

Hybrid machine learning for stock price prediction in the Moroccan banking sector

Bouzarne Itri¹, Youssfi Mohamed¹, Bouattane Omar¹, El Madani Latifa¹, Moumoun Lahcen², Oualid Adil²

¹Laboratory Computer Science, Artificial Intelligence and Cyber Security (2IACS), ENSET Mohammedia, Hassan II University of Casablanca, Casablanca, Morocco

²Laboratory Mathématiques, Informatique et Sciences de l'Ingénieur (MISI), FST Settât, Hassan 1st University in Settât, Settât, Morocco

Article Info

Article history:

Received Jun 19, 2023

Revised Nov 11, 2023

Accepted Jan 18, 2024

Keywords:

Banking stock market

Feature selection

Linear support vector regression

Machine learning

Regression algorithm

ABSTRACT

Analyzing historical stock market data using machine-learning techniques is crucial for data scientists and researchers to optimize stock price prediction models. This study uses machine learning regression algorithms and feature selection methods to optimize a simulated stock price prediction model using real historical data from Bank of Africa, a Moroccan bank. The approach compares multiple supervised regression algorithms, such as linear regression, extreme gradient boosting, ordinary least squared, random forest regressor, a linear least-squares L2-regularized, epsilon-support vector regression, and linear support vector regression. Each of these algorithms is associated with different feature selection algorithms to improve the performance of the prediction model. The analysis results revealed that hybridizing algorithms between the highest score percentiles, univariate linear regression, and linear support vector regression perform better according to the root mean squared error and R2-Score measures. This approach overcomes the problems associated with high-dimensional data by reducing the number of features and improving prediction accuracy.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Bouzarne Itri

Laboratory Computer Science, Artificial Intelligence and Cyber Security (2IACS), ENSET Mohammedia, Hassan II University of Casablanca

Casablanca, Morocco

Email: bouzarne.itri@gmail.com

1. INTRODUCTION

Many investors place their money in the stock market in the hopes of achieving high returns. They may use various methods, such as news articles or moving averages, to understand market trends and make trading decisions. It is also important to learn the fundamentals that will help build a long-term investment strategy. However, it is difficult to achieve good returns due to the non-linearity and randomness of the stock market. The deteriorating economic climate, due in part to inflation, should encourage beginner investors in the stock market to prepare for a potential recession and to educate themselves in order to create a portfolio suited for all market situations. To accurately predict trends, many investors turn to machine learning methods, which allow for a more in-depth analysis of the many external factors influencing stock prices, such as policy, economy, social issues, and investment sentiment [1].

Automated trading based using regression methods is one of the machine learning techniques that involve using historical market data to train a machine learning model to predict future price movements [2]. These predictions are then used to automatically execute trades or provide indicators that can help with

decision-making. A multitude of researchers has employed diverse machine learning models to forecast stock market trends, yielding promising outcomes, such as support vector regression (SVR) [3] and random forest (RF) [4] and multiple algorithms in comparison: random forest, extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), support vector machine regression, k-nearest neighbors (KNN), and artificial neural network (ANN) [5].

Machine learning regression techniques have shown their exceptional proficiency in various stock market domains: i) increased accuracy: machine-learning algorithms can process large amounts of data and identify patterns that may not be apparent to human traders. This can lead to more accurate predictions and better trading decisions, ii) faster decision-making: automated trading based on machine learning can execute trades much faster than human traders, allowing for quicker reactions to market conditions, and iii) reduction of emotional biases: automated trading based on machine learning is not subject to emotional biases that can affect human traders, such as fear, greed, or overconfidence. This can lead to more objective and rational trading decisions.

Automated trading algorithms based on machine learning methods offer several advantages over traditional trading methods, including greater accuracy, faster decision-making, and reduced emotional bias. However, it is important to note that these algorithms are still under development and their performance can be affected by a variety of factors, so it is important to carefully monitor and adjust automated trading strategies as necessary. Machine learning practitioners in the stock market need to carefully choose appropriate algorithms, data preprocessing techniques, and feature engineering approaches.

Feature selection involves choosing the most relevant features (i.e., variables) in a data set for training a machine learning model. This can be useful for improving model performance by reducing noise and avoiding overfitting. The study [6] highlights the importance of stock index futures prices in financial markets and the challenges of accurately predicting them, using deep learning techniques to improve the accuracy of stock index futures price predictions. Zhong and Enke [7] proposes dimensionality reduction techniques to improve the accuracy of stock market predictions using principal component analysis (PCA) to reduce the dimensionality of the predictor variables, autoregressive integrated moving average (ARIMA) models to predict the principal components, and back-propagation neural network (BPNN) model to predict the daily stock market returns, improving the accuracy of daily stock market return predictions compared to using all available predictor variables. Other studies chose feature selection techniques for the stock price and trend prediction using BPNN [8], as well as long short-term memory (LSTM) [9] which compare between gated recurrent units (GRU), ANN and SVM models.

On the other hand, feature engineering involves creating new features from existing data. This can be useful for capturing additional information in the data and improving model performance. This study highlights the importance of new features (volatility and moving average) to improve the model prediction. Yun *et al.* [10] studies illustrate the generating of technical features in stock price direction prediction using a hybrid genetic extreme gradient boosting (GA-XGBoost) algorithm to perform trend precision. Sadorsky [11] explores the role of feature volatility and economic policy uncertainty in predicting clean energy stock prices using machine-learning techniques. Wen *et al.* [12] develops the forecasting approach to improve predictive accuracy by incorporating additional features using moving average technical indicators, their results outperform traditional indicators and are more effective in predicting crude oil market returns.

