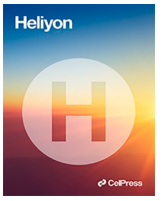
Heliyon 10 (2024) e24123

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

# How effective is machine learning in stock market predictions?

Nazif Ayyildiz [a,*], Omer Iskenderoglu [b]

[a] Harran University, Suruc Vocational School, 63800, Sanliurfa, Turkey
[b] Omer Halisdemir University, Faculty of Economics and Administrative Sciences, Department of Business, 51240, Nigde, Turkey

ARTICLE INFO

*JEL classification:*
C38
G15
G17

*Keywords:*
Financial analysis
Machine learning
Classification algorithms
Stock market indexes

ABSTRACT

In this study, it is aimed to compare the performances of the algorithms by predicting the movement directions of stock market indexes in developed countries by employing machine learning algorithms (MLMs) and determining the best estimation algorithm. For this purpose, the movement directions of indexes such as the NYSE 100 (the USA), NIKKEI 225 (Japan), FTSE 100 (the UK), CAC 40 (France), DAX 30 (Germany), FTSE MIB (Italy), and TSX (Canada) were estimated by employing the decision tree, random forest k-nearest neighbor, naive Bayes, logistic regression, support vector machines and artificial neural network algorithms. According to the results obtained, artificial neural networks were found to be the best algorithm for NYSE 100, FTSE 100, DAX 30 and FTSE MIB indices, while logistic regression was determined to be the best algorithm for the NIKKEI 225, CAC 40, and TSX indices. The artificial neural networks, which exhibited the highest average prediction performance, have been determined as the best prediction algorithm for the stock market indices of developed countries. It was also noted that artificial neural networks, logistic regression, and support vector machines algorithms were capable of predicting the directional movements of all indices with an accuracy rate of over 70 %.

## 1. Introduction

Accurate estimation of stock prices provides important information for economic planning and investment decisions. Given the trend of stock price movements, firms can plan their activities more effectively; however, investors can yield more productivity from stock trades [1]. Therefore, it is crucial for both companies and investors to be able to predict the movements in the stock markets, which are characterized by high volatility, dynamism, and complexity. On the other hand, the subject of the predictability of stock prices, which is exposed to various impacts such as macroeconomic factors, international phenomena, and human behavior, is a controversial issue in the literature [2]. In the "Efficient Markets Hypothesis" developed by Fama's (1970) it was stated that the price of a stock fully reflects available information, therefore, the price movements of stocks are independent of their past prices and predicting prices in advance is impossible [3]. The validity of the Efficient Markets Hypothesis has been debated in many studies suggesting the inefficiency of markets [4]. As an alternative to the Efficient Markets Hypothesis, "The Adaptive Market Hypothesis" was developed. In the Adaptive Market Hypothesis, it is suggested that markets are adaptive, evolve over time and therefore, they switch between efficiency and inefficiency at different times [5]. In other words, it is asserted that the efficiency of a stock exchange market is not static, but rather variable. On the other hand, it is stated that asset prices are not independent of historical information and therefore, higher returns can be yielded from the market by considering past price movements.

* Corresponding author.
*E-mail address:* nazifayyildiz@harran.edu.tr (N. Ayyildiz).

https://doi.org/10.1016/j.heliyon.2024.e24123
Received 28 October 2023; Received in revised form 16 December 2023; Accepted 3 January 2024
Available online 8 January 2024

In stock markets, a prediction model or method to be employed in order to minimize investment risk and uncertainty should be able to accurately predict the direction of the market. Two different analysis methods, namely, fundamental and technical analyses, have been conducted for predicting market prices in stock markets. In fundamental analysis, stock prices are predicted based on financial analyses of companies or industries. The purpose of fundamental analysis is to buy an asset at a low price and sell it at a high price by determining the difference between the actual value of the asset and its market price. In technical analysis, future prices are predicted using historical securities data, assuming that stock prices are determined by market forces and history tends to repeat itself [6].

Significant progress has been made in the prediction of stock returns over the course of history. In the first half of the 20th century, financial analysis, which only includes examining a company's financial and economic data, was used to predict its future performance [7]. Nevertheless, until the 1950s, investors were under the impression that they could mitigate investment risks simply by increasing the number of securities. Over time, new strategies, theories, and models have been developed in parallel with the increase in the number and variety of securities. As an alternative to passive investment strategies, which have been initially followed for buying and holding, many active investment strategies offering returns above the market average have been developed. Active investment strategies are based on the trading of securities and the constant monitoring of their performance to surpass a certain index in the market [8]. Then, Markowitz (1952) investigated the extent to which the maximum possible return rate of the securities in the investment portfolio could be obtained at certain risk levels, and the modern portfolio theory was developed [9].

In 1963, a single-index model was developed to maximize returns on alternative stock investments [10]. Subsequently, the Capital Asset Pricing Model (CAPM) was developed to calculate the cost of equity and to understand whether capital assets were over or underpriced [11–13]. Around the same time, an Arbitrage Pricing Theory was proposed, examining the relationship between risk and return, aiming to prevent any investor from indefinitely increasing wealth through arbitrage [14]. Autoregressive Integrated Moving Average (ARIMA) models have been established for time series analysis and forecasting [15]. Following this, the Markov chain model, a nonlinear time series prediction model, was developed [16]. By the late 1980s, various machine learning models, such as feedforward, backpropagation, and recurrent neural networks were introduced for predictive purposes. Since the early 2000s, the application of Machine Learning Models (MLMs) in stock predictions has enabled the analysis and prediction of large volumes of data with higher accuracy [17]. Consequently, the development of machine learning models and algorithms in this context has facilitated both effective and efficient decision-making processes, enabling instant and highly predictive outcomes [18,19].

The predictability of stock prices may provide stock markets, investors, market regulators, and business managers with various benefits [20]. Predicting the direction of stock price movements would primarily increase investment flow to the market. On the other hand, it would provide investors with crucial benefits such as protecting their savings, avoiding transaction costs, seizing investment opportunities, and foreseeing the risks that may arise in extraordinary situations [21]. It would also be beneficial for market regulators to make appropriate decisions and take corrective actions. It also provides business managers with the opportunity to act correctly in maximizing firm values [22]. In this context, estimating stock prices by employing new and different methods is as an important issue.

The aim of this study is to compare the performances of the algorithms by predicting the movement directions of stock market indexes in developed countries by employing machine learning classification algorithms and determine the best prediction algorithm. This study consists of six parts. In the second part, a literature review on the subject is presented. In the third part, the dataset used in the study is explained. In the fourth part, the machine learning algorithms used in this study are introduced. The empirical findings are discussed in the fifth part. Finally, in the sixth part, the results of the study are explained.

