Predicting Market Performance Using Machine and Deep Learning Techniques

MOHAMED EL MAHJOUBY1, MOHAMED TAJ BENNANI1, MOHAMED LAMRINI1, BADRE BOSSOUFI2, THAMER A. H. ALGHAMDI3,4, AND MOHAMED EL FAR1

1Department of Computer Science, LPAIS Laboratory, Faculty of Science Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez 30003, Morocco
2LIMAS Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez 30003, Morocco
3Wolfson Centre for Magnetics, School of Engineering, Cardiff University, CF24 3AA Cardiff, U.K.
4Electrical Engineering Department, Faculty of Engineering, Al-Baha University, Al-Baha 65779, Saudi Arabia

Corresponding authors: Mohamed El Mahjouby (mohamed.elmahjouby@usmba.ac.ma) and Badre Bossoufi (badre.bossoufi@usmba.ac.ma)

ABSTRACT Today, forecasting the stock market has been one of the most challenging issues for the “artificial intelligence” AI research community. Stock market investment methods are sophisticated and rely on analyzing massive volumes of data. In recent years, machine-learning techniques have come under increasing scrutiny to assess and improve market predictions over traditional approaches. The observation in time is due to their dependence. Their predictions are crucial tasks in data mining and have attracted great interest and considerable effort over the past decades. Tackling this challenge remains difficult due to the inherent characteristics of time series data, including its high dimensionality, large volume of data, and constant updates. Exploration of Machine Learning and Deep Learning methods undertaken to enhance the effectiveness of conventional approaches. In this document, we aim precisely to forecast the performance of the stock market at the close of the day by applying various machine-learning algorithms on the two data sets “CoinMarketCap, CryptoCurrency” and thus analyze the predictions of the architectures.

INDEX TERMS Machine learning, deep learning, LSTM, ARIMA, linear regression.

I. INTRODUCTION The stock market is one of the indicators of a country’s economy. Few people excel at correctly comprehending the evolution of the stock trend, so many people are afraid to invest in stocks. By its forecast scope, data science (data analytics) has shattered the assumed paradigm of the only engagement of economics and finance in stocks. In the financial domain, stock price is a type of time series. Financial time series variations are dynamic, selective, nonlinear, nonstationary, and noisy, making forecasting challenging. One of the most pressing concerns is predicting stock prices effectively utilizing data mining or machine learning approaches. However, because the stock price will follow a random walk pattern, it is challenging to make forecasts based on the premise of the efficient market hypothesis. Furthermore, a stationary prediction approach is not viable since investors would quickly uncover such strategies, and good forecasting rules will self-destruct. The stock market is both a difficulty and a chance for investors to benefit. However, investors are aware of the “High Risk, High Return” philosophy and are looking for ways to reduce risk while increasing earnings. Fundamental analysis and technical analysis are the two methods of stock forecasting. Fundamental analysis forecasts prices using income statements, balance sheets, and cash flows to find chances for growth, financial health, and competitive advantages. It also examines the business and national environment, the country’s economy, and business possibilities in certain industrial groups. Technical analysis, on the other hand, forecasts stock values based on price history. Stock forecasting has long been a challenging topic for statisticians and finance experts. The reason for this forecast is to acquire stocks expected to gain in value and then sell stocks expected to fall in value. In general, there are two approaches to stock market forecasting. Two approaches
are utilized for analyzing stock markets. The initial method is fundamental analysis, which depends on a company’s strategy and essential data, including its market position, expenditures, and yearly growth rates. The second method involves technical analysis, relying on historical stock prices and values. This study forecasts future prices using past charts and trends. Financial experts used to be able to foretell stock market movements. However, developments in machine learning techniques have enabled data scientists to overcome prediction difficulties. In addition, computer scientists have begun to use machine-learning algorithms to increase the performance and accuracy of forecasting models.

The next step was to employ deep learning to improve the prediction models’ performance. Stock market forecasting is fraught with difficulties, and data scientists frequently encounter challenges when attempting to produce a predictive model. The overarching question is how to anticipate whether the price of given stocks will rise or fall over the next year. It is critical to address this issue. In this paper, we employed methodologies from Machine Learning and Deep Learning. In general, different academics utilized time series to forecast prices using deep learning and machine learning methodologies. This study provides recommendations to help investors make informed choices about closing prices.