Using these two techniques, feature selection and feature engineering, can effectively process large data sets for machine learning and achieve accurate results for predicting stock prices and market trends. The dataset under study is composed of stock market indices. They are collections of stocks that represent the performance of a particular market or market segment. They are often used as a benchmark for the overall performance of a market or to track the performance of a particular sector within a market. In the field of machine learning, stock indices are used to accurately predict trends by training machine learning models on historical data and using the resulting models to make predictions about future performance.

Moroccan stock indices are calculated using a representative sample of listed stock securities. These indices allow Moroccan and foreign investors to track the performance of the Casablanca Stock Exchange using indices such as the Moroccan All Shares (MASI), the Moroccan Most Active Shares Index (MADEX), the USI 10 and other Sector indices. The Moroccan All Shares capitalization index (MASI) is a wide index that includes all listed stocks on the Casablanca Stock Exchange. It measures the overall performance of the Stock Exchange, i.e., the daily evolution of floating market capitalization due to price variations. The MADEX is a compact index composed of the most active values on the market in terms of liquidity over the previous semester. This index is particularly useful for portfolio management [13]. Sector indices measure the performance of a subset of companies with common characteristics; these indices are calculated from companies included in the subset concerned (e.g., companies in a sector of activity such as banks, and insurance companies).

Most prior research in predicting Moroccan stock market prices has primarily focused on using individual machine learning algorithms such as linear regression, support vector machines, random forests

and deep learning. These studies have made valuable contributions to understanding stock market prediction, but are limited to combining different machine learning algorithms and techniques like feature selection and feature engineering. Ouaadi and Ibourk [14] introduced recurrent neural network techniques, specifically long-term memory, to predict stock trends based on Climate change indexes in Morocco. While this is a valuable approach, it is limited to climate-related factors. In contrast, this work focuses on more comprehensive feature selection and hybrid algorithm composition. Oukhouya and El Himdi [15] explored the use of various machine learning methods like SVR, XGBoost, MLP, and LSTM to forecast daily prices of the MSI 20 index. But our research stands out because it combines feature selection and regression algorithms to enhance stock price prediction. The study [16] combines technical indicators (TIs) with the LSTM model to predict stock prices, outperforming the random forest model. While this approach is promising, it does not explicitly delve into a comprehensive feature selection method or hybrid composition of algorithms. This study, on the other hand, emphasizes these elements, which could contribute to even more accurate predictions.

Overall, research contributes to the field of stock market analysis by providing novel insights into the optimization of stock price prediction models through the integration of machine learning algorithm. However, most of this research has focused on using individual machine learning algorithms. There has been less research on combining different machine learning algorithms and other techniques. The proposed approach makes a significant contribution to the field of stock market analysis by introducing a cutting-edge solution that combines regression algorithms, feature selection techniques and feature engineering. By harnessing the power of this hybrid approach, the study aims to identify the stock indices that are most accurately predicted by machine learning models. Integrating regression algorithms and feature selection methods improves the performance and relevance of the selected features, leading to more accurate and efficient stock market predictions. This study provides valuable insights into the optimization of stock price prediction models, suggesting that a hybrid approach combining machine learning regression algorithms and feature selection techniques has the potential to improve the accuracy and efficiency of stock market predictions. This approach is more effective than using individual machine learning algorithms alone, as it can overcome the problems associated with high-dimensional data and improve prediction accuracy. Additionally, our research focuses on the Moroccan banking sector, which is a relatively understudied area in stock price prediction research. This makes this study even more valuable, as it can help to improve the understanding of the Moroccan stock market and provide valuable insights for investors.

The structure of this paper is as follows: section 1 introduces the research question and motivation, providing background information and a summary of previous studies. In section 2, we present the methodology used to create our prediction model. We provide a detailed description of the approach taken, including the optimization process, and discuss the selection of features, regression and feature selection algorithms, and evaluation criteria used to assess the model's performance. Section 3 describes the experimentation process, including the dataset used, the experimental setup, and the methodology for evaluating the model's performance. We present the results of our experiments in this section. Finally, we offer our concluding remarks and suggestions for future research in section 4.

2. METHOD

This research leverages Python and libraries like Scikit-Learn and Pandas to explore historical stock market data from the Casablanca Stock Exchange, with a focus on Bank of Africa. It utilizes various feature selection and regression algorithms to optimize a simulated stock price prediction model, presenting insights into different techniques to enhance model performance. The approach adopted in this work is based on the supervised learning method, following this process. This process uses the following feature selection and regression algorithms:

a. Feature selection algorithm:

- FS1: Select features according to the K highest scores (*SelectKBest*) [17] with univariate linear regression tests (*f_regression*) [18].
- FS2: Select features according to the K highest scores (*SelectKBest*) with estimate mutual information for a continuous target variable (*mutual_info_regression*) [19].
- FS3: Feature ranking with recursive feature elimination (RFE) [20] with Ordinary least squares linear regression (*LinearRegression*) [21].
- FS4: Select features according to a percentile of the highest scores (*SelectPercentile*) [22] with univariate linear regression tests (*f_regression*).
- FS5: Feature ranking recursive feature elimination and cross-validated (RFECV) [23] with random forest regressor (*RandomForestRegressor*) [24].
- FS6: Pearson correlation coefficient [25].

