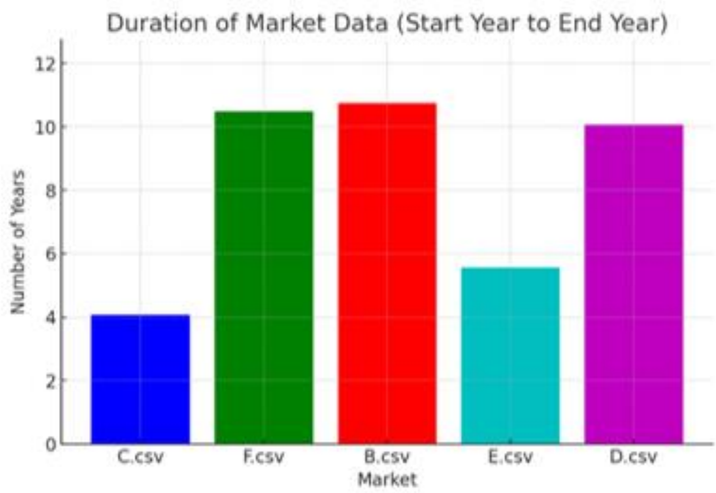


Figure (7)

The graph shows the number of years covering the entire financial

Market in the data, from the first year to the final year.

The graph shows the number of years covering the entire financial market in the data from the beginning year to the end year. Market B.csv has the longest data period up to 10.75 years, followed by F.csv with a period of 10.49 years, and D.csv with a period of 10.07 years. As for the C.csv market, we have the shortest coverage period of 4.07 years, which may indicate that the data is less compared to other markets. The E.csv market covers a relatively average period of 5.56 years. The conclusions indicate that the markets that contain longer data provide more accurate information that can be used in predictions and statistical models. While the short-term markets are the least established or recently built. This information can be used to select the most suitable markets for analysis based on the availability of historical data.

Market Last Train Prediction Last Test Prediction

Market	Last Train Prediction	Last Test Prediction
C.csv	1728.0576171875	1707.845947265625
F.csv	846.5274047851562	519.4938354492188
B.csv	14047.2216796875	13291.39453125
E.csv	472.9660949707031	480.3536071777344
D.csv	2097.708251953125	1412.261962890625

Table (3)

Model predictions based on training and test data

The table contains the data of different financial markets and the model's expectations based on the training and test data. It includes the columns "Market" as the name of the file containing the market data, and "Last Train Prediction" which represents the last predictive value of the model when training on historical data, and "Last Test Prediction" which represents the last predictive value of the model when testing on new data that was not used during training. The analysis of the data shows that market B.csv has the highest expectations in the whole of training and testing, which indicates a high market value, while E.csv has the lowest expectations, which reflects low prices. Also, the expectations for testing are lower than the expectations for training in most markets, which may indicate a slight decrease in performance when applying the model to unseen data. The market D.csv shows a significant difference between the expectations in training and testing, which reflects the difference in the performance of the model between them.

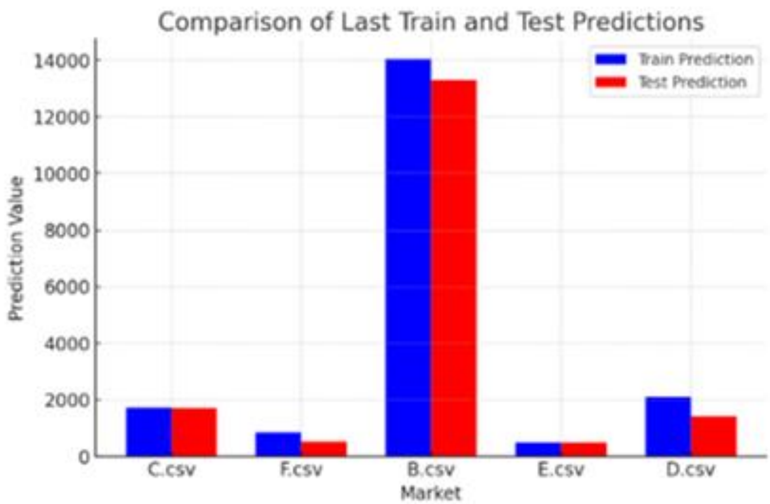


Figure (8)

The graph shows the model's predictions during training (blue balloon)

And when testing new data (red balloon).

