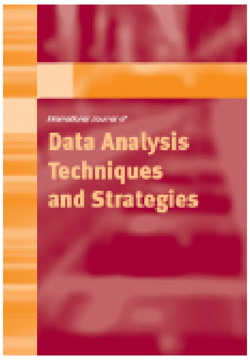


International Journal of Data Analysis Techniques and Strategies

ISSN online: 1755-8069 - ISSN print: 1755-8050
https://www.inderscience.com/ijdats

Analysis of machine learning's performance in stock market prediction, compared to traditional technical analysis indicators

Mohammed Bouasabah

DOI: 10.1504/IJDATS.2024.10061891

Article History:

Received:	08 July 2023
Last revised:	13 December 2023
Accepted:	15 December 2023
Published online:	19 March 2024

Analysis of machine learning's performance in stock market prediction, compared to traditional technical analysis indicators

Mohammed Bouasabah

Department of Business and Management,
National School of Business and Management,
Ibn Tofail University,
Kenitra, Morocco
Email: Mohammed.bouasabah@uit.ac.ma

Abstract: This study compares the performance of machine learning (ML) algorithms with traditional technical indicators in real estate, technology, and healthcare sectors. Unveiling the limitations of classical indicators, particularly their struggle to surpass the 50% threshold, the research explores the predictive capabilities of ML algorithms, focusing on AdaBoost and support vector machine (SVM). The relative strength index (RSI) emerges as a reliable performer for buy decisions but with potential oversight. Results affirm the superiority of ML algorithms in precision, recall, and F1 score, transcending traditional indicators. Sector-specific variations showcase exceptional ML efficacy, particularly in healthcare. Algorithmic evaluation spotlights AdaBoost and SVM, underscoring the importance of strategic selection. The study advocates for a nuanced approach, blending RSI with ML for refined strategies. In conclusion, this research contributes significantly to financial decision-making, exposing limitations and positioning ML algorithms as powerful tools for improved investment strategies.

Keywords: machine learning; ML; technical analysis indicators; prediction; financial market; data analysis; SMA; MACD; RSI; trading; support vector machine; SVM.

Reference to this paper should be made as follows: Bouasabah, M. (2024) 'Analysis of machine learning's performance in stock market prediction, compared to traditional technical analysis indicators', *Int. J. Data Analysis Techniques and Strategies*, Vol. 16, No. 1, pp.32–46.

Biographical notes: Mohammed Bouasabah is an academician in applied mathematics, specialising in stochastic modelling, with a particular emphasis on quantitative methods in finance. Bringing experience as a Professor at Ibn Tofail University in Kenitra, Morocco, he possesses a wealth of knowledge in data analysis and project management, adeptly navigating the intricate intersections of mathematical theory and practical applications. In addition to his academic accomplishments, he holds a degree in computer engineering and telecommunication.

1 Introduction

Despite the advancements in predictive models for stock market movements, a critical gap remains in achieving consistently accurate forecasts. Existing evaluations often fall short in capturing the intricate dynamics of financial markets, leading to suboptimal decision-making for investors and traders. This persistent challenge underscores the need for a more robust and nuanced approach to stock market prediction. This research study aims to address this gap by delving into the comparative performance of machine learning (ML) algorithms and classical technical analysis in diverse sectors, namely real estate, technology, and healthcare.

The limitations of current predictive models, especially their struggle to adapt to the dynamic and non-linear nature of financial markets, propel us towards ML algorithms. These algorithms have demonstrated effectiveness across various fields, extending beyond financial markets (Dubey and Chandani, 2022).

Previous studies have shown that ML can deliver significant improvements in stock market prediction over traditional approaches. For example, Brogaard and Zareei (2023) used ML algorithms, such as neural networks, to predict stock market trends with greater accuracy than models based on traditional technical analysis. In addition, research by Pahwa and Agarwal (2019), revealed that the use of ML techniques can help identify non-linear patterns and relationships in financial data, thereby improving the accuracy of predictions.

However, it is important to note that classical technical analysis has also proved it is worth as a tool for predicting stock market movements. Indicators such as moving averages, Bollinger bands and oscillators have been widely used by investors and traders for many years. For example, the work of Sehgal and Gupta (2007) highlights the effectiveness of these classic indicators in detecting market trends and reversals.

This study explores whether the use of ML is superior in predicting stock market movements compared to traditional methods. The focus extends beyond the accuracy of predictions to assessing their financial efficacy in trade decisions. Additionally, it evaluates the adaptability of ML tools to changes in the market, gauging their effectiveness in diverse financial situations. It is goal is to give useful insights that can help people make better decisions when it comes to investing in the stock market.

Clearly, the objective of the study is to look at different applications of ML to finance, compared with more traditional financial models. Then, to compare the performance, in terms of both predictions and proposed returns, of four ML models (random forest, AdaBoost, SVM and KNN) and four technical analysis tools (SMA, MACD, RSI and ROC). The aim is to answer the following question: 'How do ML algorithms perform against technical analysis technical analysis methods in the case of the biotech, real estate and energy sectors?'