## 2. Literature review

The research within the scope of this study focused on predicting stock market indices in Developed Countries (G-7 Economies) using machine learning algorithms. Some of the relevant studies in the literature compared the performance of algorithms in predicting the direction of the index, while others investigated the predictive performance of a single algorithm on different indices. Therefore, the pioneering and important studies found in the relevant literature are presented in two categories.

One of the pioneering studies comparing the success of machine learning algorithms in predicting the direction of stock market indices aimed to predict the direction of the Tokyo Stock Price Index (TOPIX) in the Tokyo Stock Exchange in Japan. The prediction was carried out using a total dataset spanning 33 months from 1987 to 1989, employing multiple regression and artificial neural networks (ANN) algorithms. It was determined that the prediction accuracy of ANN algorithms was higher compared to multiple regression [23]. In another study comparing the prediction performance of two different algorithms, a multilayer perceptron trained by support vector machines (SVMs) and a backpropagation algorithm were used to predict the direction of the S&P 500 index movement. The study utilized daily index price data from January 4, 1993, to December 31, 1995. It was found that the support vector machine algorithm made more successful predictions [24]. In a study predicting the direction of the S&P 500 index using four different algorithms, a daily dataset from 1976 to 1999 was employed. Based on the results obtained from using generalized regression neural networks, probabilistic neural networks, linear regression algorithms, and ANNs, it was observed that neural network-based strategies were more successful than strategies utilizing other models [25]. In a study comparing the success levels of algorithms used in predicting the NYSE index, predictions were made using ANN and ARIMA algorithms. The study used a daily dataset spanning from 1988 to 2011, and it was concluded that both models were successful in predicting the direction of the index movement, with the ANN model exhibiting better performance [26].

In one of the pioneering studies comparing the success of machine learning algorithms in predicting the direction of stock market indices, the aim was to predict the direction of the TOPIX index of the Tokyo Stock Exchange in Japan. The prediction was made using multiple regression and artificial neural networks (ANN) algorithms with a total dataset spanning 33 months from 1987 to 1989. It was determined that the ANN method had higher prediction accuracy compared to multiple regression [23]. In a study comparing the

prediction performance of two different algorithms, a multilayer perceptron trained by support vector machines (SVMs) and a backpropagation algorithm were used to predict the direction of the S&P 500 index movement. The study utilized daily index price data from January 4, 1993, to December 31, 1995. It was found that the support vector machine method made more successful predictions [24]. In a study where the direction of the S&P 500 index was predicted using four different algorithms, a daily dataset from 1976 to 1999 was used. The direction of the S&P 500 index was predicted. Based on research using neural network-based strategies, such as neural networks, generalized regression neural networks, probabilistic neural networks and linear regression algorithms, it was observed that strategies using neural networks were more successful than those using other algorithms [25]. In a study comparing the success levels of methods used in predicting the NYSE index, predictions were made using ANN and ARIMA methods. The study used a daily dataset spanning from 1988 to 2011, and according to the results obtained, both models were successful in predicting the direction of the index movement, with the ANN model showing better performance [26].

In one study comparing the success levels of machine learning algorithms in predicting index directions, the performance of YSA, SVM, naive Bayes, k-nearest neighbors, and decision trees algorithms in predicting the direction of the S&P 500 index was compared. The dataset used spanned from April 1, 2010, to December 31, 2014. It was found that the YSA algorithm performed better than classifiers like support vector machines, naive Bayes, k-nearest neighbors and decision trees [27]. In a different study, the movement directions of the DJIA and NDX indices in the US were predicted using random forest and support vector machine algorithms. The analysis conducted using daily data from 2012 to 2018 revealed that the random forest algorithm performed better for the DJIA index, while the support vector machine algorithm yielded better results for the NDX index [28]. In a study comparing the success levels of algorithms in predicting the Nikkei 225 index, the performances of YSA and support vector machine algorithms were compared. The study utilized different technical indicators derived from daily stock trading, such as opening, closing, highest and lowest indices, between January 1, 2011, and September 27, 2020. According to the results, the YSA algorithm outperformed the support vector machine algorithm in predicting the daily movement directions of indices [29]. In another study, the movement directions of the S&P 500 and FTSE100 were predicted using SVMs, k-nearest neighbors, decision trees and random forest algorithms. The research spanned from March 7, 1995, to August 28, 2018, utilizing a dataset of 6042 daily records. The study concluded that in predicting market movements in developed countries, random forest was the best algorithm [30]. Additionally, a study compared the prediction success levels of machine learning algorithms on NASDAQ, NYSE, NIKKEI and FTSE indices using daily data from March 24, 2010, to March 24, 2020. It was determined that the random forest algorithm exhibited the most successful performance [31]. Lastly, a study focusing on Nasdaq, S&P 500 and Dow Jones stock price indices compared the performance of support vector machines and random forest algorithms. The study utilized trading data calculated from January 2007 to December 2017, encompassing eleven years, incorporating ten technical indicators. The research findings indicated that the random forest algorithm exhibited superior performance [32].

One of the pioneering studies measuring the predictive performance of a single algorithm on different indices involved predicting the movement direction of Germany's DAX index along with the US's DJIA, FTSE-100, and NASDAQ indices using the YSA algorithm. The price data from all stocks serving as component shares of each market index were used as datasets spanning from April 1, 1994, to September 30, 2002. The results indicated that the YSA algorithm predicted the movement directions of the indices with an accuracy rate exceeding 60 % [33]. Similarly, in another study examining a different subject, the predictability of the weekly movement direction of the NIKKEI 225 index was investigated using SVMs algorithm. The study, utilizing a total of 676 observations between January 1, 1990, and December 31, 2002, concluded that SVMs demonstrated successful performance [34]. In another study, the movement directions of the US's DJIA, NASDAQ, and S&P500 indices were predicted using the YSA algorithm. The study utilized daily data from December 19, 1990, to June 7, 2008. The algorithm's success was evaluated using various parameters, determining that the YSA algorithm was successful in predicting the movement directions of the indices [35]. Another study predicted the movement directions of the US's NASDAQ, the UK's FTSE100, Germany's DAX and France's CAC40 indices using weekly data from 2012 to 2016 with the YSA algorithm. The algorithm used was found to be successful in predicting the movement directions of the indices [36]. In a different study, five different developed stock market indices, including the US's S&P500, Germany's DAX, Japan's Nikkei, Denmark's OMX30 and Sweden's OMX30, were predicted using YSA with a daily dataset from January 1, 2014, to December 1, 2014. It was determined that the YSA algorithm was successful in predicting all indices, and its success did not vary significantly across countries [37].