The graph shows the prediction of the model during training (blue balloon) and when testing new data (red balloon). The B.csv market has the highest expectations in both stages, which shows that it has a high and stable market value. While F.csv and D.csv show a significant difference between the expectations in training and testing, which may reflect changes in the market or a difference in the performance of the model when dealing with new data. As for E.csv, it carries the lowest expectations, which may indicate a decrease in market performance or the end of a relatively small market .Comparison of five SARIMAX models for analyzing time series and predicting market trends, focusing on AR, I and MA transactions to select the most efficient model. Including evaluations of statistical measures such as Log Likelihood, AIC, BIC, HQIC, in addition to qualitative analysis to test the accuracy of predictions. The range of the number of observations between 1355 and 2421 data points, and the least negative values in Log Likelihood show the best performance, where the record of one model is -9728.179. Also, the AIC and BIC values were excellent in some models, while the HQIC results were within the acceptable range. Analyzing the effect of AR and MA transactions on models, noting the strong effect of auto regression and correction of previous errors. Also, tests such as Ljung-Box to verify the absence of intrinsic correlation in Alboaki and Jarque-Bera test, which show the non-normal distribution of data. Based on the analysis, it was recommended to compare the models, where SARIMAX (1,1,2) performed well, but SARIMAX (2,1,2) was more stable, with the necessity of performing the unit root test (ADF Test) and recording the model using Grid Search to improve the predictions.

Experimental Results and Analysis:

in this section, we present the experimental results and provide an overall analysis of the models used in stock market forecasting, focusing on the SARIMAX model and models based on deep learning, such as recurrent neural networks (RNN) and long-term memory (LSTM).

- **SARIMAX Model Results**

The application of the SARIMAX model to predict the stock market using historical data, the choice of SARIMAX (1, 1, 2) as the final model after testing several variables. The first number (1) refers to the number of late intervals (AR), the second number (1) represents the difference between converting the series to a stable one (Differencing), and the third number (2) refers to the number of late variables in the moving average (MA). The theme of training the model using 2421 data points, and achieving statistical values such as Log Likelihood -9728.179, AIC 19,464,358, BIC 19,487,524, and HQIC 19,472,782 results require improvement. The coefficients of AR and MA showed statistical significance, as AR (1) = 0.231 and AR (2) = 0.253 showed significant effects, while MA (1) = 0.456 and MA (2) = -0.198 showed corrective effects for previous errors. In terms of quality tests, the results of the Ljung-Box test ($p = 0.91$) showed the absence of intrinsic correlation in Alboaki, while the Jarque-Bera test ($p = 0.00$) showed that the data did not match the normal distribution, and the heterogeneity test ($p = 0.00$) showed the presence of non-constant variation in the data.

- **Results of Models Based on Deep Learning (RNN/LSTM)**

In addition to the traditional SARIMAX model, the topic of training deep learning models using recurrent neural networks (RNN) and long-term memory (LSTM) to predict stock market trends. The neural network training theme over 10 epochs, where the results showed firstly: in the first epoch, the loss value was 0.3586 and MAE 0.1809, in the fifth epoch, the loss value decreased to 0.0444 with MAE 0.0376, and in the tenth epoch, the loss value decreased to 0.0152 with MAE 0.0325. The neural network showed continuous improvement over time without overfitting, which indicates the ability of the model to learn and adapt to the data actively. Also, the performance of the model gradually improved on the validation set, which indicates that LSTM is a suitable model for stock market forecasting.

- ## Comparative Analysis and Conclusions

The SARIMAX model showed a reasonable agreement with the data, as it allowed me to pick up the time dependences, but it showed signs of non-normal distribution and inhomogeneity, which indicates the need to improve the model. Also, the high values of AIC and BIC suggest that other models may be more effective. On the other hand, deep learning models such as RNN and LSTM showed better performance, as the value of loss and MAE decreased continuously during training. Especially the LSTM model, which showed a strong ability to learn and perform well in the validation set, which shows its ability to pick up complex nonlinear patterns in the data, and now I will present a diagram showing the improvement of the performance of the LSTM model during training in terms of the value of Loss and MAE across different horizons.

of Loss and MAE across different horizons.