This research explores three distinct sectors: real estate, technology, and healthcare, using specific trackers for each: RWR, XLK, and XLV respectively. Within these sectors, ML algorithms and technical analysis indicators will be assigned the task of predicting the direction of the next day's share prices (up or down). To assess their predictive efficacy, this study will use key metrics such as Precision, Recall, F1 score, and confusion matrix. The use of Python for this analysis not only facilitates a thorough examination but also allows for the efficient implementation of ML algorithms, ensuring a robust and insightful evaluation.

Moreover, these findings aim to give people in the market a clear picture of how well ML works for predicting stock market changes compared to the usual methods. Instead of just comparing them, this study wants to dig deeper and uncover the strengths and weaknesses of each approach. Specifically, the aim is to determine where teaming up ML with the regular technical analysis indicators could lead to even better results. By exploring these details, this study hopes to provide practical insights, empowering investors and traders to make smart decisions in the ever-changing world of finance with confidence and precision.

2 Literature review

2.1 Technical indicators used in comparison

Technical analysis indicators, based on the study of historical charts and market trends, have long been indispensable tools for financial analysts (Fang et al., 2014). These indicators provide valuable insights into past market movements, thereby offering clues to potential future changes. However, with the advent of more advanced technologies, ML algorithms have emerged as serious contenders in the field of stock market prediction. (Ayala et al., 2021).

Numerous technical analysis indicators are commonly employed to forecast stock market trends. Among the widely recognised ones are the simple moving average (SMA), moving average convergence divergence (MACD), relative strength index (RSI), and rate of change (ROC). These indicators play a prominent role in helping traders and analysts gain insights into market dynamics and anticipate potential shifts. Each of them brings a unique perspective to the table, contributing to a comprehensive understanding of market movements.

The SMA is a nifty indicator that works by calculating the average price over a specific period. It is a go-to tool for many because it helps spot market trends. Essentially, it gives you a smoothed-out view of the price movement, making it easier to see if the market is on an upswing, downswing, or just chilling out. Traders find it handy for catching those broader trends in the midst of all the market noise (Hansun, 2013). The MACD, or moving average convergence divergence, serves as a crucial indicator within the realm of stock market analysis. It is functionality involves the amalgamation of various moving averages, enabling the detection of shifts in trends and the provision of buy or sell signals. This intricate process enhances it is capability to identify opportune moments for market participants to initiate buying or selling actions. The MACD, with it is sophisticated approach, contributes to a nuanced understanding of market dynamics, offering a valuable tool for traders to make informed decisions within the academic framework of technical analysis (Aguirre et al., 2020). The RSI is an oscillator that operates by gauging both the strength and speed of price movements. It is primary function is to identify levels of overbought and oversold conditions within the market. In simpler terms, the RSI acts as a speedometer for price changes, helping traders discern when an asset might be reaching a point of being overbought (potentially due for a downward correction) or oversold (possibly signalling an upcoming upward correction). This makes it a valuable tool for market participants looking to understand the momentum behind price movements and make strategic decisions accordingly (Rosillo et al., 2013). The ROC is an indicator designed to measure the percentage change in price relative to a preceding period. It is primary function is to identify instances of impulse within the market. In practical terms, the ROC provides insights into the momentum of price movements by quantifying the rate at which these changes occur. This information becomes valuable for traders seeking to pinpoint moments of significant price shifts. Essentially, the ROC acts as a magnifying glass for detecting the pace at which prices are evolving, aiding market participants in making informed decisions based on the intensity of market momentum (Shynkevich et al., 2017).

2.2 Machine learning algorithms used in stock market prediction

In the ever-evolving landscape of financial markets, the quest for more sophisticated tools to predict stock movements has led to the rise of ML algorithms. Unlike traditional methods, ML leverages the power of data-driven models and computational intelligence to discern patterns, identify trends, and make predictions. In an era dominated by technological advancements, these algorithms have emerged as formidable contenders, challenging conventional approaches to stock market analysis (Gerlein et al., 2016).

ML offers a promising approach to stock market prediction, using algorithms capable of learning from data and discovering complex patterns. Various ML algorithms have been used in previous studies for stock market prediction, including random forest, AdaBoost, support vector machines (SVM) and K-nearest neighbours (KNN).

Random forest is a supervised learning algorithm based on decision trees, which combines the predictions of several trees to obtain a final prediction (Schonlau and Zou, 2020). AdaBoost is an algorithm that combines several weak classifiers to form a stronger model (Sun et al., 2011). SVM is an algorithm that uses kernel functions to transform data into a higher-dimensional space, enabling classes to be separated (Jaiwang and Jeatrakul, 2016). The SVM algorithm has demonstrated remarkable performance in various areas of classification problems and has shown it is superiority over other classification algorithms (Alweshah et al., 2017; Jain and Saxena, 2022). KNN is an algorithm that relies on the nearest training examples to predict the class of a new example (Chen and Hao, 2017).

