Machine learning portfolios for US stock prices: Directional forecasting before and during the COVID-19 pandemic

Portafolios de machine learning para precios accionarios estadounidenses: pronósticos direccionales antes y durante la pandemia del COVID-19

Adán Reyes Santiago, Jaime González Maíz Jiménez[*]

Tecnológico de Monterrey, México
Universidad de las Américas, México

Abstract

In this study, we evaluate the performance of five Machine Learning (ML) portfolios—logistic regression, random forest, decision tree, gradient boosting, and adaptive boosting—against that of the Dow Industrial Average -passive approach. We consider an active portfolio management approach, employing out-of-sample backtesting to simulate the strategy performance as a categorical approach. We employ as predictors the opening price, the highest price, the lowest price, the closing price, the Williams %R and a the 13-week T-bills. During the whole period, before COVID-19, and during the pandemic, in all cases, at least one ML portfolio beats the index. These results suggest that overall, investors obtain positive outcomes if they use ML portfolios instead of investing passively in the index, obtaining the most benefits in times of greater uncertainty, such as the peak of the pandemic.

[*] Corresponding author.
E-mail address: jaime.gonzalezmaiz@udlap.mx (J. González Maíz Jiménez).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

Resumen

En este estudio, evaluamos el rendimiento de cinco portafolios de Machine Learning (ML) (regresión logística, bosque aleatorio, árbol de decisión, aumento de gradiente y aumento adaptativo) en comparación con un enfoque pasivo del Dow Industrial Average. Consideramos un enfoque de gestión de portafolios activo, empleando pruebas de backtesting fuera de la muestra para simular el desempeño de la estrategia como un enfoque categórico. Empleamos como predictores el precio de apertura, el precio más alto, el precio más bajo, el precio de cierre, el %R de Williams y las T-bills a 13 semanas. Durante todo los escenarios considerados -incluido el periodo del COVID-19-, al menos un portafolio de ML fue superior al índice. Estos resultados sugieren que, en general, los inversores pueden obtener resultados positivos si utilizan carteras de ML, obteniendo los mayores beneficios en momentos de mayor incertidumbre, como el pico de la pandemia.

Código JEL: C63; G15; G17
Palabras clave: machine learning; COVID-19; backtesting

Introduction

According to the geometric Brownian motion model, returns on a certain stock in successive equal periods are independent and normally distributed. In 1863, Jules Regnault, a French stockbroker, observed that the longer a security is held, the higher is the gain or loss on its price variations: the price deviation is directly proportional to the square root of time (Regnault, 1863). In 1900, Louis Bachelier, a French mathematician, in his Ph.D. thesis, "The theory of speculation," made the first attempt to predict the stock market using Brownian motion (Bachelier, 1900). Moreover, in 1915, Mitchell was the first to note that distributions of price changes are too "peaked" to be relative samples from Gaussian populations (Mitchell, 1915). In 1923, Keynes stated that investors in financial markets are rewarded not for knowing better than the market what the future has in store but rather for risk bearing (Keynes, 1923). Larson (1960) applied a new method of time-series analysis and noted that the distribution of prices is "very nearly normally distributed for the central 80 percent of the data, but there is an excessive number of extreme values."

The stock market is a dynamic system that is nonstationary and irregular in nature and is affected by several factors, such as political conditions, economic uncertainty, and financial reports. According to Rossi (2018), evaluating the predictability of stock returns requires formulating equity premium forecasts based on large sets of conditioning information; however, conventional methods fail in such circumstances. Parametric models are usually unduly restrictive in terms of functional form specification and are subject to data overfitting concerns as the number of estimated parameters increases. In contrast, linear models reduce the dimensionality of the forecasting problem, although these methods do not

consider large portions of the conditioning information set, thereby reducing the accuracy of the forecasts. According to Vijh et al. (2020), two main approaches are used to predict stock prices: technical analysis, which uses historical stock prices to predict future stock prices, and fundamental analysis, which is based on external factors, such as news articles and economic information. Currently, advanced intelligent techniques use either technical or fundamental analysis to predict stock prices.

According to de Prado (2018), ML is changing virtually every aspect of our lives. Currently, ML algorithms accomplish tasks that until recently only expert humans could perform. Hence, it is an exciting time to adopt a disruptive technology that can transform how people invest for generations. Thus, this study aims to evaluate the performance of five ML portfolios against that of investing passively in the Dow Jones Index (DJI).

The contributions of this study are threefold. First, we provide a categorical perspective, applying the five machine learning tools to design portfolios using both a technical and a fundamental indicator as predictors. Second, we compute actual gains/losses in US dollars by applying backtesting, adjusting the performance of each portfolio, with the proposed compensation risk ratio. Third, we compare the performance of the ML portfolios versus investing passively in the Dow Jones Index (DJI) - benchmark-, dividing the results into four parts. The first part is the whole sample. Then we consider the period before the pandemic. Finally, we split the COVID-19 period into two subsamples: the high volatility span and the low volatility interval. The results indicate that investors can indeed reap the benefits using ML portfolios instead of investing passively in the index, mainly during times of greater uncertainty, such as the peak of the COVID-19 pandemic.

Regarding related studies, Islam and Nguyen (2020) contended that White (1988) was the first to conduct a significant study of neural network models for stock price prediction using IBM's daily common stock, although the training predictions were very optimistic. Qiu et al. (2012) developed a new forecasting model based on fuzzy time series and C-fuzzy decision trees to predict the stock index of the Shanghai Composite Index. Hassan et al. (2007) proposed a fusion model by combining the hidden Markov model, Artificial Neural Network (ANN), and genetic algorithms to forecast financial market behavior and found that the performance of this fusion was better than that of the basic model. Zhang and Wu (2008) predicted the S&P 500 index using an integrated model of improved bacterial chemotaxis optimization and a backpropagation artificial network, which was better and computationally less complex. Merh et al. (2010) applied a three-layer feed-forward neural network model and an Autoregressive Integrated Moving Average (ARIMA) model to predict the stock price value and demonstrated that the ARIMA models performed better than the ANN models. Adebiyi et al. (2014) analyzed the forecasting performance of Dell's stock price and found that the neural network model was superior to the ARIMA model. Rathnayaka et al. (2014) examined the Colombo Stock Exchange and

documented that the geometric Brownian motion model had a higher forecast accuracy than the traditional ARIMA model. Khare et al. [17] investigated the prices of 10 unique stocks listed on the New York Stock Exchange and reported that feed-forward multilayer perception performs better than long short-term memory in predicting the short-term prices of a stock. Agustini et al. (2018) built predictive models with Brownian motion using the Jakarta Corporate Index and found a mean absolute percentage error (MAPE) of less than 20%.

Using ML algorithms, Chatzis et al. (2018) provided significant evidence of interdependence and cross-contagion effects among stock, bond, and currency markets. Zhong and Enke (2019) reported that deep neural networks on the S&P 500 ETF were better than the two other standard benchmarks in predicting the daily direction of future stock returns. Kim et al. (2020) integrated time-varying effective transfer entropy to predict stock price direction. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) as criteria, Vijh et al. (2020) demonstrated that both ANN and random forest techniques are efficient in predicting stock closing prices. Bhardwaj and Bangia (2020) proposed a multivariate adaptive regression spline and M5 prime regression tree to predict significant statistics for stocks listed on the S&P BSE. Jaggi et al. (2021) applied fin ALBERT and ALBERT ML models to predict the stock prices of 25 companies.