The organization of our paper is as follows: Initially, we offer an introduction, and after an exploration of the literature review and various machine learning algorithms. Next, we feature an interview focused on deep learning long short-term memory. After that, we present and describe our methodologies. Subsequently, we present and analyze our findings. Finally, we conclude with final remarks.

II. LITERATURE REVIEW

Sun et al. [1] identified daily patterns in the Bitcoin market while examining semantics and insights connected to elements impacting Bitcoin’s price. As part of their research, they anticipated changes in Bitcoin’s daily price using both Random Forest and Bayesian regression models. Pirani et al. [2] introduced a comparative study into existing literature, examining the performance of the BiLSTM model in forecasting time series data. Their analysis primarily centered on evaluating the behaviors of ARIMA, BiLSTM, GRU, and LSTM models. Makala and Li [3] investigated gold commodity forecasting using ARIMA and a Support Vector Machine, and three types of kernel were utilized in this publication. Lee et al. [4] used three models to forecast four key currency pairs. This research compares the performance of the CNN, LSTM, and ARIMA models in forecasting currency pair prices. Each model uses EUR/USD closing price data from 1, 3, 5, and 7 years. The forecasting findings revealed that ARIMA is the best for data with fewer than seven years of closing prices, while LSTM and CNN are better suited for training with data with seven years of closing prices. Because of its success in processing vast volumes of data, LSTM is the better of the two deep learning models for currency forecasts.

Researchers typically use the Decision Tree algorithm to predict financial time series because it generates rules to understand and analyze [5]. McNally et al. [6] estimated the coin’s price with 52% accuracy using a Bayesian-optimized recurrent neural network. Rekha et al. [7] studied the use of RNN and CNN algorithms. These models’ precision is assessed by comparing their accuracy to actual stock market values in real-world settings. Another study in [8] used autoregressive (AR) properties in an LSTM network to forecast daily BTC values. Their suggested LSTM-AR model outperformed a regular LSTM model in terms of (MSE) RMSE, MAPE, and MAE. Using the hourly values for BTC, ETH, and XRP, Livieris et al. [9] employed deep learning methods to predict patterns and values of specific cryptocurrencies. They suggested an ensemble learning system that incorporated Bi-LSTM, LSTM, and CNN and discovered that it was capable of making accurate and dependable predictions.

In a different investigation [10], a combined approach that integrated LSTM and GRU networks was applied to predict the prices of LTC and Monero. This hybrid method contrasted with a sole LSTM approach, and the results indicated that the combined model demonstrated superior accuracy in forecasting the prices of the specified cryptocurrencies. In a study by Chowdhury et al. [11], the application of ensemble approaches based on machine learning, such as KNN, ANN, gradient-boosted trees, and a combined ensemble model, was examined to predict the prices of 9 prominent cryptocurrencies. The findings indicated that the ensemble-learning model exhibited the lowest prediction error.

III. MACHINE LEARNING

Machine learning is based on experience, and is a system that works from algorithms that, fed with data, tend to learn and improve automatically. The continuous improvement processes are based on experience and not through programming. Thus, learning consists of processing observations or data (examples, experience, or instructions) to find models that allow the implementation of predictions and decision-making.

A. LINEAR REGRESSION

We use linear regression (LR) to forecast cryptocurrency. In this scenario, the linear regression mathematical equation, you can find the formula in Equation (1).

\[ w = bc + y \] (1)

In this case, the dependent variable is symbolized by W, while the independent variable is marked by y. The slope is denoted by b, and the interceptor by c.

B. SGDREGRESSOR

SGD Regressor is a machine-learning algorithm that makes predictions using stochastic gradient descent. The SGDRegressor is a technique that iteratively optimizes an objective function chosen for its smoothness properties.
Its stochastic nature stems from the random selection or shuffling of samples to assess gradients. This randomness characterizes SGD as a stochastic approximation of gradient descent optimization. SGDRegressor proves to be a straightforward yet highly effective method for discriminative learning commonly useful for convex loss functions like those found in linear Support Vector Machines and Logistic Regression [12].

C. RANDOM FOREST
Random Forest (RF) [13] is a versatile ensemble learning approach for classification and regression tasks. Employing bootstrapped aggregation and random feature selection constructs decision trees within a forest. RF amalgamates the simplicity of single decision trees, yielding class mode for classification and mean prediction for regression by leveraging multiple trees. Renowned for its favorable traits such as robust generalization, simplicity, resilience, and reduced variance, RF finds extensive application.