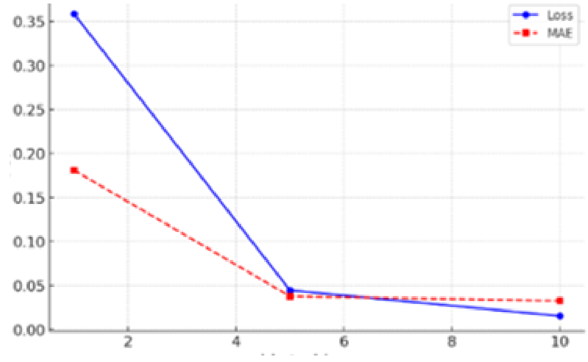


Figure (9)

The graph shows the improvement in the performance of the LSTM model as

The number of training sessions increases.

The graph shows the improvement of the performance of the LSTM model with the increase in the number of training sessions, as the value of Loss decreases from 0.3586 to 0.0152, and MAE decreases from 0.1809 to 0.0325, which indicates greater accuracy in prediction and reduction of errors. This indicates that the model gradually learned and became more efficient compared to the SARIMAX model, which despite achieving acceptable results, the high AIC and BIC values and statistical tests indicate the accuracy limit. Therefore, deep learning models such as LSTM are more suitable for analyzing and predicting financial markets because of their ability to capture complex relationships in temporal data.

Data Description:

The theme of this research is the use of historical data for the stock market to evaluate models for predicting future trends, including daily stock prices, such as opening and closing, highest and lowest prices, and trading volume, as well as a reliable source covering a specific period. Subjecting the data to cleaning and analysis processes, including removing outliers, unifying time intervals, and testing the establishment of the time series using ADF, taking differences when necessary to make them analyzable with models like SARIMAX. Also, the study of the relationship between past prices and stock performance using variables such as historical prices and technical indicators, including moving averages. The amount of data reached 2421 points, and it was divided into training and test groups to accurately evaluate the performance of the models. In order to ensure the quality of predictions, the theme of Alboaqi analysis is to realize the normal distribution, which helps to improve the accuracy of the results.

Experiment Setup:

The theme of deciding the experiment to evaluate the performance of different models in forecasting stock markets using time analysis and machine learning techniques. Including the SARIMAX model experiment, which is based on time series analysis, in addition to deep neural networks such as RNN and LSTM, which extract complex relationships between variables. The theme of experiment implementation using Python and libraries such as Scikit-learn, TensorFlow, Keras, Pandas, with data preparation through cleaning and testing deployments using ADF, then dividing into 80% for training and 20% for testing. To evaluate the performance, the theme employs criteria such as MAE, MSE, AIC, BIC, as well as statistical tests for panel analysis. The theme of recording SARIMAX parameters experimentally, while the theme of training LSTM on 10 epochs using Backpropagation. And finally, perform a comparative analysis between the performances of the models to determine the most efficient in predicting the behavior of the financial markets.

Evaluation Measures:

To evaluate the performance of the models used in this research (SARIMAX model and deep neural networks such as (RNN/LSTM), the theme of defining a set of criteria that aims to measure the accuracy of predictions and the ability of the models to pick up temporal patterns in the data. The evaluation criteria include the following:

- **Error metrics**

Mean Absolute Error (MAE): It is used to calculate the average difference between the predicted value and the actual value. The smaller the MAE value, the more accurate the model [39].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\lceil y - y|$$

- **Mean Squared Error (MSE):**

Measuring the error difference between the actual values and predictions. The consequences of big mistakes are bigger than MAE. , and the following formula [40].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\lceil y - y|$$

- **Root Mean Squared Error (RMSE):**

It is the square root of MSE and gives a better idea of the extent of error distribution. It reflects how close my predictions are to the actual values. And the following formula [41].

$$RMSE = \sqrt{MSE}$$

- **Distribution ratios**

Ljung-Box Test (Ljung-Box Test): It is used to test if it follows random distribution (i.e. there is no intrinsic correlation). The high values of the probability level indicate that the model still predicts al-Bawaqi well. In this experiment, the theme is used to determine if the remaining data show a non-interpretable temporal relationship before the model [42].

