Utilizing Deep Learning and Machine Learning Methods to Forecast Market Performance

1st Mohamed EL MAHJOUBY
Department of Computer Science Laboratory (LPAIS)
Faculty of Science Dhar El Mahraz, USMBA, Fez, Morocco
mohamed.elmahjouby@usmba.ac.ma

2nd Khalid El Fahssi
Department of Computer Science Laboratory (LISAC)
Faculty of Science Dhar El Mahraz, USMBA, Fez, Morocco
khalid.elfahssi@usmba.ac.ma

3rd Mohamed Taj Bennani
Department of Computer Science Laboratory (LPAIS)
Faculty of Science Dhar El Mahraz, USMBA, Fez, Morocco
bennani.taj@gmail.com

4th Mohamed Lamrini
Department of Computer Science Laboratory (LPAIS)
Faculty of Science Dhar El Mahraz, USMBA, Fez, Morocco
mohamed.lamrini@usmba.ac.ma

5th Mohamed EL FAR
Department of Computer Science Laboratory (LPAIS)
Faculty of Science Dhar El Mahraz, USMBA, Fez, Morocco
mohamed.elfar@usmba.ac.ma

***Abstract*—Forecasting the stock market has become one of the most challenging problems for the AI research community. Stock market investment strategies are complex and depend on analyzing amounts of data. Recently, machine learning techniques have increasingly evaluated and enhanced market predictions compared to traditional methods. This interest is driven by the time-dependent nature of stock prices, making their predictions essential in data mining and a subject of significant interest and effort over the past decades. This challenge remains difficult due to the intrinsic characteristics of time series data, such as its high dimensionality, large volume, and constant updates. Researchers have explored machine learning and deep learning methods to improve the effectiveness of traditional approaches. In this document, we aim to forecast stock market performance at the end of the trading day by applying various machine learning algorithms to two datasets, "CoinMarketCap" and "CryptoCurrency," and analyzing the prediction outcomes of these models.**

Keywords—AdaBoost, Cryptocurrency, CoinMarketCap, regression

I. INTRODUCTION

The stock market is one of the indicators of a country's economic health. However, only a few individuals excel at accurately understanding stock trends, and many hesitate when investing in stocks. Data science has broadened the scope of stock market analysis beyond the traditional domains of economics and finance. In finance, stock prices represent a type of time series. The variations in financial time series are dynamic, selective, nonlinear, nonstationary, and noisy, making accurate forecasting a significant challenge. One of the most pressing issues is effectively predicting stock prices using data mining or machine learning techniques. The efficient market hypothesis suggests that stock prices follow a random walk pattern, making it difficult to make reliable forecasts. A stationary prediction approach is not feasible because investors would quickly exploit such strategies, causing effective forecasting rules to become obsolete. The stock market presents both a challenge and an opportunity for investors to profit. Investors "High Risk, High Return" principle and seek methods to minimize risk while maximizing returns. There are two primary approaches to stock forecasting: fundamental analysis and technical analysis [1].

In this paper, we applied methodologies from Machine Learning and Deep Learning. Generally, various researchers have used time series to forecast prices with these methodologies. This study offers recommendations to assist investors in making informed decisions about closing prices.

The structure of our paper is as follows: We begin with an introduction, a literature review, and an overview of various machine learning algorithms. Next, we include a section dedicated to an interview focused on deep learning and long short-term memory. Afterward, we present and explain our methodologies. Then, we discuss and analyze our findings. Finally, we conclude with some final remarks.

II. RELATED WORK

Several scholars have forecasted time series data using different methods. An LSTM model was created in reference [2] to predict stock prices. Using CNN, RNN, and LSTM, the suggested approach anticipates agricultural productivity [3].

The authors of the reference [4] presented a novel approach to stock market trend forecasting by combining LSTM and genetic algorithm (GA) techniques. Their analysis indicates that this technique outperformed the industry standard benchmark model. The authors of reference [5] used a variety of machine learning algorithms, including bagging, Gaussian naïve Bayes, approach, extreme gradient boosting classifier, random forest classifier, logistic regression, and AdaBoost. Your goal is to forecast when buying and selling the euro relative to the dollar will be most profitable. According to research, machine learning to predict changes in the price of Bitcoin and has grown to be an effective tool in this endeavor [6]. Using candlesticks of six different currency pairings—two minor pairs, such EUR/GBP and GBP/JPY, and four pairs, like GBP/USD, EUR/USD, USD/CHF, and USD/JPY—the authors of reference [7] offered a reusable trading method that trained on historical data gathered at 4-hour intervals. In reference [8], this work proposes an adaptive boosting regression technique combined with linear regression as a base estimator for machine learning. The NASDAQ, EUR/USD, GBP/USD, and the closing values of the gold commodity are all predicted using this combination. Seven technical indicators are employed in this method to provide data for the adaptive boosting regression during training to increase accuracy.