When reviewing similar studies in the literature, it was observed that the directional movements of stock indices in developed countries are predicted using one or a few machine learning algorithms. In this study, however, the predictions of seven different stock indices are made using seven different machine learning algorithms. In this context, it is believed that this study will contribute to the literature in terms of its scope.

## 3. Dataset

In the scope of this study, G-7 economies are considered developed countries. When selecting stock market indices for each country, the primary stock market index with the highest market capitalization in each country was preferred. In this context, the study focused on the stock market indices of G-7 countries, which include the NYSE 100 (USA), NIKKEI 225 (Japan), FTSE 100 (UK), CAC 40 (France), DAX 30 (Germany), FTSE MIB (Italy), and TSX (Canada). The aim of this study was to predict the directional movements of selected stock market indices using machine learning classification algorithms, including decision trees, random forests, k-nearest neighbors, naive Bayes, logistic regression, support vector machines, and artificial neural networks.

In machine learning prediction algorithms, the use of a reasonably sized dataset is necessary for modeling. However, what constitutes a reasonable data size is uncertain and context-dependent. Typically, decision trees, naïve Bayes and support vector machine algorithms tend to perform better on small and medium-sized datasets, while random forests, k-nearest neighbors, and artificial neural

networks may perform better on larger datasets [38,39]. Similar studies in the literature have been observed to use datasets of different sizes. In this context, similar studies [29,31,40] were reviewed, and a ten-year period from January 1, 2012, to December 31, 2021, was selected as the appropriate research timeframe for analyzing stock market indices. Daily historical data containing opening prices, closing prices, highest prices and lowest prices of the examined indices during the chosen research period were obtained from https://tr.investing.com. Subsequently, technical indicators used as input variables in similar studies and holding potential to enhance prediction performance of algorithms were investigated. Technical indicators utilized as input variables were calculated based on studies focused on predicting stock index directions [41–45]. The technical indicators and their calculation methods used as input variables are presented in Table 1.

Simple Moving Average, Weighted Moving Average, Exponential Moving Average, Momentum, Stochastic K, Stochastic D, Relative Strength Index, Congruence of Moving Averages, Larry William's R, and Commodity Channel Index indicators which have been calculated according to the above-mentioned calculations are used as input variables in measuring the performance of all employed algorithms in the study. The next day's movement direction, which is expressed as "fall" or "rise" according to the closing prices of the stock market indexes, is used as the output data. Furthermore, the number of days that stock exchanges remained open during the research period, which was defined as January 1, 2012, to December 31, 2021, varied due to the differences in national and religious official holidays. In this context, the daily data of the examined stock indices, NYSE had 504, NIKKEI-225 had 495, FTSE-100 had 506, CAC-40 had 512, DAX-30 had 506, FTSE-MIB had 511, and TSX had 502 days' worth of data. These observation counts were used for the analysis, considering that there was not a significant imbalance that would potentially impact the prediction results.

Prior to the training and testing phases, the dataset is checked and it is determined that no incorrect or missing data exist in the dataset. In the machine learning application process, the dataset is divided into two parts, the training and the test datasets. The machine learns the model using the training dataset, then the predictions made with the learned model are compared with the test dataset, and the results are evaluated [46]. In similar studies in the literature, datasets have been divided into different ratios such as 75:25, 78.6:21.4, 90:10, 60:40, 80:20, 83:17, and 70:30 [18,29,30,36,41,44,45]. Although no general opinion exists about the ratio into which the data be divided, it is recommended to use the ratio of 80:20 in classification processes for large datasets [47]. Then, by dividing the dataset into the aforementioned ratios, it is tried to determine the ratio with the highest performance by trial and error. It is observed that the prediction performances of the algorithms improve when the size of the training dataset is increased from 50 to 80, but the prediction performances of most of the algorithms deteriorate when the size of the training dataset is increased from 80 to 90. The dataset has been divided into 80 % for training and 20 % for testing, considering recommendations from the literature and practical observations [46]. While running the algorithms, the features were checked and maximum iteration were made to achieve the highest prediction accuracy performance. measured. Seven different machine learning algorithms, namely Decision Tree, Random Forest, k- Nearest Neighbor, Naive Bayes, Logistic Regression, SVMs, and ANNs are used to predict the movement directions of the stock market indexes of developed countries, expressed as "rise and "fall" [41].

## 4. Methodology

In this study, decision trees, random forests, k-nearest neighbors, naive Bayes, logistic regression, support vector machines and

**Table 1**
Technical indicators and calculations.

| Technical Indicators | Calculation Method |
|---|---|
| **Simple Moving Average (MA)** | $\frac{\mathbf{C_t} + \mathbf{C_{t-1}} + \cdots + \mathbf{C_{t-30}}}{\mathbf{n}}$ |
| **Weighted Moving Average (WMA)** | $\frac{((\mathbf{n}) * \mathbf{C_t} + (\mathbf{n-1}) * \mathbf{C_{t-1}} + \cdots + \mathbf{C_{t-14}}}{(\mathbf{n} + (\mathbf{n-1}) + \cdots + \mathbf{1})}$ |
| **Exponential Moving Average (EMA)** | $\mathrm{EMA}(\mathbf{k})_t = \mathbf{EMA}(\mathbf{k})_{t-1} + \mathbf{a} * (\mathbf{C_t} - \mathrm{EMA}(\mathbf{k})_{t-1}$ |
| **Momentum (Mom)** | $\mathbf{C_t} - \mathbf{C_{t-n}}$ |
| **Stochastic K% (K%)** | $\frac{\mathbf{C_t} - \mathbf{LL_{t-n}}}{\mathrm{HH_{t-n}} - \mathrm{LL_{t-n}}} * \mathbf{100}$ |
| **Stochastic D% (D%)** | $\frac{\sum_{\mathbf{I=0}}^{\mathbf{n-1}} \mathbf{K_{t-1}\%}}{\mathbf{n}}$ |
| **Relative Strength Index (RSI)** | $\mathbf{100} - \frac{\mathbf{100}}{\mathbf{1} + (\sum_{\mathbf{I=0}}^{\mathbf{n-1}} \mathrm{Up_{t-i}}/\mathbf{n})/(\sum_{\mathbf{I=0}}^{\mathbf{n-1}} \mathrm{Dw_{t-i}}/\mathbf{n})}$ |
| **Moving Average Convergence/Divergence (MACD)** | $\mathbf{MACD}(\mathbf{n})_{t-1} + \frac{\mathbf{2}}{\mathbf{n+1}} * \mathrm{DIFF_t} - \mathbf{MACD}(\mathbf{n})_{t-1}$ |
| **Larry William's R% (LW)** | $\frac{\mathbf{H_n} - \mathbf{C_t}}{\mathbf{H_n} - \mathbf{L_n}} * \mathbf{100}$ |
| **Commodity Channel Index (CCI)** | $\frac{\mathbf{M_t} - \mathrm{SM_t}}{\mathbf{0,015 D_t}}$ |
| * $\mathbf{C_t}$: Closing Price | $\mathrm{HH_t}$: Highest of the highest within the last $\boldsymbol{t}$ days |
| $\mathbf{L_t}$: Lowest Price | $\mathbf{M_t} = (\mathbf{H_t} + \mathbf{L_t} + \mathbf{C_t})/\mathbf{3}$ |
| $\mathbf{H_t}$: Highest Price | $\mathrm{SM_t} = \sum_{\mathbf{I=0}}^{\mathbf{n}} \mathbf{M_{t-I+1}}/\mathbf{n})$ |
| $\mathbf{DIFF_t} = \mathbf{EMA(12)_t} - \mathbf{EMA(26)_t}$ | $\mathbf{D_t} = \sum_{\mathbf{I=1}}^{\mathbf{n}} \lvert \mathbf{M_{t-I+1}} - \mathrm{SM_t} \rvert/\mathbf{n}$ |
| **a**: Adjustment Factor | $\mathrm{UP_t}$: Upward price at time $t$ |
| $\mathbf{LL_t}$: Lowest of the lowest within the last $t$ days | $\mathrm{DV_t}$: Downward price at time $t$ |

