Journal of Computing and Communication Vol.3 , No.1 , PP. 88-103 , 2024

Micrtsoft_Stock_Price: An Efficient Framework For Microsoft Stock Price Prediction Using Computational Intelligence

Maged Farouk[a] , Nashwa S Ragab[a], Diaa Salama*[b,c], Omnia Elrashidy[a], Mohamed Osama[a], Mohamed Khames[a], Mohamed Mansour[a], Mohamed Abdelrazeq[a], Mohamed Ali[a], Belal Fathy[a], Reda Elazab[a]

[a]*Department of Business Information Systems, Faculty of Business, El Alamein International University, El Alamein, Egypt*

[b] *Faculty of Computers Science, Misr International University, Cairo, Egypt*

[c]*Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt*

*Corresponding Author: Diaa Salama [***diaa.salama@miuegypt.edu.eg***]

ARTICLE DATA

Article history:
Received 08 Jan 2024
Revised 13 Jan 2024
Accepted 04 Feb 2024
Available online

Keywords:
Machine Learning
Stock Price Prediction
Econometrics
AdaBoost
Linear Regression

ABSTRACT

Econometrics uses statistical methods to analyze relationships using data. While its name suggests a focus on economics, it's widely used in various social sciences and beyond. One of the challenges in predicting stock prices is data availability since obtaining data can often be quite challenging. Predicting stock prices is difficult because it involves analyzing data with various methods, but it's not always accurate due to many factors involved. These methods help understand trends but aren't foolproof for making investment decisions. In this paper, we have proposed an efficient framework for the prediction of Microsoft stock price using nine different machine learning algorithms (AdaBoost, kNN, Linear Regression, Gradient Boosting, Tree, Neural Network, SVM, Constant, Random Forest) on six different datasets. The best algorithm in the four datasets was adaboost, with the smallest percentage of errors, 0.004, and the best algorithm in the two datasets was linear regression. The best result algorithm in all datasets is AdaBoost.

1. Introduction

Econometrics refers to using methods to analyze and evaluate relationships using data. While the term may imply that these methods only apply to analysis, they are used in a broader range of fields, including social sciences. While economics has played a role in the development of econometrics, other disciplines have also contributed significantly. The central technique, regression analysis was initially developed for applications in astronomy by Legendre and Gauss in the 1800s. Econometrics helps us differentiate between competing theories and provides a framework for one. Economists employ analysis to enhance our understanding of how the economy functions at both microeconomic and macroeconomic levels or with specific objectives. In the sector, comprehending relevant markets and predictive abilities can lead to financial benefits. In the sector, evidence-based policy initiatives are recognized as having an impact, which motivates econometric analysis [1-4]. The financial industry has increasingly relied on computing technologies to stay competitive in the economy. As a result, in finance, accurately predicting stock prices using data mining techniques has become a concern. This area of research has garnered interest and is essential for improving prediction accuracy. Fuzzy logic (FL) and Artificial Neural Networks (ANN) offer exciting and promising approaches with applications in prediction. There is a growing interest in the logic computing field and the financial world to leverage logic for forecasting future changes in stock price exchange rates, commodities, and other financial time series. Fuzzy logic enables concluding imprecise information. Artificial Neural Network, as an accepted data mining technique in business settings, excels at learning and detecting relationships among variables. It outperforms regression models. Allows for deeper analysis of large datasets, particularly those prone to short-term fluctuations.

In this paper, we explore the potential of using Fuzzy logic and multilayer perceptron (MLP), a type of neural Network (ANN), to address the challenge of forecasting stock prices in financial time series. We conducted experiments using price data from companies obtained from Yahoo Finance. Additionally, we compared approaches to gain insights [5-6].

Machine learning tackles the challenge of creating computers that can improve themselves through experience. It is a growing field that combines computer science and statistics, playing a role in artificial intelligence and data science. The development of new learning algorithms and theories, the increasing availability of online data, and affordable computing power have fueled the advancements in machine learning. These data-driven machine-learning techniques are widely used in healthcare, manufacturing, education, finance, law enforcement, and marketing, leading to informed decision-making based on evidence [7-9].

Predicting stock market returns accurately is a task because financial stock markets are volatile and nonlinear. However, with the advancements in intelligence and computational capabilities, programmed prediction methods have proven more effective in anticipating stock prices. In this study we have employed Artificial Neural Network and Random Forest techniques to forecast the closing price of five companies from sectors. We have utilized data such as stocks' Open, High, Low and Close prices to create input variables for our models. The performance of these models has been evaluated using indicators like RMSE and MAPE. The model's ability to achieve values for these indicators demonstrates their efficiency in predicting stock closing prices [10-11].

The main contribution of this paper can be summarized as follows: we used nine algorithms on six different datasets to predict Microsoft stock prices, and we used cross-validation with the number of folds equal to ten and random sampling Training 80% and Test 20 %. The best algorithm of all these algorithms is AdaBoost.

The rest of the paper can be organized as follows: Section Two is the related work in which we summarize six papers related to stock price prediction using machine learning topic. Section Three is the methodology in which we describe datasets, explain the algorithms used, and describe the performance metrics. Section Four, is the results after applying cross-validation with ten folds and random sampling with 80% training and 20% testing on six different datasets. Section Five is the conclusion, and Section Six is the acknowledgment.