D. BAGGING
Bagging [14] is an abbreviation for Bootstrap Aggregation, one of the most well-known and successful ensemble learning approaches. Breiman popularized bagging. The basic concept of bagging [15] is straightforward. Bagging intends to generate multiple classifiers in parallel and then aggregate them together. As a result, it chooses a basic classifier algorithm to train the base classifiers on randomly distributed training datasets. The primary concept of bagging [16] is to randomly sample from a training dataset using a weak learning algorithm. This sampling process involves selecting n samples from the original training set, with some samples potentially being chosen multiple times and others not being selected at all. Each subset of the training data is then assigned a base classifier. When faced with an unknown dataset, the findings of each base classifier are considered as votes, and the final classification outcomes are established by aggregating the votes.

E. ADABOOST
This gives more weight [17] to the observation that was misclassified to the last weak learner usually it is a decision stub which means a level one decision tree, which also means the tree is based on decision variables and contains a root and the child connected to it. At each iteration, a model is learned and the samples are re-weighted so that subsequent classifiers focus on samples that were incorrectly predicted by the previous classifier. Correctly predicted samples have their weights reduced. The weights of samples that were inaccurately predicted are increased. The final prediction is a weighted average of model projections.

IV. DEEP LEARNING
A deep neural network is a parallel processing architecture composed of linked neurons arranged in layers. A deep neural network’s fundamental structure is made up of three crucial components: input, output, and one or more hidden layers of hidden units [18]. Deep learning handles vast amounts of data by organizing intricate data structures in manners beyond the capability of an artificial neural network (ANN) [19]. A prototype of data processing that works in the same way as the human nervous system, it analyzes and processes information in a similar way to our brain this prototype contains a lot of elements connected and working to solve different problems the main reason to build a neural network is its similarity with the human brain, they wanted to test the different possible manipulations that can happen to our brain, so they presented it as a group of connected elements that create a neural network or as called neural. We should know that our brain has over a billion neurons and each one is connected to a million cells.

Deep learning [20], a subset of machine learning techniques, focuses on discerning distinctions within data through various architectural approaches. Artificial neural networks serve as the foundation for deep learning methods, executing the learning process through multi-layered neural networks.

The objective is to create abstract models that effectively handle data, encompassing diverse deep learning architectures such as Recurrent Neural Networks, Forward Neural Networks, and Convolutional Neural Networks. This deep learning has made a revolution in the field of voice and image recognition, financial prediction, handwritten character recognition, and data compression, and also in the field of auto-driving such as the automatic detection of traffic lights or cuts, without forgetting the medical field which allows the scientist to automatically analyze cancer cells, also in the field of safety, especially in factories, he can detect the safety distance between employers and machines.

A. LSTM
Long short-term memories, commonly referred to as “LSTM” [21], are a type of recurrent neural network [22] that excels at capturing long-term dependencies. Initially proposed by Hochreiter and Schmidhuber in 1997, these concepts have undergone continuous refinement and gained widespread attention through subsequent research by various scholars. They are currently frequently utilized and function quite well in a wide range of situations. Long Short-Term Memory (LSTM) [23] represents a specialized form of Recurrent Neural Network (RNN) tailored to efficiently handle and forecast data characterized by recurrent time series. Its utility extends across diverse fields like natural language processing, speech recognition, and time series forecasting. Unlike conventional RNN architectures, LSTM demonstrates superior performance particularly with lengthy data sequences, mitigating challenges such as vanishing and exploding gradients. Additionally, LSTM stands out in capturing extended patterns and dependencies within the data over time. LSTMs (Long Short-Term Memory networks) [24] also possess a similar chain-like structure, but the
repeating module differs. Instead of employing a solitary neural network layer, four layers interact distinctively.

V. MATERIALS AND METHODS

A. DATA

Our dataset includes the following attributes: Open, High, Low, and Close, representing the daily stock prices of cryptocurrency and CoinMarketCap. The cryptocurrency data spans from September 17, 2014, to December 8, 2023, while the CoinMarketCap data covers the period from December 1, 2013, to December 11, 2023, without any information gaps.