Ellington et al. (2022) used ML methods and found that oil, telecommunications, and finance were leading indicators for other industries. Grayling et al. (2022), using data from Twitter and applying three different ML algorithms, found that emotions derived from stock market-related tweets were significant predictors of stock market movements. Posatto (2022) created long–short investment strategies by applying ML out-of-sample predictions, achieving an annual return of 26.4% with a Sharpe ratio of 0.50. Rubesam (2022) concluded that ML models successfully forecast stock returns in the Brazilian market and proposed a method to combine ML portfolios while balancing their risk contributions. Researchers have usually tried to predict the direction of the stock price using sophisticated approaches, including Zhong and Enke (2019) and Kim et al. (2020), but they applied RMSE and MAPE to measure forecast accuracy. This is the first study of this kind that compares index-weighted ML portfolios with US shares versus investing passively in the DJI by applying the compensation risk ratio as a metric, discerning results between the periods before and during the COVID-19 pandemic, dividing the latter into two subsamples: low volatility and high volatility, and using both a technical and a fundamental indicator as predictors. Our findings indicate that investors can indeed reap benefits if they invest actively in ML portfolios instead of only investing passively in the index. Rubesam (2022) used an equal risk contribution approach for portfolios, and we apply index-weighted portfolios. On the other hand, Grayling et al. (2022) used a categorical approach, similar to the present study, but did not apply backtesting or use the compensation risk ratio, as we do. Furthermore, Liu et al. (2023) analyze SMEs using a similar approach

of that of our paper applying RF, DNN, GBDT and Adaboost models forecasting the next day's closing price, although they employ R^2, RMSE, and MAPE as performance metrics. Alsayed (2023) using a Developed ML algorithm named elastic-net regression, find that both the COVID-19 pandemic and Russian invasion have an impact on the Turkish Stock Market.

The remainder of this paper is organized as follows. The next section presents the materials and methods. The third section describes the results. Finally, we present our discussion in the fourth section.

Materials and methods

Logistic regression represents a traditional model used in statistics and thus represents a good comparison to evaluate versus the other ML models. On the other hand, even though decision tree is a basic ML model, it is worth comparing its performance versus the other more complex ML models. Furthermore, random forests represent the "ensemble learning" approach through "bagging." Finally, gradient boosting and adaptive boosting both represent "ensemble learning" but through the "boosting" method, where the former learns through "errors," and the latter modifies the weights on the data points.

Logistic regression

Classification techniques are an essential part of machine learning and data mining applications. Approximately 70% of problems in data science are classification problems. There are many classification problems that are available, but logistic regression is common and is a useful regression method for solving binary classification problems. Logistic regression is one of the simplest and most commonly used ML algorithms for two-class classification. It is easy to implement and can be used as the baseline for any binary classification problem.

Logistic regression predicts the probability of occurrence of a binary event utilizing a logit function (formulas 1 and 2), where the dependent variable follows a Bernoulli distribution, and the estimation is done through maximum likelihood. In our study, the independent variables are both the technical and the fundamental indicators; whereas the dependent variable is the closing price of the next day.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

(1)

where y is a dependent variable and X_1, X_2.....and X_n are explanatory variables.

Apply the sigmoid function to the linear regression:

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2+x_2+\cdots+\beta_n+x_n)}}$$

(2)

Decision Tree

The goal of machine learning is to decrease uncertainty or disorders from the dataset, and for this, we can use decision trees. Decision trees are upside down, which means the root is at the top, and then this root is split into various nodes. Decision trees are a bunch of if-else statements in layman terms. It checks if the condition is true and if it is, then it goes to the next node attached to that decision (Fig. 1). The algorithm stops splitting using the metric called "entropy," which is the amount of uncertainty in the dataset. In a decision tree, the output is mostly "yes" or "no" (Saini, 2022), and specifically applied to our research is whether stock prices in the next period go "up" or "down".

The formula (3) for entropy is:

$$E(S) = -p_{(+)}\log p_{(+)} - p_{(-)}\log p_{(-)}$$

(3)

where $p_{(+)}$ is the probability of positive cases (prices going up)

$p_{(-)}$ is the probability of negative cases (prices going down)
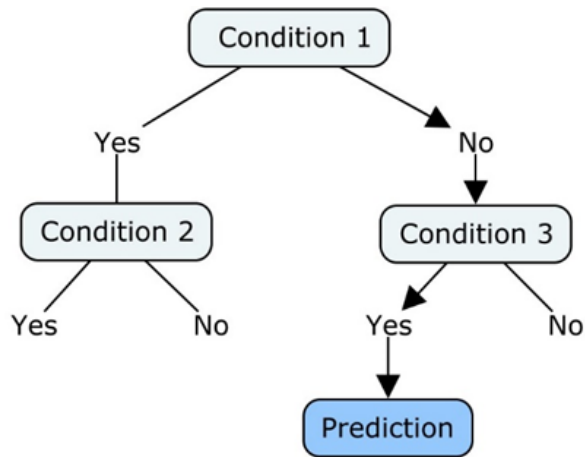
S is the subset of the training dataset

Figure 1. Decision tree model
Source: own creation

Random forest

Random forest is a supervised learning algorithm. It can be used for both classification and regression. It is also a flexible and easy to use algorithm. A forest is composed of trees. It is said that the more trees there are, the more robust a forest is. Random forests create decision trees and randomly selected data samples, obtain predictions from each tree and select the best solution by means of voting (Fig. 2). It also provides a fairly good indicator of the feature importance (Navlani, 2022).
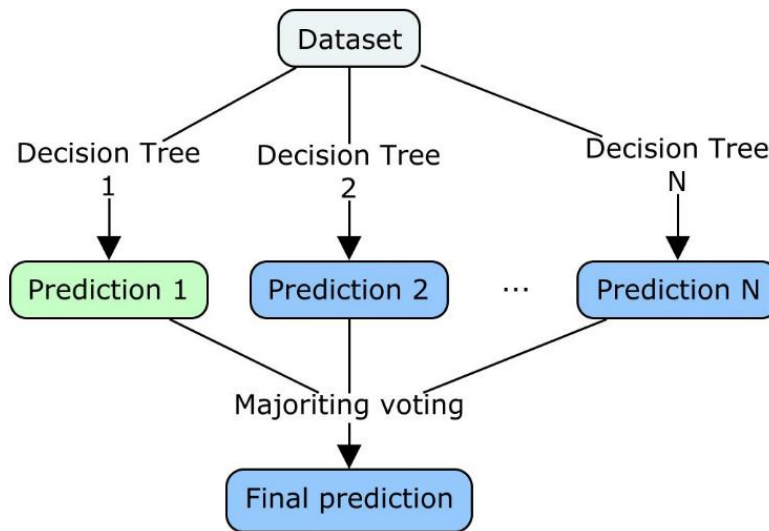