2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS) | 979-8-3503-5120-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDS62089.2024.10756293

979-8-3503-5120-0/24/$31.00 ©2024 IEEE

III. MACHINE AND DEEP LEARNING

This section describes the different methods (bagging, Adaboost, random forest, LSTM, and linear regression) used to predict market performance.

A. Random Forest

Random Forest (RF) is a method for ensemble learning for classification and regression tasks [9]. Random feature selection and bootstrap augmentation allow decision trees to be inside forests. RF uses many decision trees to combine their simplicity to produce a mean forecast for regression and a class mode for classification. RF is widely used and is well-known for its beneficial qualities, which include robust generalization, simplicity, durability, and low variation.

B. Bagging

Bootstrap Aggregation, one of the most well-known and effective ensemble learning strategies, is shortened to "bagging" [10]. Breiman made bagging. The fundamental idea behind bagging [11] is simple. The goal of bagging is to build several classifiers and combine them. Consequently, it selects a simple classifier technique and uses randomly distributed training datasets to train the base classifiers. Using a weak learning algorithm, bagging [12] primarily involves taking random samples from a training dataset.

C. AdaBoost

AdaBoost, short for "Adaptive Boosting," is a machine learning algorithm used to improve the performance of weak classifiers. The observation incorrectly assigned to the final poor learner [13] becomes more significant. Usually, a level one decision tree matches a decision stub, indicating that the tree has a root and a connected child predicated on decision variables. For every iteration, we resample the samples and develop a new model, directing the attention of subsequent classifiers toward the examples that the preceding classifier failed to correct the forecast. Weight loss occurs in accurately anticipated samples. High sample weights were incorrect. Weighted analysis of the model's predictions results in the final prediction.

D. Linear regression

To forecast the market, we employ linear regression (LR). Equation (1) contains the formula for this situation, the mathematical equation for linear regression. In this instance, y represents the independent variable, and w is the dependent variable. c is the interceptor, the slope by b.

$$w = bc + y \qquad (1)$$

E. LSTM

RNN modules also come in LSTM (Long Short Term Memory) modules. After being developed and made famous by numerous researchers, LSTM was first developed by Hochreiter & Schmidhuber (1997) [14]. The recurrently consistent modules that make up the LSTM network are similar to those of the RNN. The link between the hidden layers of an RNN is different in an upgraded version of an RNN called LSTM.

IV. DATA AND METHODOLOGY

A. Dataset

In this study, we use datasets that represent the daily stock values of cryptocurrencies and CoinMarketCap and have the following attributes: Open, High, Low, and Close. Whereas the CoinMarketCap data is complete, it covers the period from December 1, 2013, to December 11, 2023, whereas the cryptocurrency data is available from September 17, 2014, to December 8, 2023.

B. Methodology

Following the dataset download, we employed a variety of machine learning and deep learning techniques to forecast the closing prices for CoinMarketCap and CryptoCurrency. These techniques included linear regression, stochastic gradient descent regressor, random forest regressor, adaptive boosting regressor, and long short-term memory. Subsequently, we juxtaposed the outcomes derived from diverse algorithms employed in this investigation, offering suggestions to assist investors in making knowledgeable decisions on closing prices. We compared several results using the same accuracy metric. Figure 1 outlines our approach.