2.3 Previous studies comparing the performance of traditional indicators and machine learning

Numerous studies have delved into the comparison between traditional technical analysis indicators and ML models in predicting stock market movements. These investigations share a common goal: evaluating the predictive accuracy of the models and assessing their efficacy in terms of trading profitability. Going beyond mere accuracy, these studies also focus on examining how well these models adapt to varying market conditions. Through this multifaceted exploration, these research endeavours contribute valuable insights to the ongoing discussion on the optimal utilisation of predictive tools in the dynamic and evolving financial markets. Kumar et al. (2018) developed and compared five supervised ML techniques for stock price prediction in order to overcome such difficulties. Zhang et al. (2018) compared the performance of MACD, RSI and ROC with random forest and SVM. Their study revealed that random forest achieved higher predictive accuracy than conventional indicators, thanks to it is ability to model non-linear relationships and capture complex patterns in financial data.

Bhattacharjee and Bhattacharja (2019) conducted a comparative study examining various methods for predicting stock prices. The research evaluated traditional statistical approaches, including SMA, weighted moving average, exponential smoothing, and the naive approach, alongside ML techniques such as linear regression, lasso, ridge, KNN, SVM, random forest, single layer perceptron, multi-layer perceptron, and long short-term memory. The results indicated that, after a thorough analysis, ML approaches, particularly neural network models, demonstrated superior accuracy in stock price prediction compared to traditional statistical methods.

Misra et al. (2018) explores stock market prediction, addressing challenges in accuracy due to evolving market patterns and missing data fields. The wordistinguish between predictions based on historical data and those incorporating social media information. Their paper categorises existing predictive analytics methods across domains, identifying shortcomings. Emphasising the need for data cleansing, the authors propose improvements for enhanced accuracy in these approaches.

Vijh et al. (2020) tackle the formidable challenge of accurately predicting stock market returns, acknowledging the volatile and non-linear nature of financial markets. Leveraging the power of artificial intelligence and enhanced computational capabilities, the researchers employ artificial neural network and random forest techniques to forecast the next day's closing price for five companies across diverse sectors. The work consists of creating new variables using financial data such as open, high, low, and close prices, serving as inputs to the models. Evaluation metrics, including RMSE and MAPE, demonstrate the efficiency of the models in predicting stock closing prices, as indicated by their low values.

Mokhtari et al. (2021) compared SMA, RSI and ROC with KNN in stock market prediction. Their study showed that KNN outperformed classical indicators in terms of profitability of completed trades, identifying trading opportunities more effectively and limiting potential losses.

In another study, Himanshu and Sopan (2018) compared the performance of the SMA, MACD, RSI and ROC with that of several ML models, including SVM and Random Forest. Their results indicated that SVM achieved higher predictive accuracy than conventional indicators, but random forest outperformed all models in predicting short-term price movements.

These studies underline the advantage of ML models in stock market prediction over traditional technical analysis indicators. ML algorithms are capable of processing complex datasets, discovering non-linear patterns and adapting to different market conditions. This enables them to better capture the dynamics and subtle relationships between financial variables, which can lead to better predictions and investment decisions.

However, it should be noted that the performance of ML models can vary depending on a number of factors, such as data quality and availability, study period and specific model parameters. A rigorous comparative performance evaluation is therefore essential to select the most suitable approaches for stock market prediction purposes.

3 Methodology

3.1 Data

As highlighted in the introduction, the research narrows it is focus to three distinct sectors, each tracked by RWR, KLV, and XLV, respectively. To initiate this analysis, the daily historical data of these trackers is gathered from Yahoo Finance. The dataset encapsulates a comprehensive thirteen-year period, extending from April 1, 2010, to April 1, 2023. Within this extensive timeframe, the dataset used in this study encompasses six key variables that form the basis of the investigation. These variables serve as crucial elements for assessing and understanding the dynamics within the selected sectors, providing a robust foundation for the subsequent analyses and findings:

- open: the price at the opening of a specific date.
- high: the highest price at which it was traded on a given day
- low: the lowest price at which it was traded on a given day
- close: the price at the close of a given day
- volume: the quantity of shares traded on a given date
- Adj.Close: the adjusted price at the close of a given day, taking into account the dividend distribution.

For the model training and testing, this study opted for a division of the dataset into two segments: 80% allocated for training the model and the remaining 20% reserved for rigorous testing. This strategic partitioning allows us to leverage a substantial portion of the data to train the model, enabling it to learn and discern patterns effectively. The reserved 20% serves as a dedicated testing set, serving as a critical assessment ground for evaluating the model's generalisability and performance on unseen data. This methodological choice aims to strike a balance between robust model training and a comprehensive evaluation of it is predictive capabilities.

3.2 Variables

Utilising the dataset, three key variables are derived to capture significant aspects of market dynamics. Firstly, the difference between the opening and closing prices for each trading day is computed, providing insight into daily price movements. The second variable reflects the range between the highest and lowest prices within a given trading day, offering a measure of intra-day volatility. Lastly, the third variable quantifies the disparity between the traded volume on the next day and the current day's volume. It is important to note that this last variable cannot be computed for the final date in the sample, resulting in the exclusion of this particular data point from the analysis to maintain accuracy and consistency in the calculations.