Figure 2. Random Forest model.
Source: Author's own

Gradient boosting

In machine learning, there are two types of supervised learning problems: classification and regression. Classification refers to the task of providing machine learning algorithm features. Features are the inputs that are given to the machine learning algorithm, and the inputs are used to calculate an output value. In our study, the features given are the directions of the opening price, the highest price, the lowest price, and the closing price of the previous day to predict the direction of the closing price of the following day. In particular, the predictions are made through majority vote (weights), with the instances being classified according to which class receives the most votes (weights) (Fig. 3). The objective of gradient boosting

classifiers is to minimize the loss, or the difference between the actual class value of the training dataset and the predicted class value (Nelson, 2022).
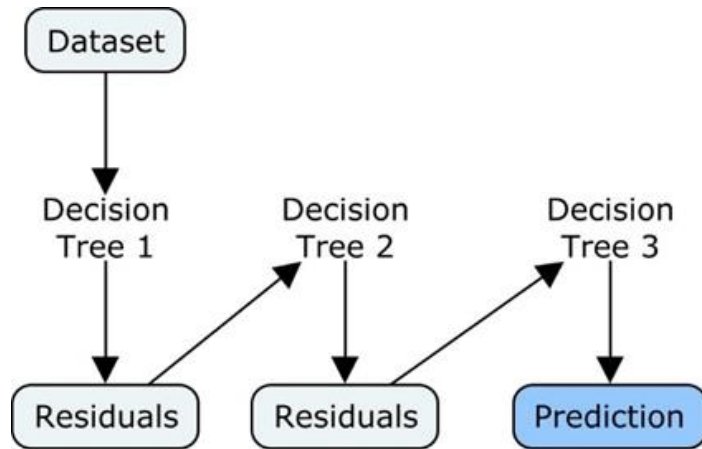

Figure 3. Gradient boosting model
Source: Author's own

AdaBoost

Ada Boost is short for Adaptive Boosting, which is an ensemble algorithm that combines many weak learners (decision trees) and turns it into one strong learner. Thus, its algorithm leverages begging and boosting methods to develop and enhance predictors.

AdaBoost is similar to random forest in the sense that the predictors are taken from many decision trees. However, there are three main differences that make AdaBoost unique. First, AdaBoost creates a forest of stumps rather than trees. A stump is a tree that is made of only one node and two leaves (Fig. 4). Second, the stumps that are created are not equally weighted in the final decision (final prediction). Stumps that create more error will have less say in the final decision. Third, the order in which the stumps are made is important because each stump aims to reduce the errors that the previous stump(s) made (Shin, 2020).
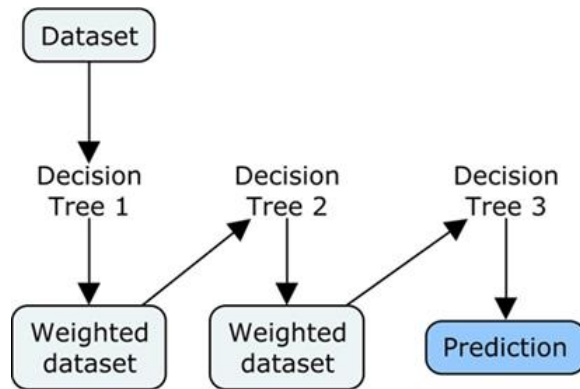
Figure 4. AdaBoost model
Source: Author's own

Backtesting

According to de Prado (2018), backtesting is the historical simulation of how a strategy would have performed had it been run over a past period of time. As such, it is a hypothetical analysis and by no means an experiment. First, we employ backtesting of the ML portfolios to analyze the behavior of the portfolios before and during the pandemic comparing the performance versus investing passively in the DJI. Second, we apply the Mann–Whitney test to check whether, in a pair of portfolios, one is significantly better than another.

Fundamental indicator

Fundamental analysis evaluates stocks by attempting to measure their intrinsic value. Fundamental analysts study everything from the overall economy and industry conditions to the financial strength and management of individual companies. Specifically, interest rates affect the value of all assets according to their appropriate level of systematic or nonsystematic risk. In particular, we use the daily price changes of the 13-week Treasury Bill as our fundamental indicator predictor.

Technical indicator

Technical analysis differs from fundamental analysis in that traders attempt to identify opportunities by looking at statistical trends, such as movements in a stock's price and volume. Technical analysts do not attempt to measure a security's intrinsic value. Instead, they use stock charts to identify patterns and trends that suggest what a stock will do in the future (Majaski, 2022). In particular, we apply the Williams % R

to this research, which was developed by Larry Williams and compares a stock's closing price to the high-low range over a specific period; in our case, we use an optimal 11-day period, and employing a modified version of the quotient where if the result is above 0.50 the stock is oversold; and if it is below 0.50 the stock is overbought. (Formula (4)).

$$Williams\%R = \frac{Highest\ High - Close}{Highest\ High - Lowest\ Low}$$

(4)

where

Highest High = Highest price in the lookback

Close = Most recent closing price

Lowest low = Lowest price in the lookback

Our study considers only two outcomes—prices either increase or decrease—and allows both long and short positions. It includes six predictor variables—the opening price, the highest price, the lowest price, the closing price, the Williams % R as a technical indicator and the daily price change of the 13-week T-bills as a fundamental indicator—and one predicted variable—the closing price of the following day. In other words, if the price actually increases (long position) the following day and the ML algorithm predicted it correctly, it is a positive outcome; if the price decreases (short position) the following day and the ML algorithm predicted it correctly, it is a positive outcome as well. In particular, we examine the 30 Dow component stocks (Table 1). On March 11, 2020, COVID-19 was declared a pandemic by the World Health Organization [37]. We consider four scenarios for stock prices: i) the whole period from April 1, 2019 to October 1, 2020; ii) a pre-COVID sample from April 1, 2019, to September 30, 2019; iii) a high volatility subsample of the COVID period from October 1, 2019, to March 30, 2020; and finally, iv) a low volatility subsample of the COVID period from April 1, 2020, to September 30, 2020. For the whole period there are a total of 380 observations, and for the remaining scenarios there are 126 observations for each one. Partitioning the data into 75/25 (training/test), out-of-sample forecast performance is measured considering the last 95 observations -for the whole sample-, and 31 observations for the remaining periods, which amount to 25% of the total. Concerning backtesting, for scenario i) the period goes from May 18 of 2020 to September 30 of 2020; for scenario ii) backtesting was carried out from August 15, 2019, to September 30, 2019; for scenario iii) it goes from February 18 of 2020 to March 31 of 2020, and for scenario iv) from August 19 of 2020 to September 30 of 2020. We consider 10,000 US dollars as the initial investment for each ML portfolio and the investment in the index. We build DJI-weighted portfolios (Table 1) for each ML model and compare the performance of these ML portfolios with that of investing passively in the DJI throughout the test period using the compensation risk ratio (5) as a performance metric by backtesting.

Compensation risk ratio

This ratio is somewhat related to the Sharpe ratio – which was introduced by (Sharpe 1966) and was a measure for the performance of mutual funds and proposed the term reward-to-variability ratio, which is the quotient of dividing the average of asset's return by the standard deviation of the same asset for a determined period of time, and the interpretation is straightforward, the greater the quotient the better, that is, the more reward the investor gets for bearing the asset's risk. Following the same line of reasoning, the compensation risk ratio (formula 5), instead of having in the numerator the average return of the asset, it has the average cumulative gains of the portfolio, and the denominator remains the same. Hence, as with the Sharpe ratio, the greater the quotient the better, but instead of applying it to historical data, we are employing the measure to gauge the out-of-sample performance of portfolios.