2. Related Work

In [11], the authors used Machine learning to predict stock market prices. They used neural Networks and random forests, and they applied this data to three algorithms,ANN is the best model because it has the smallest numbers; results show that the best values obtained by the ANN model give RMSE (0.42), MAPE (0.77) and MBE (0.013) and The comparative analysis based on RMSE, MAPE and MBE values indicate that ANN gives better prediction of stock prices as compared to RF and Random Forest (RF) is an ensemble machine learning technique and ANN, is one of the intelligent data mining techniques.

In [12], The authors of this paper used Machine Learning techniques to predict electron industry stock prices. They tested their models on four training and five different testing datasets using both Artificial Neural Network (ANN) and decision tree Machine Learning models. According to their results, ANN provides an average prediction accuracy of 59.016%, while decision tree provides an average prediction accuracy of 65.415%. However, when the authors combined decision tree and ANN models, they achieved a much higher prediction accuracy of 77.1985%. The authors also combined two Decision Tree Models, giving an average prediction accuracy of 66.8515%.

In [13], This paper depends on using 5 ML algorithms, which are the K- Nearest Neighbors (KNN) algorithm, Linear Regression (LR) algorithm, Support Vector Regressor (SVR) algorithm, Decision Tree Regressor (DTR) algorithm, and Long Short-Term Memory algorithm (LSTM). And they used 12 different datasets for an Indian company. The best ML algorithm is the Long Short-Term Memory algorithm (LSTM) with SMAPE (1.59), R2 (-0.11), and RMSE (22.55), And the second best ML algorithm is the Support Vector Regressor (SVR) with a SMAPE value (5.59), R2 value (-1.69) and RMSE (46.36).

In [14].The authors of this paper utilized Machine Learning to predict the performance of the Karachi Stock Exchange. They tested their models on two datasets, each containing 100 instances. The training data set contained 70 instances, while the testing data set contained 30 instances. They applied four machine learning algorithms to these data sets: single-layer perceptron, multilayer perceptron, radial basis function, and support vector machine. The single-layer perceptron demonstrated an accuracy of 83% when tested on the training data set. The RBF algorithm produced 63% accuracy when tested on the test data set and 61% on the training data set. The SVM algorithm achieved 100% accuracy on the training data set but only 60% on the test data set. Finally, the MLP model demonstrated an accuracy of 77% when tested on the test data set and 67% on the training set.

In [15], This paper depends on using 4 ML. Algorithms, Random Forest, SVM, Gradient Boosting, and AdaBoost, for stock prediction, that has the Accuracy that is 68.4% for 90% training data and 73% for 70% training data for(Random Forest), 78.95% for 90% training data and 73% for 70% training data For(SVM), 78.95 for 90% training data and 73.2% for 70% training data for (Gradient Boosting), 78.95% for 90% training data and 77% for 70% training For (AdaBoost), AdaBoost shows the highest Accuracy of 76.79% for 70% training data and 75% for untrained data, and that by using one Dataset.

In [16], This paper used Logistic Regression and Support Vector Machines (SVM) for stock price prediction in China, showcasing their effectiveness with SVM notably achieving an annual return of 17.13%, surpassing the market index. It compares different machine learning models, emphasizing the importance of model selection. The study suggests that SVM outperforms Logistic Regression in predicting stock movements, highlighting the potential for higher returns. It concludes by proposing future research avenues, including exploring additional indicators various machine learning models, and applying these strategies in cryptocurrency investment for improved outcomes.

This paper depends on using 4 ML. Algorithms, Random Forest, SVM, Gradient Boosting, and AdaBoost for stock prediction, that has the Accuracy that is 68.4% for 90% training data and 73% for 70% training data for(Random Forest), 78.95% for 90% training data and 73% for 70% training data For(SVM), 78.95 for 90% training data and 73.2% for 70% training data for (Gradient Boosting), 78.95% for 90% training data and 77% for 70% training For (AdaBoost), AdaBoost shows the highest Accuracy of 76.79% for 70% training data and 75% for untrained data, and that by using one Dataset [15].

This paper uses Logistic Regression and Support Vector Machines (SVM) for stock price prediction in China, showcasing their effectiveness with SVM notably achieving an annual return of 17.13%, surpassing the market index. It compares different machine learning models, emphasizing the importance of model selection. The study suggests that SVM outperforms Logistic Regression in predicting stock movements, highlighting the potential for higher returns. It concludes by proposing future research avenues, including exploring additional indicators various machine learning models, and applying these strategies in cryptocurrency investment for improved outcomes [16].

3. Methodology

The proposed methods for predicting Microsoft stock prices are illustrated in this section.

3.1. Datasets Descriptions

The Dataset of the MAANG HISTORICAL STOCK MARKET DATA(2001-2023) consists of 6 features.

TABLE 1
FEATURES OF MAANG HISTORICAL STOCK MARKET DATA(2001-2023)

Features	Types	Values
Open	Numerical	From 15.2 to 345
High	Numerical	From 15.6 to 350
Low	Numerical	From 14.9 to 342
Adj close	Numerical	From 11.4 to 339
Volume	Numerical	From 7.43 to 591
date	Numericial	2001-2023

Microsoft Monthly and Weekly Stock Prize Dataset dataset consists of 6 features.