- **Jarque-Bera test (Jarque-Bera Test):**

The test is used to determine if the data (signatures) follow the normal distribution. Small values (P-value < 0.05) indicate that the data do not follow normal distribution. In this experiment, the results showed that alboaqi does not follow the normal distribution, which indicates the necessity of improving the model or changing it **[43].**

- **Quality measures (Model Selection Criteria)**

AIC (Akaike Information Criterion): scale used to choose the best model based on the balance between model accuracy and complexity. The lower the AIC, the better the model. And the following formula

$$AIC = 2k - \ln(L)\,2$$

- **Prediction accuracy**

The topic of employing prediction accuracy in general to measure the model's ability to accurately predict future market values. The prediction accuracy of the measurements using the mentioned measures such as MAE and MSE in addition to the visual analysis of the graphs.

Results Analysis:

The evaluation of the models used in this research (SARIMAX model and deep neural networks such as RNN/LSTM) using a set of metrics based on errors and distribution and predictions. Below is an analysis of the results obtained from the models:

- Analysis of the results of the SARIMAX model

The SARIMAX (1,1,2) model shows good results in the analysis of time series of the capital markets, but it faces some challenges regarding the normal distribution of the residuals and establishing variance over time. When analyzing the performance measures, it was found that the value of Log Likelihood reached -9728.179, which indicates the ability of the model to interpret the data in an acceptable manner, despite the existence of room for improvement. Also, the AIC and BIC values were recorded at 19464.358 and 19487.524, respectively, which indicates the possibility of improving the model by recording the parameters or employing other more suitable models. Regarding the statistical coefficients, the autoregressive coefficients (AR) showed a strong effect of the previous variables on the predictions, which reflects the strong temporal correlation in the data. Also, moving average (MA) transactions showed a significant effect in correcting previous errors, which indicates the model's ability to adapt to fluctuations and temporal corrections. As for the quality tests, the Ljung-Box test showed that the variables do not contain intrinsic correlation, which indicates that the model captures the important temporal patterns actively. However, the Jarque-Bera test showed that the data does not follow a normal distribution, which calls for consideration of model optimization using non-normal data fitting techniques. In addition, the Heteroskedasticity test detects the existence of non-constant variation over time, which reflects the fluctuation in the data and may require additional adjustments to the model to achieve higher accuracy in predictions.

- Analysis of the results of deep neural network models (RNN/LSTM)

The RNN/LSTM model shows excellent performance in learning and improving predictions through training steps. During the implementation of 10 training periods, the results showed significant improvement in prediction accuracy, which reflects the ability of the model to learn and adapt to the data actively. Performance metrics: In the first phase, the value of loss was about 0.3240, while MAE was at 0.1441, which indicates that the model has started to learn patterns, but the errors are still relatively high. After Khums al-Haqb, the loss decreased to 0.0444 and the MAE improved to 0.0376, which indicates the development of the model and the increase in the accuracy of predictions. When reaching the tenth stage, the loss reached 0.0152, while the MAE value was 0.0325, which indicates that the model learned the patterns well and became able to predict more accurately. Learning and development: The results show a gradual decrease in total loss and MAE across different directions, which reflects that the model continues to learn without overfitting. Also, the performance on validation data shows continuous improvement, which shows the ability of the model to adapt to different data and predict patterns more accurately. Comparison between the two models: The SARIMAX model shows a good ability to predict temporal trends, but it faces challenges related to distribution problems and non-establishment of variability. On the other hand, the superiority of the RNN/LSTM model in improving the accuracy of predictions through training sessions, which makes it more efficient in dealing with non-linear data and adapting to complex patterns better than SARIMAX. and we can say that improving the SARIMAX model using techniques such as ARIMA or GARCH, as these models can better deal with the inhomogeneous variation in the data, which improves the accuracy of predictions. As for the improvement of deep neural networks, this can be done by recording the parameters more accurately using techniques such as Grid Search or Random Search, which helps in finding the optimal values that enhance the performance of the model. In order to choose the best model, the performance between SARIMAX and deep neural networks should be compared on more data, which allows determining the most efficient model according to the nature of the data studied.