The more positive the quotient of this ratio, the greater the compensation for risk. The schematic representation of portfolio creation using ML techniques is shown in Fig. 5.

$$C = \frac{\frac{\sum_1^n G/L}{n}}{\sigma}$$

(5)

where $\sum$ is out-of-sample cumulative gains/losses for each portfolio and σ is the standard deviation of cumulative gains/losses for each portfolio.
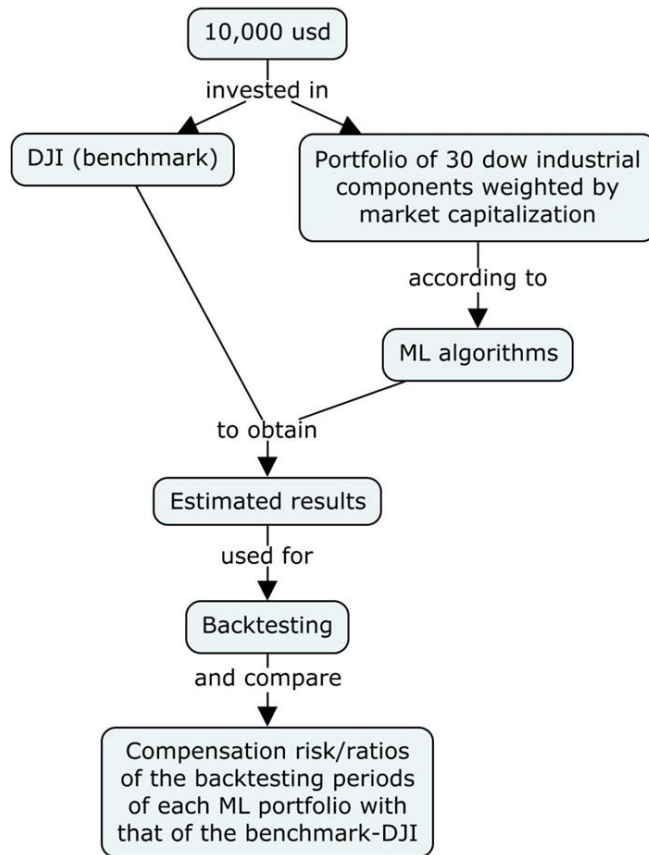
Figure 5. Schematic representation of portfolio creation
Source: Author's own

Table 1
Stocks' weights into Dow Jones

Ticker	Weights (%)	Ticker	Weights (%)	Ticker	Weights (%)
AAPL	2.84	GS	7.36	MRK	2.1
AMGN	5.48	HD	6.27	MSFT	4.88
AXP	3.02	HON	4.17	NKE	2.13
BA	3.36	IBM	2.86	PG	2.86
CAT	4.52	INTC	0.57	TRV	3.62
CRM	2.82	JNJ	3.43	UNH	10.29
CSCO	0.96	JPM	2.61	V	4.16
CVX	3.5	KO	1.22	VZ	0.73
DIS	1.89	MCD	5.24	WBA	0.79
DOW	0.98	MMM	2.41	WMT	2.94

Source: Wikipedia (2022)

Results

Figures 6 through 9 show the percentage returns of the Dow Jones Index for the four periods considered in this study, where the red line splits the graphs into two parts, the training part (75% of data) and the test part (25% of data) (right-hand side), where backtesting occurs. Taking into consideration the whole sample, the standard deviation of the percentage returns for the backtesting period is 1.44% (Fig. 6). In the precovid period, the standard deviation of the percentage returns of the backtesting period is 0.80% (Fig. 5), whereas during the high volatility subsample associated to COVID-19, the standard deviation explodes to 5.50% (Fig. 6), going back to "normal" levels of 1.20% during the low volatility subsample (Fig. 7). On the other hand, Figs. 10-to 13 show the results of backtesting, including the index, and by visual inspection only. It can be seen that in all four scenarios, the DJI is the most volatile, being at its most volatile during the high volatility COVID-19 subsample. Meanwhile, Table 2 shows the cumulative gains, the average of cumulative gains, standard deviation, and compensation risk ratio of each portfolio for all scenarios. Furthermore, in Table 3, we show both the mean and p values for the Mann–Whitney test.
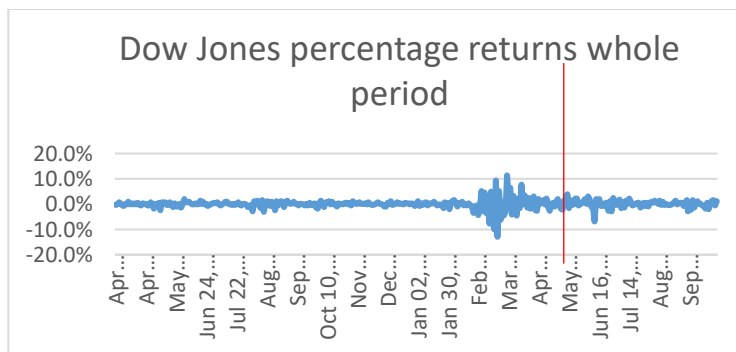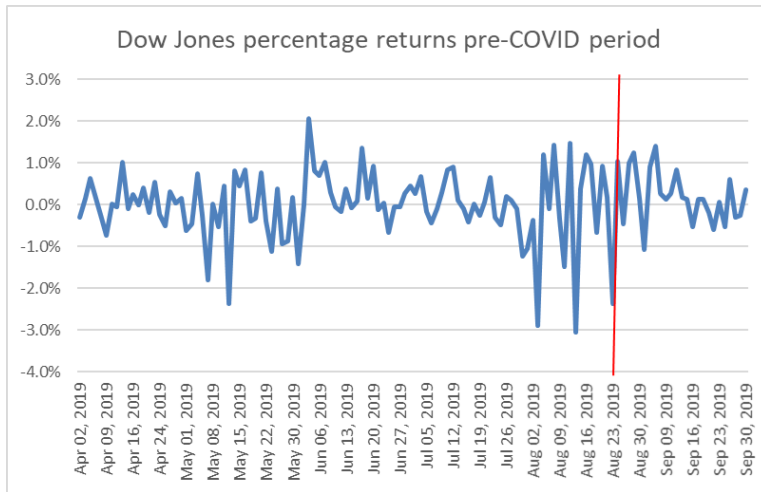