Reference: [41].

artificial neural networks algorithms are used to predict the direction of movement of stock market indexes in developed countries. The algorithms used within the scope of this study have been explained in this section.

### 4.1. Decision trees algorithm

Decision trees are one of the algorithms used to predict the direction of an index. Based on historical data, a tree-like structure is created to classify the upward or downward trends of the index. In the decision tree algorithm, the values of the decision boundaries in a univariate decision tree are estimated experimentally from the training data. In the case of continuous data, logical test $X_i > C$ is performed at each internal node from the training data, where $X_i$ indicates a feature in the data space, whereas C is a threshold value in the observed range of $X_i$. The threshold value C can be determined by fulfilling certain conditions, such as maximizing the differences or minimizing the similarity in the descent nodes. Upon considering that a data set consists of several classes in the form of $C_1, C_2 \ldots, C_n$ and T denotes class values, probability $P_i = C_i/T$ of a class and entropy of classes are calculated. Upon considering that the T class values are divided into subsets in the form of $T_1, T_2 \ldots, T_n$ according to the B attribute values in the dataset, there is a gain ratio to be obtained as a result of splitting the T class values by using the B attribute values. Using the gain ratio criterion, the training set T is iteratively split so that the gain ratio would be maximum at each node of the tree. The process is repeated until each leaf node contains observation values belonging to only one class. To calculate the gain ratio, the operations specified in equations (1)–(4) are performed sequentially [48,49].

$$\text{Entropy}(T) = -\sum_{i=1}^{n} (p_i \, log_2(P_i)) \tag{1}$$

$$\text{Gain}(B, T) = \text{Entropy}(T) - \sum_{i=1}^{n} \frac{|T_i|}{|T|} (\text{Entropy}(T_i) \tag{2}$$

$$\text{Splitting Criterion } (B) = -\sum_{i=1}^{k} \frac{|T_i|}{|T|} \text{Log}_2 \left( \frac{|T_i|}{|T|} \right) \tag{3}$$

$$\text{Gain Ratio} = \frac{\text{Gain } (B, T)}{\text{Splitting Criterion } (B)} \tag{4}$$

### 4.2. Random forest algorithm

The Random Forest algorithm creates multiple decision trees using subsets of the dataset and combines the predictions of these trees to obtain the final result in predicting the upward and downward movements of the index. The process of constructing a random forest is simple. In the random forest algorithm, firstly, the number of decision trees (n) to be created is determined according to the characteristics of the dataset [50]. Then, m variables are randomly selected at each node in the created decision trees, and the best branch is determined by calculating with the Gini index. Then, the determined best branch is divided into two sub-branches. This process is sustained until there is only one class left in each node, that is, until the Gini index equals zero. Finally, the class with the highest number of votes among the predictions made by n decision trees is selected as the prediction class. T: The entire dataset, $p_j$: the square of the division of each data in the dataset by the number of elements lower than and higher than itself, and c: the selected data and the Gini index is calculated as in equation number 5 [51,52].

$$I_G = \sum_{j=1}^{e} p_j^2 \tag{5}$$

### 4.3. K-nearest neighbors algorithm

The K-Nearest Neighbors (KNN) algorithm is one of the classification methods used to predict stock market index movements. In the K-nearest neighbor algorithm, the known class category is performed through example-based classification based on the distance measurement of the examples in the training data (North, 2016). In the k-nearest neighbor algorithm, sample-based classification is made based on the distance measure of the samples in the training dataset whose class category is known. In the n-data sample $x_1^t = (x_1, \ldots, x_{1p}), \ldots, x_1^t = (x_1, \ldots, x_{np})$ defined by p attributes, each sample represents a point in the p-dimensional vector space. The distance between the sample points *i* and *j* is denoted by $d(x_i, x_j)$ and when all attributes are expressed in numerical terms, the algorithm calculates the Euclidean distance using the formula in equation (6) [53].

$$d(x_i * x_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - y_{jk})^2} \tag{6}$$

In order to determine the class to which a new sample would belong, the distance from this calculated point to all points in the training dataset is calculated, and the calculated distances are ranked from smallest to largest.