TABLE 2
FEATURES OF Microsoft Monthly and Weekly Stock Price

Features	Types	Values
opening values	Numericial	From 16 to 339
Highest Values	Numericial	From 18.9 to 367
Lowest Values	Numericial	From 14.9 to 327
Adjusted# Closin Values	Numericial	From 27.7604 to 327.26
Volumes of Stocks	Numericial	From 105m to 3.04b
Date	Numericial	1999-2023

Stock Market Analysis Data Dataset Consists Of 6 features.

TABLE 3
FEATURES OF Stock Market Analysis Data

Features	Types	Values
open	Numericial	From 89.5 to 372
High	Numericial	From 90.1 to 374
Low	Numericial	From 88.9 to 362
Volume	Numericial	From 2.66m to 113m
Date	Numericial	2023-02-07 to 2023-05-05
Adj close	Numericial	From 89.3 to 367

Microsoft Stock Price (All-Time) Dataset consists of 6 features.

TABLE 4
FEATURES OF Microsoft Stock Price (All-Time)

Features	Types	Values
opening values	Numerical	From 0.09 to 305
Highest Values	Numerical	From 0.09 to 306
Lowest Values	Numerical	From 0.09 to 302
# Adjusted Closing Values	Numerical	From 0.06 to 305
Volumes of Stocks	Numerical	From 2.30m to 1.03b
Date	Numerical	1986-2021

Amazon Stock Price (All-Time) Dataset consists of 6 features.

TABLE 5
FEATURES OF Amazon Stock Price (All-Time)

Features	Types	Values
opening values	Numerical	From 1.41 to 3.74k
Highest Values	Numerical	From 1.45 to 3.77k
Lowest Values	Numerical	From 1.31 to 3.7k
# Adjusted Closing Values	Numerical	From 1.4 to 3.73k
Volumes of Stocks	Numerical	From 487k to 104m
Date	Numerical	1997-2021

Microsoft | Stock Market Analysis | Founding Years (1986-2022) data set has six features.

TABLE 6: FEATURES OF Microsoft | Stock Market Analysis | Founding Years (1986-2022)

Features	Types	values
opening values	Numerical	From 0.09 to 345
Highest Values	Numerical	From 0.09 to 350
Lowest Values	Numerical	From 0.09 to 342
#Adjusted Closing Values	Numerical	From 0.06 to 341
Volumes of Stocks	Numerical	From 2.30m to 789m
Date	Numerical	1986-2022

3.2. Used Algorithms

We used nine Machine Learning Algorithms: SVM, Linear Regression, Tree, Random Forest, AdaBoost, Neural network, Constant, KNN, and Gradient boosting.

SVM:

Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive Accuracy while automatically avoiding over-fit to the data. Support Vector machines can be defined as systems that use a hypothesis space of a linear function in a high-dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory [17].

Equation:

$$w.x+b=0$$

The Linear Regression:

Linear regression finds a linear relationship between one or more predictors. The linear regression has two types: simple and multiple regression (MLR). This paper discusses various works by researchers on linear and polynomial regression and compares their performance using the best approach to optimize prediction and precision. Almost all of the articles analyzed in this review are focused on datasets; to determine a model's efficiency, it must be correlated with the actual values obtained for the explanatory variables [18].

Equation:

$$Y = mX + b$$

AdaBoost:

AdaBoost is one of the most excellent Boosting algorithms. It has a solid theoretical basis and has succeeded in practical applications. AdaBoost can boost a weak learning algorithm with an accuracy slightly better than random guessing into an arbitrarily accurate strong learning algorithm, bringing a new method and a new design idea to the learning algorithm design [19].

Equation:

$$E_t = \sum_i E[F_{t-1}(x_i) + \alpha_t h(x_i)]$$

Random forest:

The approach, which combines several randomized decision trees and aggregates their predictions by averaging, has shown excellent performance in settings where the number of variables is much larger than the number of observations. Moreover, it is versatile enough to be applied to large-scale problems, easily adapted to various ad hoc learning tasks, and returns variable importance measures [20].

Equation:

$$m = \sqrt{p}$$

Tree:

Tree algorithms have been shown to produce very accurate results. These classifiers are in the form of a majority vote over several decision trees; unfortunately, these classifiers are often large, complex, and difficult to interpret [21].

Equation:

if feature≤threshold: go left
else: go right

Gradient boosting:

Gradient boosting machines are a family of powerful machine-learning techniques with considerable success in many practical applications. They are highly customizable to the application's particular needs, like being learned with respect to different loss functions [22].

Equation:

$$h_m(x_i) = y_i - F_m(x_i)$$

KNN:

The kNN (k Nearest Neighbors) algorithm is a non-parametric, instance-based, or lazy method and has been regarded as one of the simplest methods in data mining and machine learning. The principle of the kNN algorithm is that the most similar samples belonging to the same class have a high probability. Generally, the kNN algorithm finds k nearest neighbors of a query in the training dataset and then predicts the query with the major class in the k nearest neighbors. Therefore, it has recently been selected as one of the top 10 algorithms in data mining [23].

Equation:

$$S_x \subseteq D \text{ s.t. } |S_x| = k \text{ and } \forall (x', y') \in D \setminus S_x$$

Neural Network:

Neural Network is a type of artificial intelligence that attempts to imitate how a human brain works. Rather than using a digital model, in which all computations manipulate zeros and ones, a neural network works by creating connections between processing elements, the computer equivalent of neurons. The organization and weights of the connections determine the output [24].