Figure 6
Source: Own elaboration
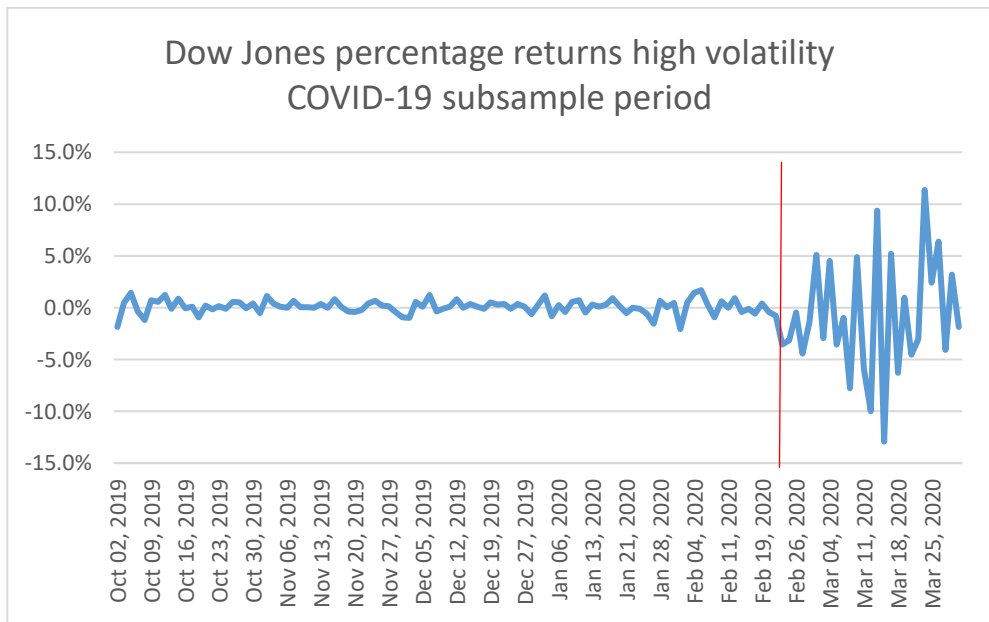
Figure 7
Source: Own elaboration
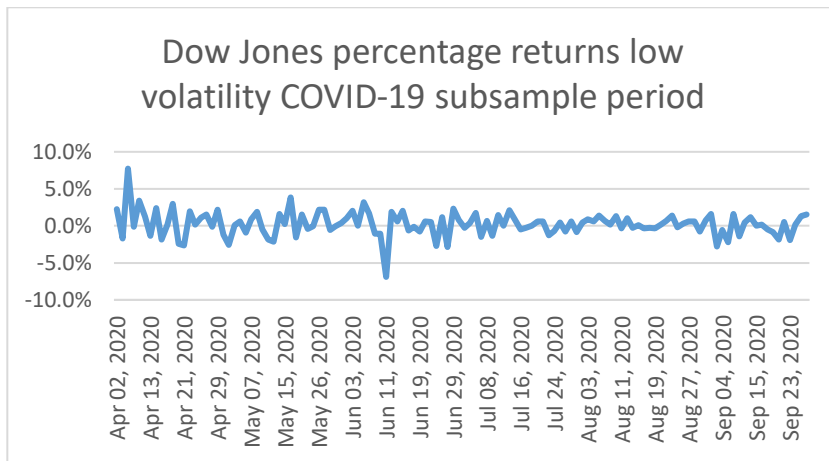

Figure 8
Source: Own elaboration
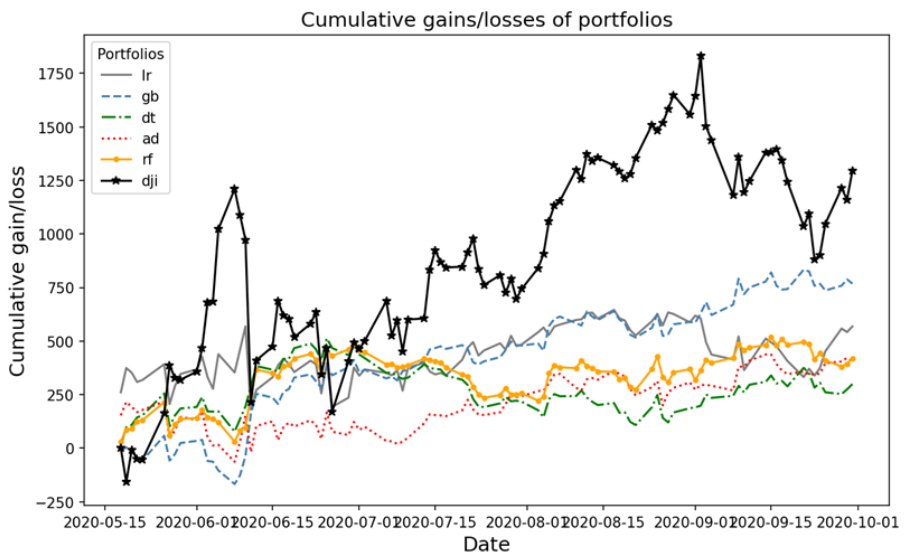
267

Figure 9
Source: Own elaboration

Figure 10
Source: Our own estimations using Python version 3.9
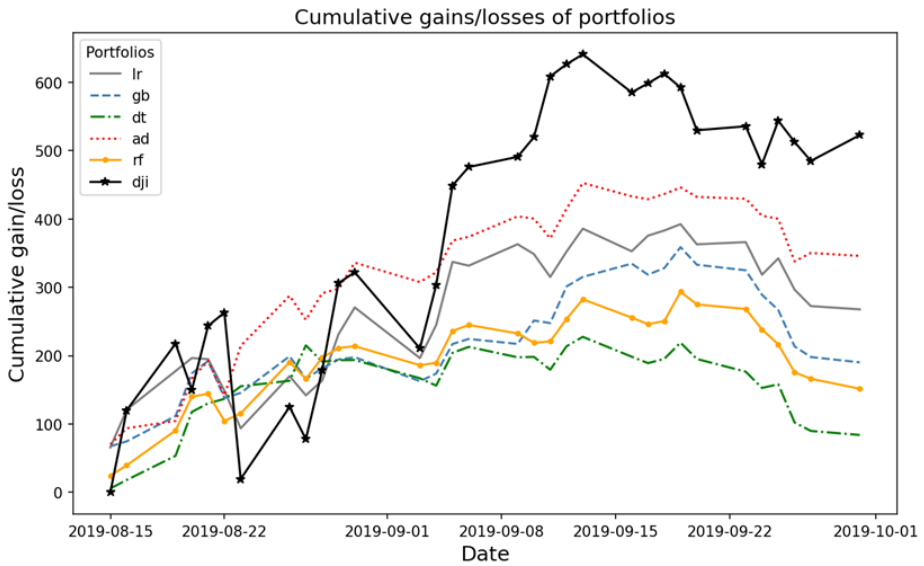
Figure 11

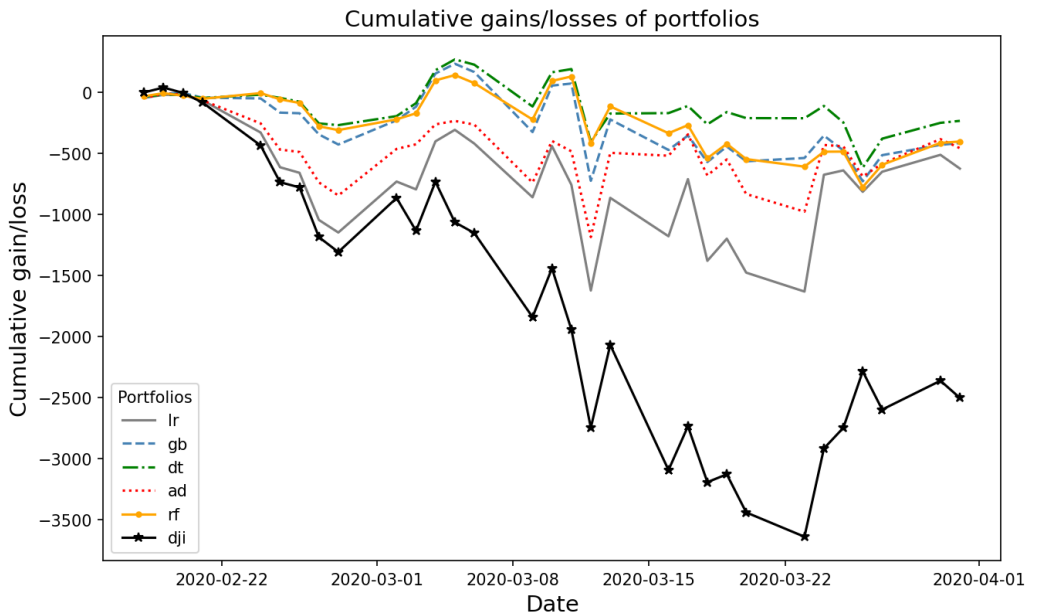Source: Our own estimations using Python version 3.9

