

Article

Decision Support Using Machine Learning Indication for Financial Investment

Ariel Vieira de Oliveira ¹, Márcia Cristina Schiavi Dazzi ¹, Anita Maria da Rocha Fernandes ^{1,*}, Rudimar Luis Scaranto Dazzi ¹, Paulo Ferreira ^{2,3} and Valderi Reis Quietinho Leithardt ^{2,4,*}

¹ School of Sea, Science, and Technology, University of Vale do Itajaí, R. Uruguai, 458, Itajaí 88302-901, Brazil

² VALORIZA, Research Center for Endogenous Resources Valorization, Instituto Politécnico de Portalegre, 7300-555 Portalegre, Portugal

³ Department of Economic Sciences and Organizations, Polytechnic Institute of Portalegre, 7300-555 Portalegre, Portugal

⁴ COPELABS, Lusófona University of Humanities and Technologies, Campo Grande 376, 1749-024 Lisboa, Portugal

* Correspondence: anita.fernandes@univali.br (A.M.d.R.F.); valderi@ipportalegre.pt (V.R.Q.L.)

Abstract: To support the decision-making process of new investors, this paper aims to implement Machine Learning algorithms to generate investment indications, considering the Brazilian scenario. Three artificial intelligence techniques were implemented, namely: Multilayer Perceptron, Logistic Regression and Decision Tree, which performed the classification of investments. The database used was the one provided by the website Oceans14, containing the history of Fundamental Indicators and the history of Quotations, considering BOVESPA (São Paulo State Stock Exchange). The results of the different algorithms were compared to each other using the following metrics: accuracy, precision, recall, and F1-score. The Decision Tree was the algorithm that obtained the best classification metrics and an accuracy of 77%.

Keywords: financial investment; machine learning; artificial intelligence

Citation: Oliveira, A.V.d.; Dazzi, M.C.S.; Fernandes, A.M.d.R.; Dazzi, R.L.S.; Ferreira, P.; Leithardt, V.R.Q. Decision Support Using Machine Learning Indication for Financial Investment. *Future Internet* **2022**, *14*, 304. <https://doi.org/10.3390/fi14110304>

Academic Editor: Maria Gabriella Xibilia

Received: 19 September 2022

Accepted: 21 October 2022

Published: 25 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The financial investment business has been growing in recent years with new people interested in the subject, mainly by individuals. According to the Official Brazilian Stock Exchange [1], the number of individuals making financial investments grew from 557,109 in 2015, to 3,173,411 on 30 November 2020. The growth in the number of new investors is considerable, but the low financial education in Brazil makes it difficult to make the first investments, precisely because it is a market that demands dedication and in-depth studies to obtain good results. Therefore, the lack of adequate knowledge ends up keeping new investors away.

Within this context, it can be noticed that new investors need some help to start their activities in the investment business. Therefore, this paper seeks to contribute in the financial investment field using Machine Learning algorithms to assist in the decision-making process of new investors, where it was sought to answer the following research questions: how to qualitatively classify shares of companies listed on Bovespa to assist in the decision-making process of new investors? Which Machine Learning algorithms have greater accuracy for this purpose?

Although the literature presents similar works to this one, it is noteworthy that the relevance of this research lies in the fact that it considers peculiarities of the Brazilian scenario, such as economic crises. In addition, the Selic rate is considered as one of the parameters, which is changed according to the needs of the Brazilian economy. In this sense,

peculiar situations can happen, which are not portrayed in the bases used in the other works.

Machine Learning algorithms that serve as financial investment advisors for new investors may be a solution to these questions. Based on these needs, this paper has the following objectives: identification of market data and investor profile that are needed to generate reliable investment suggestions; perform pre-processing and database assembly; define algorithms that will be used; implement Machine Learning algorithms to solve the problem at hand; Conduct validation and performance tests with the used model. The use of machine learning models to support decision making has been used in various fields, such as medicine [2,3], security [4,5], electricity [6–8], and in financial investments [9].

2. Theoretical Background

This section will present concepts and definitions relevant to the development of the solution proposed by this paper.

2.1. Financial Market Investment Types

There are different types of financial investments that can be made in different types of applications, such as: stocks, savings, commodities, and foreign currencies, among others. However, a financial investment is basically characterized as an application of capital in some financial application aiming at future income [10].

Financial investments can also be divided into two basic groups: fixed-income investments and variable income investments. In fixed-income investments, the investor can know the amount of profitability that will be earned when the term of his investment is completed. Some of the main examples of fixed-income investments are savings accounts and direct treasury, among others. As for variable-income investments, the profitability value is not known in advance by the investor due to the variations that it undergoes over time; due to this characteristic, they are investments with a higher risk but with greater profitability. Some of the main examples of variable income investments are stocks and stock funds, among others [11].

The implementation of the present work is focused on shares. A share is the smallest portion of the capital stock of publicly traded companies, which gives the owners of the share's rights and duties equal to those of any other partner in the company, limited by the number of shares the owner owns [12]. The owner of the shares is called a shareholder, who, from the moment he or she acquires a certain share, becomes a partner or co-owner of the company. Stocks are considered a variable-income investment because they have a return that can vary over time. They are considered a relatively high-risk investment because they belong to a highly volatile market [13].

2.2. Fundamental Analysis

Fundamental Analysis determines the appropriate stock prices using the earnings and dividends of a given company, expectations of future interest rates, and risk assessment of the company [14].

The basic objective of the valuation of a company is to obtain a fair value, which reflects the expected return on future performance projections consistent with the reality of the company evaluated. Since it is based on projections, the valuation is subject to errors and uncertainties, mainly since the analysis of external variables is not controlled by the company in question. Within this context, the result obtained through an evaluation is not an exact estimate of the value of the evaluated company [15].

To be able to analyze companies, it is important to use variables and indicators, known as Fundamental Indicators, that impact a company. In the present work, the most used indicators were used according to Bered and Rosa [16], having as a basis the theoretical foundation presented to obtain each indicator. The Fundamental Indicators that were used are: Net Margin (ML), Earnings per Share (EPS), Price/Earnings (P/E), Book

Value per Share (NAV), Price/Equity Value (P/EV), Return on Equity (ROE) and Ebitda Margin (EM).

2.3. Investor Profile

Each investor active in the financial investment market has different characteristics, i.e., each investor has a unique way of making his investments, and the result of the investments is directly related to his way of acting and thinking about his decisions. Thus, it is necessary that each investor makes investments that match his profile, considering mainly three fundamental points: the types of risks he is willing to face; how much he is willing to lose; and what is the desired financial return. To learn more about the types of investors, the following presents the characteristics and differences of the three profiles: conservative, moderate, and riskier [17].

Conservative: Their main characteristic is a low tolerance for risk and, consequently, a low return when compared to the other investor profiles. In general, they have little knowledge about investments and do not like to take risks; that is, they prefer to exchange the possibility of reaching a high profitability for a higher level of security. They commonly aim at preserving their assets by making fixed-income investments or investments that make it possible to withdraw their resources in a short period of time [18].

Moderate: Moderate profile investors are people who show interest in higher returns, take more risks, and seek to increase their wealth in the medium term through investments. In general terms, they have some knowledge about investments and tend to be people who are not willing to take high levels of risk but are also not extremely conservative to the point of fitting into the conservative profile. For wanting a higher return and still wanting to protect the security in their investments, it is commonly indicated to make investments with different degrees of risk, diversifying their investments among various products, so that they can reach a higher return than the conservative profile and still maintain a moderate level of security in their investments [19].

Riskier: Investors with a riskier profile are characterized by their goal of obtaining higher returns on their investments and are therefore willing to take more risks. They have a high level of knowledge about investments, are usually advised by qualified professionals who guide their investments and use variable income to achieve their financial goals, even accepting the loss of capital, due to the high risk involved in the operations [20].

2.4. Machine Learning Techniques Used in Research

To develop this paper, different Artificial Intelligence methods were used, which will be presented further on.