B. METHODOLOGY

After downloading the datasets, we used different machine learning and deep learning techniques such as linear regression, SGDRegressor (stochastic gradient descent regressor), random forest regressor, adaptive boosting regressor, LSTM (Long short-term memory), and ARIMA (Autoregressive integrated moving average) to predict the closing prices for CoinMarketCap and CryptoCurrency. After that, we compared the results obtained from various algorithms used in this study, which provides recommendations to help investors make informed choices about closing prices. We utilized one measure of accuracy for comparing different results. Figure 1 describes our methodology.

![Diagram of methodology](image)

1) Dataset: We have collected our datasets consisting of daily prices for CoinMarketCap and CryptoCurrency.
2) Applied various machine learning and deep learning techniques: In this phase, we used different machine learning and deep learning techniques such as linear regression, SGDRegressor (stochastic gradient descent regressor), random forest regressor, adaptive boosting regressor, LSTM (Long short-term memory), and ARIMA (Autoregressive integrated moving average) to predict the closing prices for CoinMarketCap and CryptoCurrency.
3) Predict closing prices: In this part, we have forecasted closing prices for CoinMarketCap and CryptoCurrency.
4) Comparing different prices: We compared different forecasted closing prices.

C. EVALUATION MEASURE

One metric, including $R^2$, is used for comparing various machine learning and deep learning approaches mentioned above. You can find the formula in Equation (2). $\bar{y}$ is the mean of the metric, and $\hat{y}_i$ is the forecasted outcome.

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

VI. RESULTS AND DISCUSSION

A. ANALYSIS OF RESULTS

This part contains an analysis of predictions of the following architectures: LSTM, ARIMA, and some machine learning methods applied to the two databases CoinMarketCap and CryptoCurrency. For our study, we will apply the architectures on the two databases and proceed by analyzing the results of the predictions of each architecture and then be able to compare them.

In our application, we apply several models that will analyze the stock market quotes in real-time and be able to make a prediction of the future and assess the future risk through artificial intelligence. Thus, the system will be able to make the appropriate decision at a given moment to invest or to withdraw from the market. This paragraph presents the different architectures LSTM and ARIMA [25] and the different Machine Learning algorithms applied to these databases. Finally, we will analyze the predictions and the results obtained to compare the LSTM, Linear Regression, SGDRegressor, and ARIMA approaches. This research allows investors to make informed decisions on closing prices. Figures 2 through 11 demonstrate the application of diverse machine learning and deep learning algorithms.

As you can see in Figure 2, and Figure 3, we have showed the forecasted price with actual price using ARIMA model.

As you can see, in Figure 4, the accuracy of LSTM reaches 98.49% for CoinMarketCap, and in Figure 5, the accuracy of LSTM is 98.35% for CryptoCurrency.

In Figure 6, the adaptive accuracy stands at 98.35%, bagging achieves an accuracy of 99.25%, and Random Forest demonstrates an accuracy rate of 99.17%.

In Figure 7, the adaptive accuracy stands at 97.62%, bagging achieves an accuracy of 98.68%, and Random Forest demonstrates an accuracy of 98.51%.

Figure 8 displays the performance of Linear Regression on CoinmarketCap, achieving an accuracy of 98.81%. Predicted
values are depicted by the orange line, and the actual trends are represented by the blue line.

In Figure 9, the SGDRegressor model applied to CoinMarketCap demonstrates a precision rate of 98.82%. The orange line illustrates the projected values, while the blue line depicts the observed trends.

In Figure 10, Linear Regression achieves an accuracy of 99.82% in predicting cryptocurrency values. The orange line illustrates the predicted values, while the blue line corresponds to the actual trends.

In Figure 11, we employed the SGDRegressor model for Cryptocurrency analysis achieving a precision rate of 99.72%. We represented the anticipated values by the orange line and the observed trends by the blue line. Figures 2 through 11 demonstrate the application of diverse machine learning and deep learning algorithms. The previous results show that the forecasts of our models are well predicted by the stock market actions and follow the market trend whether it is bullish or bearish.
As you see, the results are different by contributions to the base of données coin market cap database and more voluminous than the crypto_currency impacts the results of no models ‘a good result by contribution to the data set coin market cap’. A voluminous data set helps our models to train well and gives good predictions.

Random forest gives the good results among the other models because The main reason random Forest gives

us the best results is the fact that it makes trees grow randomly, and alternately of looking for the best characteristics when a division of nodes occurs, it looks for the best advantage from a random subset of forest characteristics, which gives a great diversity. More neural networks have become increasingly used in forecasting the time series because they correct some defects of ‘classical’ methods such as ARIMA.