Consistency Examination:

Consistency checking is an essential element in any analytical experiment to ensure the agreement of results and accuracies across different models. In this research, the topic of analyzing the performance of SARIMAX model and deep neural network model (RNN/LSTM) in terms of consistency across several key aspects. The SARIMAX (1,1,2) model showed acceptable consistency in predictions but faced some problems in the normal distribution of errors and non-stationary variances, while the RNN/LSTM model achieved a gradual improvement in accuracy with a continuous decrease in Loss and MAE over time, which reflects continuous learning and improvement in performance. When comparing the two models, it is shown that SARIMAX is able to pick up temporal patterns but suffers from challenges in establishing long-term performance, while RNN/LSTM shows higher predictive accuracy and continuous improvement in performance, which makes it more stable. Also, the ability of the models to provide consistent performance over different time periods was tested, where SARIMAX showed some challenges due to inhomogeneous variance and lack of normal distribution, while neural networks showed continuous improvement over time, which shows the ability to predict consistently. To ensure the accuracy of the models, the theme of employing tests such as Ljung-Box and Jarque-Bera, as the first discovered the absence of intrinsic correlation in the SARIMAX values, while the second showed that the values in all the models do not follow the normal distribution, which calls for the optimization of the distribution to obtain more accurate results. In future forecasting, SARIMAX faces challenges in long-term stability due to statistical inhomogeneity, while RNN/LSTM excels due to its ability to learn patterns and adapt to non-linear data, making it more efficient in predicting financial markets. Based on these results, it can be concluded that RNN/LSTM enjoys higher consistency across different time periods compared to SARIMAX, which faces difficulties in dealing with normal distribution and establishing variance, which makes deep neural networks a more accurate and effective option in analyzing financial markets.

Conclusion

During the comprehensive analysis of the data and samples used in this study, the following conclusions can be reached Effectiveness of the SARIMAX model: Despite some challenges related to non-normal distribution and inhomogeneous variation in the data, the SARIMAX model provides a powerful tool for time series analysis. and predict future trends in the financial markets. The results show that the model can pick up general time trends, which makes it an effective tool in financial data analysis. Improving the performance of deep neural networks (RNN/LSTM): The introduction of deep neural networks has promising results in improving the accuracy of predictions with the review of training courses. The decrease in the value of loss and Mae with the improvement of the performance on the verification data indicates the ability of the network to learn patterns actively. However, it is important to compare the performance of this model with the SARIMAX model to determine the best. Analysis of warnings and future challenges: The warnings related to scikit-learn library updates indicate the need to update the code to match future versions. Although these warnings do not affect the results of the analysis now, ignoring them may affect the employability of the model in the future. The importance of testing the stability of the model: the model needs to verify the stability of the data and make sure that the remaining errors are randomly distributed. This helps in improving the performance of the models and guarantees accuracy in forecasting. Improvement horizons and practical applications: It is possible to improve existing models by using techniques such as online search to select the best parameters. Also, the results indicate that advanced models such as SARIMAX and RNN/LSTM provide institutions with powerful tools to improve investment strategies, identify financial risks, and investigate competitive advantages in markets

Based on these results, these models can greatly contribute to the improvement of corporate strategies in various fields such as financial analysis, investment decision-making, and management.

References

Ajiga, D., Adeleye, R. A., Asuzu, O. F., & Ndubuisi, N. L. (2024). Review of AI techniques in financial forecasting: Applications in stock market analysis. Finance & Accounting Research Journal, 6(2), 125-145. https://doi.org/10.51594/farj.v6i2.784

Bengani, V. (2024). Hybrid learning systems: Integrating traditional machine learning with deep learning techniques. ResearchGate. https://doi.org/10.13140/RG.2.2.10461.22244/1

Latif, S., Javaid, N., Aslam, F., Aldegheishem, A., Alrajeh, N., & Bouk, S. H. (2024). Enhanced prediction of stock markets using a novel deep learning model PLSTM-TAL in urbanized smart cities. Heliyon, 10(6), e27747. https://doi.org/10.1016/j.heliyon.2024.e27747

GAO, H., Kou, G., Liang, H., Zhang, H., Chao, X., Li, C.-C., & Dong, Y. (2024). Machine learning in business and finance: A literature review and research opportunities. Financial Innovation, 10, Article 86. https://doi.org/10.1186/s40854-024-00277-1

Rane, N., Paramesha, M., Choudhary, S. P., & Rane, J. (2024). Machine learning and deep learning for big data analytics: A review of methods and applications. Zenodo. https://doi.org/10.5281/zenodo.12271006

