

Forecasting Stock Prices with an Integrated Approach Combining ARIMA and Machine Learning Techniques ARIMAML

Ali Abdulhafidh Ibrahim , Bilal N. Saeed, Marwa A. Fadil

Department of Economics of Banking Management, Al-Nahrain University, Baghdad, Iraq
Email: dr_ali9@yahoo.com, belal@eco.nahrainuniv.edu.iq, marwaomjana@gmail.com

How to cite this paper: Ibrahim, A.A., Saeed, B.N. and Fadil, M.A. (2023) Forecasting Stock Prices with an Integrated Approach Combining ARIMA and Machine Learning Techniques ARIMAML. *Journal of Computer and Communications*, 11, 58-70.

<https://doi.org/10.4236/jcc.2023.118005>

Received: June 20, 2023

Accepted: August 12, 2023

Published: August 15, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Stock market prediction has long been an area of interest for investors, traders, and researchers alike. Accurate forecasting of stock prices is crucial for financial decision-making and risk management. This paper presents a novel approach to predict stock prices by integrating Autoregressive Integrated Moving Average (ARIMA) and Exponential smoothing and Machine Learning (ML) techniques. Our study aims to enhance the predictive accuracy of stock price forecasting, which can significantly impact investment strategies and economic growth in this research paper implement the ARIMAML proposed method to predict the stock prices for Investment Bank of Iraq.

Keywords

Stock Prediction, ARIMA Model, Exponential Smoothing Model, Machine Learning, ARIMAML Model

1. Introduction

The stock market is a complex and dynamic system that is affected by various factors such as economic conditions, investor sentiment, and global events. Accurate and timely prediction of stock prices can lead to profitable investment decisions and optimal allocation of financial resources. Traditional statistical methods such as ARIMA and exponential smoothing have been widely used for time series analysis and forecasting. However, these methods have certain limitations in capturing the nonlinear relationships and complex patterns found in stock market data.

To address these challenges, this paper proposes an integrated approach that combines the strengths of exponential smoothing and ML techniques. ARIMA is

a well-established method for predicting linear time series, and forms the basis of our model. On the other hand, ML algorithms including neural networks, support vector machines, and decision trees have shown great potential in dealing with nonlinear patterns and improving prediction accuracy. [1]

Our approach includes developing preliminary ARIMA exponential smoothing models to capture linear and seasonal trends in stock price data. Then, the residual errors from the ARIMA exponential smoothing models are fed into the ML algorithm to capture nonlinear relationships and improve predictions. This two-stage process allows us to take advantage of both ARIMA and ML exponential smoothing techniques, resulting in more accurate and robust stock price predictions. [2]

The remainder of this paper is organized as follows: Section 1 presents Introduction and survey of the literature. Section 2 is the methodology of ARIMA exponential smoothing models and their application in forecasting stock prices. Section 3 results and discussions, while Section 4 concludes the paper with future research directions.

By combining ARIMA and ML exponential smoothing techniques, this study aims to contribute to the current body of research on stock price forecasting and provide valuable insights to investors, traders, and policymakers alike.

1.1. Literature Survey

Stock price forecasting is a critical aspect of investment decision-making and risk management in financial markets. Various methodologies have been proposed to predict stock prices, ranging from traditional statistical models to advanced machine learning techniques. Among these approaches, Autoregressive Integrated Moving Average (ARIMA) models have been extensively employed due to their ability to capture linear relationships and inherent simplicity. This literature survey aims to review the existing body of research on stock price forecasting using ARIMA models, highlighting key findings, methodologies, and challenges ARIMA models have been widely adopted in time series analysis to forecast univariate data, including stock prices. Several studies have investigated the effectiveness of ARIMA models in predicting stock prices across different financial markets, time horizons, and stock categories. The following subsections provide an overview of key studies and their findings in the domain of stock price forecasting using ARIMA models. [1]

1.2. Early Studies on ARIMA-Based Stock Price Forecasting

The application of ARIMA models in stock price prediction can be traced back to the seminal work of Box and Jenkins (1970). Early studies focused on the application of ARIMA models to predict stock price indices and individual stock prices. For instance, researchers like Fama (1970) and Fama and Blume (1966) explored the effectiveness of ARIMA models for stock price forecasting and found mixed results. While these early studies laid the groundwork for future

research they also highlighted the limitations of ARIMA models in capturing non-linear patterns in stock price data. [2] [3] [4].