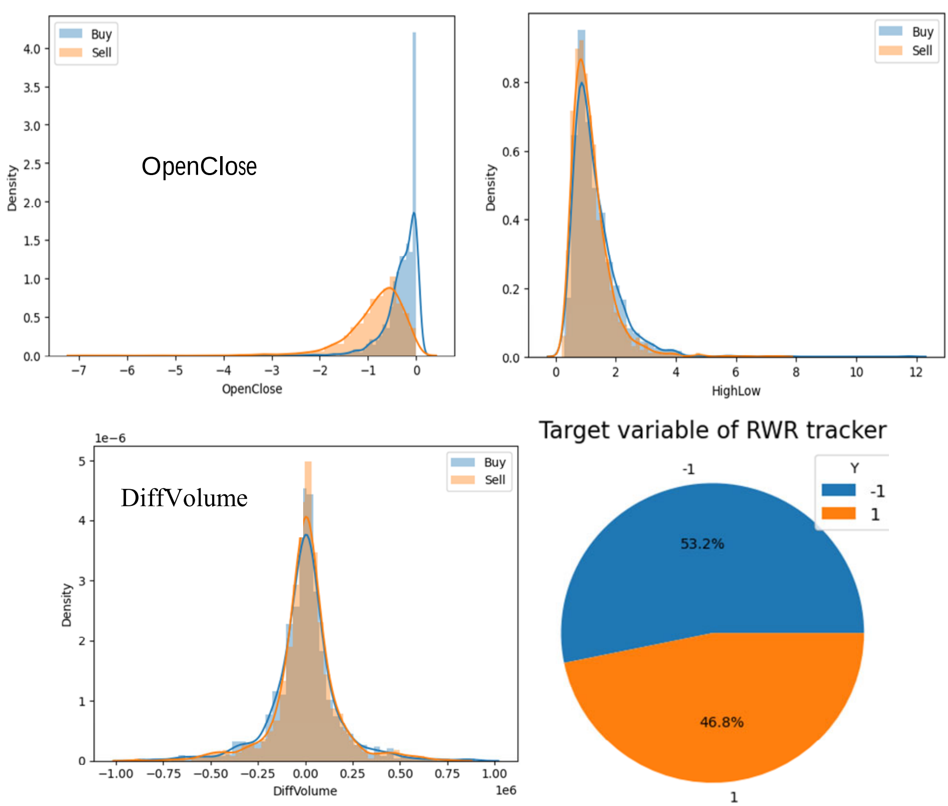
The dependent variable, called Y in this study, corresponds to the label attached to the data, and to what the model will try to predict on the basis of the independent variables. This variable defines whether the next day's stock price will close higher or lower, and takes either the value 1 or –1. A value of 1 means a buy signal for the period concerned, while a value of –1 means to sell it. To do this, the returns are calculated, in percent, based on the Adjusted closing price. Then the variable is created, equal to 1 if the return is positive and –1 if the return is negative. The last line of the database was deleted, as it was not possible to calculate the yield at that date, nor the volume difference.

In the remainder, this study will detail each of the models developed on the basis of RWR financial data. The process is identical for the other two sectors. As mentioned above, four different algorithms were tested, all of them are classification and supervised algorithms, i.e., the different possible classes that the variable Y can take are provided to the algorithm, i.e., it can be equal to 1 or –1.

3.3 *Exploratory data analysis*

The analysis of the RWR tracker's feature and target variables is plotted in Figure 1 by separating the data into two sets (buy and sell) according to the dependent variable. It should be noted that the HighLow and DiffVolume variables behave in the same way on both sets, while the OpenClose variable behaves differently on both sets. Also the two sets are almost balanced (53.2% for the sell decision and 46.8% for the buy decision).

Figure 1 Graph of features and target variables for both buy and sell datasets (see online version for colours)



The various ML algorithms are implemented using Jupyter Notebook, a popular interactive development environment for Python that lets you run code, visualise results and display graphs.

3.4 Machine learning algorithms

The scikit-learn (sklearn) Python library is used to implement ML algorithms such as random forest, Adaboost, SVM and KNN in the Jupyter Notebook environment. The utilisation of scikit-learn is instrumental in this study, as it furnishes straightforward and efficient implementations of these algorithms (Hackeling, 2017). Choosing this library makes it easier to build models and blend different ML methods seamlessly. This helps make the analysis strong and adaptable. Using scikit-learn simplifies the process of working with various algorithms and ensures a standardised and efficient way to evaluate and compare models.

First off, for each algorithm, a corresponding object is created:

For random forest, scikit-learn's random forest classifier class, is used tweaking details like the number of trees we want in the forest. Once the algorithm has been configured, it is trained on the training data using the fitting method. After the training session, the model becomes a predictive assistant. It can be used to predict the results of new data using the prediction method. This step-by-step process ensures that the algorithms are well prepared and ready to interpret new information.

Finally, the quality of each algorithm is verified using measures such as precision, recall and F1 score. Scikit-learn facilitates this part with functions that help to measure and understand algorithm performance.

To initialise AdaBoost, scikit-learn's AdaBoostClassifier is employed. Key parameters, such as the number of weak learners (base estimators) and the learning rate, are fine-tuned to achieve optimal performance. After configuration, the algorithm undergoes a training phase on the provided dataset using the fit method. This equips the model with the ability to make predictions on new data using the predict method, thereby enhancing it is predictive capabilities.