Figure 12

Source: Our own estimations using Python version 3.9
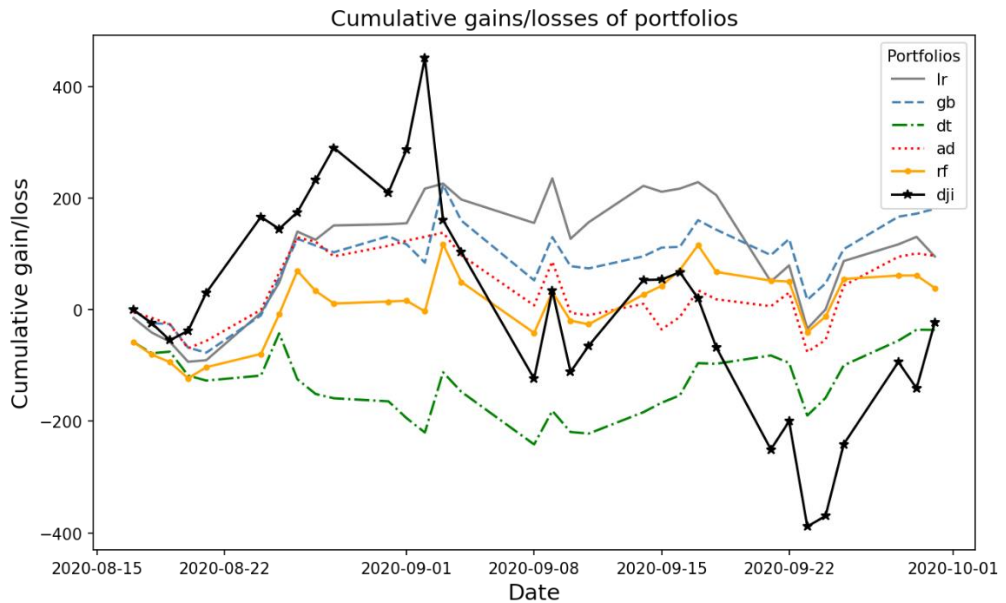
Figure 13
Source: Our own estimations using Python version 3.9

Table 2
Cumulative gains/losses, average cumulative gains/losses, standard deviation, and compensating risk ratio of ML portfolios and DJI

Metric/ Portfolio	LR	GB	DT	RF	AB	DJI
Whole Period						
CG/L	568	768	299	420	403	1,295
ACG/L	439	424	265	331	212	886
SD	116	266	105	123	124	453
CRR	3.77	1.60	2.52	2.69	1.71	1.95
Before Covid						
CGL	428	233	24	-117	258	523
ACG/L	367	183	18	-74	237	398
SD	150	48	34	70	55	187
CRR	2.45	3.85	0.53	-1.06	4.30	2.13
During high volatility of COVID						
CG/L	-1,834	735	875	866	129	-2,502
ACG/L	-1,302	-158	-9	-100	-421	-1,804
SD	793	568	508	545	438	1,079
CRR	-1.64	-0.28	-0.02	-0.18	-0.96	-1.67
During low volatility of COVID						
CG/L	74	-146	-246	-270	41	32
ACG/L	63	-157	-229	-237	11	66

SD	84	64	69	84	55	193
CRR	0.76	-2.46	-3.31	-2.82	0.19	0.34

Table 2. Cumulative gains/losses, average cumulative gains/losses, standard deviation, and compensating risk ratio of ML portfolios and DJI

CG/L = Cumulative gains/losses; ACG/L = Average Cumulative gains/losses; SD = Standard Deviation; CRR = Compensation Risk Ratio; LR = Logistic Regression; GB = Gradient Boosting; DT = Decision Tree; RF = Random Forest; AB = Adaptive Boosting; DJI = Dow Jones Index

Source: Our own estimations using Python version 3.9

Table 3

Means and p values of the Mann–Whitney test for the difference of distribution of cumulative gains/losses

Model		Whole Period			Before COVID-19			Low Volatility		
1	2	M1	M2	M-W	M1	M2	M-W	M1	M2	M-W
LR	GB	439	424	p = 0.49	367	183	p < 0.01	106	84	p = 0.24
LR	DT	439	265	p < 0.01	367	18	p < 0.01	106	-134	p < 0.01
LR	RF	439	331	p < 0.01	367	-74	p < 0.01	106	9	p < 0.01
LR	AB	439	212	p < 0.01	367	237	p < 0.01	106	35	p < 0.01
LR	DJI	439	886	p < 0.01	367	398	p = 0.37	106	66	p = 0.27
GB	DT	424	265	p < 0.01	183	18	p < 0.01	84	-134	p < 0.01
GB	RF	424	331	p < 0.01	183	-74	p < 0.01	84	9	p < 0.01
GB	AB	424	212	p < 0.01	183	237	p < 0.01	84	36	p < 0.01
GB	DJI	424	886	p < 0.01	183	398	p < 0.01	84	66	p < 0.01
DT	RF	265	331	p < 0.01	18	-74	p < 0.01	-134	9	p < 0.19
DT	AB	265	212	p < 0.01	18	237	p < 0.01	-134	84	p < 0.01
DT	DJI	265	886	p < 0.01	18	398	p < 0.01	-134	66	p < 0.01
RF	AB	331	212	p < 0.01	-74	237	p < 0.01	9	35	p < 0.01
RF	DJI	331	886	p < 0.01	-74	398	p < 0.01	9	66	p = 0.15
AB	DJI	212	886	p < 0.01	237	398	p < 0.01	35	66	p = 0.57

Source: Our own estimations using Minitab version 20

In particular, during the whole period, in terms of cumulative gains/losses (CG/L) (Table 2) the DJI is the best with 1,295 US dollars, followed by gradient boosting with 768 US dollars, the logistic regression with 568 US dollars, the random forest with 420 US dollars, the adaptive boosting with 403 US dollars, and the last place is occupied by the decision tree with 299 US dollars. While, Fig. 10 shows that the Index is more volatile than the ML models, which is established when calculating the standard deviation (SD) of cumulative gains/losses. The standard deviation of the DJI is the greatest at 453 US dollars, followed by the gradient boosting at 266 US dollars, representing 59% of the index volatility; the adaptive boosting model at 124 US dollars, representing 27% of the index risk; the random forest at 123, representing 27% of the index variability; then the logistic regression at 116 US dollars, representing 26% of the index volatility; the least volatile is the decision tree model at 105 US dollars with 23% of the variability represented by the DJI. Although, when adjusting the cumulative gains/losses with the compensation risk ratio -Table2-, the logistic regression model is the best, with 3.77 units, followed by the random forest with 2.69 units, the decision tree model with 2.52 units, the DJI with 1.95 units, the adaptive boosting with 1.71 units; the worst is the gradient boosting with 1.60 units. That is, even though

271

the logistic regression model does not have the greatest cumulative gains, is the best compensating the risk for the investor. Moreover, we use the Mann-Whitney test, which does not require the normal distribution assumption. The results in Table 3 indicate that portfolio managers can indeed benefit from difference in terms of performance using different portfolios, except the logistic regression-gradient boosting pair of portfolios that are not independent from each other.