1.3. Comparative Studies of ARIMA and Alternative Models

Several studies have compared the forecasting performance of ARIMA models with other statistical and econometric models, such as GARCH, EGARCH, and VAR. For example, Engle (1982) introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model was later extended to the Generalized ARCH (GARCH) model by Bollerslev (1986). These models have been compared to ARIMA in various studies, with mixed results regarding their relative performance (Brooks, 2008; Poon & Granger, 2003). Some studies have found ARIMA models to be competitive with more advanced models, while others have reported superior performance from alternative models. In comparative studies, ARIMA (Auto Regressive Integrated Moving Average) and alternative models, including Exponential Smoothing, have been assessed for time series forecasting. ARIMA excels in capturing linear dependencies and trends but struggles with nonlinear patterns and requires data stationarity. Hybrid approaches, combining ARIMA and Exponential Smoothing, attempt to leverage both strengths, enhancing forecasting accuracy and robustness. Evaluating multiple datasets using appropriate metrics is vital to select the most suitable model for specific forecasting tasks [5] [6].

1.4. Recent Advances in ARIMA-Based Stock Price Forecasting

In recent years, researchers have continued to explore ARIMA-based stock price forecasting, with some studies reporting improved accuracy through innovations in model selection and parameter estimation. For instance, Adeyemo *et al.* (2019) applied an optimized ARIMA model to forecast stock prices in the Nigerian stock market, reporting enhanced prediction accuracy compared to traditional ARIMA models. Additionally, Kumar and Thenmozhi (2020) investigated the use of ARIMA models for intraday stock price forecasting, demonstrating the applicability of ARIMA models for high-frequency financial data? [3] [4]

2. Models Methodology

2.1. Exponential Smoothing Method

It is one of the methods for forecasting in the short term (one time period in the future). This method is used to refine the time series data when the long-term direction of the series is unknown. Exponential smoothing is also known as one of the recognized techniques for predicting a time series. It gives results with high efficiency, as the exponential smoothing reduces the missing value. It is also known that it is a prediction method that relies on the weighting of the data, meaning that the new data has a greater weight than the old data. This method is distinguished from others by its accuracy as a result. Its reliance on the error in forecasting in the previous period [2] [7] [8]

$$Y_t = Y_{t-1} + \beta(F_t - F_{t-1}) \quad (1)$$

Y_t : Trend for period t ;

Y_{t-1} : Trend for the last period;

F_t : The predicted value for period t ;

F_{t-1} : The predicted value for the last period;

β : The smoothing constant ranges between (0 - 1), and this is the error rate.

2.2. ARIMA Model Methodology

Time series analysis is a vital tool for understanding and predicting temporal patterns in various domains, including economics, finance, and meteorology. Among the numerous methods employed in time series forecasting Autoregressive Integrated Moving Average (ARIMA) models have gained considerable popularity due to their simplicity and effectiveness ARIMA, an acronym for Autoregressive Integrated Moving Average, is a linear model used for forecasting univariate time series data. The ARIMA model is a combination of three components:

- Autoregressive (AR) component;
- Integrated (I) component;
- Moving Average (MA) component. [9] [10] [11]

These components help capture different aspects of the time series data, such as trends, seasonality, and noise. The ARIMA model is denoted as ARIMA(p, d, q), where p represents the order of the AR component, d represents the degree of differencing, and q represents the order of the MA component.

$$x_t = c + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2)$$

where y is the time series, c is a constant, φ 's are the autoregressive coefficients, θ 's are the moving average coefficients, a is the error term, and p and q are the orders of autoregression and moving average, respectively.

2.3. ML Technique and Their Relevance to Stock Market

Time series analysis is a statistical technique that involves analyzing time-based data to identify trends, patterns, and relationships in the data. It is often used in finance, economics, and other fields where data is collected over time [12].

In the context of stock market analysis, time series analysis can be used to analyze stock prices over time to identify trends and patterns that can be used to make predictions about future stock prices [10].

Machine learning (ML) can be used to enhance time series analysis by providing more sophisticated methods for modeling time-series data. For example, ML algorithms such as artificial neural networks (ANNs) and recurrent neural networks (RNNs) can be used to model complex relationships in time-series data. These algorithms can be trained on historical stock price data to make predictions about future stock prices [13].