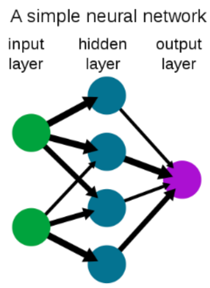

Fig 1. Neural Network demonstration

3.3. Performance Metrics

Mean absolute percentage error (**MAPE**):

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{y_i - x_i}{y_i}\right|$$

MAPE is another performance metric for regression models, having a very intuitive interpretation of relative error. Due to its definition, its use is recommended in tasks where it is more important to be sensitive to relative than absolute variations. However, it has several drawbacks, the most critical ones being the restriction of its use to strictly positive data by definition and being biased towards low forecasts, making it unsuitable for predictive models where large errors are expected [25].

Coefficient of determination (R2 or R-squared):

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(Y - Y_i)^2}$$

(worst value = −∞; best value = +1)

The coefficient of determination can be interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables [25].

4. Results

The following results are from MAANG HISTORICAL STOCK MARKET DATA(2001-2023) (Microsoft) Dataset.

TABLE 7

Statistics of Algorithms with cross-validation no. folds 10

Model	MAPE	R2
AdaBoost	0.004	1.000
Random forest	0.004	1.000
Linear regression	0.007	1.000
Tree	0. 005	1.000
Gradient boosting	0.008	1.000
kNN	0.030	0.9298
Neural Network	22.603	-442.659
constant	1.131	-0.000
svm	3.325	-1.345

According to the lowest value of MAPE, The Best model is (AdaBoost) the second best model is (Random Forest), the third best is (Tree) and the worst model is (Neural Network).

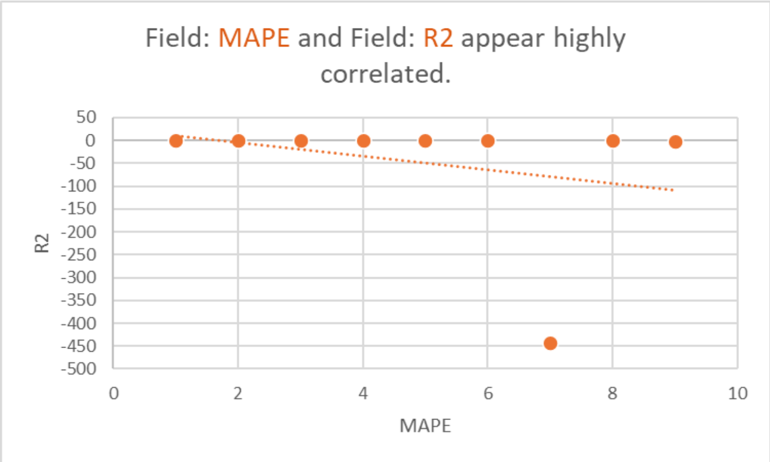

Fig 2. Algorithms chart for First Dataset with cross-validation no. folds 10

TABLE 8
Statistics Of Algorithms with Random Sampling Train 80 % and Test 20%

Model	MAPE	R2
AdaBoost	0.004	1.000
Random forest	0.005	1.000
Linear regression	0.007	1.000
Tree	0.006	1.000
Gradient boosting	0.008	1.000
kNN	0.033	0.998
Neural Network	1.167	0.188
constant	1.129	-0.000
SVM	3.284	-1.259

According to the lowest value of MAPE, The Best model is (AdaBoost) the second best model is (Random Forest), the third best is (Tree) and the worst model is (SVM).

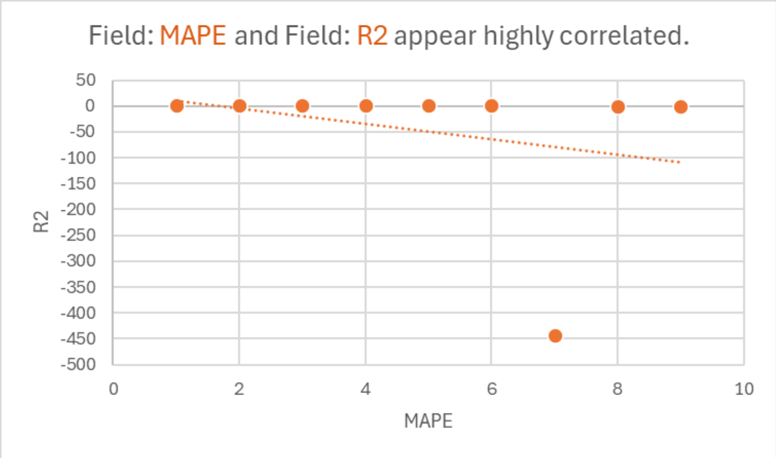

Fig 3. Algorithms chart for First Dataset with Data split 80/20.

The following results are from the Microsoft Monthly and Weekly Stock Price dataset.

TABLE 9
Statistics Of Algorithms with cross-validation no. folds 10

Model	MAPE	R2
AdaBoost	0.012	1.000
Random Forest	0.013	1.000
Gradient Boosting	0.015	1.000
Tree	0.016	1.000
Linear Regression	0.016	0.999
kNN	0.088	0.971
Neural Network	1.179	0.199
constant	1.182	-0.001
SVM	2.373	-0.443

According to the lowest value of MAPE, The Best model is (AdaBoost) the second best model is (Random Forest), the third best is (Gradient Boosting) and the worst model is (SVM).