In the pre-COVID-19 period, in terms of cumulative gains (Table 2). The index performs best with 523 US dollars, followed by the logistic regression model with 428 US dollars, the adaptive boosting with 258 US dollars, the gradient boosting with 233 US dollars, the decision tree model with 24 US dollars; and the random forest model with a loss of 117 US dollars. Meanwhile, Fig. 11 shows that the Index is more volatile than the ML models, which is confirmed when computing the standard deviation of cumulative gains/losses. The standard deviation of the index is largest at 187 US dollars, followed by the logistic regression model at 150 US dollars, representing 80% of the index volatility; the random forest model at 70 US dollars, representing 37% of the index risk; the adaptive boosting at 55 US dollars, representing 29% of the index variability; then the gradient boosting at 48 US dollars, representing 25% of the index volatility; the least volatile is the decision tree model at 34 US dollars with 18% of the volatility represented by the index. Nevertheless, when we adjust the cumulative gains/losses using the proposed metric of compensating risk ratio -Table 2-, the adaptive boosting performs best, with 4.30 units, followed by the gradient boosting, with 3.85 units, the logistic regression, with 2.45 units, the DJI, with 2.13 units, and the decision tree model, with 0.53 units; the worst performer is the random forest model, with -1.06 units. In other words, although the adaptive boosting model is not with the greatest cumulative gains, is the best in terms of compensating the risk for the investor. Furthermore, the Mann-Whitney test indicates that all portfolios are independent form each other, except the DJI-logistic regression model pair.

During the high volatility subsample of the pandemic, concerning cumulative gains (Table 2), the best performer is the decision tree model with a cumulative gain of 875 US dollars, followed by the random forest model with 866 US dollars, the gradient boosting with 735 US dollars, the adaptive boosting with 129 US dollars, the logistic regression model with a loss of 1,834 US dollars; the worst performer is the DJI with a cumulative loss of 2,502 US dollars. Fig. 12 clearly shows that the Index is more volatile than the ML models. This finding is confirmed when computing the standard deviation of cumulative gains/losses. The standard deviation of the DJI is 1,079 US dollars, followed by the logistic regression model with 793 US dollars, representing 74% of the index volatility; the gradient boosting model with 568 US dollars, representing 53% of the DJI volatility; the random forest model with 545 US dollars, representing 51% of the index volatility; the decision tree model with 508 US dollars, representing 47% of the volatility represented by the DJI; and the least volatile is the adaptive boosting model with 438 US dollars, i.e. 41% of the index's volatility. Furthermore, when we adjust the cumulative gains/losses using

the proposed metric of compensation risk ratio -Table 2-, the decision tree model performs best, with -0.02 units, followed by the random forest with -0.18 units, the gradient boosting with -0.28 units, the adaptive boosting with -0.96 units, and the logistic regression model with -1.64 units; the worst performer is the DJI with -1.67 units. In other words, the decision tree model is the best in both cumulative gains and the best in terms of compensating the risk for the investor. Moreover, by applying the Mann-Whitney test (Table 3), indicate that portfolio managers can indeed benefit from difference in terms of performance using different portfolios, except the following three pairs: decision tree-random forest, gradient boosting-adaptive boosting, and random forest-gradient boosting, which are not independent from each other.

Throughout the low volatility span of the COVID-19 period, regarding cumulative gains, the best performer is the logistic regression model with a cumulative gain of 74 US dollars, followed by the adaptive boosting with 41 US dollars, the DJI with 32 US dollars, the gradient boosting with a loss of 146 US dollars, and the decision tree with a cumulative loss of 246 US dollars; the worst performer is the random forest with a loss of 270 US dollars. Fig. 13 again shows that the DJI has more variability than the ML models. This is affirmed when computing the standard deviation. The highest belong to the Index with 193 US dollars, followed by the random forest and logistic regression models both with 84 US dollars, representing 44% of the DJI volatility; the decision tree model with 69 US dollars, representing 36% of the index volatility; the gradient boosting with 64 US dollars, representing 33% of the DJI variability; and the least volatile is the adaptive boosting with 55 US dollars, with only 29% represented by the index. On the other hand, when adjusting the cumulative gains/losses using the compensation risk ratio, the logistic regression model performs best, getting 0.76 units, followed by the DJI with 0.34 units, the adaptive boosting with 0.19 units, the gradient boosting with -2.46 units, the random forest model with -2.82 units, and the decision tree model with -3.31 units. That is, during the span of low volatility of COVID-19, the logistic regression model is the best in terms of compensating the risk for the investor. Furthermore, when applying the Mann-Whitney test (Table 3), portfolio managers can reap benefits from difference regarding performance of portfolios except the following pairs: decision tree-random forest, DJI-logistic regression, and DJI-adaptive boosting.

Discussion

In light of Rossi's (2018) observation that investors contend with an ever-expanding deluge of information, data, and statistics, the task of forecasting stock returns becomes increasingly challenging. The necessity to formulate equity premium forecasts based on extensive sets of conditioning information underscores the complexity of this endeavor. In this study, we present a model with a simplified interpretative framework, focusing on just two outcomes: whether prices rise or fall. Our analysis encompasses the

periods before and during the COVID-19 pandemic, providing a comparative lens.

We assess the performance of actively investing in the 30 down components, weighted by market capitalization, across five machine learning portfolios. This evaluation is juxtaposed against passive investment in the Dow Jones Industrial Average (DJI). Employing backtesting and computing the compensation risk ratio during the test period, we unveil noteworthy insights.

The findings of this study underscore the resilience and superiority of the machine learning portfolios in all four examined time spans, consistently outperforming the DJI. This phenomenon is most pronounced during the high volatility phase of COVID-19, suggesting that in times of heightened uncertainty, employing more sophisticated investment strategies, such as the machine learning portfolios in this study, can be advantageous.

One notable departure from recent research is our use of the compensating risk ratio as a performance metric, as opposed to the conventional RMSE and MAPE. This approach enhances interpretability and practicality for both academics and practitioners, with direct implications for risk management practices and portfolio investment decisions.

However, it's essential to acknowledge certain limitations within this study. Firstly, it is confined to the US stock market. Secondly, the comparative analysis is limited to five market capitalization-weighted machine learning portfolios versus the DJI. Thirdly, the study focuses on a single time horizon, specifically before and during the pandemic, with the COVID-19 period bifurcated into two distinct subsamples. There remains considerable room for extending this methodology to other global financial markets and contexts, opening doors to further exploration and insight.

References

Adebiyi, A.A., Adewumi, A.O., Ayo C.K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics., 1–7. https://doi.org/10.1155/2014/614342.