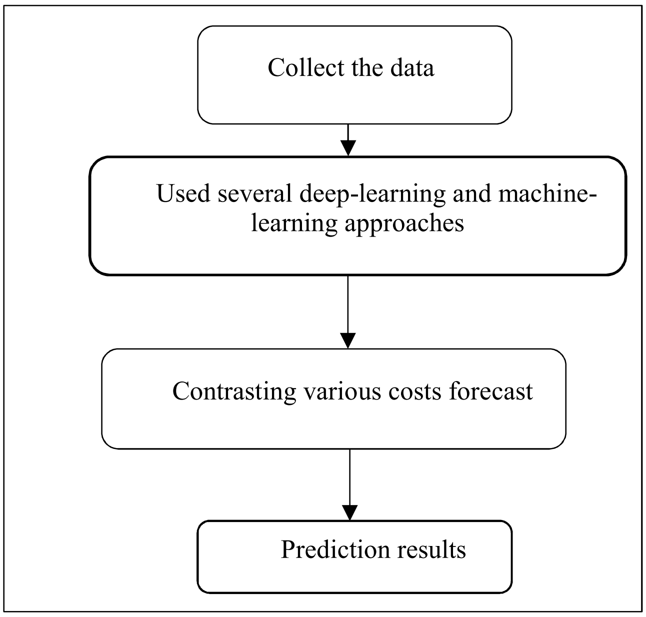

Fig. 1. Prediction method (our methodologies)

C. Performance evaluation

The many machine learning and deep learning techniques discussed above are compared using a single statistic, which includes R^2. Equation (2) has the formula. $\hat{y_i}$Represents the expected result, and $\bar{y}$ is the metric's mean.

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

V. RESULTS AND DISCUSSION

This section includes a prediction analysis of the long short-term memory and several machine learning techniques applied to the CoinMarketCap and CryptoCurrency databases. To conduct our study, we will use the architectures of the two databases, analyze the outcomes of each architecture's predictions, and then compare them.

We use several algorithms that can use artificial intelligence to anticipate the future and evaluate future risks by analyzing real-time stock market quotes. As a result, the system decides when to enter the market and when to exit it. The many LSTM architectures and machine learning methods on these databases. Evaluate the LSTM, Linear

Regression, and SGD Regressor, and analyze the predictions and the generated results. Investors can choose closing prices with knowledge thanks to this research. Use various machine learning and deep learning methods in Figures 2 through 7.

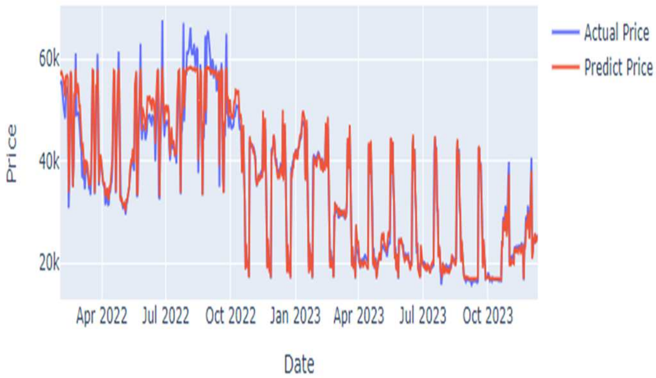

Fig. 2. Price prediction with LSTM (CoinMarketCap)

Fig. 3. Price prediction with LSTM (Cryptocurrency)

As you can see, Figure 2 shows that the LSTM accuracy for CoinMarketCap is 98.49%, and Figure 3 shows that the LSTM accuracy for CryptoCurrency is 98.35%.

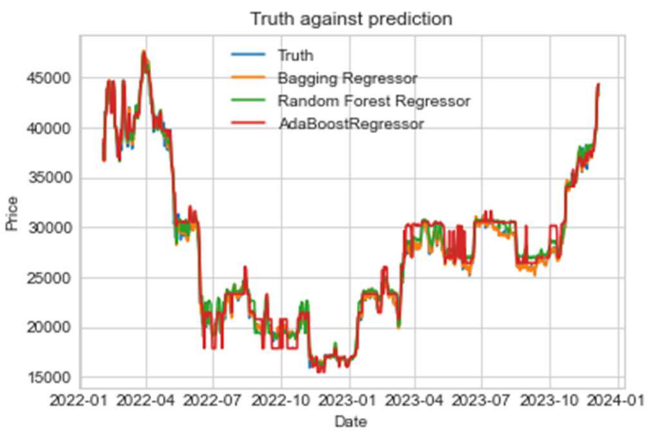

Fig. 4. Cryptocurrency prediction versus truth

Figure 4 shows that the adaptive accuracy is 98.35%, the bagging accuracy is 99.25%, and the random forest accuracy is 99.17%.