Considering the predetermined $k$ number of neighbor parameters, $k$ amount of points with the closest distance are selected. Generally, majority voting or weighted voting methods are employed to determine the class of new observation. The votes of the $k$ amount of $x_m$; m = 1, …, k nearest neighbors in the training dataset are inversely proportional to the distance to the new sample point $x_z$. The algorithm performs this calculation using the formula of equation (7) [54].

$$\text{vote}(x_m) = \begin{cases} \infty, & \text{if } d(x_{m,} x_z) = 0 \\ \dfrac{1}{d(x_{m,} x_z)}, & \text{otherwise} \end{cases} \tag{7}$$

### 4.4. Naive Bayes algorithm

The Naive Bayes Algorithm is one of the probability-based binary classification methods used to predict the upward and downward movements of stock market indices. The algorithm learns which class the samples from the training dataset belong to and makes class predictions for test data. In the naive Bayes algorithm, the fundamental approach used in predicting the direction of a stock market index is based on statistical probabilities, specifically utilizing Bayes' theorem. The Bayes theorem, expressed as P(C/X) which gives the probability of event C happening when event X has occurred, is provided in equation (8) [55,56].

$$P\left(\frac{C}{X}\right) = \frac{P(C) * \left(\frac{C}{X}\right)}{P(X)} \tag{8}$$

### 4.5. Logistic regression algorithm

The logistic regression algorithm is one of the supervised machine learning algorithms used in classification problems [57]. Logistic regression, a statistical model, is used for classification when the dependent variable belongs to two categorical classes. When predicting the direction of the index, the dependent variable is typically divided into two categories, such as 'upward' or 'downward,' and a logistic regression model based on historical data is created to estimate the probability of the index rising or falling. Assuming that the dependent variable has two categories in the logistic regression algorithm, Y being the dependent variable and p being the number of vectors of independent variables $X^T = (X_1, \ldots . X_p)$: Y's category is usually encoded as Y = 0 and Y = 1 and the probability of assuming the value of 1 for Y, when $X^T = (X_1, \ldots . X_p)$ is known, is shown as $\pi(X) = P(Y = 1|X = x)$. The multiple logistic regression model is expressed in equation (9) [58,59].

$$\pi(X) = P(Y = 1|X = x) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 x_1 + \ldots . + \beta_p x_p\right)}} \tag{9}$$

When the logit transformation, denoted by $\text{logit}[\pi(x)] = g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right)$ is applied to the multiple logistic regression model, The formula is transformed into the linear model in equation (10):

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x_1 + \ldots . + \beta_p x_p\right. \tag{10}$$

Depending on the value set of x, g(x) may be continuous, and the values of g(x) may range between $-\infty$ and $+\infty$. By using the model's parameters $(\beta^T = \beta_0 + \beta_1 + \ldots . + \beta_p)$ and the likelihood function $L(\beta) = \prod_{i=1}^{n} (x_i)^{y_i} (1 - \pi(x_i)^{1-y_i}$, the prediction would be made employing the maximum likelihood method [58,59].

### 4.6. Support vector machines algorithm

Support Vector Machines (SVMs) are one of the commonly used algorithms for predicting the direction of stock market indices. SVMs is particularly effective in machine learning for classification problems. It works by creating an optimized separating hyperplane to assign data points to specific classes. The SVMs algorithm is based on the principle of structural risk minimization, which aims to minimize the upper bound of the generalization error. Given that, w is the weight vector, b is the constant term, $y_i \in -1, 1$ is the label values, and $x_i \in \Re^d$ is the features vector, in the case where the dataset consisting of l elements to be used for training is assumed to be $\{x_i, y_i\}, i, = 1, 2, \ldots, l$ the training points called support vectors are expressed by equations (11)–(13) [58,59]

$$\text{for } y_i = 1, w * x_i + b \geq 1 \tag{11}$$

$$\text{for } y_i = -1, w * x_i + b \geq -1 \tag{12}$$

$$y_i = w * x_i + b \geq 1 \forall i \tag{13}$$

### 4.7. Artificial neural networks

Artificial Neural Networks (ANNs) are one of the machine learning classification algorithms commonly used for binary classification problems like predicting the direction of stock market indices. ANNs draw inspiration from the structure of neural systems and can model complex relationships. When designing an ANN model, a three-layer structure is often used. These layers consist of an input layer, one or more hidden layers, and an output layer. The input layer is composed of nodes corresponding to each feature in the dataset. Hidden layers are used to detect complex patterns in the data. The output layer, consisting of a single node, makes predictions for the classes of test data. The training of an ANN is a process aimed at improving the model's performance on the dataset. During this process, the error (loss) between the predicted classes and the actual classes is minimized. In the ANNs algorithm, a fitness function consisting of the sum of the squares of the error signals of each neuron in the output layer of the ANN is defined. In the kth iteration of the training, when the output value $y_i$ of the ith nerve in the output layer of the ANN is denoted by the value $d_i$ that is wanted to be assigned from this nerve, the error signal of the ith nerve is obtained as $e_i = d_i - y_i(k)$. The fitness function, which is the sum of the squares of the error signals of each neuron in the output layer of the ANN, is expressed as in equation (14) [60].

$$E = \frac{1}{2} i e_i^2 \; k = \frac{1}{2} i^{(d_i - y_i \; k)^2} \tag{14}$$

In the ANN algorithm, the backpropagation algorithm is used for error backpropagation, and the fitness function is minimized. The fitness function depends on the weight values of ANNs, and the most accurate way to update their weights is calculated using the gradient descent method. The equation of the gradient reduction method, where η is the learning coefficient, is expressed as in equation (15) [34,61].

$$\Delta w_{ij} = -\eta \frac{\partial E(w)}{\partial w_{ij}} \tag{15}$$

The backpropagation algorithm consists of two main stages: the forward pass and the backward pass. In the forward pass stage, the outputs are calculated and compared with the desired outputs. Then, the error rate is computed by comparing the desired and actual outputs. In the backward pass, the weights in the network are adjusted based on the errors calculated in the forward pass. The forward and backward pass processes are repeated until the error reaches a sufficiently low level [58].

## 5. Findings and discussion

Within the scope of this study, the performances of machine learning algorithms such as Decision Trees (DT), Random Forest (RF), k-Nearest Neighbor (KNN), Naive Bayes (NB), Logistic Regression (LR), Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) in predicting the movement directions of stock market indexes are investigated. The accuracy ratios obtained from the analyses are presented in Table 2.

Upon examining Table 2; it is seen that ANNs with an average accuracy ratio of 83.43 % stand out as the most successful algorithm. The index of which the algorithm best predicts the movement direction is the FTSE 100 index, which is predicted with 93.48 % accuracy. This result is also the highest achievable prediction accuracy ratio in the analysis. The index predicted by the ANNs algorithm with the lowest accuracy (81.01 %) is the NIKKEI 225. Therefore, it can be claimed that the ANN algorithm successfully predicts the movement directions of the stock market indexes of developed countries with a prediction accuracy ratio of over 80 %, followed by logistic regression (82.56 %), SVMs (79.43 %), Naive Bayes (62.60 %), random forest (59 %), decision trees (56.63 %), and k-nearest neighbor (50.26 %) algorithms, respectively.

The obtained results align with some studies comparing the success levels of various machine learning algorithms in predicting stock market indices of developed countries [23,25,27,29]. These studies respectively compared the success levels of different machine learning algorithms on Japan's TOPIX, the US's S&P 500, and Japan's Nikkei 225 indices, determining artificial neural networks as the most successful algorithm. However, these findings tend to contradict findings in certain studies [30,31]. In studies with contrasting results, random forest was identified as the most successful prediction algorithm for the US's S&P500 and NASDAQ, UK's FTSE100, and Japan's NIKKEI 225 indices.