2.4.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) are inspired by the biological neural networks existing in the human brain, which can perform complex tasks automatically, quickly, and simultaneously [21]. These structures served as the basis for the development of the models, which seek to simulate the learning capacity of the brain [22]. In other words, ANNs are non-linear mathematical systems that have neurons connected by connections that are associated with weights [23]. ANNs can recognize patterns, detecting relationships, performing operations with imprecise data, and predicting time series, among other functions [24].

In addition to time series prediction, ANN shows promise for pattern classification [25–27], control [28,29], and optimization. Optimization gains space in this context because of the need to improve the components necessary to maintain the system's operation [30,31], especially considering the expansive growth of communication systems [32,33], Internet of Things [34–36], the need for sustainability [37] and [38], technology development [39], and data privacy [40,41]. Figure 1 shows a representation of an artificial neuron from a Perceptron Neural Network.

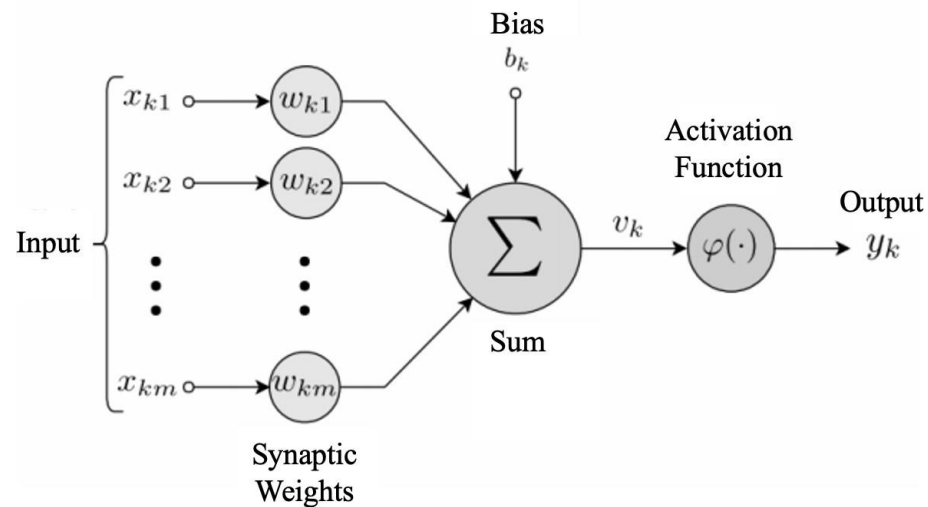


Figure 1. Perceptron.

The output of the neuron is the result of the computation involving the inputs X_k , bias B_k and synaptic weights called W_k . Here, the inputs X_k are multiplied by a synaptic weight W_k added to a bias B_k , which activates or not the neuron through the activation function. The activation function is responsible for controlling the activation of a neuron, and therefore, they are fundamental to the correct operation of the ANN [42]. Commonly, the Sigmoid and ReLu functions are used, which was the one used in the present work. Vanishing Gradient is used with the backpropagation algorithm.

The interconnections between neurons in a neural network is what makes it possible to perform complex tasks [43]. The Multilayer Perceptron architecture organizes perceptrons into multiple layers, which are divided into three parts: the Input Layer, Hidden Layers, and Output Layer [44]. Networks with intermediate layers can implement continuous functions and can perform function approximation by using two intermediate layers [45], which served as the basis for the present work.

2.4.2. Logistic Regression

The Logistic Regression method basically performs a binary classification, which returns the probability that the input data belong to a certain class or not: that is, the estimated probability of a given output (y) for an input (x). The dependent variable is usually binary (nominal or ordinal), and the independent variables can be categorical or continuous. The Logistic Regression model is based on the Sigmoid function, where its output varies between 0 and 1. For this reason, it is a widely used model to describe the probability of something happening or not happening based on the input variables [46].

2.4.3. Decision Tree

A Decision Tree is a classification/regression model, where its structure is in tree form and consists basically of nodes and arcs (also known as branches) [47]. Decision Trees are widely used for a few reasons, such as: it has support for diversified features (categorical and numerical), it has a representation of acquired knowledge that is easily understood, and it is relatively fast to perform the entire training and learning process compared to other algorithms such as ANNs [48].

Each internal node of the tree represents a test on a feature of an instance. The arcs represent the result of each test performed. The outer nodes, also called end nodes or leaf nodes, represent the classification classes. To classify an instance, the tree is run from top to bottom, traversing the nodes and arcs by performing the tests on each node until it arrives at a leaf node, which contains the new classification of the instance. An example of a Decision Tree is shown in Figure 2.

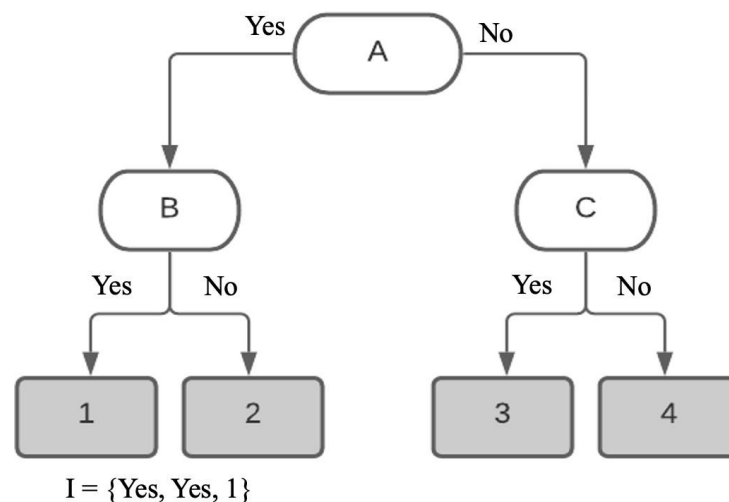


Figure 2. Example of a Decision Tree.

The example in Figure 2 succinctly demonstrates how instance I contains two characteristics: yes and yes. In each node of the tree, a test is executed where a comparison is performed with the features of the instance. The class to which the instance belongs is the third characteristic, which in the case of the example is class 1. By performing the tests, the instance is labeled according to the class of the leaf-node.

2.5. Related Works

Romani's work [49] aims to increase investors' returns by presenting investment products that fit their investment profile through the training of an ANNs. The neural network was implemented in Python and trained using data from investors with the highest returns in their investment portfolios within each investor profile, analyzing investments made in the month of September 2016. In this way, investors who have similar profiles can benefit from the knowledge of the investors with the highest profitability through the trained neural network. According to the author, through the simulations, the total profitability of the investment portfolio increased in most cases. Approximately 61% of the tests had a profitability gain, 4% maintained profitability, and 35% lost profitability.

The research of Lins [12] aimed to analyze Machine Learning techniques to estimate future variations of financial investments in three fund characteristics: conservative, moderate, and aggressive. The data used were collected through the daily share value of three investment funds: Western Asset Investment Fund Equity BDR Level I, JGP Strategy Fund of Investment in Quotas of Multimarket Investment Funds and Daycoval Classic *Fundo de Investimento Renda Fixa Crédito Privado*. With the three funds, data were obtained from May 2014 to September 2020. A comparison was made between the results obtained with neural networks and Linear Regression implemented using the Weka tool. According to the author, the application proved to be effective for forecasting financial investment funds. The ANN proved to be more effective for forecasting, considering various metrics and different databases. To evaluate the results, the following metrics were used: MAE, MSE, RMSE, and MAPE.

The work of Vilela, Penedo and Pereira [50] intended to develop a model with ANNs to forecast the prices of shares traded on the BM&FBovespa, using traditional indicators of profitability, liquidity, and debt. The database used was *Econômática*, with quarterly series referring to 371 companies, from 31 March 2012 to 31 March 2017. The implementation was performed in Matlab software. The type of neural network used was the Multilayer Perceptron. Regarding the results obtained, in cases where no sharp variations occurred, the result obtained was extremely close to the values observed. However, the neural network was not able to generate satisfactory predictions in relation to sudden variations in the trend, which are generally linked to factors external to the indicators.