Utilising scikit-learn's SVM implementation, specifically the support vector classification (SVC) class, the SVM algorithm is configured. Parameters like the type of kernel, regularisation parameter (C), and kernel coefficients are adjusted to enhance it is performance. Following the configuration, the algorithm undergoes a training session on the training data using the fit method. Once trained, the SVM model is primed to make predictions on new data using the predict method, contributing to it is predictive prowess.

For KNN, scikit-learn's K-neighbours classifier class is instantiated, with essential parameters like the number of neighbours (K) being fine-tuned. After the configuration phase, the algorithm is trained on the provided dataset using the fit method. Post-training, the KNN model is well-prepared to offer predictions for new data points using the predict method, enriching it is predictive capabilities.

In Jupyter Notebook, the code is executed, checking the results and using graphs and tables to evaluate the performance of each algorithm. It is a kind of visual guide that helps us experiment and improve the ML models (Garreta and Moncecchi, 2013).

3.5 Technical analysis indicators

To implement technical analysis indicators such as SMA, MACD, RSI and ROC in Jupyter Notebook, popular libraries such as Pandas, NumPy and Matplotlib in Python are used.

Once the data has been loaded, Pandas library is used to perform data manipulation operations, such as transformation, cleaning or solving missing data problems (Nti et al., 2020). Next, technical indicators using functions provided by Pandas or by implementing them manually is calculated.

For the SMA, Panda's rolling() function is used to calculate the average price over a specific time window. Similarly, when dealing with the MACD, the subtraction of the short-term exponential moving average (EMA) from the long-term EMA is done, following the common practice in technical analysis.

After obtaining these technical indicators, the next step involves comparing their decisions with the actual decisions marked by the dependent variable Y. To ensure a fair evaluation, the same metrics are used as those used to evaluate ML algorithms, enabling us to assess the extent to which technical indicators align with real results. This process provides a basis for understanding the effectiveness of these indicators in making informed decisions in a given context.

3.6 Metrics

To gauge and compare the effectiveness of ML algorithms against traditional technical analysis indicators, specific metrics are used: precision, recall, f1 score, and accuracy. These metrics serve as objective measures to evaluate how well each method performs in terms of making accurate predictions. This approach allows for a balanced and neutral assessment, helping us discern which method, whether algorithms or technical indicators, proves more reliable and suitable for the analysis.

- Precision: Precision is a measure of a model's ability to classify positive instances correctly (a buy decision is considered positive, a sell decision negative). It is calculated by dividing the number of true positives by the sum of true positives and false positives. In other words, precision indicates the proportion of positive observations predicted correctly compared to all observations predicted as positive.
- Recall: Recall, also known as sensitivity, measures a model's ability to identify all positive instances. It is calculated by dividing the number of true positives by the sum of true positives and false negatives. Recall indicates the proportion of correctly predicted true positive observations to all true positive observations. A high recall indicates that the model has a low propensity to miss true positives (false negatives).
- F1 score: The F1 score is a measure that combines precision and recall into a single value. It represents the harmonic mean of precision and recall, and is calculated using the formula: $2\times\frac{precision\times recall}{precision+recall}$. The F1 score is particularly useful when you want to strike a balance between precision and recall. It reaches it is maximum value of one when both precision and recall are perfect.
- Accuracy: Accuracy is a global measure of performance that evaluates the proportion of correctly classified observations in relation to the total number of observations. It is calculated by dividing the number of true positives and true negatives by the total number of observations. Accuracy provides an overall view of model performance, but can be misleading if classes are unbalanced, i.e., if one class is much more frequent than the other. It should be noted that in this work, the accuracy metric is relevant because the two classes of decisions are almost balanced. When evaluating the performance of both ML algorithms and traditional technical analysis indicators,

these metrics will allow us to assess their effectiveness in making accurate predictions and decisions.

4 Results

In this section, the results obtained will be presented and the answer to the question on which the study is based, i.e., 'How do ML algorithms perform compared with technical analysis methods in the case of the technology, real estate and healthcare sectors?' will be detailed.

Table 1 Metric values for the random forest algorithm with RWR tracker data

	Precision	*Recall*	*F1 score*	*Support*
–1	76%	76%	76%	334
1	74%	75%	75%	313
Accuracy	75%	647		

Notes: The calculation of metric values for ML algorithms is based on the test set (647 values).

Table 2 Metric values for the AdaBoost algorithm with RWR Tracker data

	Precision	*Recall*	*F1 score*	*Support*
–1	79%	81%	80%	334
1	79%	77%	78%	313
Accuracy	79%	647		

Table 3 Metric values for the SVM algorithm with RWR tracker data

	Precision	*Recall*	*F1 score*	*Support*
–1	83%	80%	81%	334
1	79%	83%	81%	313
Accuracy	81%	647		

Table 4 Metric values for the AdaBoost algorithm with RWR Tracker data

	Precision	*Recall*	*F1_score*	*Support*
–1	80%	78%	79%	334
1	77%	79%	78%	313
Accuracy	78%	647		

Tables 1 to 4 show the values of the four metrics: Accuracy, Recall, F1 Score and Precision obtained for each algorithm applied to RWR Tracker financial data, specifying the metric value for buy and sell decisions separately. Table 5 summarises the values of these same metrics for the ML algorithms and also for the technical indicators, giving the value of the metric for both buy and sell decisions. Table 6 and Table 7 give the values of the same metrics for the technology and health sectors respectively.