According to the analysis findings, the ANN algorithm identified as the best algorithm was not universally applicable across all

**Table 2**
Prediction accuracy ratios of machine learning algorithms.

| STOCK INDEXES | DT | RF | KNN | NB | LR | SVMs | ANNs |
|---|---|---|---|---|---|---|---|
| **NYSE100** | 0.5337 | 0.6409 | 0.5298 | 0.6409 | 0.8214 | 0.7222 | [a]**0.8373** |
| **NIKKEI225** | 0.5414 | 0.5717 | 0.4848 | 0.5899 | [a]0.8162 | 0.8000 | **0.8101** |
| **FTSE100** | 0.5889 | 0.5929 | 0.5119 | 0.6067 | 0.8498 | 0.8162 | [a]**0.9348** |
| **CAC40** | 0.5957 | 0.6270 | 0.5098 | 0.6406 | [a]0.8359 | 0.8105 | **0.8301** |
| **DAX30** | 0.5474 | 0.5059 | 0.4960 | 0.6225 | 0.8083 | 0.8162 | [a]**0.8182** |
| **FTSEMIB** | 0.5656 | 0.6282 | 0.4834 | 0.6380 | 0.8219 | 0.8004 | [a]**0.8513** |
| **TSX** | 0.5916 | 0.5637 | 0.5438 | 0.6434 | [a]0.8645 | 0.8466 | **0.8586** |
| **AVARAGE** | 0.5663 | 0.5900 | 0.5026 | 0.6260 | 0.8256 | 0.7943 | [a]**0.8343** |

[a] The accuracy ratio of the algorithm that predicts the movement direction of the index with the highest accuracy.

indices, meaning it did not exhibit the highest accuracy across all indices. Accordingly, while ANNs appear as the best predictors for NYSE 100, FTSE 100, DAX 30, and FTSE MIB indices, logistic regression seems to be the best prediction algorithm for NIKKEI 225, CAC 40, and TSX indices. These results align with findings from some studies investigating the location dependence of the ANN algorithm [33,35,36]. Similar studies reaching comparable conclusions predicted the movement directions of indices such as DJIA, S&P500, FTSE-100, and NASDAQ for the USA; FTSE100 for the England; DAX for Germany; and CAC40 for France using the ANN algorithm, revealing varying predictive accuracy among countries. However, these results contradict the findings of one study [37]. In that particular study, the ANN algorithm was found successful in predicting the movement directions of stock market indices of developed countries such as S&P500 for the USA, DAX for Germany, Nikkei 225 for Japan, OMX 30 for Denmark, and OMX30 for Sweden with no significant variation in predictive success among countries.

## 6. Conclusions

In this study, it is aimed to compare the performances of the algorithm by predicting the movement directions of the stock market indexes of developed countries with machine learning classification algorithms and to determine the best prediction algorithm. For this purpose, an analysis is conducted on the main stock market indexes of the G-7 countries, known as developed countries, such as the NYSE 100 (USA), NIKKEI 225 (Japan), FTSE 100 (UK), CAC 40 (France), DAX 30 (Germany), FTSE MIB (Italy), and TSX (Canada). Decision Tree, Random Forest, k-Nearest Neighbor, Naive Bayes, Logistic Regression, SVMs, and ANNs, which are machine learning classification algorithms, are used to predict the movement directions of the mentioned indexes. In the analyses, the daily dataset obtained over the period 01.01.2012–12.31.2021 and the Moving Average, Weighted Moving Average, Exponential Moving Average, MACD, CCI, RSI, Stochastic %K, Stochastic %D, William's %R, and Momentum technical indicators were used as the input data. The successful performances of the machine learning classification algorithms in predicting the movement directions of the indexes are examined with the established prediction model.

According to the results obtained from the conducted analyses, ANNs are determined as the best methods for predicting the movement directions of the stock market indexes of developed countries. It has been determined that the logistic regression and support vector machine algorithms, combined with the ANNs method, predicted the movement direction of all indexes with an accuracy ratio of over 70 %. Furthermore, it is determined that the ANNs, which is the best algorithm, is not valid for all indexes, in other words, it is not the algorithm with the highest ratio of accuracy performance in all indexes. Accordingly, although the ANNs algorithm is the best predictor of stock market indexes such as the NYSE 100, FTSE 100, DAX 30, and FTSE MIB; logistic regression is determined as the best estimation method for NIKKEI 225, CAC 40, and TSX. It is known that the examined exchanges have unique characteristics such as volatility, market structure, and economic conditions. Therefore, it is thought that this situation arises from the different dynamics of the stock market.

During the period under examination, artificial neural networks, logistic regression and support vector machine algorithms have been found to be successful in predicting the movements of stock market indices. The results obtained can serve as a significant guiding factor for future financial forecasts. From this perspective, it is possible to provide a set of recommendations to investors, portfolio managers, companies, and economic policymakers. Investors can optimize their investment strategies for future periods by using artificial neural networks, logistic regression, and support vector machine algorithms, leading to more informed investment decisions. Portfolio managers can enhance their client services by integrating these methods into their portfolio management processes. Additionally, presenting market trend forecasts alongside stock predictions can offer a more comprehensive approach to clients. Companies, on the other hand, can utilize these prediction methods to comprehend stock performance and trends or evaluate future growth potentials and risks. Finally, economic policymakers can leverage the successful methods identified to measure the effectiveness of economic policies or anticipate the direction of the economy.

The results obtained within the scope of this study are subject to various constraints. Firstly, despite algorithms such as deep learning, gradient boosting, multi-layer perceptions, and ensemble learning methods being commonly used for classification purposes, they were not included in this study. This decision has restricted the diversity of the compared methods and the extent of the analysis. Furthermore, the study only examined a ten-year time period, which may lead to variations in prediction results that could occur over shorter or longer periods. In addition, the analysis did not incorporate the use of macroeconomic data, which is often interrelated with the performance of financial markets, including economic indicators, interest rates, and other macroeconomic factors. The exclusion of these factors from the analysis could impose inherent limitations on the accuracy of the predictions. Finally, another limitation of the study is the number and selection of technical indicators used. The absence of certain selected technical indicators may lead to an incomplete representation of the dataset in the analysis. All these limitations should be considered when interpreting the results of the present study and when planning future research. In this context, conducting further research on algorithms and input variables that can be employed to predict the directions of stock market indices can be beneficial in achieving higher prediction accuracy. This, in turn, will shape the focus of future studies.