The work of Aydin and Cavdar [51] aimed to develop an early warning system to predict financial crises in Turkey. The type of neural network used was Multilayer Perceptron. The tests were performed in JAVA language and 298 months of data of 7 key macroeconomic and financial indicators of the Turkish economy between January 1990 and September 2014 were used, obtained from the websites of the Electronic Data Distribution System of the Turkish National Bank (EDDS) and the World Bank—The World Bank. In the conclusion, the author points out that the data obtained with ANNs are impressive; however, it is still a great challenge to perform forecasting in a complex economic environment influenced by external factors, such as crises in other countries and political disturbances, which may influence the reliability of the forecast.

The work by Dingli and Fournier [52] had the objective of developing a system to forecast financial time series using Convolutional Neural Networks to predict the direction of the next period relative to the current price of a stock. Data from Yahoo Finance, which provides historical daily prices, were used to fetch the data from 2003 to 2016. The creation of the neural network was performed using the open-source library TensorFlow. In the results, 65% accuracy was achieved in predicting next month's price direction and 60% accuracy in predicting next week's price direction. In addition to the models that are based on deep layers [53,54] and hybrid models [55], other models are gaining space, such as ensemble learning methods [56–58], neuro-fuzzy systems [59,60], and group method of data handling [61]. In Table 1, a comparison is made between some of these presented related works.

Table 1. Comparison of Related Works.

Authors	Objective Tool	Language Algorithm	Method	Database
Aydin and Cavdar [51]	Predicting financial crises in Turkey	Java	Multilayer Perceptron	Electronic Data Distribution System of the Turkish National Bank and World Bank
Dingli and Fournier [52]	Predicting financial time series	Python	Convolutional Neural Networks	Yahoo Finance
Romani [49]	Increasing the performance of investors	Python	Multilayer Perceptron	Not mentioned by the author
Lins [12]	Estimating variances of financial investments	Weka	Neural Networks and Linear Regression	Investment Funds: Western Asset, JGP and Daycoval Classic
Vilela, Penedo, and Pereira [50]	Predict stock prices	Matlab	Multilayer Perceptron	Económica Database

3. Project

In this project, Machine Learning algorithms capable of evaluating a certain stock and based on this evaluation indicate whether or not to invest in the stock in question were implemented. For the development of the work, different techniques were used, namely: ANNs, Logistic Regression and Decision Tree. The database used is the one provided by Oceans14 [62], containing the history of Fundamental Indicators for each stock. In addition, the historical quotes of each stock obtained through the Yahoo Finance API [63] were used.

Each algorithm performed the stock evaluation, classifying them as indicated or not indicated as a good financial investment. Finally, the results obtained with each technique were compared to identify the technique with the highest accuracy and classification metrics, namely: accuracy, recall and F1-score. Each of the steps will be presented below in more detail, with sections divided into: Database, Algorithms and Validation.

3.1. Database

The database used was the one provided by the Oceans14 website [62], containing the history of Fundamentalists Indicators and the history of Quotes. The most commonly used indicators were used based on the theoretical foundation presented to obtain each indicator. The Fundamentalists Indicators that were used are: ML, EPS, P/E, NAV, P/EV, ROE, and EM. Companies belonging to the financial sector were excluded from the database, according to the Oceans14 [62] classification of sectors. The exclusion was made because the companies in this sector present very distinct characteristics if compared to the companies in other sectors, impairing comparability and consequently the training of algorithms. The tests were performed with data samples from two distinct periods, namely: 1998 to 2019, and 2014 to 2019. The data sample composed of the period from 2014 to 2019 has 1021 records and 13 variables, as follows: share, company, sector, subsector, segment, ML, EPS, P/E, Equity Value per Share, P/EV, ROE, EM, and year.

In the 2014–2019 database, the data prior to the year 2014 were not used due to the fact that the economic recession process started from 2014 [64]. The data from 2020 onwards were also not used, because the COVID-19 pandemic affected the market in an unexpected way, and as the objective of the work is to work with fundamentalist analysis, adding the data from the pandemic period to the database used in the project would only hinder the training, generating inconsistencies that totally interfere with the reliability of the results obtained with the work.

As the Oceans14 website [62] contains data records since 1998, tests were also performed with the data from 1998 to 2019, containing 2627 records with the purpose of identifying which database (1998–2019 or 2014–2019) obtained the best results for the present work. To use the data, it was necessary to perform the classification of each stock as indicated or not indicated as a good financial investment, where “1” represents a stock classified as a good investment and “0” represents a stock that is not considered a good investment. This classification was performed by comparing the variation of each share’s quotation with the variation of the Selic Rate in the same period.

The variation in the price of each share was obtained by comparing the price of each share on the date recorded in the database, with the price of the same share, but referring to 5 years after the recorded date, based on the return time and the profitability references of fixed income investments of the main Benchmarks demonstrated by Araujo [64], such as: savings, IPCA, and CDI, among others. The variation of the Selic Rate was obtained through the History of Basic Interest Rates presented by the site of the Central Bank of Brazil.

3.2. Algorithms

In this paper, three different algorithms were implemented: Multilayer Perceptron, Logistic Regression, and Decision Tree, with the purpose of performing the tests and comparing which one generates better investment indications based on Fundamentalist Analysis. The type of Neural Network used was the Multilayer Perceptron. The Multilayer Perceptron is formed by a set of source nodes that form the input layer (Input Layer), hidden layers (Hidden Layers) consisting of neurons, which can be one or more, and an output layer (Output Layer) [21].

The learning method used in the ANN was Feed Forward Backpropagation, in which the error is calculated by performing the reverse path, that is, from the last layers to the first layers of the network [44]. The number of neurons for the input layer is equal to the number of variables present in the database. The number of neurons in the hidden layer was defined empirically by performing tests with different numbers of neurons in order to select the number of neurons that presents the best accuracy.

3.3. Validation

To perform the performance comparison of the different models, evaluation metrics were used that are calculated through the Confusion Matrix. The metrics Accuracy, Precision, Recall and F1-score were used to more accurately evaluate the results obtained with the classifications of each algorithm, to compare them and obtain the one that achieved the best results for the tests performed in this paper, which is given by:

$$\text{accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predctions made}} \quad (1)$$

$$\text{precision} = \frac{tp}{tp + fp} \quad (2)$$

$$\text{recall} = \frac{tp}{tp + fn} \quad (3)$$

$$\text{F1 - Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

where tp is true positive, fn is false negative, and fp is false positive.

4. Development

This section will present the steps taken to develop this work, which are: database preparation, Neural Network, Decision Tree, and Logistic Regression. The first step of the development was to prepare the database for the algorithms that would be implemented later. The data preparation was performed in two steps, namely: exclusion of unnecessary data and transforming the data format to suit the algorithms. At first, the columns that were unnecessary for the implementation of the algorithms and that are not part of the fundamental indicators, which are the main basis of Fundamental Analysis, on which this work is structured, were removed. Thus, the columns removed were: "VAR PRICE", "SUBSECTOR", "SEGMENT", "STOCK", "COMPANY" and "YEAR".

In addition to removing unnecessary columns from the database, it was necessary to transform the columns that had their data in String form, to Int. That is, the "SECTOR" column was transformed into Int using Sklearn's "LabelEncoder" function, which uses an integer to represent each class of the column in question, so that the data are represented only in numeric format. After all the data preparation, the splitting of the training and test sets was performed using Sklearn's "train_test_split" function, where the data were split into 30% for training and 70% for testing.