- b. Regression algorithm:
- Linear regression (LR) [26],
 - Ordinary least squared (OLS) [21],
 - Extreme gradient boosting regressor (XGBR) [27],
 - Random forest regressor (RFR) [24],
 - Linear least squares with l2 regularization (RIDGE) [28],
 - Epsilon-support vector regression (SVR) [29],
 - Linear support vector regression (LSVR) [30].

2.1. Data collection

Data collection is the first primordial phase of our study, during which we collected important features of the Casablanca Stock Exchange to train the model and draw conclusions. The features were extracted from the official website of the Casablanca Stock Exchange [13]. These features include historical data from Bank of Africa's asset (session, closing price, evolution, quantity exchanged), indices (session, value, variation) that measure the general performance of the market (*MASI*), Investment (*Companies and Other Finance*), and finally the index that measures the performance by banking sector (*BANKS*). Table 1 describes the characteristics of the dataset under study.

The dataset comprises 2,737 instances with 13 features, corresponding to the trading day from 02/01/2009 to 31/12/2019. The target variable is represented by the price of the next day's closing price ($d+1$). The training and testing sessions are described in Table 2.

Table 1. Feature description

Index and Instrument	Feature	Description
BOA	Session	The trading day
	Close_Price	The last price of the last transacted price during a regular trading session before the market officially closes
	Evolution	The variation between value index of the previous session and the current one
MASI	Quantity_Exchange	The number of shares traded on this security at a session
	Session	The trading day
	ValueM	The MASI value index
INVEST	ChangeM	The variation between value MASI index of the previous session and the current one
	Session	The trading day
	ValueI	The investment value index
BANKS	ChangeI	The variation between value INVEST index of the previous session and the current one
	Session	The trading day
	ValueB	The BANKS value index
Moving average	ChangeB	The variation between value BANKS index of the previous session and the current one
	MM20	20-day moving average
	MM8	8-day moving average
Volatility	VOL20	20-day volatility
	VOL8	8-day volatility

Table 2. Data training and testing periods

	Data set	Training sample	Testing sample
Time frame	02/01/2009 - 31/12/2019	02/01/2009 - 31/12/2018	02/01/2019 - 31/12/2019

2.2. Data pre-processing

Following data collection, we conducted preprocessing by merging the indices based on their respective dates (sessions) and effectively handled missing data. Furthermore, we standardized and scaled the features using the mean and standard deviation [31], as demonstrated in the formula:

$$\frac{x_i - \text{mean}(x)}{\text{stdev}(x)}$$

For each feature (i), x_i is the feature training sample, $\text{mean}(x)$ is the mean of the training samples and $\text{stdev}(x)$ is the standard deviation.

2.3. Feature engineering

One of the techniques that plays a role in improving the input data for machine learning algorithms is feature engineering. In fact, algorithms will need certain features and characteristics that may be hidden in

raw data to work correctly. Highlighting certain features or transforming them dramatically improves the performance of the trained model. In this sense, we added four features: the 20-day and 8-day moving averages and the 20-day and 8-day volatilities.

- The moving average [32] is a widely used statistical indicator in the stock market that provides insight into the direction and evolution of a stock's price and volatility. By taking the average of a stock's prices over a specified number of days, the moving average helps investors predict future price movements and make informed investment decisions. We denote P_t the price at the time (t); the moving average M_t at the time (t) over (n) days is calculated using the following formula:

$$M_t = \frac{P_t + P_{t-1} + \dots + P_{t-n}}{n}$$

- Volatility [33] is a crucial statistic in finance that measures the fluctuations of a stock's price over time. It is defined as the standard deviation of price returns over a given period. It is widely used in the stock market to assess the risk associated with a particular investment. Volatility is calculated as the standard deviation of the daily returns of a stock's price over a specified period. If we consider P_t the price at a given time (t), and \bar{P}_t the average price over the preceding n days, the volatility σ_t at a given time (t) over (n) days is calculated using the following formula:

$$\sigma_t = \sqrt{\frac{1}{n} \sum_{t=1}^n (P_t - \bar{P}_t)^2}$$

2.4. Feature selection

Feature selection [34] is the process of identifying a relevant subset of features, with a size of K , from the dataset after feature engineering transformations, as shown in Figure 1. The size of K is gradually increased until all features are considered. At each iteration, the following five feature selection algorithms are compared: FS1, FS2, FS3, FS4 and FS5. As a result, we obtain K relevant features selected by each feature selection algorithm.

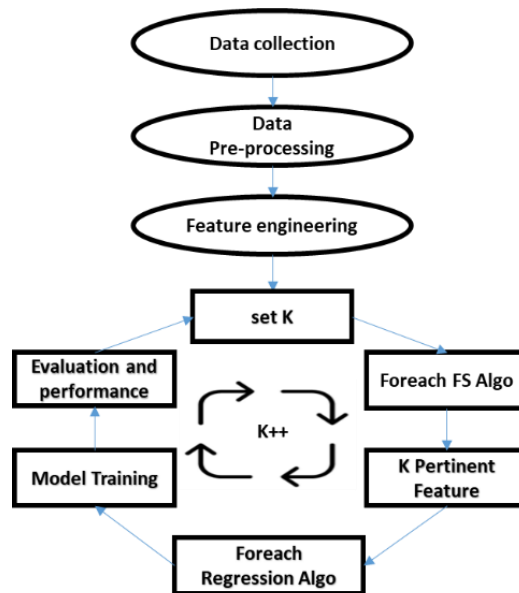


Figure 1. Proposed approach

2.5. Regression algorithm

After the feature selection step, we will have five subsets of K features from the dataset. For each of these subsets, we train the model using seven supervised regression algorithms: LR, OLS, XGBR, RFR, RIDGE, SVR, and LSVR. This results in 35 (5×7) training iterations, each using a different subset of K features.