Yaqoob, A., Aziz, R. M., & Verma, N. K. (2023). Applications and techniques of machine learning in cancer classification: A systematic review. Human-Centric Intelligent Systems, 3, 588–615. https://doi.org/10.1007/s44230-023-00149-1

Shah, J., Vaidya, D., & Shah, M. (2022). A comprehensive review on multiple hybrid deep learning approaches for stock prediction. Intelligent Systems with Applications, 16, 200111. https://doi.org/10.1016/j.iswa.2022.200111

Bakar, S., & Ng Chui Yi, A. (2016). The impact of psychological factors on investors' decision making in Malaysian stock market: A case of Klang Valley and Pahang. Procedia Economics and Finance, 35, 319-328. https://doi.org/10.1016/S2212-5671 (16)00040-

Pan, E. (2024). Machine learning in financial transaction fraud detection and prevention. Transactions on Economics Business and Management Research, 5, 243-249. https://doi.org/10.62051/16r3aa10

Martinez-Castillo, C., Astray, G., Mejuto, J. C., & Simal-Gandara, J. (2019). Random forest, artificial neural network, and support vector machine models for honey classification. EFood, 1(1), 69-76. https://doi.org/10.2991/efood.k.191004.001

Adebowale, M., & Lwin, K. (2019, August). Deep learning with convolutional neural network and long short-term memory for phishing detection. Proceedings of the 2019 IEEE International Conference on Software, Knowledge, Information Management and Applications (SKIMA). https://doi.org/10.1109/SKIMA47702.2019.8982427

Chaudhry, M., Shafi, I., Mahnoor, M., & Ashraf, I. (2023). A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective. Symmetry, 15(9), 1679. https://doi.org/10.3390/sym15091679

Udeh, C. A., Orieno, O. H., Daraojimba, O. D., & Oriekhoe, O. I. (2024). Big data analytics: A review of its transformative role in modern business intelligence. Computer Science & IT Research Journal, 5(1), 219-236. https://doi.org/10.51594/csitrj.v5i1.718

Kolambe, M. (2024). Forecasting the future: A comprehensive review of time series prediction techniques. Journal of Electrical Systems, 20(2s), 575-586. https://doi.org/10.52783/jes.1478

Sasabella. (2024). Correlation analysis between financial ratios and stock prices of real estate companies listed on the Indonesia Stock Exchange. Jurnal Ilmiah Ekonomi Global Masa Kini, 15(1), 24-31. https://doi.org/10.36982/jiegmk.v15i1.3905

Nugroho, H. (2023). A review: Data quality problem in predictive analytics. IJAIT (International Journal of Applied Information Technology), 7(2), 79. https://doi.org/10.25124/ijait.v7i02.5980

The Pecan Team. (2024, June 13). Mastering model complexity: Avoiding underfitting and overfitting pitfalls. Pecan Blog. Retrieved from https://www.pecan.ai/blog/mastering-model-complexity-avoiding-underfitting-and-overfitting-pitfalls

Chakraborty, S., Tomsett, R., Raghavendra, R., & Gurram, P. (2017, August). Interpretability of deep learning models: A survey of results. In Proceedings of the UIC-ATC 2017 Conference. https://doi.org/10.1109/UIC-ATC.2017.8397411.

Xie, Y. (2023). Stock price forecasting: Traditional statistical methods and deep learning methods. Highlights in Business Economics and Management, 21, 740-745. https://doi.org/10.54097/hbem.v21i.14754

Alamu, O. S., & Siam, M. K. (2024). Stock price prediction and traditional models: An approach to achieve short-, medium- and long-term goals. Journal of Intelligent Learning Systems and Applications, 16(4), 363–383. https://doi.org/10.4236/jilsa.2024.164018

Chang, V., Xu, Q. A., Anyamele, C., & Wang, H. (2024). Predicting economic trends and stock market prices with deep learning and advanced machine learning techniques. Electronics, 13(17), Article 3396. https://doi.org/10.3390/electronics13173396

Mintarya, L. N., Halim, J. N. M., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. Procedia Computer Science, 216, 96-102. https://doi.org/10.1016/j.procs.2022.12.115

Jabed, M. I. K. (2024). Stock market price prediction using machine learning techniques. American International Journal of Sciences and Engineering Research, 7(1), 1-6. https://doi.org/10.46545/aijser.v7i1.308