4.1. ANN Implementation

The ANN was implemented using Sklearn's MLPClassifier, which is a neural network that is intended to work with classification problems. The Activation Function used in the Hidden Layers is ReLu, which has no negative part, presenting a set of outputs between zero and infinity. Here, zero represents the neuron as deactivated, and values greater than zero represent activation of the neuron [65]. To define the configuration of the Artificial Neural Net parameters, tests were performed using Sklearn's GridSearchCV function, which is basically used to automate the process of adjusting the parameters of a given algorithm, performing several combinations of parameters, and evaluating each configuration to obtain the one that presents the best results.

The number of neurons in the two hidden layers was defined through the parameter "hidden_layers_sizes", the Learning Rate was defined through the parameter "learning_rate_init", and the Number of epochs was defined through the parameter "max_iter". After running GridSearchCV, the best parameters were obtained through the method "grid.best_params", being them: 4 neurons in the first hidden layer, 3 neurons in the second hidden layer, learning rate of 0.01 and number of epochs of 1000.

After defining the settings of the ANNs, the tests were performed using the 2014–2019 database. The Accuracy obtained with the model was 74%. Data classified as “1” obtained an accuracy of 0.74, recall of 0.96 and F1-score of 0.84. However, data classified as “0” obtained a precision of 0.72, recall of 0.22, and F1-score of 0.34. In other words, among the data that should be classified as “0”, only 22% was actually classified as “0”. Considering that the label “0” represents a Stock that is not indicated as a good investment option, having a recall of only 0.22 indicates that the algorithm is classifying 78% of the stocks that should not be indicated as a good investment option as being a good investment option. This negatively influences the decision making of an investor who might use the algorithm to help decide, causing him to invest in a stock that should not be classified as a good investment choice.

At first, a database containing 1021 stock records from 2014 to 2019 was used. However, the Oceans14 website [62], from where the data were collected, contains records since 1998. Therefore, to perform further tests and to test the algorithms, tests were performed using the database containing the Fundamentalist Stock Indicators since 1998. With the insertion of data since 1998, the database has 2627 records, of which 1231 have the label “0” and 1396 have the label “1”. After training and testing the ANNs using the database with records from 1998 to 2019, a report was generated with the main classification metrics. Figure 3 shows the metrics of the ANNs using the database from 1998 to 2019.

	precision	recall	f1-score	support
0	0.66	0.68	0.67	365
1	0.72	0.69	0.71	424
accuracy			0.69	789
macro avg	0.69	0.69	0.69	789
weighted avg	0.69	0.69	0.69	789

Figure 3. Metrics of the Artificial Neural Network.

In the metrics presented in Figure 3, the Accuracy obtained with the model was 69%. Data classified as “1” obtained an accuracy of 0.72, recall of 0.69 and F1-score of 0.71. Data classified as “0” obtained a precision of 0.66, recall of 0.68 and F1-score of 0.67. Figure 4 shows the Confusion Matrix of the ANNs.

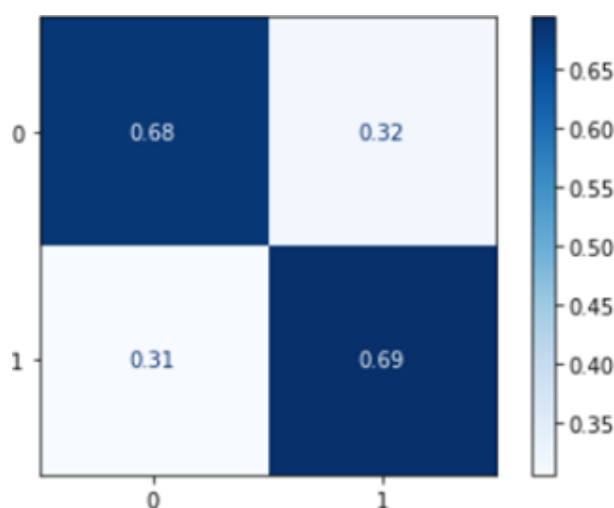


Figure 4. Confusion Matrix of the Artificial Neural Network.

In the Confusion Matrix presented in Figure 4, it is shown that data labeled as “1” are 69% correctly classified. The data labeled as “0” obtained 68% correct classification.

4.2. Decision Tree

The Decision Tree was implemented using Sklearn’s DecisionTreeClassifier, which is a Decision Tree that is intended to work with classification problems. To set the parameters of the algorithm, the GridSearchCV was used, which was also used in the ANNs. Three parameters were chosen to be tested with GridSearchCV: class_weight, criterion and max_depth. The class_weight is a parameter used to work with databases that are unbalanced, which is the case of this work, where the database contains more data labeled as “1” (stocks considered as being a good investment) than labeled as “0” (stocks considered as not being a good investment option). The criterion is a parameter that basically defines which rule/criterion the Decision Tree will use to generate the decision and may vary depending on the application. The max_depth is basically the maximum height of the Decision Tree, which if too small can generate the problem of underfitting and if too large can generate overfitting.

After running GridSearchCV, the best parameters for the Decision Tree were generated, which are: class_weight set to “balanced”, criterion set to “gini” and max_depth set to 13. After defining the Decision Tree settings, the tests were performed using the 2014–2019 database. The Accuracy obtained with the model was 76%. The data classified as “1” obtained an accuracy of 0.83, recall of 0.81 and F1-score of 0.82. However, data classified as “0” had an accuracy of 0.62, recall of 0.66, and F1-score of 0.64. After performing the training and testing using the 2014–2019 database, training and testing was performed using the 1998–2019 database, which obtained the classification metrics shown in Figure 5.

	precision	recall	f1-score	support
0	0.79	0.73	0.76	395
1	0.75	0.80	0.78	394
accuracy			0.77	789
macro avg	0.77	0.77	0.77	789
weighted avg	0.77	0.77	0.77	789

Figure 5. Decision Tree Classification Metrics.

In the metrics presented in Figure 5, it can be seen that the Accuracy obtained with the model was 77%. Data classified as “1” had an accuracy of 0.75, recall of 0.80, and F1-score of 0.78. Data classified as “0” obtained an accuracy of 0.79, recall of 0.73, and F1-score of 0.76. Similarly to the ANNs, when performing the comparison of the metrics obtained with the 1998–2019 database, compared to those obtained with the 2014–2019 database, the improvement of the metrics of the data classified as “0” is noted. Figure 6 shows the Confusion Matrix obtained with the Decision Tree.

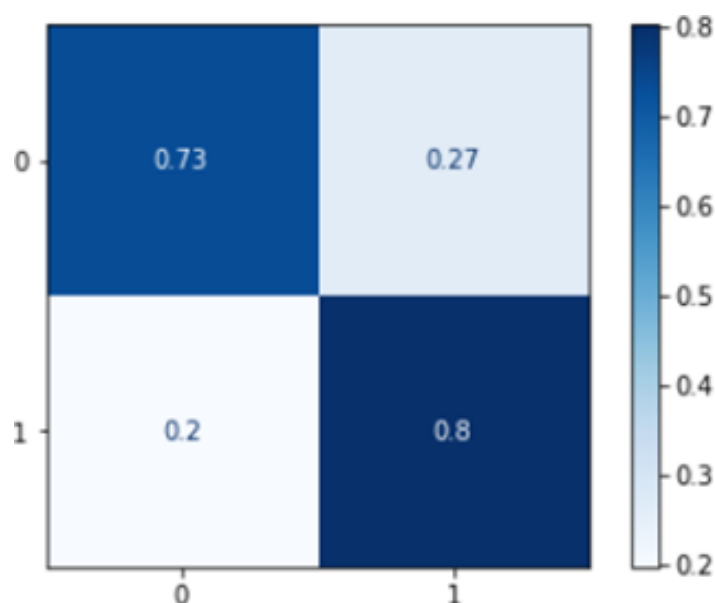


Figure 6. Decision Tree Confusion Matrix.

In the Confusion Matrix presented in Figure 6, it is shown that the data labeled as “1” are 80% correctly classified. The data labeled “0” have a 73% correct classification.