One common application of time series analysis in stock market analysis is trend analysis, which involves identifying the direction of a stock's price over

time. This can be done by fitting a regression model to the stock price data and using it to make predictions about future prices. Another application of time series analysis in stock market analysis is seasonality analysis, which involves identifying recurring patterns in stock prices over time. This can be done by decomposing the time-series data into its trend, seasonal, and residual components [14] [15] [16].

2.4. Proposed Integrated Approach between ARIMA & ML (ARIMAML)

The proposed integrated approach between ARIMA (Auto Regressive Integrated Moving Average) and ML (Machine Learning) aims to combine the strengths of both methods to improve forecasting accuracy and handle complex time series data more effectively. ARIMA is a traditional time series forecasting method, while ML refers to a set of techniques that allow machines to learn patterns and relationships from data without being explicitly programmed

The main advantage of this integrated approach is its ability to leverage ARIMA's ability to capture the linear components of the time series and ML models' capacity to capture more complex patterns and relationships. This combination can be particularly useful when dealing with time series data that exhibits non-linear or irregular patterns, making the forecasting more robust and accurate.

This proposal aims to develop a hybrid framework that combines machine learning (ML) algorithms with the Autoregressive Integrated Moving Average (ARIMA) model on the one hand and with the exponential smoothing model on the other hand to improve time series prediction. The proposed methodology aims to harness the strengths of both ML and ARIMA & exponential smoothing techniques to increase prediction accuracy and effectively manage complex, nonlinear patterns in data. Comparing the two models ML models will be developed following a two-step process [17] (see **Figure 1**) training and testing.

3. Results and Discussions

3.1. Data

The dataset for this research paper contains the close stock price for Iraqi Banks:

Investment Bank of Iraq. (see **Table 1**)

3.2. Implement Investment Bank of Iraq Data

A) Exponential smoothing model:

When implementing the exponential smoothing model through machine learning, quick and accurate results emerged for the benefit of investors, as the error square appeared with a very small percentage of 0.0144. The practical results showed a clear decline in stock prices in the predicted years.

The data in this research paper contains 60 records ($n = 60$), and by implement ML processing upon this data it divides the data for both two groups Banks as follows:

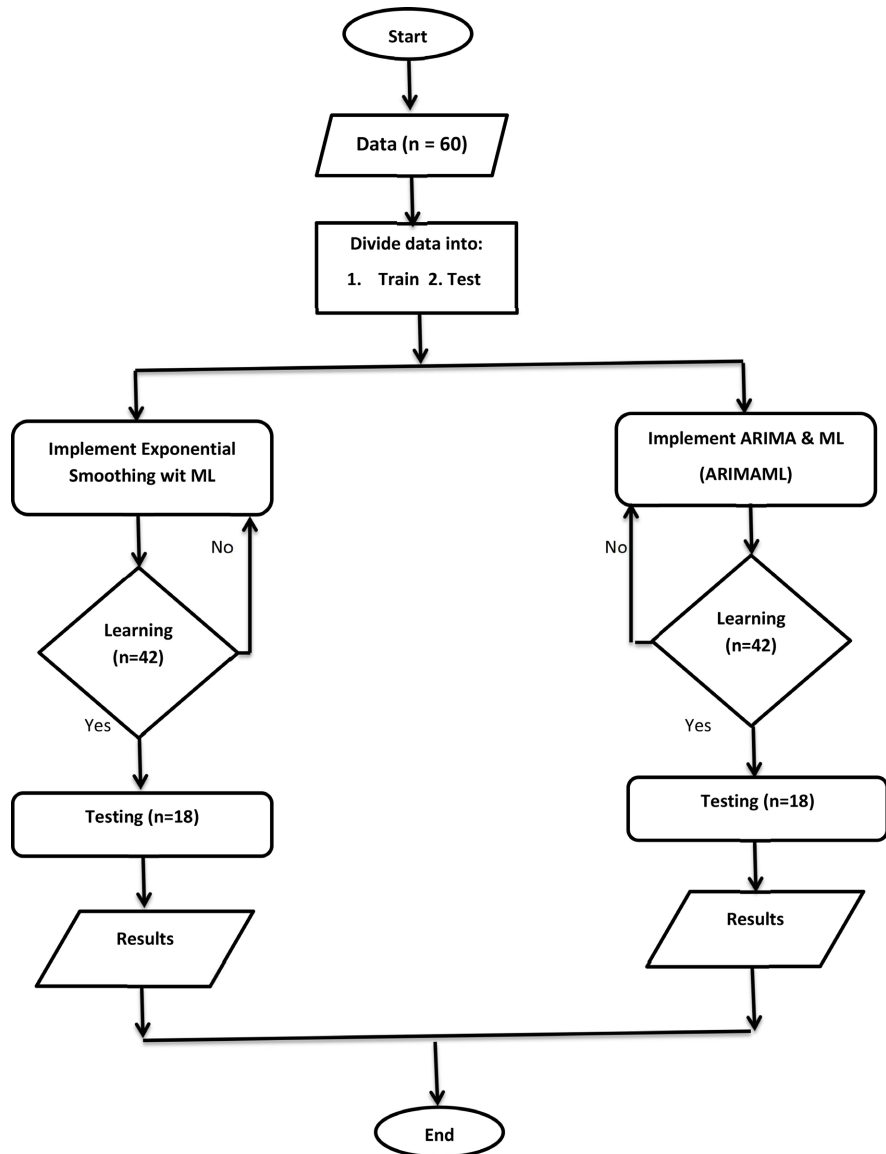


Figure 1. Flowchart of ARIMA and exponential smoothing using machine learning.

Table 1. Data file contains stock prices for investment bank of Iraq. [18]

	Date	Investment Bank of Iraq
1.	2017/1	0.660
2.	2017/2	0.680
3.	2017/3	0.600
4.	2017/4	0.550
5.	2017/5	0.490
6.	2017/6	0.540
7.	2017/7	0.460
8.	2017/8	0.460

Continued

9.	2017/9	0.460
10.	2017/10	0.410
11.	2017/11	0.410
12.	2017/12	0.420
13.	2018/1	0.460
14.	2018/2	0.480
15.	2018/3	0.470
16.	2018/4	0.440
17.	2018/5	0.410
18.	2018/6	0.390
19.	2018/7	0.370
20.	2018/8	0.340
21.	2018/9	0.300
22.	2018/10	0.300
23.	2018/11	0.280
24.	2018/12	0.280
25.	2019/1	0.270
26.	2019/2	0.210
27.	2019/3	0.200
28.	2019/4	0.190
29.	2019/5	0.270
30.	2019/6	0.240
31.	2019/7	0.230
32.	2019/8	0.240
33.	2019/9	0.230
34.	2019/10	0.260
35.	2019/11	0.300
36.	2019/12	0.270
37.	2020/1	0.250
38.	2020/2	0.240
39.	2020/3	0.240
40.	2020/4	0.230
41.	2020/5	0.220
42.	2020/6	0.210
43.	2020/7	0.250
44.	2020/8	0.260
45.	2020/9	0.260

Continued

46.	2020/10	0.250
47.	2020/11	0.230
48.	2020/12	0.220
49.	2021/1	0.230
50.	2021/2	0.260
51.	2021/3	0.260
52.	2021/4	0.300
53.	2021/5	0.270
54.	2021/6	0.270
55.	2021/7	0.270
56.	2021/8	0.290
57.	2021/9	0.290
58.	2021/10	0.270
59.	2021/11	0.260
60.	2021/12	0.300

1) Train data: It is trained from part of the data within the time series for the selected period with size equal to 42 records.

2) Test data: After training the data, the remaining part of the data with size 18 records is tested to pave the way for the prediction process.

Implement Python code of exponential smoothing model (see **Figure 1**) using the data (**Table 1**), getting the following results (see **Figure 2**).

The exponential smoothing time series mode plot using ML displays the IBI real data, projected data, and forecast values (see **Figure 3**).

```
[ ] mean_squared_error(test_data['Investment Bank of Iraq'],test_predictions)
0.01445921155005051

[ ] np.sqrt(mean_squared_error(test_data['Investment Bank of Iraq'],test_predictions)
0.12024646169454846
```

B) ARIMA:

The data in this research paper contains 60 records (n = 60), and by implement ML processing upon this data it divide the data for both the two Banks as follows:

1) Train data: It is trained from part of the data within the time series for the selected period with size equal to 42 records.

2) Test data: After training the data, the remaining part of the data with size 18 records is tested to pave the way for the prediction process.

Using python program to implement RAIMA using ML, starts with reading data (see **Figure 4**).