*This article is derived from the doctoral thesis titled 'Prediction of Stock Market Index Movements Using Machine Learning Methods: An Application on the Stock Markets of Developed and Developing Countries' completed under the supervision of Prof. Dr. Ömer ISKENDEROĞLU. Bibliographic information for the aforementioned doctoral thesis is provided below:

N. Ayyildiz, 'Prediction of Stock Market Index Movements Using Machine Learning Methods: An Application on the Stock Markets of Developed and Developing Countries,' PhD thesis, Nigde Omer Halisdemir University, Social Sciences Institute, Department of Business Administration (2023).

## CRediT author statement

Omer Iskenderoglu: Writing – review & editing, Validation, Data curation. Nazif Ayyildiz: Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation

## Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data associated with your study has not been deposited into any publicly available repository. It will however be made available upon request.

## References

[1] E. Akyildirim, A.F. Bariviera, D.K. Nguyen, A. Sensoy, Forecasting high-frequency stock returns: a comparison of alternative methods, Ann. Oper. Res. 313 (2022) 639–690, https://doi.org/10.1007/s10479-021-04464-8.
[2] R.R. Varghese, B.R. Mohan, Study on the sentimental influence on Indian stock price, Heliyon 9 (2023) e22788, https://doi.org/10.1016/j.heliyon.2023.e22788.
[3] E.F. Fama, Efficient capital markets: a review of theory and empirical work, J. Finance 25 (1970) 383–417, https://doi.org/10.1111/j.1540-6261.1970.tb00518.x.
[4] L. Nevasalmi, Forecasting multinomial stock returns using machine learning methods, The Journal of Finance and Data Science 6 (2020) 86–106, https://doi.org/10.1016/j.jfds.2020.09.001.
[5] A.W. Lo, The adaptive markets Hypothesis, J. Portfolio Manag. 30 (2004) 15–29, https://doi.org/10.3905/jpm.2004.442611.
[6] D. Spahija, S. Xhaferi, Fundamental and technical analysis of the stock price, International Scientific Journal Monte 1 (2018), https://doi.org/10.33807/monte.1.201904160.
[7] Z. Shang, The Research of Financial Forecasting and Valuation Models, 2021, https://doi.org/10.2991/aebmr.k.210601.012.
[8] R.M.C. Gopwani, Active vs Passıve Investment. The Optımal Dıversıfıcatıon Effect, Universidad Pontificia, Master, 2019.
[9] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 77–91, https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.
[10] W.F. Sharpe, A simplified model for portfolio analysis, Manag. Sci. 9 (1963) 277–293, https://doi.org/10.1287/mnsc.9.2.277.
[11] W.F. Sharpe, Capital asset prices: a theory of market equilibrium under conditions of risk, J. Finance 19 (1964) 425–442.
[12] J. Lintner, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Rev. Econ. Stat. 47 (1965) 13, https://doi.org/10.2307/1924119.
[13] J. Mossin, Equilibrium in a capital asset market, Econometrica 34 (1966) 768, https://doi.org/10.2307/1910098.
[14] S.A. Ross, The arbitrage theory of capital asset pricing, J. Econ. Theor. 13 (1976) 341–360, https://doi.org/10.1016/0022-0531(76)90046-6.
[15] G.E.P. Box, G.M. Jenkins, Time Series Analysis, Forecasting and Control, San Francisco, 1970, https://doi.org/10.1057/9781137291264_6.
[16] J.D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57 (1989) 357, https://doi.org/10.2307/1912559.
[17] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175, https://doi.org/10.1016/S0925-2312(01)00702-0.
[18] A. Nõu, D. Lapitskaya, M.H. Eratalay, R. Sharma, Predicting stock return and volatility with machine learning and econometric models: a comparative case study of the Baltic stock market, SSRN Electron. J. (2021), https://doi.org/10.2139/ssrn.3974770.
[19] T. Rothman, Machine Learning versus Fundamental Investment Analysis: A Meta-Analysis, 2021. https://purl.utwente.nl/essays/87928. (Accessed 23 January 2023).
[20] X. Yin, X. Zhang, H. Li, Y. Chen, W. He, An interpretable model for stock price movement prediction based on the hierarchical belief rule base, Heliyon 9 (2023) e16589, https://doi.org/10.1016/j.heliyon.2023.e16589.
[21] E. Filiz, S. Akogul, H.A. Karaboga, Classification of BIST-100 index change direction with machine learning algorithms using major world indexes, Bitlis Eren University Journal of Science and Technology 10 (2021) 432–441, https://doi.org/10.17798/bitlisfen.889007.
[22] M. Mallikarjuna, R.P. Rao, Evaluation of forecasting methods from selected stock market returns, Financial Innovation 5 (2019) 40, https://doi.org/10.1186/s40854-019-0157-x.
[23] T. Kimoto, K. Asakawa, M. Yoda, M. Takeoka, Stock market prediction system with modular neural networks, 1990 IJCNN International Joint Conference on Neural Networks, IEEE 1 (1990) 1–6, https://doi.org/10.1109/IJCNN.1990.137535.
[24] L. Cao, F.E.H. Tay, Financial forecasting using support vector machines, Neural Comput. Appl. 10 (2001) 184–192, https://doi.org/10.1007/s005210170010.
[25] D. Enke, S. Thawornwong, The use of data mining and neural networks for forecasting stock market returns, Expert Syst. Appl. 29 (2005) 927–940, https://doi.org/10.1016/J.ESWA.2005.06.024.
[26] A. Adebiyi, C. Ayo, A. Marion, O. Sunday O, Stock Price Prediction Using Neural Network with Hybridized Market Indicators, 2012, p. 3. http://www.cisjournal.org.
[27] R. Dash, P.K. Dash, A hybrid stock trading framework integrating technical analysis with machine learning techniques, The Journal of Finance and Data Science 2 (2016) 42–57, https://doi.org/10.1016/j.jfds.2016.03.002.
[28] S. Aktaş, Predıctıon of the Dow Jones Industrial Average and the Nasdaq 100 Indices Usıng Machine Learning, Master Thesis, Social Sciences, 2019. https://polen.itu.edu.tr:8443/server/api/core/bitstreams/2918a720-5dad-4db9-8eeb-ad582bbcb19d/content.
[29] M. Ali, D.M. Khan, M. Aamir, A. Ali, Z. Ahmad, Predicting the direction movement of financial time series using artificial neural network and support vector machine, Complexity 2021 (2021) 1–13, https://doi.org/10.1155/2021/2906463.
[30] R. Dimingo, J.W. Muteba Mwamba, L. Bonga-Bonga, Prediction of Stock Market Direction: Application of Machine Learning Models, vol. 74, Economia Internazionale/International Economics, 2021. http://www.iei1946.it/upload/rivista_articoli/allegati/425_bonga-et-al-predictionricfinal.pdf. (Accessed 23 January 2023).