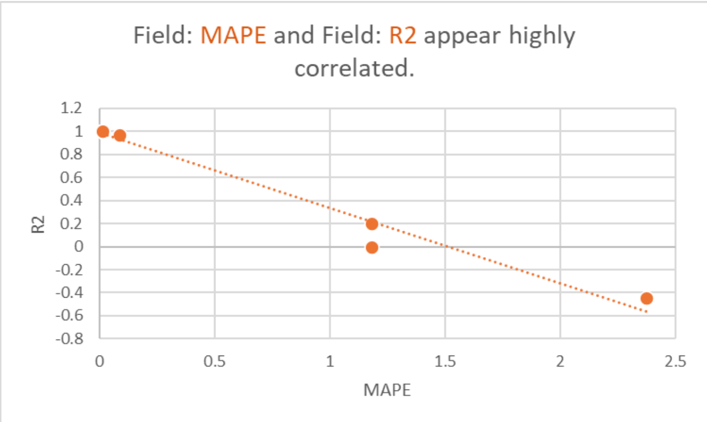

Fig 4. Algorithms chart for Second Dataset with Data split 80/20.

TABLE 10
Statistics Of Algorithms Random Sampling Training 80%Testing 20%

Model	MAPE	R2
AdaBoost	0.013	1.000
Gradient Boosting	0.016	1.000
Random Forest	0.014	1.000
Linear Regression	0.016	1.000
Tree	0.017	0.999
kNN	0.090	0.968
Neural Network	1.187	0.200
Constant	1.199	-0.001
SVM	2.315	-0.369

According to the lowest value of MAPE, The Best model is (AdaBoost) the second best model is (Random Forest), the third best is (Gradient Boosting) and the worst model is (SVM).

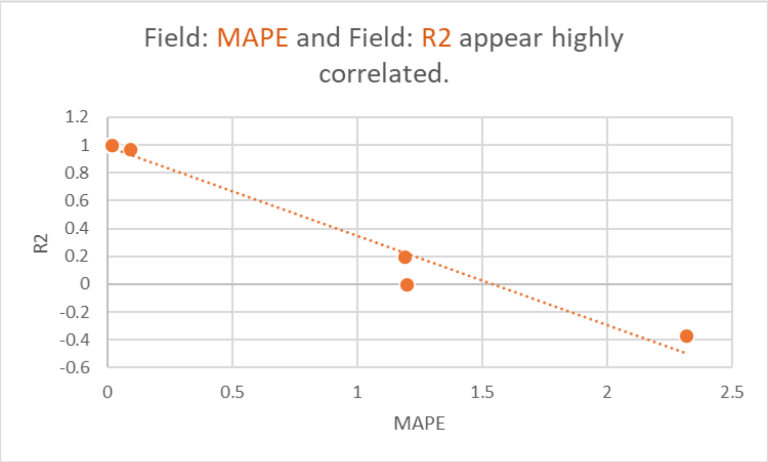

Fig 5. Algorithms chart for Second Dataset with Data split 80/20.

The following results are from the Stock Market Analysis Data dataset.

TABLE 11
Statistics Of Algorithms with cross-validation no. folds 10

Model	MAPE	R2
Linear regression	0.000	1.000
Random forest	0.005	1.000
Tree	0.007	0.999
kNN	0.323	0.427
Constant	0.519	-0.008
SVM	0.486	-0.025

According to the lowest values MAPE, The Best model is (Linear regression) the second best model is (Random Forest), the third best is (Tree) and the worst model is (Constant).

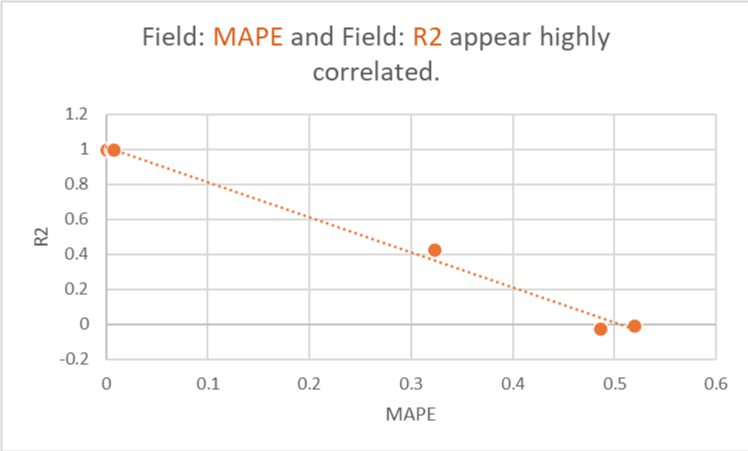

Fig 6. Algorithms chart for Third Dataset with cross-validation no. folds 10

TABLE 12

Statistics Of Algorithms with Random Sampling Training:80%Testing:20%

Model	MAPE	R2
Linear regression	0.000	1.000
Random forest	0.006	1.000
Tree	0.008	0.999
kNN	0.340	0.358
Constant	0.524	-0.016
SVM	0.556	-0.277

According to the lowest no. MAPE, The Best model is (Linear regression) the second best model is (Random Forest), the third best is (Tree) and the worst model is (SVM).