Alsayed, A. (2023). Turkish stock market from pandemic to Russian invasion, evidence from developed machine learning algorithm. Computational Economics Journal. 62, 1107-1123 https://doi.org/10.1007/s10614-022-10293-z

Agustini WF, Affianti I.R., Putri E.R.M. (2018). Stock price prediction using geometric Brownian motion. J Phys Conf Ser., 974:012047. https://doi.org/10.1088/1742-6596/974/1/012047

Bachelier, L. (1900). Théorie de la spéculation. Ann Sci École Norm Super, 17, 21–86. https://doi.org/10.24033/asens.476

Bhardwaj, R., Bangia, A. (2020) Assessment of stock prices variation using intelligent machine learning techniques for the prediction of BSE. In: Dutta D, Mahanty B, editors. Numerical optimization in engineering and sciences. Advances in intelligent systems and computing. Singapore: Springer, pp 159–66. https://doi.org/10.1007/978-981-15-3215-3_15

Brown, R. (2009) XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Philos Mag 4:161-173. doi:10.1080/14786442808674769.

Chatzis, S.P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert Systems Applications, 112:353–71. https://doi.org/10.1016/j.eswa.2018.06.032

de Prado, M.L. (2018) Advances in financial machine learning. Hoboken, New Jersey: John Wiley & Sons. ISBN: 978-1-119-48208-6.

Ellington, M., Stamatogiannis, M.P., Zheng, Y. (2022). A study of cross-industry return predictability in the Chinese stock market. International Review of Financial Analysis., 83:1–15. https://doi.org/10.1016/j.irfa.2022.102249.

Grayling, T., Rossouw, S., Dinitri, S. (2022). The prediction of intraday stock market movements in developed & emerging markets using sentiment and emotions from Twitter. Finance India, 36:907-939.

Hassan, M.R., Nath, B., Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. Expert Systems with Applications., 33:171–80. https://doi.org/10.1016/j.eswa.2006.04.007

IBM (2022) What is logistic regression? https://www.ibm.com/topics/logistic-regression. Accessed 5 Oct 2022.

Islam, M.R., Nguyen, N. (2020) Comparison of financial models for stock price prediction. J Risk Financial Management., 13:181. https://doi.org/10.3390/jrfm13080181

Jaggi, M., Mandal, P., Narang, S., Naseem, U., Khushi, M. (2021). Text mining of stocktwits data for predicting stock prices. Applied System Innovation., 4(1), 13. https://doi.org/10.3390/asi4010013.

Keynes, J.M. (1923). A tract on monetary reform. London; MacMillan and Co., Limited, 1923.

Khare, K., Darekar, O., Gupta, P., Attar, V.Z. (2017, May). Short term stock price prediction using deep learning. Paper presented at the 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), IEEE, India, 19-20.

Kim, S., Ku, S., Chang, W., Song, J.W. (2020). Predicting the direction of us stock prices using effective transfer entropy and machine learning techniques. IEEE Access, 8:111660–82. https://doi.org/10.1109/access.2020.3002174.

Larson, A.B. (1960). Measurement of a random process in futures prices. In: [editors]. Proceedings of the annual meeting (Western farm economics association). London: Palgrave Macmillan UK, pp 101–12. http://dx.doi.org/10.22004/ag.econ.136584

Liu, W., Suzuki, Y., Du, S. (2023). Forecasting the stock price listed innovative SMEs using machine learning methods based on Bayesian optimization: evidence from China. Computational Economics Journal. https://doi.org/10.1007/s10614-023-10393-4

Majaski, C. (2022). Fundamental vs. technical analysis: what's the difference? Investopedia. https://www.investopedia.com/ask/answers/difference-between-fundamental-and-technical-analysis/. Accessed 30 Dec 2022.

Merh, N., Vinod, P., Kamal, R. (2010). A comparison between hybrid approaches of ANN and Arima for Indian stock trend forecasting. Business Intelligence Journal., 3:23–43.

Mitchell, W.C. (1915). The making and using of index numbers. Bull US Bur Labor Stat, 173.

Navlani, A. (2019). Understanding random forest classifiers in python tutorial. Datacamp. https://www.datacamp.com/tutorial/understanding-logistic-regression-python/. Accessed 21 Dec 2022.

Nelson, D. (2022). Gradient boosting classifiers in python with scikit-learn. StackAbuse. https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/. Accessed 21 Dec 2022.

Prasad, V.V., Gumparthi, S., Venkataramana, L.Y., Srinethe, S., Sree, R.M.S., Nishanthi, K. (2021) Prediction of stock prices using statistical and machine learning models: a comparative analysis. The Computer Journal. 65:1338-1351. doi:10.1093/comjnl/bxab008.

Posatto, A.B. (2022) Painting the black box white: interpreting an algorithm-based trading strategy. Revista Brasileira de Financas, 20:105–38.

Qiu, W., Liu, X., Wang, L. (2012). Forecasting shanghai composite index based on fuzzy time series and improved C-fuzzy decision trees. Expert Systems Applications., 39:7680–9. https://doi.org/10.1016/j.eswa.2012.01.051

Regnault, J. (1863). Calcul des chances et philosophie de la bourse. Paris: Mallet-Bachelier et Castel, Piloy.

Rathnayaka, R.M.K.T., Jianguo, W., Seneviratna, D.M.K.N. (2014, October). Geometric brownian motion with Ito's lemma approach to evaluate market fluctuations: a case study on colombo

stock exchange. Paper presented at the 2014 international conference on behavioral, economic, and socio-cultural computing (BESC2014), IEEE, Shangai, China.

Rossi, A.G. (2018). Predicting stock market returns with machine learning. https://www.semanticscholar.org/paper/Predicting-Stock-Market-Returns-with-Machine-Rossi/fceba41a7694d2fd434f2987ea209a735c9c2f89. Accessed 20 Dec 2022.

Rubesam, A. (2022). Machine learning portfolios with equal risk contributions: evidence from the Brazilian market. Emerging Markets Review, 51:100891. https://doi.org/10.1016/j.ememar.2022.100891

Saini, A. (2021) Decision tree algorithm - a complete guide. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/. Accessed 21 Dec 2022.

Sharpe, W.F. (1966) Mutual fund performance. The Journal of Business, 39, 119-138. http://dx.doi.org/10.1086/294846

Shin, T. (2020). A mathematical explanation of AdaBoost in 5 minutes. Towards Data Science. https://towardsdatascience.com/a-mathematical-explanation-of-adaboost-4b0c20ce4382. Accessed 27 Dec 2022.

Vijh, M., Chandola, D., Tikkiwal, V.A., Kumar, A. (2020). Stock closing price prediction using machine learning techniques. Procedia Computer Science., 167:599-606. https://doi.org/10.1016/j.procs.2020.03.326

White, H. (1988, July). Economic prediction using neural networks: the case of IBM daily stock returns. Paper presented at the IEEE international conference on neural networks, IEEE, San Diego, CA, 24-27.

Wikipedia. Dow Jones industrial average. https://en.wikipedia.org/wiki/Dow_Jones_Industrial_Average. Accessed 22 Dec 2022.

World Health Organization. (2020). Director-general's opening remarks at the media briefing on COVID-19. WHO, Geneva, Switzerland.

Yiu, T. (2019) Understanding random forest. https://towardsdatascience.com/understanding-random-forest-58381e0602d2. Accessed 22 Dec 2022.

Zhang ,Y., Wu, L. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. Expert Systems Applications. 36, 8849–54. https://doi.org/10.1016/j.eswa.2008.11.028

Zhong, X., Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financial Innovation., 5, 24. https://doi.org/10.1186/s40854-019-0138-0