2.6. Evaluation and model performance

After the training phase, the model generated for each combination of feature selection and regression algorithms is tested and evaluated. The performance of the model is calculated between the predicted closing prices and the actual closing prices. The root mean square error (RMSE) [35] and the coefficient of determination (R2-score) [36] are the performance measures we adopt to compare the model's performances.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{P}_t - P_t)^2}$$

$$R2 - Score = 1 - \frac{\sum_{t=1}^n (\hat{P}_t - P_t)^2}{\sum_{t=1}^n (P_t - \bar{P})^2}$$

where P_t the actual price, \hat{P}_t is the model's predicted price and \bar{P} represents the mean actual price.

The higher the RMSE value, the worse the model. On the other hand, the value of R2-Score is between 0 and 1; the closer the value is to 1, the better the model. In practice, we consider that an R2-Score is high when it is between 0.85 and 1; this score means that all movements of stock are fully explained by the movements of the index (or the features).

2.7. Hybrid algorithm

In this section, we propose a hybrid feature selection and regression algorithm for predicting BOA stock prices developed in Python. The algorithm first reads the BOA dataset into a single *DataFrame* and removes any rows that contain missing values. Then, it splits the *DataFrame* into two *DataFrames*: *X* and *Y_Class*. The *X DataFrame* contains all of the features that will be used for stock price prediction, while the *Y_Class DataFrame* contains the target variable, which is the closing price of BOA shares on the next day.

As shown in Figure 1, after data pre-processing and feature engineering, we set k equal to "3" at the initial iteration. The algorithm iterates over the feature selection methods and regression algorithms in a nested for loop to identify the k most relevant features and the best-performing regression model. The algorithm tracks the best-performing composition of feature selection method and regression algorithm, as well as the corresponding maximum performance. After the nested for loop has finished executing, the algorithm returns the best-performing composition and the corresponding maximum performance. The following provides a detailed description of the proposed algorithm. We denote:

- $D = \{X_i, i = 1 \dots n\}$: A list of features X_i .
- $F = \{F_j, j = 1 \dots q\}$: A list of q feature selection method, we have $F = \{FS1, FS2, FS3, FS4, FS5, FS6\}$.
- $R(D_n) = \{R_k, k = 1 \dots r\}$: A list of r machine learning regression, applied to the training data set D_n with the n selected features. We have $R = \{LR, OLS, XGBR, RFR, RIDGE, SVR, LSVR\}$.
- $W_{j,k}(D_i) = F_j \circ R_k(D_i)$: A composition of algorithms to constitute a hybrid algorithm feature selection and regression for i feature selected. W_o is noted as the best composition offering the best performance.
- $Perf(D_i)_{j,k}$: Calculate the performance of the composition $W_{j,k}(D_i)$ after training. $MaxPerf$ is the best performance found.

Input: Dataset, R, F, D

Output: W_o , $MaxPerf$

Begin

For $i=3$ to n // n is the total number of features ($n=13$).

For F_j in F // j is the index of feature selection algorithm in the list F .

$D_i \leftarrow F_j(i)$ // apply feature selection algorithm F_j to select i most relevant features.

For R_k in R // k is the index of regression algorithm method in list R .

$Perf(D_i)_{j,k} \leftarrow Evaluate[Training(R_k(D_i))]$ // training dataset by R_k with D_i features and evaluate the model for each iteration.

If $Perf(D_i)_{j,k} > MaxPerf$ // $MaxPerf$ initialized by the first iteration (i, j, k).

Then

$MaxPerf \leftarrow Perf(D_i)_{j,k}$

$W_o \leftarrow W_{j,k}(D_i)$

End // the best composition of feature selection and regression algorithms W_o , as well as the best performance achieved $MaxPerf$.

The algorithm explores various combinations of feature selection methods and regression algorithms to identify the best-performing composition (Wo) and the corresponding maximum performance (*MaxPerf*), aiming to identify the most accurate model for predicting BOA stock prices.

3. EXPERIMENTATION AND DISCUSSION

In the first phase of experimentation, after the pre-processing and feature engineering phases, we proposed a comparative analysis between the different feature selection algorithms. The results are based on the correlation between the features and the closing price. In Figure 2, we present the correlation matrix between all the characteristics we study and the closing price (class). The Pearson correlation coefficient method [25] is used to measure the weight (strength, direction) of the linear relationship between the variables. The features with a coefficient greater than 0.5 are assigned a weight of (Strong, Positive); these features are *Close_Price*, MM8, and MM20. The features with a coefficient value of 0.3 to 0.5 are assigned a weight of (Moderate, Positive); these features are *ValueM*, VOL20, and VOL8. The features *Evolution*, *Quantity_Exchange*, *ChangeM*, *ChangeB*, *ValueI*, and *ChangeI* have a weight of (Weak, Positive). The values of these features are between 0 and 0.3. Finally, the weakest weight (weak, negative) is assigned to the feature *ValueB*, which has a negative value.