ARIMA also gives a good result, called robust and efficient in forecasting financial time series. A statistical technique that uses historical market data to predict its performance. The results show a performance comparison of various machine-learning methods. Experiment analysis shows that linear regression outperforms SGDRegressor, LSTM, ARIMA (Autoregressive Integrated Moving Average) in words of accuracy.

Table 1 illustrates our results, which show that the accuracy of the linear regression model outperforms other algorithms. The findings in Table 1 indicate that the linear regression model achieves a prediction accuracy of 99.82% for the closing price of the cryptocurrency, followed by the SGDRegressor algorithm in second place with an accuracy of 99.72%, and bagging regressor algorithm in third place with an accuracy of 99.25%.

### TABLE 1. Comparison between different algorithms.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>99.82%</td>
</tr>
<tr>
<td>SGDRegressor</td>
<td>99.72%</td>
</tr>
<tr>
<td>LSTM</td>
<td>98.35%</td>
</tr>
<tr>
<td>RandomForestRegressor</td>
<td>99.17%</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>98.35%</td>
</tr>
<tr>
<td>Bagging</td>
<td>99.25%</td>
</tr>
</tbody>
</table>

**VII. CONCLUSION**

This thesis contributes to the research theme by comparing different Machine Learning and Deep Learning architectures
for predicting stock market trends. The main targeted application of this work is to study the existing state of the art to apply it to a database and thus make predictions and analyze them. The latter allowed us to evaluate our models to know their strong and weak points and thus to extract the advantages of each to make a better prediction. We have proposed to improve prediction by a method that joins the trend predicted by different algorithms. To understand the different aspects of the architectures, we applied modifications to the architectures and examined the change in the predictions to understand the effect of the last on the result. In this article, we have used distinct algorithms to predict closing prices for CoinMarketCap and cryptocurrency. In our future work, we recommend using algorithms on other stock markets.

REFERENCES


MOHAMED EL MAHJOUBY received the master’s degree in information systems engineering (MSID) from Sidi Mohamed Ben Abdellah University, Fez, Morocco, in 2020. His thesis was on “stock market predictive system analysis and modeling.” His research interests include artificial intelligence and machine learning.

MOHAMED TAJ BENNAJNI received the master’s and Ph.D. degrees in computer science and networking from the Faculty of Science and Technology of Tangier, Tangier, in 2011 and 2019, respectively. At present, he is working as a Professor with the Faculty of Science Dhar El Mahraz, Fez, since 2019. He has been a qualified university professor since June 2023.
MOHAMED LAMRINI received the Ph.D. degree from the Claude Bernard University Lyon 1, in 1993. He is currently a Professor of computer science with USMBA, Fez, where he is also a member of the LPAIS Laboratory. His research interests include software quality (CMMI, Six Sigma, ISO 9001,...) and industrial engineering (methods and statistical tools).

BADRE BOSSOUFI was born in Fez, Morocco, in 1985. He received the Ph.D. degree in electrical engineering from the Faculty of Sciences, Sidi Mohammed Ben Abdellah University, Morocco, and the joint Ph.D. degree from the Faculty of Electronics and Computer, University of Pitesti, Romania, and the Montefiore Institute of Electrical Engineering, Liège, Belgium, in 2013. He was an Assistant Professor of electrical engineering at the Higher School of Technology, Oujda, Morocco. His research interests include static converters, electrical motor drives, power electronics, smart grids, renewable energy, and artificial intelligence.

THAMER A. H. ALGHAMDI received the B.Sc. degree from Al-Baha University, Al-Baha, Saudi Arabia, in 2012, the M.Sc. degree from Northumbria University, Newcastle upon Tyne, U.K., in 2016, and the Ph.D. degree from Cardiff University, Cardiff, U.K., in 2023. He was a Power Distribution Engineer at Saudi Electricity Company (SEC), in 2013. He was a Lecturer Assistant at Al-Baha University, from 2016 to 2018, where he is currently an Assistant Professor of electrical power engineering. His main research interests include power systems, power quality, the integration of renewables, and AI applications in electrical engineering.

MOHAMED EL FAR is currently a Professor with Sidi Mohamed Ben Abdellah University, Fez, Morocco. He teaches computer science. His research interests include data science, big data, and artificial intelligence. He has published many researches.

* * *