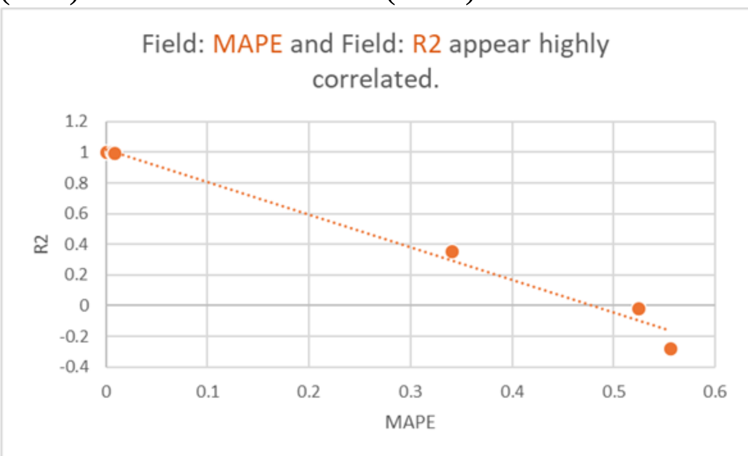

Fig 7. Algorithms chart for Third Dataset with Data split 80/20.

The following results are from the Microsoft Stock Price (All-Time) Dataset.

TABLE 13

Statistics Of Algorithms with cross-validation no. folds 10

Model	MAPE	R2
Neural Network	14.731	0.251
Random Forest	0.004	1.000
kNN	0.042	0.998
Linear Regression	0.019	1.000
AdaBoost	0.004	1.000
Gradient Boosting	0.025	1.000
Tree	0.005	1.000
Constant	19.642	-0.000
SVM	66.456	-2.883

According to the lowest no. MAPE, The Best model is (Adaboost) the second best model is (Random Forest),the third best is (Tree) and the worst model is (SVM).

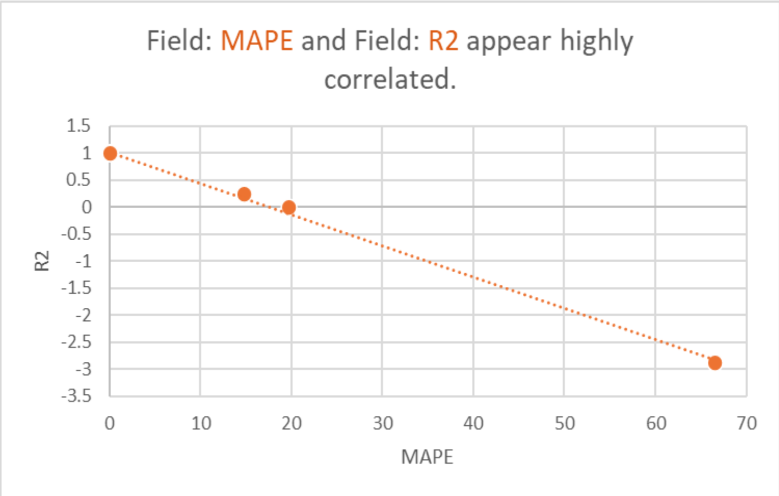

Fig 8. Algorithms chart for Fourth Dataset with cross-validation no. folds 10

TABLE 14

Statistics Of Algorithms with Random Sampling Train 80% Test 20%

Model	MAPE	R2
Neural Network	51.027	-16.304
Random Forest	0.004	1.000
kNN	0.043	0.998
Linear Regression	0.018	1.000
AdaBoost	0.004	1.000
Gradient Boosting	0.025	1.000
Tree	0.005	1.000
Constant	19.417	-0.000
SVM	64.484	-2.706

According to the lowest no. MAPE, The Best model is (Adaboost) the second best model is (Random Forest), the third best is (Tree) and the worst model is (SVM).

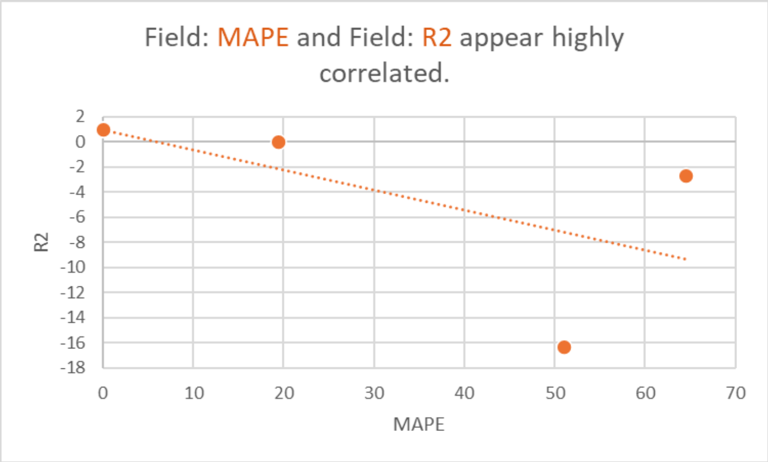

Fig 8. Algorithms chart for Fourth Dataset with Data split 80/20.

The following results are from Amazon Stock Price (All Time) Dataset.