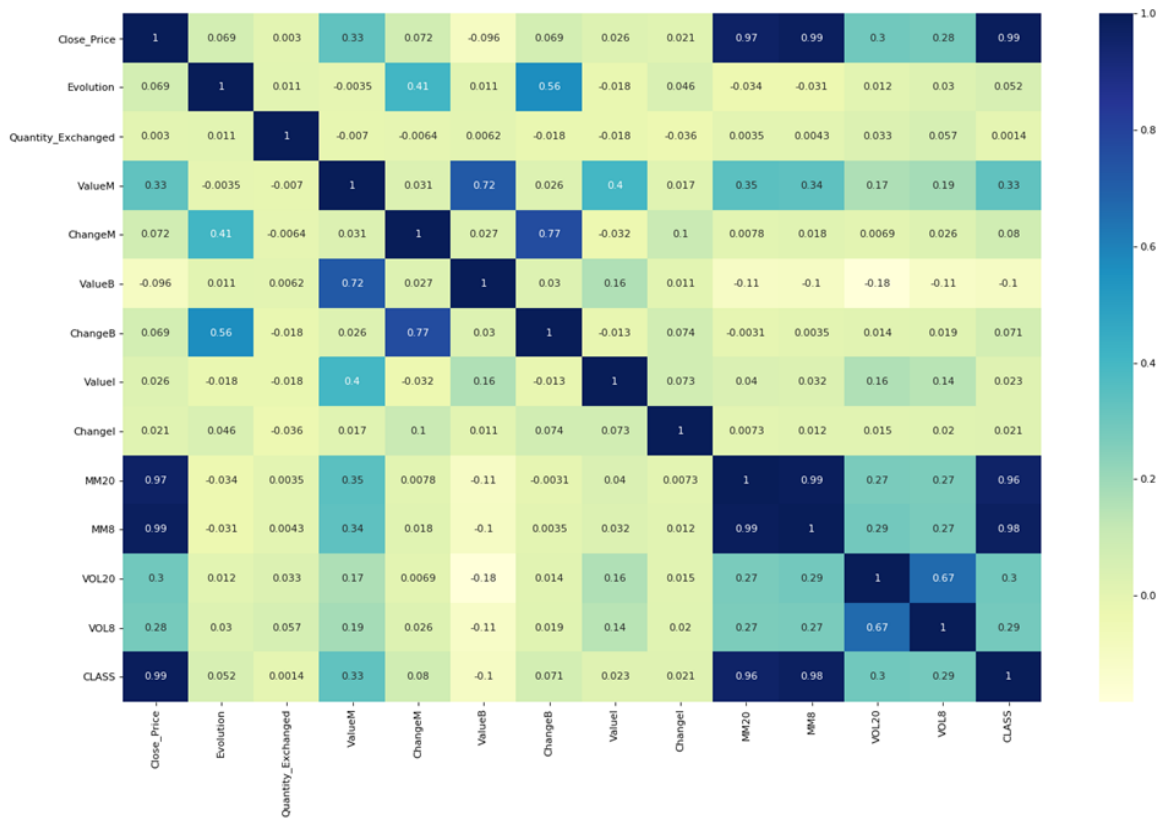


Figure 2. Pearson correlation coefficient between input features and price closing

In the second phase, we trained the regression model using the seven regression algorithms under study, following the weights (Moderate, Positive) and (Weak, Positive), the test results are listed in Table 3 and Table 4. Some scores regressed after increasing the number of features by adding those (Weak, Positive) to the (Moderate, Positive), this score concerns the models generated by the SVR, RFR, XGBR algorithms justified by the ranking score. However, other scores were slightly improved for some regression algorithms, such as LR and OLS, the other algorithms stagnated. We can conclude that the result of the feature selection algorithm based on the Pearson correlation coefficient depends on the regression algorithm used to train the model with the selected input features.

This latest observation led us to the second phase of experimentation, following the approach outlined in Figure 1, where each feature selection method (FS1, FS2, FS3, FS4, FS5) is combined with

several regression algorithms (LR, OLS, XGBR, RFR, RIDGE, SVR, LSVR) by iterating through the feature set. Each iteration fixes a determined number of features based on the highest score; this number is incremented until all features under study (13 features) are reached.

Figure 3 shows the best "R2 Score" obtained during the model evaluation step. We observe that the performance of the model and the relevant features selected vary based on the hybridization of the feature selection algorithms with the regression algorithms. For example, following FS1, the best score of 92.16% was obtained with the LSVR algorithm using 10 features; however, for FS5, the best score of 92.17% was obtained with the same LSVR algorithm, but the number of relevant features was reduced to 6. It can be concluded that the hybridization of the algorithms in our approach influences the performance of the trained model and the relevance of the features.

Table 3. Performance of the trained model by the regression algorithm under selected features with (moderate, positive) weight

Regression Algorithm	RMSE	R2 Score
LR	1,618	91.94%
OLS	1,618	91.94%
XGBR	2,156	85.69%
RFR	2,161	85.62%
RIDGE	1,637	91.75%
SVR	1,625	91.87%
LSVR	1,602	92.09%

Table 4. Performance of the trained model by regression algorithm under selected features with (moderate, positive) and (weak, positive) weights

Regression Algorithm	RMSE	R2 Score
LR	1,605	92.07%
OLS	1,605	92.07%
XGBR	2,415	82.04%
RFR	3,100	70.41%
RIDGE	1,616	91.96%
SVR	2,819	75.53%
LSVR	1,605	92.07%