Tran, P., Pham, T. K. A., Phan, H. T., & Nguyen, C. V. (2024) Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam. Humanities and Social Sciences Communications, 11, Article 393. https://doi.org/10.1057/s41599-024-01933-7

Ayyildiz, N., & Iskenderoglu, O. (2024). How effective is machine learning in stock market predictions? Heliyon, 10(2), e24123. https://doi.org/10.1016/j.heliyon.2024.e24123

Sheth, D., & Shah, M. (2023). Predicting stock market using machine learning: Best and accurate way to know future stock prices. International Journal of System Assurance Engineering and Management, 14, 1–18. https://doi.org/10.1007/s13198-023-01745-9

Khan, A. H., Shah, A., Ali, A., Shahid, R., Zahid, Z. U., Sharif, M. U., Jan, T., & Zafar, M. H. (2023). A performance comparison of machine learning models for stock market prediction with novel investment strategy. PLOS ONE, 18(9), e0286362. https://doi.org/10.1371/journal.pone.0286362

Yinka-Banjo, C. O., Akinyemi, M. I., & Er-rabbany, B. (2023). Stock market prediction using a hybrid of deep learning models. International Journal of Financial Studies Economics and Management, 2(2), 1–16. https://doi.org/10.61549/ijfsem.v2i2.111

Long, W., GAO, J., Bai, K., & Lu, Z. (2024). A hybrid model for stock price prediction based on multi-view heterogeneous data. Financial Innovation, 10(48). https://doi.org/10.1186/s40854-024-00548-9

Gharooni, M., Idris, N., & Madjlesi, Z. (2023). Enhancing LSTM predictions: A novel approach to stock market efficiency evaluation with retrainable LSTM model. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4782069

Sharma, V., Rakhra, M., & Mathur, G. (2023). Hybrid approaches for stocks prediction and recommendation system. E3S Web of Conferences, 453, 01047. https://doi.org/10.1051/e3sconf/202345301047

Luo, X., Yin, J., & Wei, D. (2024). Complex-systems analysis of the CSI 300 index: Evolution, resilience, and prediction in stock correlation network. Systems, 12(8), 285. https://doi.org/10.3390/systems12080285

Ozupek, O., Yilmaz, R., Ghasemkhani, B., Birant, D., & Kut, R. A. (2024). A novel hybrid model (EMD-TI-LSTM) for enhanced financial forecasting with machine learning. Mathematics, 12(17), 2794. https://doi.org/10.3390/math12172794

John, A., & Latha, T. (2023). Stock market prediction based on deep hybrid RNN model and sentiment analysis. Automatika, 64(4), 981-995. https://doi.org/10.1080/00051144.2023.2217602

Balasubramanian, P., Chinthan, P., Badarudeen, S., & Sriraman, H. (2024). A systematic literature survey on recent trends in stock market prediction. PeerJ Computer Science, 10(4), e1700. https://doi.org/10.7717/peerj-cs.1700

Lin, C. Y., & Marques, J. A. L. (2024). Stock market prediction using artificial intelligence: A systematic review of systematic reviews. Social Sciences & Humanities Open, 9, 100864. https://doi.org/10.1016/j.ssaho.2024.100864

Ajiga, D., Adeleye, R. A., Asuzu, O. F., Ndubuisi, N. L., & [Author(s) 5-6]. (2024). Review of AI techniques in financial forecasting: Applications in stock market analysis. Finance & Accounting Research Journal, 6(2), 125-145. https://doi.org/10.51594/farj.v6i2.784

Muhammad, D., Ahmed, I., Naveed, K., & Bendechache, M. (2024). An explainable deep learning approach for stock market trend prediction. Heliyon, 10(21), e40095. https://doi.org/10.1016/j.heliyon.2024.e40095

Chai, T., & Draxler, R. R. (2014). Root means square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. Journal of the Operational Research Society, 66(3). https://doi.org/10.1057/jors.2014.124

Hodson, T. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development, 15(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Enders, Walter (2010). Applied Econometric Time Series (Third ed.). New York: Wiley. pp. 69–70. ISBN 978-0470-50539-7.

Glinskiy, V. V., Ismayilova, Y., Khrushchev, S., & Zaykov, K. (2024). Modifications to the Jarque–Bera test. Mathematics, 12(16), 2523. https://doi.org/10.3390/math1216252.