Table 5 Metric values for the machine learning algorithms and technical indicators for RWR tracker

Strategy	*Precision*	*Recall*	*F1_score*	*Accuracy*
Random forest	75%	75%	75%	75%
AdaBoost	79%	79%	79%	79%
SVM	81%	81%	81%	81%
KNN	78%	78%	78%	78%
SMA	51%	51%	51%	51%
MACD	50%	53%	40%	53%
RSI	96%	47%	62%	47%
ROC	46%	49%	49%	49%

Table 6 Metric values for the machine learning algorithms and technical indicators for XLK Tracker

Strategy	*Precision*	*Recall*	*F1_score*	*Accuracy*
Random forest	78%	77%	78%	78%
AdaBoost	83%	83%	83%	84%
SVM	81%	79%	79%	80%
KNN	77%	76%	76%	77%
SMA	54%	52%	53%	52%
MACD	54%	56%	44%	56%
RSI	99%	44%	61%	44%
ROC	44%	50%	50%	50%

Table 7 Metric values for the machine learning algorithms and technical indicators for XLV tracker

Strategy	*Precision*	*Recall*	*F1_score*	*Accuracy*
Random forest	80%	79%	80%	80%
AdaBoost	83%	82%	83%	83%
SVM	82%	81%	82%	82%
KNN	81%	80%	80%	80%
SMA	51%	49%	50%	49%
MACD	50%	55%	42%	55%
RSI	99%	45%	61%	45%
ROC	42%	49%	49%	49%

5 Discussion

The objective of this study was to compare the performance of ML algorithms with that of traditional technical indicators in the real estate, technology and healthcare sectors. The results highlighted several significant findings.

Firstly, by examining the technical indicator metrics, all indicators, with the exception of the ROC indicator, exceed the 50% threshold. This indicates that these indicators are able to provide some predictive capability although their performance is limited. The ROC indicator, on the other hand, failed to cross this threshold, suggesting that it may not be as effective for the analysis. At the same time, the 'Precision' metric of the RSI indicator shows very high values, reaching 96%, 99% and 99% respectively for the real estate, technology and healthcare sectors. This high value demonstrates that the RSI is an excellent indicator for buying decisions (positive decisions) and predicts these decisions with a very high degree of certainty. These results confirm the reliability of RSI-based purchasing decisions in all three sectors. Although the RSI can accurately predict positive cases with a high degree of accuracy, it is important to note that the RSI only manages to identify 47%, 44% and 45% of positive cases for the real estate, technology and healthcare sectors respectively. Consequently, it is essential to combine the use of RSI with other indicators or to adopt a more holistic approach, such as the use of ML algorithms, which demonstrated superior performance in this study.

Secondly, of the four ML algorithms trained, namely random forest, AdaBoost, SVM and KNN. The AdaBoost and the SVM algorithms outperformed the other two algorithms. These results underline the importance of carefully selecting the appropriate ML algorithm to maximise predictive performance.

Given that the datasets have balanced buy and sell classes for all three trackers, the use of the 'Accuracy' metric is relevant. However, it is necessary to take into account a combination of metrics such as 'Precision', 'Recall' and 'F1 score' to assess overall model performance (Torgo and Ribeiro, 2009). These metrics provide a more comprehensive measure of performance, taking into account both correct predictions and Type I and II errors. It is also important to note that the performance of ML algorithms can be further optimised by adjusting hyperparameters. This offers a significant advantage over technical indicators, which are often static and cannot be optimised in the same way (Singh et al., 2021). By fine-tuning the hyperparameters of ML models, it is possible to achieve higher scores and further improve predictive performance.

What's more, when the results across the three sectors are compared, the finding shows that the metrics for ML algorithms were higher in the healthcare sector, even exceeding 80%. This suggests that ML algorithms have a stronger predictive capability in the healthcare sector than in the other sectors studied. It is possible to go further in studying the performance of ML algorithms applied to stock market prediction, and to see whether the sector has an impact on model performance but the study focuses only on the comparison between ML and technical analysis.

As a perspective for this research study, it would be interesting to widen the range of sectors studied, to see if the sector really has an impact on the efficiency of the models. As far as model robustness is concerned, an analysis covering a wider time period, with an increase in the number of input variables, would offer better models and enable better results to be obtained. A final suggestion would be to broaden the number of models studied. Indeed, there is an abundance of algorithms that have not been seen in this work, and it would be interesting to examine their effectiveness.