After implement the python program the following results (see **Figure 5**).


```
[ ] test_predictions
```

```
2020-07-01    0.189224
2020-08-01    0.185062
2020-09-01    0.177548
2020-10-01    0.169502
2020-11-01    0.171852
2020-12-01    0.169472
2021-01-01    0.173240
2021-02-01    0.174045
2021-03-01    0.160706
2021-04-01    0.149219
2021-05-01    0.140797
2021-06-01    0.142812
2021-07-01    0.127821
2021-08-01    0.125009
2021-09-01    0.119934
2021-10-01    0.114499
2021-11-01    0.116086
2021-12-01    0.114479
Freq: MS, dtype: float64
```

Figure 2. Prediction of investment bank stock prices.

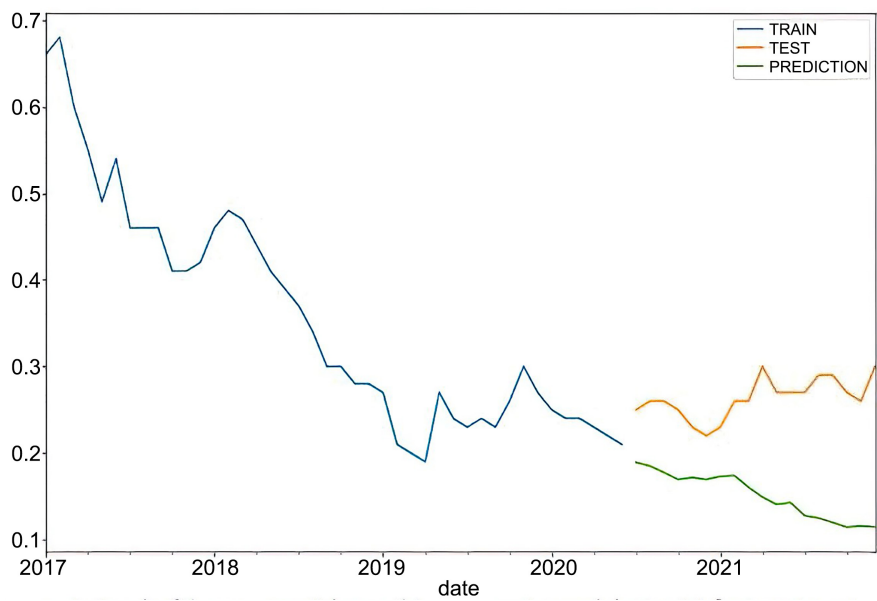


Figure 3. Graph of the exponential smoothing time series model using ML for investment bank of Iraq.

```
import pandas as pd
import numpy as np
from pmdarima import auto_arima
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

# Read the data from the CSV file
data = pd.read_csv('/content/3. Bank.csv', parse_dates=['date'],
                  index_col='date') # Replace 'date_column' with the name of

# Replace 'column_name' with the name of the column containing the time series
time_series_data = data['Investment Bank of Iraq']
```

Figure 4. Reading data files using Python code.

```

stepwise_fit.summary()
Best model: ARIMA(0,1,0)(0,0,0)[0]
Total fit time: 0.323 seconds
                ARIMA Results
Dep.Variable: y                No. Observations: 60
Model:    ARIMA(0,1,0) Log Likelihood 102.390
                                AIC      -202.780
                                BIC      -200.703

```

Figure 5. The output of Python program for investment bank of Iraq.

4. Discussions

Exponential smoothing model:

Exponential smoothing is a widely used time series forecasting method with significant applicability in finance, particularly for stock price prediction. Its suitability stems from its ability to adapt to trends and seasonality often present in stock prices. By emphasizing recent data while diminishing the impact of older observations, exponential smoothing captures short-term fluctuations and dynamically adjusts to changing market conditions. Moreover, the optimization of the smoothing parameter allows the model to be fine-tuned using historical data, enhancing its forecasting accuracy and providing valuable insights into potential future price trends in the stock market.

ARIMA Depending on the above results, we get the following discussions,

The ARIMA(0, 1, 0) (0, 0, 0) [0] model can be interpreted as follows:

$p = 0$: This means that there is no autoregression component in the model. This means that the model is not using lagged values of the time series as predictors.

$d = 1$: This means that the model uses first-order differencing. This means that the model is using the difference between consecutive values in the time series as the input data. This can be used to remove trends or seasonality from the time series, making it easier to model.

$q = 0$: This means that there is no moving average component in the model. This means that the model is not using a weighted average of past errors as a predictor.