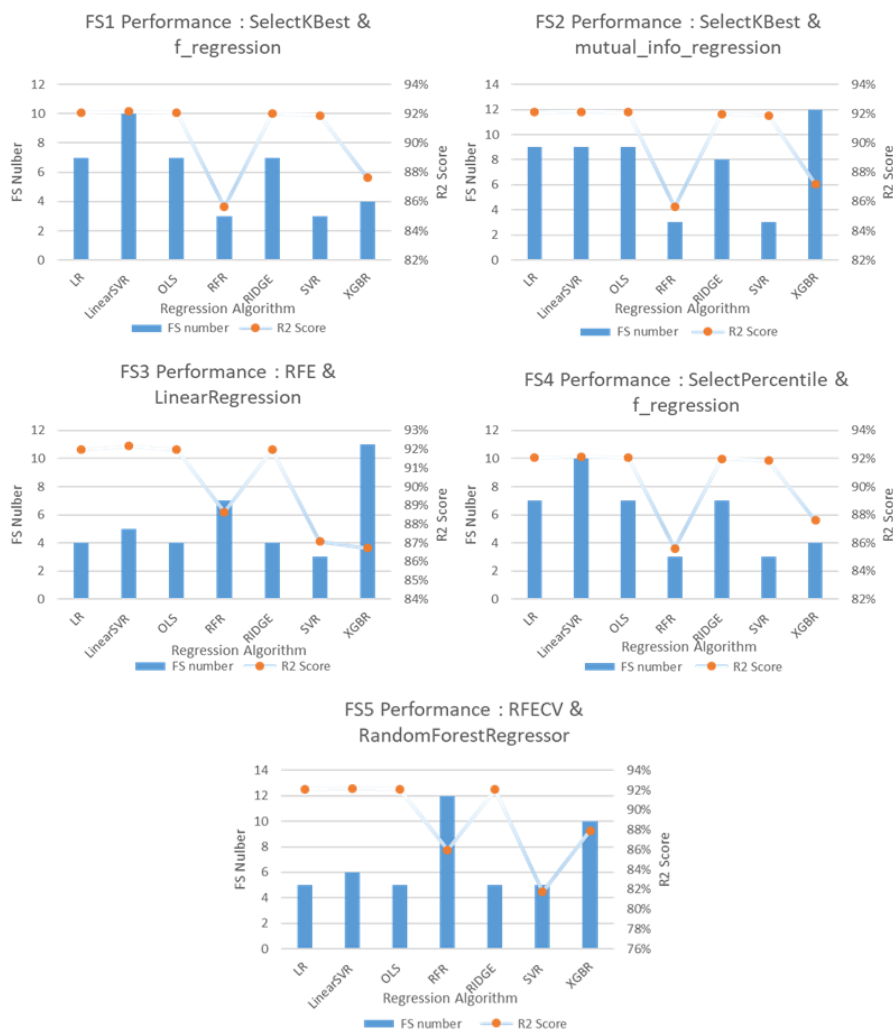


Figure 3. Performance and relevance of features through feature selection and regression algorithms

Table 5 shows the optimal RMSE and R2 scores for each combination of feature selection and training regression algorithms. While LSVR consistently achieves the highest rankings for both RMSE and R2-Score, the effectiveness of feature selection varies depending on the specific algorithm used. However, the most remarkable performance is attained when using FS4 in combination with LSVR, particularly when taking into account the number (5) of relevant features selected.

Table 5. The best R2 score and RMSE rankings after algorithm hybridization

FS Algorithm	Regression Algorithm	RMSE	R2 Score	FS Number
FS1	LinearSVR	1,596	92.16%	10
FS2	LinearSVR	1,599	92.12%	9
FS3	LinearSVR	1,594	92.18%	5
FS4	LinearSVR	1,598	92.14%	10
FS5	LinearSVR	1,595	92.17%	6

4. CONCLUSION

This study highlights the significance of the model training process in stock price prediction, emphasizing the crucial role of feature selection algorithms in determining model performance. The findings demonstrate that the hybrid algorithm of feature selection and regression has a significant impact on the relevance and performance of the features that are selected. The best performance, as measured by RMSE and R2 score, is achieved with the combination of Select Percentile with univariate linear regression tests (FS4) and linear support vector regression (LSVR). This combination resulted in the selection of only 5 relevant indices out of the 13 indices under study.