4.3. Logistic Regression

The regression method basically performs the prediction of Y_i from the knowledge of x_i . There are several regression techniques that estimate the relationship between variables, for example: logistic, linear, non-linear, and ridge, among others [46]. Linear regression refers to the relationship between variables, where the conditional expectation of Y , given $X = x$, is a linear function of x [66]. In a model based on simple or multiple linear regression, the dependent variable Y is a variable of a continuous nature. However, it can also be qualitative, where it is represented by two or more categories. In this case, the least squares optimization method does not provide a reasonable estimator. The ridge model solves a regression model where the loss function is the linear least square’s function and regularization is given by the l2-norm and has built-in support for multi-variate regression [67]. Therefore, logistic regression offers a more appropriate approximation, allowing the use of regression models to calculate or predict the probability of a given event [68].

The Logistic Regression algorithm was implemented using Sklearn’s LogisticRegression. To define the parameters of the algorithm, GridSearchCV was used, which was also used for the same purpose in the ANNs and Decision Tree. Three parameters were chosen to be tested with GridSearchCV, namely: class_weight, max_iter and solver. The class_weight is a parameter used to work with databases that are unbalanced, which is the case in this work. The max_iter is basically the number of “epochs”, like the parameter defined in the ANNs. The solver is basically the algorithm to be used for optimization.

After running GridSearchCV, the best parameters for the Logistic Regression were generated, being class_weight set to “None”, max_iter set to 100 and solver set to “newtoncg”. After setting the Logistic Regression settings, the tests were performed using the 2014–2019 database. The Accuracy obtained with the model was 64%. Data classified as “1” obtained an accuracy of 0.67, recall of 0.89 and F1-score of 0.76. However, data classified as “0” obtained an accuracy of 0.41, recall of 0.14, and F1-score of 0.21. After performing the training and testing using the 2014–2019 database, training and testing was performed using the 1998–2019 database, which obtained the classification metrics shown in Figure 7.

	precision	recall	f1-score	support
0	0.62	0.64	0.63	361
1	0.69	0.67	0.68	428
accuracy			0.66	789
macro avg	0.65	0.66	0.65	789
weighted avg	0.66	0.66	0.66	789

Figure 7. Logistic Regression Classification Metrics.

In the metrics presented in Figure 7, one can see that the Accuracy obtained with the model was 66%. Data classified as “1” obtained an accuracy of 0.69, recall of 0.67 and F1-score of 0.68. Data classified as “0” obtained an accuracy of 0.62, recall of 0.64 and F1-score of 0.63. Similarly to the ANNs and Decision Tree, when performing the comparison of the metrics obtained with the 1998–2019 database, compared to those obtained with the 2014–2019 database, the improvement in the metrics of the data classified as “0” is noted. Figure 8 shows the Confusion Matrix obtained with the Decision Tree.

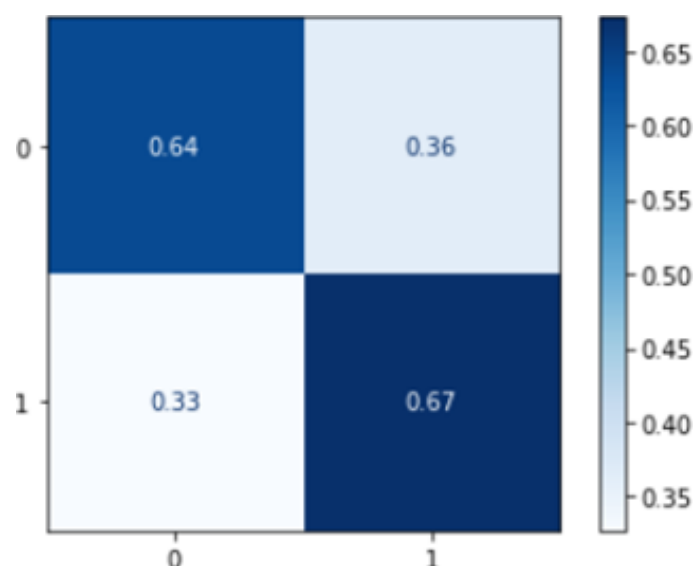


Figure 8. Confusion Matrix of the Logistic Regression.

In the Confusion Matrix presented in Figure 8, it is shown that the data labeled as “1” are 67% correctly classified. The data labeled as “0”, on the other hand, have a 64% correct classification.

5. Results

This section will present the results obtained with the algorithms implemented in this work. First, a comparison will be presented regarding the accuracy obtained with each algorithm using the 1998–2019 database, as shown in Figure 9.

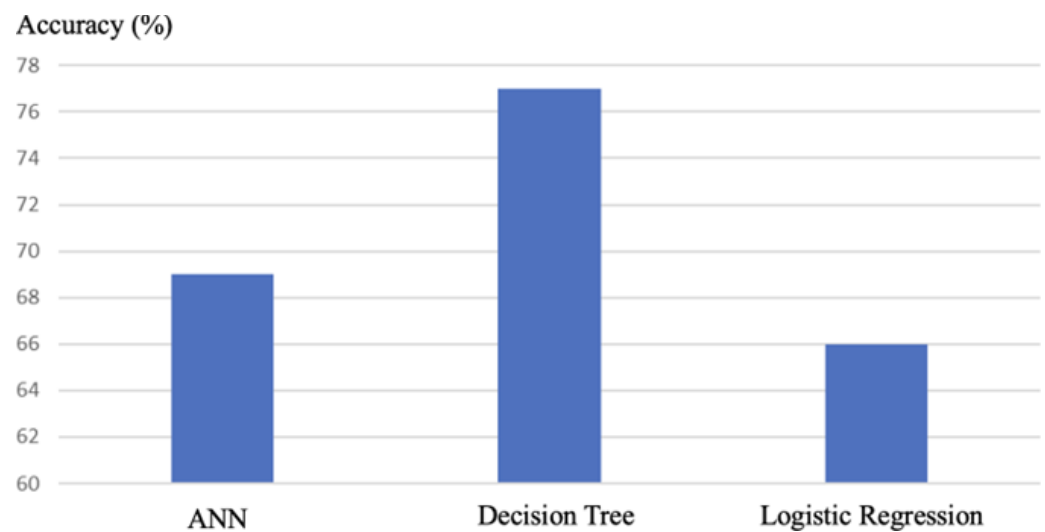


Figure 9. Comparison of algorithm Accuracy.

In the graph shown in Figure 9, the Decision Tree has the highest accuracy, with 77%. This is followed by 69% accuracy for the ANNs and 66% for the Logistic Regression. However, to analyze the results in a deeper and more adequate way, it is necessary to analyze the classification metrics, as presented in Figure 10.

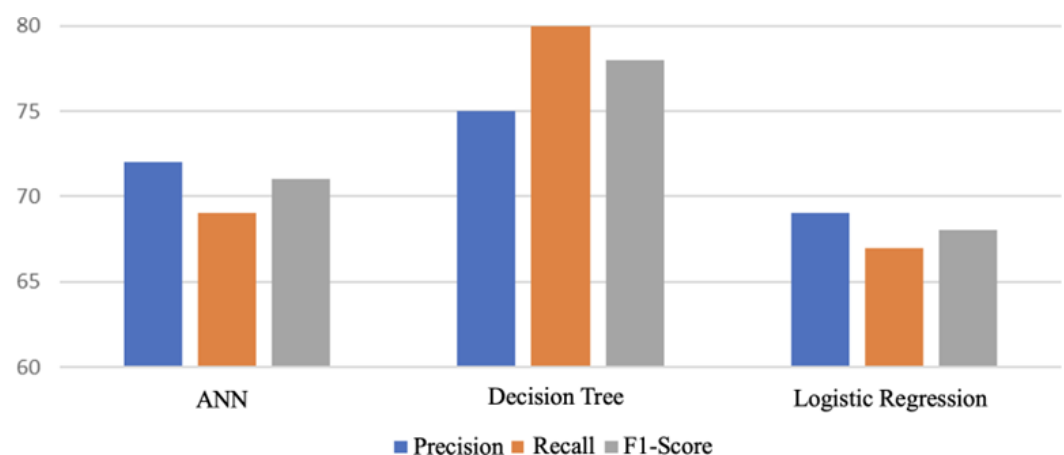


Figure 10. Comparison of Metrics for Class “1” Classification.