TABLE 15
Statistics Of Algorithms with cross-validation no. folds 10

Model	MAPE	R2
Linear regression	0.000	1.000
AdaBoost	0.002	1.000
Random forest	0.002	1.000
Tree	0.002	1.000
Gradient Boosting	0.020	1.000
kNN	0.040	0.999
Neural Network	11.954	0.157
SVM	47.518	- 1.770

According to the lowest value of MAPE, The Best model is (Linear regression) the second best model is (Adaboost), the third best is (Random forest) and the worst model is (SVM).

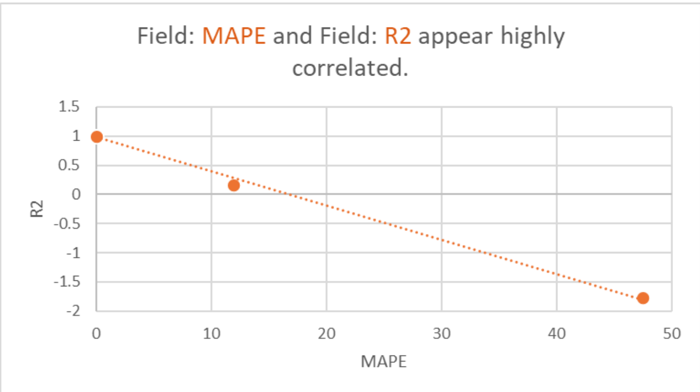

Fig 10. Algorithms chart for Fifth dataset with cross-validation no. folds 10

TABLE 16
Statistics Of Algorithms with Random Sampling Training: 80% Testing: 20%

Model	MAPE	R2
Linear regression	0.000	1.000
Random forest	0.002	1.000
AdaBosst	0.002	1.000
Tree	0.019	1.000
Gradient Boosting	0.003	1.000
kNN	0.041	0.999
Neural Network	14.345	0.153
SVM	47.691	-1.764

According to the lowest no. MAPE, The Best model is (Linear regression) the second best model is (Adaboost), the third best is (Gradient Boosting), and the worst model is (SVM).

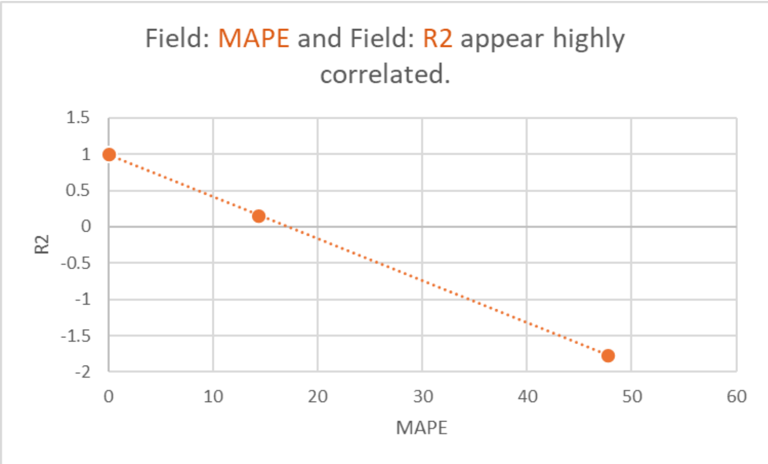

Fig 11. Algorithms chart for FifthDataset with Data split 80/20.

The following results are from the Microsoft | Stock Market Analysis | Founding Years (1986-2022) dataset.

TABLE 17
Statistics Of Algorithms with cross-validation no. folds 10

Model	MAPE	R2
AdaBoost	0.004	1.000
Constant	22.547	-0.000
Gradient Boosting	0.029	1.000
kNN	0.040	0.998
Linear Regression	0.030	1.000
Neural Network	10.154	-0.168
Random Forest	0.004	1.000
SVM	77.696	-2.631
Tree	0.005	1.000

According to the lowest value of MAPE, The Best model is (Adaboost) the second best model is (Random Forest),the third best is (Tree) and the worst model is (SVM).

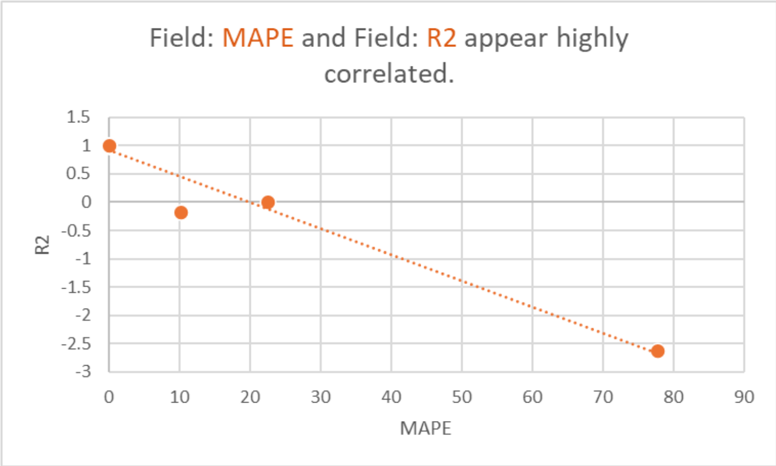

Fig 12. Algorithms chart for the sixth Dataset with cross-validation no. folds 10

TABLE 18
Statistics Of Algorithms with Random Sampling Testing 20% and Training 80%

Model	MAPE	R2
AdaBoost	0.004	1.000
Random Forest	0.004	1.000
Tree	0.005	1.000
Linear Regression	0.029	1.000
Gradient Boosting	0.030	1.000
kNN	0.042	0.998
Constant	23.063	-0.000
SVM	78.727	-2.521
Neural Network	155.174	-87.674

According to the lowest value of MAPE, The Best model is (AdaBoost) the second best model is (Random Forest), the third best is (Tree) and the worst model is (Neural Network).