To conclude, this study highlights several important findings. Classical technical indicators have limited predictive capacity, although some achieve metrics above 50%. Among the ML algorithms evaluated, AdaBoost and SVM proved to be the best performers, outperforming the others. Furthermore, the performance of ML algorithms varies by sector, with particularly strong performance in the healthcare sector. It is essential to use an appropriate combination of metrics to evaluate model performance, taking into account both correct predictions and errors. Finally, the flexibility of ML algorithms to adjust hyperparameters enables performance to be further optimised, offering a significant advantage over technical indicators.

In summary, the study unequivocally demonstrates the superior performance of ML algorithms over traditional technical indicators in the real estate, technology, and healthcare sectors. Notably, ML, particularly with AdaBoost and SVM, outperforms other algorithms like random forest and KNN, emphasising the critical need for a thoughtful selection of ML algorithms tailored to the unique characteristics of each sector. While classical indicators exhibit some predictive ability, with the RSI indicator excelling in accurate buying decisions, this research underscores the imperative of integrating these traditional indicators with ML algorithms for a more comprehensive approach. The study highlights the flexibility of ML algorithms in parameter adjustment for enhanced results, a feature absent in traditional indicators. Future research exploring diverse indicators, ML algorithms, and sectors will provide a deeper understanding of comparative performance. Overall, the findings affirm that ML algorithms surpass traditional indicators, opening avenues for advancements in financial prediction and decision-making, ultimately enhancing investment and business strategies.

6 Conclusions

In summary, this study clearly shows that ML algorithms outperform traditional technical indicators across real estate, technology, and healthcare sectors. ML, especially with AdaBoost and SVM, takes the lead over other algorithms like random forest and KNN. This highlights the need to carefully pick the right ML algorithm for the best predictions, considering the unique characteristics of each sector.

Classic indicators do have some predictive ability, with the RSI indicator shining in accurate buying decisions. However, it is vital to combine these traditional indicators with ML algorithms for a more well-rounded approach.

The study emphasises the flexibility of ML algorithms in adjusting parameters for better results, unlike traditional indicators. Future research exploring different indicators, ML algorithms, and sectors will provide a deeper understanding of comparative performance. In conclusion, it is important to consider potential avenues for future research. One of these could be to further investigate the impact of the sector on algorithm performance. Future studies could aim to confirm or refute this influence, providing valuable information for a more nuanced understanding of algorithmic behaviour in different sectors.

References

Aguirre, A.A.A., Medina, R.A.R. and Méndez, N.D.D. (2020) 'Machine learning applied in the stock market through the moving average convergence divergence (Macd) indicator', *Investment Management and Financial Innovations*, Vol. 17, No. 4, pp.44–60, DOI: 10.21511/imfi.17(4).2020.05.

Alweshah, M., Rashaideh, H., Hammouri, A.I., Tayyeb, H. and Ababneh, M. (2017) 'Solving time series classification problems using support vector machine and neural network', *International Journal of Data Analysis Techniques and Strategies*, Vol. 9, No. 3, p.237, DOI: 10.1504/IJDATS.2017.086634.

Ayala, J., García-Torres, M., Noguera, J.L.V., Gómez-Vela, F. and Divina, F. (2021) 'Technical analysis strategy optimization using a machine learning approach in stock market indices', *Knowledge-Based Systems*, Vol. 225, p.107119, DOI: 10.1016/j.knosys.2021.107119.

Bhattacharjee, I. and Bhattacharja, P. (2019) 'Stock price prediction: a comparative study between traditional statistical approach and machine learning approach', *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, IEEE, Khulna, Bangladesh, pp.1–6, DOI: 10.1109/EICT48899.2019.9068850.

Brogaard, J. and Zareei, A. (2023) 'Machine learning and the stock market', *Journal of Financial and Quantitative Analysis*, Vol. 58, No. 4, pp.1431–1472, DOI: 10.1017/S0022109022001120.

Chen, Y. and Hao, Y. (2017) 'A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction', *Expert Systems with Applications*, Vol. 80, pp.340–355, DOI: 10.1016/j.eswa.2017.02.044.

Dubey, R. and Chandani, A. (2022) 'Application of machine learning in banking and finance: a bibliometric analysis', *International Journal of Data Analysis Techniques and Strategies*, Vol. 14 No. 3, p.219, DOI: 10.1504/IJDATS.2022.128268.

Fang, J., Qin, Y. and Jacobsen, B. (2014) 'Technical market indicators: an overview', *Journal of Behavioral and Experimental Finance*, Vol. 4, pp.25–56, DOI: 10.1016/j.jbef.2014.09.001.

Garreta, R. and Moncecchi, G. (2013) *Learning Scikit-learn: Machine Learning in Python: Experience the Benefit is of Machine Learning Techniques by Applying them to Real-World Problems Using Python and the Open Source Scikit-Learn Library*, Packt Publishing Ltd, Birmingham, UK.

Gerlein, E.A., McGinnity, M., Belatreche, A. and Coleman, S. (2016) 'Evaluating machine learning classification for financial trading: an empirical approach', *Expert Systems with Applications*, Vol. 54, pp.193–207, DOI: 10.1016/j.eswa.2016.01.018.