In the graph shown in Figure 10, the metrics for the three algorithms are presented. The ANN has an accuracy of 0.72, recall of 0.69 and F1-score of 0.71. The Decision Tree has an accuracy of 0.75, recall of 0.80 and F1-score of 0.78. Finally, the Logistic Regression has an accuracy of 0.69, recall of 0.67 and F1-score of 0.68. Decision Tree obtained higher values for its classification metrics compared to the other algorithms. It is worth noting that the Recall of the Decision Tree was higher than the Precision and F1-score. This indicates that the algorithm obtained a low number of false negatives for Class “1”. Figure 11 shows the metrics obtained with the three algorithms for Class “0”.

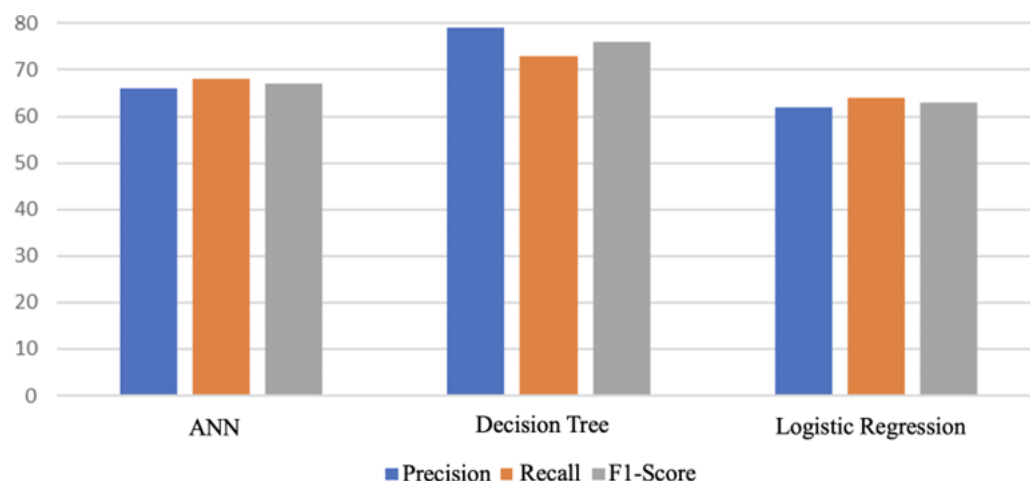


Figure 11. Comparison of Metrics for Class “0” Classification.

In the graph shown in Figure 11, the metrics of the three algorithms for Classes “0” are presented. The ANN has an accuracy of 0.66, recall of 0.68, and F1-score of 0.67. The Decision Tree has an accuracy of 0.79, recall of 0.73 and F1-score of 0.76. Finally, the Logistic Regression has an accuracy of 0.62, recall of 0.64 and F1-score of 0.63.

Through the numbers presented in each metric of the graphs presented, it is noted that the Decision Tree presented higher values for all metrics both for Class “1” and for Class “0”. It is worth noting that in the application of this work, bad stocks (Class “0”) classified as good (Class “1”) generate worse consequences for the investor than good stocks (Class “1”) classified as bad (Class “0”).

Making an investment in a bad stock generates losses for the investor. However, not making an investment in a good stock does not generate profit, but it also does not generate loss. Therefore, to select the best algorithm for the application of this work, it is necessary to analyze the metrics of the Class “0” classifications (stocks considered bad for investment). Thus, Decision Tree was really the algorithm that presented the best values for precision, recall and F1-score for the application of this work.

6. Conclusions

This paper was dedicated to implementing Machine Learning algorithms with the ability to act as financial investment advisors for new investors. The proposed objectives were achieved, namely: identify market and investor profile data that are needed to generate reliable investment suggestions; perform preprocessing and database assembly; define algorithms that will be used. implement Machine Learning algorithms; perform validation and performance tests.

The Fundamental Analysis was used as a basis for evaluating and generating investment indications, using the Fundamental Indicators of each stock as variables in the Database. The Fundamental Indicators used in the Database were obtained from the Oceans14 website [62]. However, the Fundamental Indicators alone were not enough for the implementation. It was necessary to label each registered stock as good (Class “1”) or bad (Class “0”) to invest in. Thus, it was necessary to compare the variation of each share’s price with the variation of the Selic Rate in the same period. If the stock variation was greater than the Selic Rate variation, the stock in question was labeled as Class “1”; otherwise, it was labeled as Class “0”. To perform the calculations, it was to use the Yahoo Finance quotation history and the Brazilian Central Bank’s Basic Interest Rate history. In addition, preparation was carried out in the database, removing unnecessary variables for the implementation of the work and stocks from companies in the financial sector.

The Machine Learning techniques used were: Multilayer Perceptron, Decision Tree, and Logistic Regression. The implementation was completed in Python language, using Google Colaboratory and the Machine Learning library Sikit-learn. The settings for each

algorithm were defined using the GridSearchCV function to find the best settings for the implementation of this work. With the parameter settings of each Artificial Intelligence technique defined, the algorithms were implemented. At first, only the database with records from 2014 to 2019 would be used, but the results obtained with records from only this period were not satisfactory, because the classification metrics for Classes “0” were very low. For this reason, new tests were subsequently performed with the Database containing records from 1998 to 2019, which in turn obtained better accuracy, precision, recall, and F1-score.

The results obtained with each algorithm were compared using the metrics: accuracy, precision, recall, and F1-score. The algorithm that obtained the best classification metrics was Decision Tree. The accuracy obtained with Decision Tree was 0.77. The metrics for Class “0” were accuracy of 0.79, recall of 0.73, and F1-score of 0.76. Metrics for Class “1” were accuracy of 0.75, recall of 0.80, and F1-score of 0.78.

Regarding the Decision Tree accuracy, the results lead us to consider that the value achieved was not better due to the market behavior itself, which has numerous variables that exceed what is “predicted” by the indicators of Fundamental Analysis, for example: economic crises, wars, natural disasters, etc. As the work was based on Fundamental Analysis, the forecast made by the models is limited to the information provided by the indicators (used as inputs to the models).

Future Works

For the future development of the research in the area of the present work, some points can be developed, such as using a database with more records, such as the one provided by Economática, which is paid, but very complete. This would help in the training of algorithms and consequently in the results achieved by them. It could also be beneficial to use more or different Fundamental Indicators. In this work we have used some of the main indicators, but there are many others that if well chosen, in conjunction with other indicators can assist in the analysis of stock valuation by the algorithms. Lastly, use the yield of other investments instead of the Selic Rate as a criterion for labeling the stocks in the database is an area that needs further research.