REFERENCES

- [1] H. Bourezk, A. Raji, N. Acha, and H. Barka, "Analyzing Moroccan stock market using machine learning and sentiment analysis," Apr. 2020, doi: 10.1109/IRASET48871.2020.9092304.
- [2] C. Manjunath, B. Marimuthu, and B. Ghosh, "Analysis of Nifty 50 index stock market trends using hybrid machine learning model in quantum finance," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 3, pp. 3549–3560, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3549-3560.
- [3] E. Guresen, G. Kayakutlu, and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10389–10397, Aug. 2011, doi: 10.1016/j.eswa.2011.02.068.
- [4] X. Lin, Z. Yang, and Y. Song, "Short-term stock price prediction based on echo state networks," *Expert Systems with Applications*, vol. 36, no. 3 PART 2, pp. 7313–7317, Apr. 2009, doi: 10.1016/j.eswa.2008.09.049.
- [5] J. Behera, A. K. Pasayat, H. Behera, and P. Kumar, "Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets," *Engineering Applications of Artificial Intelligence*, vol. 120, Art. no. 105843, Apr. 2023, doi: 10.1016/j.engappai.2023.105843.
- [6] W. L. Yan, "Stock index futures price prediction using feature selection and deep learning," *North American Journal of Economics and Finance*, vol. 64, p. 101867, Jan. 2023, doi: 10.1016/j.najef.2022.101867.
- [7] X. Zhong and D. Enke, "Forecasting daily stock market return using dimensionality reduction," *Expert Systems with Applications*, vol. 67, pp. 126–139, Jan. 2017, doi: 10.1016/j.eswa.2016.09.027.
- [8] H. Hu, L. Tang, S. Zhang, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with Google Trends," *Neurocomputing*, vol. 285, pp. 188–195, Apr. 2018, doi: 10.1016/j.neucom.2018.01.038.
- [9] A. Sethia and P. Raut, "Application of LSTM, GRU and ICA for stock price prediction," in *Smart Innovation, Systems and Technologies*, vol. 107, Springer Singapore, 2019, pp. 479–487.
- [10] K. K. Yun, S. W. Yoon, and D. Won, "Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process," *Expert Systems with Applications*, vol. 186, Art. no. 115716, Dec. 2021, doi: 10.1016/j.eswa.2021.115716.
- [11] P. Sadorsky, "Using machine learning to predict clean energy stock prices: How important are market volatility and economic policy uncertainty?," *Journal of Climate Finance*, vol. 1, p. 100002, Dec. 2022, doi: 10.1016/j.jclimf.2022.100002.
- [12] D. Wen, L. Liu, Y. Wang, and Y. Zhang, "Forecasting crude oil market returns: Enhanced moving average technical indicators," *Resources Policy*, vol. 76, p. 102570, Jun. 2022, doi: 10.1016/j.resourpol.2022.102570.
- [13] Bourse de Casablanca, "Casablanca stock exchange," *Casablanca-bourse.com*, 2020. <https://www.casablanca-bourse.com/bourseweb/index.aspx> (accessed Dec. 16, 2022).
- [14] I. Ouadi and A. Ibourk, "Deep learning and stock market trends: an experiment with food companies in Morocco," in *Lecture Notes in Networks and Systems*, vol. 771 LNNS, Springer Nature Switzerland, 2023, pp. 194–205.
- [15] H. Oukhouya and K. El Himdi, "Comparing machine learning methods—SVR, XGBoost, LSTM, and MLP— for forecasting the Moroccan stock market," in *IOCMA 2023*, Apr. 2023, doi: 10.3390/iocma2023-14409.
- [16] A. Ifleh, A. Bilal, and M. El Kabbouri, "Moroccan stock price prediction using trend technical indicators: a comparison study," in *Lecture Notes in Networks and Systems*, vol. 646 LNNS, Springer Nature Switzerland, 2023, pp. 380–386.
- [17] X. Wu et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, Dec. 2008, doi: 10.1007/s10115-007-0114-2.
- [18] D. Pregibon, "Logistic regression diagnostics," *The Annals of Statistics*, vol. 9, no. 4, Jul. 2007, doi: 10.1214/aos/1176345513.
- [19] A. Kraskov, H. Stögbauer, and P. Grassberger, "Erratum: Estimating mutual information (Phys. Rev. E (2004) 69 (066138)),," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 1, Jan. 2011, doi: 10.1103/PhysRevE.83.019903.
- [20] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002, doi: 10.1023/A:1012487302797.




- [21] T. Hastie, J. Friedman, and R. Tibshirani, "The elements of statistical learning," *International Statistical Review*, vol. 77, no. 3, Art. no. 482, Oct. 2001, doi: 10.1007/978-0-387-21606-5.
- [22] H. El-Amir and M. Hamdy, "Feature selection and feature engineering," in *Deep Learning Pipeline*, Apress, 2020, pp. 233–276.
- [23] M. A. A. Matin, A. Triayudi, and R. T. Aldisa, "Comparison of principal component analysis and recursive feature elimination with cross-validation feature selection algorithms for customer churn prediction," in *Lecture Notes in Electrical Engineering*, vol. 1008, Springer Nature Singapore, 2023, pp. 203–218.
- [24] M. A. Mithu, K. M. Rahman, M. R. A. Razu, M. Riajuliislam, S. I. Momo, and A. Sattar, "Gold price forecasting using regression techniques for settling economic and stock market inconsistency," Jul. 2021, doi: 10.1109/ICCCNT51525.2021.9579755.
- [25] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Springer Topics in Signal Processing*, vol. 2, Springer Berlin Heidelberg, 2009, pp. 1–4.
- [26] M. Rahman, M. T. Ahmed, S. Nur, and A. Z. M. Touhidul Islam, "The prediction of coronavirus disease 2019 outbreak on Bangladesh perspective using machine learning: a comparative study," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 4276–4287, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4276-4287.
- [27] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-August-2016, pp. 785–794, doi: 10.1145/2939672.2939785.
- [28] X. Wang, X. Wang, B. Ma, Q. Li, and Y. Q. Shi, "High precision error prediction algorithm based on ridge regression predictor for reversible data hiding," *IEEE Signal Processing Letters*, vol. 28, pp. 1125–1129, 2021, doi: 10.1109/LSP.2021.3080181.
- [29] K. Mostafaei, S. Maleki, M. Z. A. Mahmoudi, and D. Knez, "Risk management prediction of mining and industrial projects by support vector machine," *Resources Policy*, vol. 78, p. 102819, Sep. 2022, doi: 10.1016/j.resourpol.2022.102819.
- [30] J. Y. Hsia and C. J. Lin, "Parameter selection for linear support vector regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5639–5644, Dec. 2020, doi: 10.1109/TNNLS.2020.2967637.
- [31] G. W. Snedecor, "Statistical methods: Applied to experiments in agriculture and biology," *The Iowa State College Press*, 1956.
- [32] O. Chantarakasemchit, S. Nuchitprasitchai, and Y. Nilsiam, "Forex rates prediction on EUR/USD with simple moving average technique and financial factors," in *17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2020*, Jun. 2020, pp. 771–774, doi: 10.1109/ECTI-CON49241.2020.9157907.
- [33] G. M. Ngene and A. N. Mungai, "Stock returns, trading volume, and volatility: The case of African stock markets," *International Review of Financial Analysis*, vol. 82, p. 102176, Jul. 2022, doi: 10.1016/j.irfa.2022.102176.
- [34] M. A. Muslim and Y. Dasril, "Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5549–5557, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5549-5557.
- [35] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.
- [36] A. Di Bucchianico, "Coefficient of determination (R²)," *Encyclopedia of Statistics in Quality and Reliability*. Wiley, Dec. 2007, doi: 10.1002/9780470061572.eqr173.