The brackets (0, 0, 0) specify the parameters for the seasonal component of the model, if there is one. In this case, there is no seasonal component, as indicated by the (0, 0, 0) parameters.

The [0] at the end of the ARIMA specification indicates that the model has no exogenous variables, or external predictors, that can be used to improve the model's accuracy.

In conclusion, the ARIMA(0, 1, 0) (0, 0, 0) [0] model is a simple time series model that uses first-order differencing to remove trends or seasonality, but does not use lagged values or past errors as predictors, and does not include any external predictors.

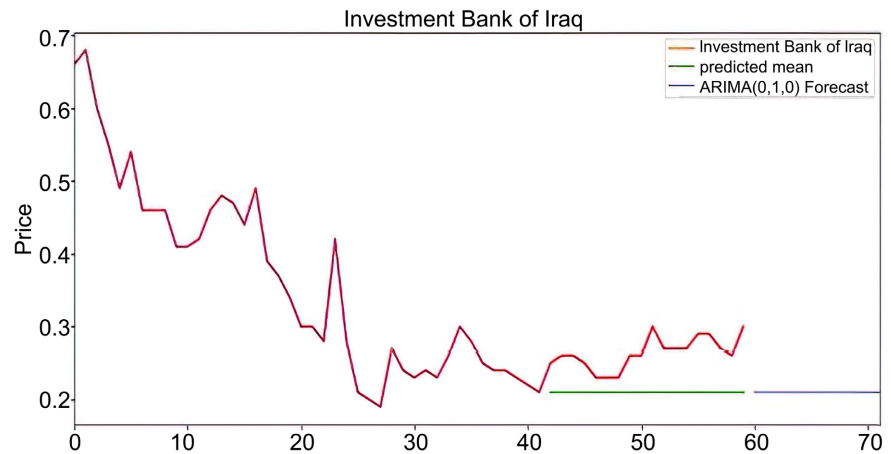


Figure 6. The plot of time series mode ARIMAML for Investment Bank of Iraq.

The log-likelihood is a measure of how well the model fits the data. In general, a higher log-likelihood indicates a better fit. The log-likelihood for the ARIMA model is given as 102.390.

The AIC is a measure of the trade-off between the goodness of fit of the model and the complexity of the model. A lower AIC value indicates a better trade-off between goodness of fit and complexity, and is therefore preferred. The AIC for the ARIMAML model is given as -202.78 .

The BIC is similar to the AIC, but places a greater emphasis on model complexity. Like the AIC, a lower BIC value indicates a better trade-off between goodness of fit and complexity. The BIC for the ARIMAML model is given as -200.703 .

Mean Square Error (MSE) = 0.00335.

The plot of time series mode ARIMAM Through fast, accurate and reliable results in the field of prediction through the use of Python program in prediction and evaluation, the log-likelihood, AIC, and BIC are three measures that can be used to evaluate the performance of **ARIMAML** models. The log-likelihood measures the goodness of fit, the AIC measures the trade-off between goodness of fit and complexity, and the BIC places a greater emphasis on model complexity. The values of 102.390, -202.78 , and -200.703 indicate that the ARIMA model has a good fit to the data and a good trade-off between goodness of fit and complexity, which means increased confidence in the use of the program.

5. Conclusions

In this paper, we proposed an integrated approach combining Autoregressive Integrated Moving Average (ARIMA) and Machine Learning (ML) techniques for forecasting stock prices. Our study aimed to enhance the predictive accuracy of stock price forecasting, recognizing the importance of accurate predictions for financial decision-making and risk management.

By integrating the strengths of both ARIMA and ML methods, we were able to leverage the time series analysis capabilities of ARIMA along with the flexibility

and adaptability of ML algorithms. This integration allowed us to capture both the linear and nonlinear patterns in stock price data, resulting in improved forecasting performance.

Through our empirical analysis for the stock prices of Investment Bank of Iraq, we demonstrated the effectiveness of our integrated approach. The results showed that our model outperformed traditional ARIMA models and standalone ML models in terms of prediction accuracy. The integrated approach provided more robust and reliable forecasts, enabling investors and traders to make informed decisions and manage risks effectively.

Our research contributes to the existing body of knowledge in stock market prediction by offering a practical and effective approach that combines traditional time series analysis techniques with modern ML algorithms. The integration of ARIMA and ML provides a comprehensive framework for stock price forecasting, which can have significant implications for investment strategies and economic growth.