Author Contributions: Writing—original draft, methodology, software, validation, and formal analysis, A.V.d.O.; Writing—review and editing, M.C.S.D.; Writing—review and editing, A.M.d.R.F.; Writing—review and editing, R.L.S.D.; Writing—review and editing, supervision, P.F.; Supervision, and project administration, V.R.Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by national funds through the Foundation for Science and Technology, I.P. (Portuguese Foundation for Science and Technology) by the project UIDB/05064/2020 (VALORIZA—Research Center for Endogenous Resource Valorization), and Project UIDB/04111/2020, ILIND—Lusophone Institute of Investigation and Development, under project COFAC/ILIND/COPELABS/3/2020.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. B3. Brazil Stock Exchange. Available online: <https://www.b3.com.br/> (accessed on 14 September 2022).
2. Fernandes, F.; Stefenon, S.F.; Seman, L.O.; Nied, A.; Ferreira, F.C.S.; Subtil, M.C.M.; Klaar, A.C.R.; Leithardt, V.R.Q. Long short-term memory stacking model to predict the number of cases and deaths caused by COVID-19. *J. Intell. Fuzzy Syst.* **2022**, *6*, 6221–6234. <https://doi.org/10.3233/JIFS-212788>.
3. Salazar, L.H.; Fernandes, A.; Dazzi, R.; Garcia, N.; Leithardt, V.R. Using different models of machine learning to predict attendance at medical appointments. *J. Inf. Syst. Eng. Manag.* **2020**, *5*, 0122. <https://doi.org/10.29333/jisem/8430>.
4. Vieira, J.C.; Sartori, A.; Stefenon, S.F.; Perez, F.L.; de Jesus, G.S.; Leithardt, V.R.Q. Low-Cost CNN for Automatic Violence Recognition on Embedded System. *IEEE Access* **2022**, *10*, 25190–25202. <https://doi.org/10.1109/ACCESS.2022.3155123>.

5. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Yow, K.C.; Mariani, V.C.; dos Santos Coelho, L.; Seman, L.O. Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam. *Electr. Power Syst. Res.* **2022**, *202*, 107584. <https://doi.org/10.1016/j.epsr.2021.107584>.
6. Sopelsa Neto, N.F.; Stefenon, S.F.; Meyer, L.H.; Ovejero, R.G.; Leithardt, V.R.Q. Fault Prediction Based on Leakage Current in Contaminated Insulators Using Enhanced Time Series Forecasting Models. *Sensors* **2022**, *22*, 6121. <https://doi.org/10.3390/s22166121>.
7. Stefenon, S.F.; Singh, G.; Yow, K.C.; Cimatti, A. Semi-ProtoPNet Deep Neural Network for the Classification of Defective Power Grid Distribution Structures. *Sensors* **2022**, *22*, 4859. <https://doi.org/10.3390/s22134859>.
8. Corso, M.P.; Perez, F.L.; Stefenon, S.F.; Yow, K.C.; Ovejero, R.G.; Leithardt, V.R.Q. Classification of Contaminated Insulators Using k-Nearest Neighbors Based on Computer Vision. *Computers* **2021**, *10*, 112. <https://doi.org/10.3390/computers10090112>.
9. Lee, T.K.; Cho, J.H.; Kwon, D.S.; Sohn, S.Y. Global stock market investment strategies based on financial network indicators using machine learning techniques. *Expert Syst. Appl.* **2019**, *117*, 228–242. <https://doi.org/10.1016/j.eswa.2018.09.005>.
10. Cardozo, T.; Modesto, N.L.P.; Magalhães, N.P.; Fonseca, R.V.S.; Policarpo, R.V.S. Análise do Perfil de Investidores Brasileiros. In Proceedings of the IX Congresso Brasileiro de Engenharia de Produção, Ponta Grossa-Paraná, Brazil, 4–6 December 2019; Volume 4, pp. 1–12.
11. Paiva, R.T.; Silva, H.A.; de Souza, J.C.M.; Novôa, N.F.; de Araújo Pereira, C.M.M. O perfil do investidor individual no mercado financeiro. *Rev. Vianna Sapiens* **2020**, *11*, 30. <https://doi.org/10.31994/rvs.v11i2.694>.
12. Lins, R.N.F. Previsão de Fundos de Investimentos com o Uso de Machine Learning. Bachelor's Thesis, Universidade Tecnológica Federal do Paraná, Curitiba, Brazil, 2020.
13. Lim, K.; Halim, A.; Lu, T.S.; Ashworth, A.; Chong, I. Klotho: A major shareholder in vascular aging enterprises. *Int. J. Mol. Sci.* **2019**, *20*, 4637. <https://doi.org/10.3390/ijms20184637>.
14. Gava Gastaldo, N.; Rediske, G.; Donaduzzi Rigo, P.; Brum Rosa, C.; Michels, L.; Mairesse Siluk, J.C. What is the profile of the investor in household solar photovoltaic energy systems? *Energies* **2019**, *12*, 4451. <https://doi.org/10.3390/en12234451>.
15. Razali, M.N.; Jalil, R.A.; Achu, K.; Ali, H.M. Identification of Risk Factors in Business Valuation. *J. Risk Financ. Manag.* **2022**, *15*, 282. <https://doi.org/10.3390/jrfm15070282>.
16. da Rosa, M.R.; Bered, R. A Importância da Análise Fundamentalista para Avaliar o Preço das Ações de Companhias Listadas na Bolsa de Valores (B3). *Rev. Eletrônica De Ciências Contábeis* **2018**, *7*, 124–150.
17. De Bortoli, D.; da Costa, N., Jr.; Goulart, M.; Campara, J. Personality traits and investor profile analysis: A behavioral finance study. *PLoS ONE* **2019**, *14*, e0214062. <https://doi.org/10.1371/journal.pone.0214062>.
18. Dickason, Z.; Ferreira, S. Establishing a link between risk tolerance, investor personality and behavioural finance in South Africa. *Cogent Econ. Financ.* **2018**, *6*, 1519898. <https://doi.org/10.1080/23322039.2018.1519898>.
19. Hapsoro, D.; Husain, Z.F. Does sustainability report moderate the effect of financial performance on investor reaction? Evidence of Indonesian listed firms. *Int. J. Bus.* **2019**, *24*, 308–328.
20. Vasudevan, E.V. Some Gains Are Riskier Than Others: Volatility Changes, Belief Revisions, and the Disposition Effect. *Belief Revis. Dispos. Eff.* **2019**, *1*, 1–54. <https://doi.org/10.2139/ssrn.3363580>.
21. Sopelsa Neto, N.F.; Stefenon, S.F.; Meyer, L.H.; Bruns, R.; Nied, A.; Seman, L.O.; Gonzalez, G.V.; Leithardt, V.R.Q.; Yow, K.C. A Study of Multilayer Perceptron Networks Applied to Classification of Ceramic Insulators Using Ultrasound. *Appl. Sci.* **2021**, *11*, 1592. <https://doi.org/10.3390/app11041592>.
22. Stefenon, S.F.; Corso, M.P.; Nied, A.; Perez, F.L.; Yow, K.C.; Gonzalez, G.V.; Leithardt, V.R.Q. Classification of insulators using neural network based on computer vision. *IET Gener. Transm. Distrib.* **2021**, *16*, 1096–1107. <https://doi.org/10.1049/gtd2.12353>.
23. Salazar, L.H.A.; Leithardt, V.R.Q.; Parreira, W.D.; da Rocha Fernandes, A.M.; Barbosa, J.L.V.; Correia, S.D. Application of Machine Learning Techniques to Predict a Patient's No-Show in the Healthcare Sector. *Future Internet* **2022**, *14*, 3. <https://doi.org/10.3390/fi14010003>.
24. Salazar, L.H.; Fernandes, A.M.R.; Dazzi, R.; Raduenz, J.; Garcia, N.M.; Leithardt, V.R.Q. Prediction of Attendance at Medical Appointments Based on Machine Learning. In Proceedings of the 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), Sevilla, Spain, 24–27 June 2020; pp. 1–6. <https://doi.org/10.23919/CISTI49556.2020.9140973>.
25. Stefenon, S.F.; Seman, L.O.; Sopelsa Neto, N.F.; Meyer, L.H.; Nied, A.; Yow, K.C. Echo state network applied for classification of medium voltage insulators. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107336. <https://doi.org/10.1016/j.ijepes.2021.107336>.
26. Leithardt, V. Classifying garments from fashion-MNIST dataset through CNNs. *Adv. Sci. Technol. Eng. Syst. J.* **2021**, *6*, 989–994.
27. Stefenon, S.F.; Silva, M.C.; Bertol, D.W.; Meyer, L.H.; Nied, A. Fault diagnosis of insulators from ultrasound detection using neural networks. *J. Intell. Fuzzy Syst.* **2019**, *37*, 6655–6664. <https://doi.org/10.3233/JIFS-190013>.
28. dos Santos, G.H.; Seman, L.O.; Bezerra, E.A.; Leithardt, V.R.Q.; Mendes, A.S.; Stefenon, S.F. Static Attitude Determination Using Convolutional Neural Networks. *Sensors* **2021**, *21*, 6419. <https://doi.org/10.3390/s21196419>.
29. da Silva, L.D.L.; Pereira, T.F.; Leithardt, V.R.Q.; Seman, L.O.; Zeferino, C.A. Hybrid Impedance-Admittance Control for Upper Limb Exoskeleton Using Electromyography. *Appl. Sci.* **2020**, *10*, 7146. <https://doi.org/10.3390/app10207146>.
30. Stefenon, S.F.; Furtado Neto, C.S.; Coelho, T.S.; Nied, A.; Yamaguchi, C.K.; Yow, K.C. Particle swarm optimization for design of insulators of distribution power system based on finite element method. *Electr. Eng.* **2022**, *104*, 615–622. <https://doi.org/10.1007/s00202-021-01332-3>.
31. Stefenon, S.F.; Seman, L.O.; Pavan, B.A.; Ovejero, R.G.; Leithardt, V.R.Q. Optimal design of electrical power distribution grid spacers using finite element method. *IET Gener. Transm. Distrib.* **2022**, *16*, 1865–1876. <https://doi.org/10.1049/gtd2.12425>.