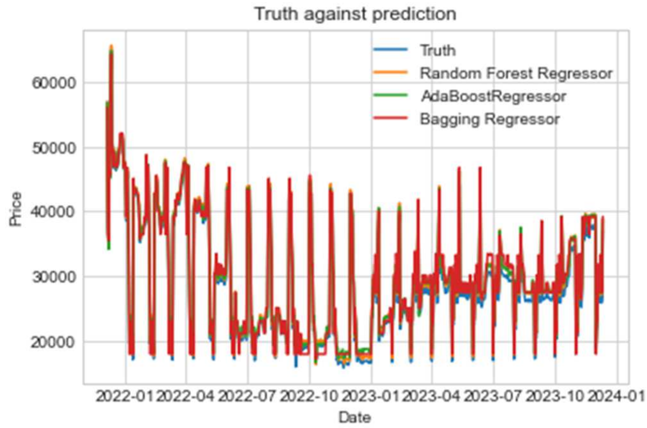

Fig. 5. Coin Market Cap prediction versus truth

Figure 5 shows that the adaptive accuracy is 97.62%, the bagging accuracy is 98.68%, and the random forest accuracy is 98.51%.

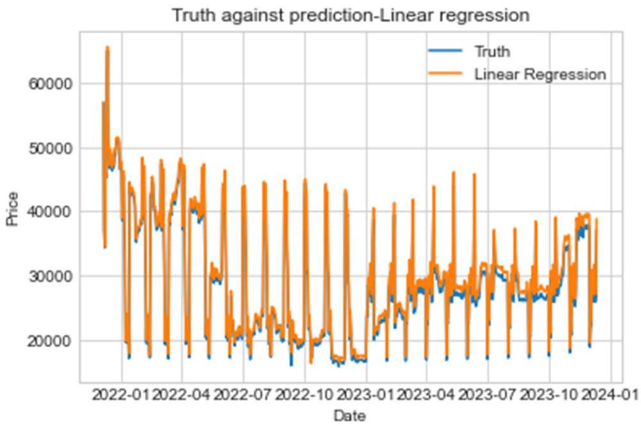

Fig. 6. Coin Market Cap (Linear regression)

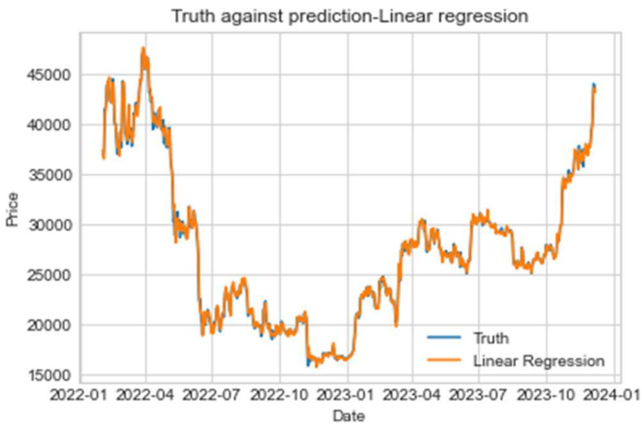

Fig. 7. Cryptocurrency (Linear regression)

The performance of Linear Regression on CoinmarketCap is in Figure 6, with an accuracy of 98.81%. Figure 7, the prediction accuracy of Linear Regression for Cryptocurrency values is 99.82%. The blue line represents the True trends, whereas the orange line shows the expected values. Previous findings demonstrate that our models' forecasts accurately predict stock market behavior and track the general trend of the market, whether it is bearish or bullish.

1. As you can see, variations in the results are due to contributions made to the database. CoinMarketCap is larger and more comprehensive than cryptocurrency and influences the outcomes of no models by adding value to CoinMarketCap. A large amount of data facilitates our models' training and prediction-making.
2. Random forest produces the best results because it allows trees to grow randomly. Additionally, when a node divides, instead of searching for the best characteristics, it searches for the best advantage from a random subset of forest characteristics, which results in a great deal of diversity.
3. In forecasting financial time series, linear regression also produces good results that are robust and efficient. A statistical method that forecasts market performance using past data. The outcomes compare the effectiveness of many machine-learning techniques. The results of the experiment analysis in terms of accuracy linear regression performs better than LSTM and other algorithms.

Table 1 presents our findings, which demonstrate that the linear regression model performs better accuracy than alternative algorithms. As can be seen from Table 1, the linear regression model predicts the closing price of the cryptocurrency with an accuracy of 99.82%, and the bagging regressor algorithm comes in second with an accuracy of 99.25%.