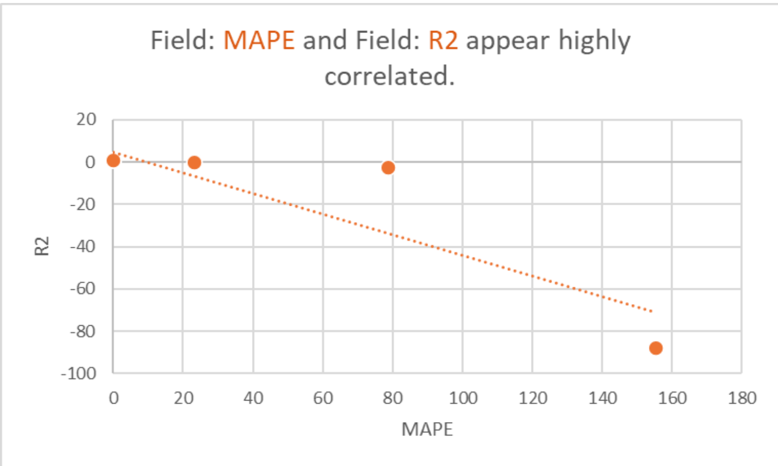

Fig 13. Algorithms chart for Fifth Dataset with Data split 80/20.

5. Conclusion

In This Paper, we used the following Machine Learning Algorithms: (AdaBoost, kNN, Linear Regression, Gradient Boosting, Tree, Neural Network, SVM, Constant, and Random Forest) on six different datasets, and the best algorithm is AdaBoost according to the lowest MAPE it was the best in most of the datasets. Finally, we could predict Microsoft stock prices using Machine Learning. As a future study topic, stock price prediction is very interesting.

References

[1] Dougherty, C. (2011). Introduction to econometrics. Oxford university press, USA.
[2] Hayashi, F. (2011). Econometrics. Princeton University Press.
[3] Durlauf, S. N., Johnson, P. A., & Temple, J. R. (2005). Growth econometrics. Handbook of economic growth, 1, 555-677.
[4] Andrews, D. W. (1994). Empirical process methods in econometrics. Handbook of econometrics, 4, 2247-2294.
[5] Khuat, T. T., & Le, M. H. (2017). An application of artificial neural networks and fuzzy logic on the stock price prediction problem. JOIV: International Journal on Informatics Visualization, 1(2), 40-49.
[6] Lee, J. W. (2001, June). Stock price prediction using reinforcement learning. In ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570) (Vol. 1, pp. 690-695). IEEE.
[7] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.
[8] El Naqa, I., & Murphy, M. J. (2015). What is machine learning? (pp. 3-11). Springer International Publishing.
[9] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., ... & Zdeborová, L. (2019). Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), 045002.
[10] Leung, C. K. S., MacKinnon, R. K., & Wang, Y. (2014, July). A machine learning approach for stock price prediction. In Proceedings of the 18th International Database Engineering & Applications Symposium (pp. 274-277).
[11] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. Procedia computer science, 167, 599-606.
[12] Tsai, C. F., & Wang, S. P. (2009, March). Stock price forecasting by hybrid machine learning techniques. In Proceedings of the international multiconference of engineers and computer scientists (Vol. 1, No. 755, p. 60).
[13] Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock market prediction with high Accuracy using machine learning techniques. Procedia Computer Science, 215, 247-265.
[14] Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016, August). Stock market prediction using machine learning techniques. In 2016 3rd international conference on computer and information sciences (ICCOINS) (pp. 322-327). IEEE.
[15] Kohli, P. P. S., Zargar, S., Arora, S., & Gupta, P. (2019). Stock prediction using machine learning algorithms. In Applications of Artificial Intelligence Techniques in Engineering: SIGMA 2018, Volume 1 (pp. 405-414). Springer Singapore.
[16] Wang, H. (2020, July). Stock price prediction based on machine learning approaches. In Proceedings of the 3rd International Conference on Data Science and Information Technology (pp. 1-5).
[17] Jakkula, V. (2006). Tutorial on support vector machine (svm). School of EECS, Washington State University, 37(2.5), 3.
[18] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.
[19] Ying, C., Qi-Guang, M., Jia-Chen, L., & Lin, G. (2013). Advance and prospects of AdaBoost algorithm. Acta Automatica Sinica, 39(6), 745-758.
[20] Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25, 197-227.
[21] Freund, Y., & Mason, L. (1999, June). The alternating decision tree learning algorithm. In icml (Vol. 99, pp. 124-133).
[22] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.
[23] Zhang, S., Cheng, D., Deng, Z., Zong, M., & Deng, X. (2018). A novel kNN algorithm with data-driven k parameter computation. Pattern Recognition Letters, 109, 44-54.
[24] Islam, M., Chen, G., & Jin, S. (2019). An overview of neural Network. American Journal of Neural Networks and Applications, 5(1), 7-11.
[25] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623.