BIOGRAPHIES OF AUTHORS






Bouzgarne Itri    holds a PhD in artificial intelligence from ENSET Mohammedia of Hassan II University, Morocco. Received a state degree in computer science engineering from the National Institute of Statistics and Applied Economics INSEA, RABAT/Morocco in 2003. Since, he has managed several projects in the modernization, innovation and transformation of information systems for large financial structures. Currently, Occupies the Transversal Department Manager role in Technology Officer at IT service company Eurafri-Information, subsidiary of the banking group BOA. In parallel, he is a member of the Computer Science, Artificial Intelligence and Cyber Security Laboratory (2IACS) and he has published several papers on artificial intelligence applied to financial fields. Furthermore, He serves as a professor teaching machine learning in the Computer Engineering Department at the National School of Applied Sciences (ENSA) in Berrechid. He can be contacted at email: bouzgarne.itri@gmail.com.






Youssfi Mohamed    born on 1970 in Ouarzazat, Morocco. He serves as researcher professor in Department of Mathematics and Computer Science, ENSET, Hassan II University of Casablanca since 1993. He received his PhD (Doctorat de 3^{ème} Cycle) from The Mohammed V University of Rabat, Morocco in 1993 in computer science and his PhD (Doctorat d'Etat) from The Mohammed V University of Rabat, Morocco in 2015 in parallel and distributed systems. Director of the research laboratory: computer science, artificial intelligence and cybersecurity. His scientific research are focuses on computational intelligence, parallel and distributed systems, multi agent systems, web semantic, high-performance computing, big data, and artificial intelligence. More than 200 videos for teaching different technologies (Java programming, middleware, distributed systems, XML, multi agents systems, web services, web and mobile development), are published in his video channel: <http://youtube.com/mohamedYoussfi>. He can be contacted at email: med@youssfi.net.






Bouattane Omar    has received his PhD from the University Hassan II of Casablanca, in parallel computing and image processing (2001). He currently serves as a full professor in the Department of Electrical Engineering at ENSET of Mohammedia. His topic is in various domains of high-performance computing, parallel image processing and electrical engineering, particularly in renewable energy and smart grids. Since 2012, He was the head of the Laboratory of Signals, Distributed Systems and Artificial Intelligence. He has registered 07 international PCT patents and published over 70 papers in distinguished scientific journals and more than 60 papers in international conferences, he supervised more than 15 PhD theses. Pr. Bouattane is actively involved in international and European Union funded research H2020 programs. He led a powerful program of linkage for entrepreneurship achievement program for industrial innovation. His research activities are drive usually with international collaborations and very closed to industries. Actually, his new research projects interest to integrate Medical imaging and pattern recognition in the Cyber-Physical Systems for Pedagogical Rehabilitation in Special Education. He can be contacted at o.bouattane@enset-media.ac.ma.






El Madani Latifa    Born in 1981 in Errachidia, Morocco, she is a researcher specializing in artificial intelligence applied to cybersecurity. She is a member of the Computer Science, Artificial Intelligence and Cybersecurity Laboratory (2IACS) at ENSET, Hassan II University of Casablanca. She obtained her State Engineer diploma in 2004 from the Mohammedia School of Engineers, Mohammed V University in Rabat. Additionally, she holds the position of IT Manager at the Presidency of Hassan II University of Casablanca, where she oversees several strategic and innovative projects. She can be reached at latifa.elmadani-etu@etu.univh2c.ma.



Moumoun Lahcen    is a professor at Hassan 1st University in Settat (Morocco). He holds a degree from the École Normale Supérieure d'Enseignement Technique in Mohammedia (Morocco), a Diploma of Advanced Studies from the École Nationale Supérieure d'Électricité et de Mécanique in Casablanca (Morocco), a PhD in computer science, and a Habilitation to Supervise Research (HDR) from Hassan 1st University in Settat (Morocco). His research within the 'Mathematics, Computer Science, and Engineering Sciences (MISI)' laboratory at the Faculty of Sciences and Techniques in Settat focuses on 3D data analysis, visual semantics, and artificial intelligence, particularly in machine learning, self-supervised learning, and federated learning. He can be contacted at lahcen.moumoun@uhp.ac.ma.



Oualid Adil    born on April 3, 1988, in Casablanca, Morocco, is a researcher specializing in the application of Artificial Intelligence in fintech, particularly in credit risk management. With an IT background in web development, he has transitioned into the field of AI, demonstrating a profound commitment to innovating financial technologies. His research aims to revolutionize credit risk evaluation by integrating advanced AI methodologies, thereby enhancing accuracy and efficiency. His contributions are reflected in my published works where he shares insights on bridging the gap between theoretical AI and practical fintech applications. He can be contacted at email a.oualid@uhp.ac.ma.