Future research can explore further refinements and extensions of the integrated approach, such as incorporating additional predictors, refining the model architecture, or exploring different ML algorithms. Additionally, applying the proposed approach to different stock markets or financial instruments would provide valuable insights into its generalizability and robustness.

In conclusion, our study highlights the benefits of an integrated approach combining ARIMA and ML techniques for forecasting stock prices. The proposed approach shows promise in improving prediction accuracy, thereby assisting investors, traders, and financial institutions in making more informed decisions and managing risks effectively. When measuring the square of the error for both sampling exponential smoothing and ARIMA when executed the Python program, the error square is shown

The Mean square error of the exponential smoothing model,

Mean Square Error (MSE) = 0.0144.

The square error of ARIMA,

Mean Square Error (MSE) = 0.00335.

It appears from the comparison of the two models that the more accurate is the model.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Fataliyev, K., Chivukula, A., Prasad, M. and Liu, W. (2021) Text-Based Stock Market Analysis: A Review. University of Technology Sydney, Sydney. ArXiv: 2106.12985v2.
- [2] Ibrahim, A.A. and Merhej, S.A.K. (2019) Forecasting the Bank of Baghdad Using the Box-Jenkins Methodology. *Journal of Dinars*, 15, 440-460.

- [3] Janeski, M. and Kalajdziski, S. (2010) Forecasting Stock Market Prices. In: Gusev, M., Ed., *ICT Innovations 2010 Web Proceedings*, ICT-ACT, Skopje, 107-116.
- [4] Filbeck, G., Baker, H.K. and Kiyamaz, H. (2020) *Equity Markets, Valuation, and Analysis*. Wiley, Hoboken, 80.
- [5] Marwala, L.R. (2008) *Forecasting the Stock Market Index Using Artificial Intelligence Techniques*. University of the Witwatersrand, Johannesburg, 48.
- [6] Hyndman, R.J. and Athanasopoulos, G. (2016) *Forecasting: Principles and Practice*. Second Edition, Monash University, Australia.
- [7] Berenson, M.L., Levine, D.M. and Krehbiel, T.C. (2009) *Basic Business Statistics: Concepts and Applications*. Pearson International Edition, London, 765.
- [8] Al-Shammari, I.H. and Al-Bayati, N.A. (2014) *Time Series Analysis—A Quantitative Method Using SPSS, Minitab & Reviews*. Al Jazeera Publishing and Publishing, Baghdad, 193.
- [9] Karasan, A. (2022) *Machine Learning for Financial Risk Management with Python*. O'Reilly Media, Sebastopol, 65.
- [10] Peixeiro, M. (2022) *Time Series Forecasting in Python*. Manning Publications Co., Shelter Island, 51.
- [11] Reinert, Time Series, Hilary Term. University in Oxford, England.
<http://www.stats.ox.ac.uk/~reinert2010.5>
- [12] Khalifa, A.R., Battal, A.H. and Hamad, A.A. (2019) Use Time Series Methods to Forecast Trading Prices for the Iraq Stock Exchange Duration (2005-2018). *AL-Anbar University Journal of Economic and Administration Sciences*, **11**, 238-259.
- [13] Abu Al-Yazid, M. (2020) *Machine Learning as an Indicator in the Future of Industrial Design*. College of Applied Arts, Badr University, Cairo.
- [14] Bontempi, G., Ben Taieb, S. and Le Borgne, Y.-A. (2013) Machine Learning Strategies for Time Series Forecasting. In: Aaufaure, M.-A. and Zimányi, E., Eds., *Business Intelligence. eBISS 2012. Lecture Notes in Business Information Processing*, Vol. 138, Springer, Berlin, 62-77. https://doi.org/10.1007/978-3-642-36318-4_3
- [15] Combs, A. and Roman, M. (2019) *Python Machine Learning*. 2 Edition. Packt Publishing Ltd., Birmingham.
- [16] Auffarth, B. (2021) *Machine Learning for Time-Series with Python: Forecast, Predict, and Detect Anomalies with State-of-the-Art Machine Learning Methods*. Packt Publishing, Birmingham, 18.
- [17] Lantz, B. (2015) *Machine Learning with R*. 2nd Edition, Packt Publishing, Birmingham, 31.
- [18] Iraq Stock Exchange.
<http://www.isx-iq.net/isxportal/portal/homePage.html?currLanguage=ar.com>