TABLE I. COMAPARISON TABLE

Algorithms	Accuracy
LSTM	0.9835
Linear regression	0.9982
Random forest	0.9917
Bagging Regressor	0.9925
Adaptive Boosting	0.9835

VI. CONCLUSION

This thesis advances the subject by contrasting LSTM and several Machine Learning architectures for trend prediction in the stock market. Studying the current state of the art and applying it to a database to generate and assess predictions is the primary intended use of this work. The latter enabled our models to identify their advantages and disadvantages and improve our predictions. In this article, we have predicted closing prices for CoinMarketCap and cryptocurrency using several algorithms. We propose to use algorithms on additional stock exchanges in our future work.

REFERENCES

[1] M. E. L. Mahjouby, M. T. Bennani, M. Lamrini, M. E. L. Far, B. Bossoufi, and T. A. H. Alghamdi, “Predicting Market Performance Using Machine and Deep Learning Techniques,” *IEEE Access*, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3408222.

[2] M. Roondiwala, H. Patel, and S. Varma, “Predicting Stock Prices Using LSTM,” *Int. J. Sci. Res.*, vol. 6, no. September, pp. 2319–7064, 2015, doi: 10.21275/ART20172755.

[3] T. Mallika, S. D. Kavila, B. Pydi, and B. Rajesh, “A Novel Method for Prediction of Crop Yield Using Deep Neural Networks,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 4, pp. 308–315, 2023.

[4] H. Chung and K. S. Shin, “Genetic algorithm-optimized long short-term memory network for stock market prediction,” *Sustain.*, vol. 10, no. 10, 2018, doi: 10.3390/su10103765.

[5] M. El Mahjouby, M. T. Bennani, M. Lamrini, M. El Far, B. Bossoufi, and T. A. H. Alghamdi, “Machine Learning Algorithms for Forecasting and Categorizing euro-to-dollar Exchange Rates,” *IEEE Access*, vol. 12, no. May, pp. 74211–74217, 2024, doi: 10.1109/ACCESS.2024.3404824.

[6] A. Muminov, “Enhanced Bitcoin Price Direction Forecasting With DQN,” *IEEE Access*, vol. 12, no. February, pp. 29093–29112, 2024, doi: 10.1109/ACCESS.2024.3367719.

[7] M. Fisichella and F. Garolla, “Can Deep Learning Improve Technical Analysis of Forex Data to Predict Future Price Movements?,” *IEEE Access*, vol. 9, pp. 153083–153101, 2021, doi: 10.1109/ACCESS.2021.3127570.

[8] M. El Mahjouby and M. Elfar, “Time Series Forecasting Using Ensemble Model,” in *Colloquium in Information Science and Technology, CIST*, 2023, pp. 155–159. doi: 10.1109/CiSt56084.2023.10409946.

[9] H. H. Htun, M. Biehl, and N. Petkov, “Survey of feature selection and extraction techniques for stock market prediction,” *Financ. Innov.*, vol. 9, no. 1, 2023, doi: 10.1186/s40854-022-00441-7.

[10] R. Richman and M. V. Wüthrich, “Nagging predictors,” *Risks*, vol. 8, no. 3, pp. 1–26, 2020, doi: 10.3390/risks8030083.

[11] W. Wu and M. Peng, “A Data Mining Approach Combining K-Means Clustering with Bagging Neural Network for Short-Term Wind Power Forecasting,” *IEEE Internet Things J.*, vol. 4, no. 4, pp. 979–986, 2017, doi: 10.1109/JIOT.2017.2677578.

[12] X. D. Zeng, S. Chao, and F. Wong, “Optimization of bagging classifiers based on SBCB algorithm,” *2010 Int. Conf. Mach. Learn. Cybern. ICMLC 2010*, vol. 1, no. July, pp. 262–267, 2010, doi: 10.1109/ICMLC.2010.5581054.

[13] A. Baharin, A. Abdullah, and S. N. M. Yousoff, “Prediction of bioprocess production using deep neural network method,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 2, pp. 805–813, 2017, doi: 10.12928/TELKOMNIKA.v15i2.6124.

[14] Ferdiansyah, S. H. Othman, R. Zahilah Raja Md Radzi, D. Stiawan, Y. Sazaki, and U. Ependi, “A LSTM-Method for Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market,” *ICECOS 2019 - 3rd Int. Conf. Electr. Eng. Comput. Sci. Proceeding*, no. June, pp. 206–210, 2019, doi: 10.1109/ICECOS47637.2019.8984499.