[31] A. Subasi, F. Amir, K. Bagedo, A. Shams, A. Sarirete, Stock market prediction using machine learning, Procedia Comput. Sci. 194 (2021) 173–179, https://doi.org/10.1016/j.procs.2021.10.071.
[32] M. Khanna, M. Kulshrestha, L.K. Singh, S. Thawkar, K. Shrivastava, Performance evaluation of machine learning algorithms for stock price and stock index movement prediction using trend deterministic data prediction, Int. J. Appl. Metaheuristic Comput. (IJAMC) 13 (2022) 1–30, https://doi.org/10.4018/IJAMC.292511.
[33] P.K.H. Phua, Xiaotian Zhu, Haur Chung, Koh, Forecasting stock index increments using neural networks with trust region methods, in: Proceedings of the International Joint Conference on Neural Networks, IEEE, 2003, pp. 260–265, https://doi.org/10.1109/IJCNN.2003.1223354, 2003.
[34] H. Li, S. Huang, Research on the prediction method of stock price based on RBF neural network optimization algorithm, E3S Web of Conferences. 235 (2021) 03088, https://doi.org/10.1051/e3sconf/202123503088.
[35] Z. Liao, J. Wang, Forecasting model of global stock index by stochastic time effective neural network, Expert Syst. Appl. 37 (2010) 834–841, https://doi.org/10.1016/j.eswa.2009.05.086.
[36] A. Ozer, S. sercan Sari, E.E. Başakın, Stock market index prediction with artificial neural networks and fuzzy logic: example of developed and developing countries, Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 11 (2018), https://doi.org/10.17218/hititsosbil.390490.
[37] S. Karlsson, M. Nordberg, Stock Market Index Prediction Using Artificial Neural Networks Trained on Foreign Markets, 2015. https://www.diva-portal.org/smash/get/diva2:812198/FULLTEXT01.pdf. (Accessed 23 January 2023).
[38] A.R. Ajiboye, R. Abdullah-Arshah, H. Qin, H. Isah-Kebbe, Evaluating the effect of dataset size on predictive model using supervised learning technique, Int. J. Comput. Syst. Sci. Eng. 1 (2015) 75–84, https://doi.org/10.15282/ijsecs.1.2015.6.0006.
[39] H. Sug, Performance of machine learning algorithms and diversity in data, MATEC Web of Conferences 210 (2018) 04019, https://doi.org/10.1051/matecconf/201821004019.
[40] W. Huang, Y. Nakamori, S.-Y. Wang, Forecasting stock market movement direction with support vector machine, Comput. Oper. Res. 32 (2005) 2513–2522, https://doi.org/10.1016/j.cor.2004.03.016.
[41] M. Kumar, M. Thenmozhi, Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest, SSRN Electronic Journal, 2006, https://doi.org/10.2139/ssrn.876544.
[42] Y. Kara, M. Acar Boyacioglu, Ö.K. Baykan, Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul Stock Exchange, Expert Syst. Appl. 38 (2011) 5311–5319, https://doi.org/10.1016/j.eswa.2010.10.027.
[43] J. Patel, S. Shah, P. Thakkar, K. Kotecha, Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques, Expert Syst. Appl. 42 (2015) 259–268, https://doi.org/10.1016/j.eswa.2014.07.040.
[44] M. Qiu, Y. Song, Predicting the direction of stock market index movement using an optimized artificial neural network model, PLoS One 11 (2016) e0155133, https://doi.org/10.1371/journal.pone.0155133.
[45] İ. Kara, F. Ecer, Comparison of performances of classificatıon methods in estimation of stock exchange index in BIST, The Journal of Academic Social Sciences 83 (2018) 514–524, https://doi.org/10.16992/ASOS.14460.
[46] O. Campesato, Artificial Intelligence, Machine Learning, and D, eep Learning, USA, 2020. https://books.google.com.tr/books?id=pqnNDwAAQBAJ&printsec=frontcover&hl=tr&pli=1#v=onepage&q&f=false. (Accessed 23 January 2023).
[47] A. Rácz, D. Bajusz, K. Héberger, Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification, Molecules 26 (2021) 1111, https://doi.org/10.3390/molecules26041111.
[48] J. Han, M. Kamber, J. Pe, Data Mining, Third, Elsevier, USA, 2012, https://doi.org/10.1016/C2009-0-61819-5.
[49] J.R. Qinlan, C4.5 Programs for Machine Learning, Elsevier, 1993, https://doi.org/10.1016/C2009-0-27846-9.
[50] Z. Tan, Z. Yan, G. Zhu, Stock selection with random forest: an exploitation of excess return in the Chinese stock market, Heliyon 5 (2019) e02310, https://doi.org/10.1016/j.heliyon.2019.e02310.
[51] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[52] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, CRC Press, 1984.
[53] T. Mitchell, Mavhine Learning, McGraw-Hill Science/Engineering/Math, 1997. https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf. (Accessed 23 January 2023).
[54] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theor. 13 (1967) 21–27, https://doi.org/10.1109/TIT.1967.1053964.
[55] D. Jurafsky, J.H. Martin, Speech and Language Processing, Third, Prentice Hall, 2023. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf. (Accessed 23 January 2023).
[56] M. Minsky, Steps Woward Artificial İntelligence, 1960. https://web.media.mit.edu/~minsky/papers/steps.html. (Accessed 23 January 2023).
[57] E. Bisong, Logistic regression, in: Building Machine Learning and Deep Learning Models on Google Cloud Platform, Apress, Berkeley, CA, 2019, pp. 243–250, https://doi.org/10.1007/978-1-4842-4470-8_20.
[58] J.M. Hilbe, Practical guide to logistic regression, J. Stat. Software 71 (2016), https://doi.org/10.18637/jss.v071.b03.
[59] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, Wiley Series in Probability and Statistics, Third, 2013, https://doi.org/10.1002/9781118548387.
[60] R. Rojas, Neural Networks A Systematic Introduction, Springer Verlag, 1996. https://link.springer.com/book/10.1007/978-3-642-61068-4. (Accessed 23 January 2023).
[61] S. Haykin, Neural Networks and Learning Machines, Third, Pearson Education, 2009. https://cours.etsmtl.ca/sys843/REFS/Books/ebook_Haykin09.pdf. (Accessed 23 January 2023).