Hackeling, G. (2017) *Mastering Machine Learning with Scikit-Learn: Learn to Implement and Evaluate Machine Learning Solutions with Scikit-learn*, Second Edition., Packt Publishing, Birmingham, UK.

Hansun, S. (2013) 'A new approach of moving average method in time series analysis', *2013 Conference on New Media Studies (CoNMedia)*, IEEE, Tangerang, Indonesia, pp.1–4, DOI: 10.1109/CoNMedia.2013.6708545.

Himanshu, H. and Sopan, A. (2018) 'Comparative analysis of stock market prediction system using SVM and ANN', *International Journal of Computer Applications*, Vol. 182, No. 1, pp.59–64, DOI: 10.5120/ijca2018917444.

Jain, M. and Saxena, V. (2022) 'An ECOSVS-based support vector machine for network anomaly detection', *International Journal of Data Analysis Techniques and Strategies*, Vol. 14, No. 1, p.32, DOI: 10.1504/IJDATS.2022.121513.

Jaiwang, G. and Jeatrakul, P. (2016), 'A forecast model for stock trading using support vector machine', *2016 International Computer Science and Engineering Conference (ICSEC)*, IEEE, Chiang Mai, Thailand, pp.1–6, DOI: 10.1109/ICSEC.2016.7859923.

Kumar, I., Dogra, K., Utreja, C. and Yadav, P. (2018) 'A comparative study of supervised machine learning algorithms for stock market trend prediction', *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, IEEE, Coimbatore, pp.1003–1007, DOI: 10.1109/ICICCT.2018.8473214.

Misra, M., Yadav, A.P. and Kaur, H. (2018) 'Stock market prediction using machine learning algorithms: a classification study', *2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering (ICRIEECE)*, IEEE, Bhubaneswar, India, pp. 2475–2478, doi: 10.1109/ICRIEECE44171.2018.9009178.

Mokhtari, S., Yen, K.K. and Liu, J. (2021) 'Effectiveness of artificial Intelligence in stock market prediction based on machine learning', *International Journal of Computer Applications*, Vol. 183, No. 7, pp.1–8, DOI: 10.5120/ijca2021921347.

Nti, I.K., Adekoya, A.F. and Weyori, B.A. (2020), 'A systematic review of fundamental and technical analysis of stock market predictions', *Artificial Intelligence Review*, Vol. 53 No. 4, pp. 3007–3057, doi: 10.1007/s10462-019-09754-z.

Pahwa, K. and Agarwal, N. (2019), 'Stock Market Analysis using Supervised Machine Learning', 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, Faridabad, India, pp. 197–200, doi: 10.1109/COMITCon.2019.8862225.

Rosillo, R., De La Fuente, D. and Brugos, J.A.L. (2013) 'Technical analysis and the Spanish stock exchange: testing the RSI, MACD, momentum and stochastic rules using Spanish market companies', *Applied Economics*, Vol. 45, No. 12, pp.1541–1550, DOI: 10.1080/00036846.2011.631894.

Schonlau, M. and Zou, R.Y. (2020), 'The random forest algorithm for statistical learning', *The Stata Journal: Promoting Communications on Statistics and Stata*, Vol. 20, No. 1, pp.3–29, DOI: 10.1177/1536867X20909688.

Sehgal, S. and Gupta, M. (2007) 'Tests of Technical Analysis in India', *Vision: The Journal of Business Perspective*, Vol. 11, No. 3, pp.11–23, DOI: 10.1177/097226290701100303.

Shynkevich, Y., McGinnity, T.M., Coleman, S.A., Belatreche, A. and Li, Y. (2017) 'Forecasting price movements using technical indicators: investigating the impact of varying input window length', *Neurocomputing*, Vol. 264, pp.71–88, DOI: 10.1016/j.neucom.2016.11.095.

Singh, A., Gupta, P. and Thakur, N. (2021) 'An empirical research and comprehensive analysis of stock market prediction using machine learning and deep learning techniques', *IOP Conference Series: Materials Science and Engineering*, Vol. 1022, No. 1, p.12098, DOI: 10.1088/1757-899X/1022/1/012098.

Sun, J., Jia, M. and Li, H. (2011) 'AdaBoost ensemble for financial distress prediction: an empirical comparison with data from Chinese listed companies', *Expert Systems with Applications*, Vol. 38, No. 8, pp.9305–9312, DOI: 10.1016/j.eswa.2011.01.042.

Torgo, L. and Ribeiro, R. (2009) 'Precision and recall for regression', in *Discovery Science*, Vol. 5808, Springer Berlin Heidelberg, Berlin, Heidelberg, pp.332–346, DOI: 10.1007/978-3-642-04747-3_26.

Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A. (2020) 'Stock closing price prediction using machine learning techniques', *Procedia Computer Science*, Vol. 167, pp.599–606, DOI: 10.1016/j.procs.2020.03.326.

Zhang, C., Ji, Z., Zhang, J., Wang, Y., Zhao, X. and Yang, Y. (2018) 'Predicting Chinese stock market price trend using machine learning approach', *Proceedings of the 2nd International Conference on Computer Science and Application Engineering*, ACM, Hohhot China, pp.1–5, DOI: 10.1145/3207677.3277966.