32. da Cruz, F.C.; Stefenon, S.F.; Furtado, R.G.; Rocca, G.A.D.; Ferreira, F.C.S. Financial Feasibility Study for Radio Installation Link on the Mobile Telephone Network. *Rev. GEINTEC-Gestão Inovação E Tecnol.* **2018**, *8*, 4447–4460.
33. Righez, F.O.; Dela Rocca, G.A.; Arruda, P.A.; Stefenon, S.F. Analysis of Technical and Financial Viability of a Fixed Site Internet Broadband. *Rev. Gestão Inovação E Tecnol.* **2016**, *6*, 3537–3552.
34. Leithardt, V.; Santos, D.; Silva, L.; Viel, F.; Zeferino, C.; Silva, J. A Solution for Dynamic Management of User Profiles in IoT Environments. *IEEE Lat. Am. Trans.* **2020**, *18*, 1193–1199. <https://doi.org/10.1109/TLA.2020.9099759>.
35. Siddiqui, S.; Nesbitt, R.; Shakir, M.Z.; Khan, A.A.; Khan, A.A.; Khan, K.K.; Ramzan, N. Artificial Neural Network (ANN) Enabled Internet of Things (IoT) Architecture for Music Therapy. *Electronics* **2020**, *9*, 2019. <https://doi.org/10.3390/electronics9122019>.
36. Viel, F.; Silva, L.A.; Valderi Leithardt, R.Q.; Zeferino, C.A. Internet of Things: Concepts, Architectures and Technologies. In Proceedings of the 2018 13th IEEE International Conference on Industry Applications (INDUSCON), São Paulo, Brazil, 11–14 November 2018; pp. 909–916. <https://doi.org/10.1109/INDUSCON.2018.8627298>.
37. Muniz, R.N.; Stefenon, S.F.; Buratto, W.G.; Nied, A.; Meyer, L.H.; Finardi, E.C.; Kühl, R.M.; Sá, J.A.S.d.; Rocha, B.R.P.d. Tools for Measuring Energy Sustainability: A Comparative Review. *Energies* **2020**, *13*, 2366. <https://doi.org/10.3390/en13092366>.
38. Leithardt, V.R.Q.; Rolim, C.; Rosseto, A.; Geyer, C.; Dantas, M.A.R.; Silva, J.S.; Nunes, D. Percontrol: A pervasive system for educational environments. In Proceedings of the 2012 International Conference of Computing. Networking and Communication. (ICNC), Maui, HI, USA, 30 January–2 February 2012; pp. 131–136. <https://doi.org/10.1109/ICCNC.2012.6167396>.
39. Pinto, H.; Américo, J.; Leal, O.; Stefenon, S. Development of Measurement Device and Data Acquisition for Electric Vehicle. *Rev. GEINTEC* **2021**, *11*, 5809–5822.
40. Silva, L.A.; Leithardt, V.R.Q.; Rolim, C.O.; González, G.V.; Geyer, C.F.R.; Silva, J.S. PRISER: Managing Notification in Multiples Devices with Data Privacy Support. *Sensors* **2019**, *19*, 3098. <https://doi.org/10.3390/s19143098>.
41. Lopes, H.; Pires, I.M.; Sánchez San Blas, H.; García-Ovejero, R.; Leithardt, V. PriADA: Management and Adaptation of Information Based on Data Privacy in Public Environments. *Computers* **2020**, *9*, 77. <https://doi.org/10.3390/computers9040077>.
42. Medeiros, A.; Sartori, A.; Stefenon, S.F.; Meyer, L.H.; Nied, A. Comparison of artificial intelligence techniques to failure prediction in contaminated insulators based on leakage current. *J. Intell. Fuzzy Syst.* **2022**, *42*, 3285–3298. <https://doi.org/10.3233/JIFS-211126>.
43. Suzin, J.C.; Zeferino, C.A.; Leithardt, V.R.Q. Digital Statelessness. In *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*; de Paz Santana, J.F., de la Iglesia, D.H., López Rivero, A.J., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 178–189. https://doi.org/10.1007/978-3-030-87687-6_18.
44. Stefenon, S.F.; Branco, N.W.; Nied, A.; Bertol, D.W.; Finardi, E.C.; Sartori, A.; Meyer, L.H.; Grebogi, R.B. Analysis of training techniques of ANN for classification of insulators in electrical power systems. *IET Gener. Transm. Distrib.* **2020**, *14*, 1591–1597. <https://doi.org/10.1049/iet-gtd.2019.1579>.
45. Stefenon, S.F.; Grebogi, R.B.; Freire, R.Z.; Nied, A.; Meyer, L.H. Optimized Ensemble Extreme Learning Machine for Classification of Electrical Insulators Conditions. *IEEE Trans. Ind. Electron.* **2020**, *67*, 5170–5178. <https://doi.org/10.1109/TIE.2019.2926044>.
46. Christodoulou, E.; Ma, J.; Collins, G.S.; Steyerberg, E.W.; Verbakel, J.Y.; Van Calster, B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **2019**, *110*, 12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>.
47. Meza, J.K.S.; Yepes, D.O.; Rodrigo-Illarri, J.; Cassiraga, E. Predictive analysis of urban waste generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks. *Heliyon* **2019**, *5*, e02810. <https://doi.org/10.1016/j.heliyon.2019.e02810>.
48. Aghaei, S.; Azizi, M.J.; Vayanos, P. Learning optimal and fair decision trees for non-discriminative decision-making. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA, 27 January 27–1 February 2019; Volume 33, pp. 1418–1426. <https://doi.org/10.1609/aaai.v33i01.33011418>.
49. Romani, L.F. *Aplicação de Redes Neurais Artificiais na Sugestão de Investimentos*; Universidade de Brasília: Brasília, Portugal, 2017; pp. 1–94.
50. Vilela, E.H.P.; Penedo, A.S.T.; Pereira, V.S. Aplicação de Redes Neurais Artificiais na Predição de Preços de Ações por Indicadores Financeiros. *Desafio Online* **2018**, *6*, 2.
51. Aydin, A.D.; Cavdar, S.C. Prediction of financial crisis with artificial neural network: An empirical analysis on Turkey. *Int. J. Financ. Res.* **2015**, *6*, 36.
52. Dingli, A.; Fournier, K.S. Financial time series forecasting-a deep learning approach. *Int. J. Mach. Learn. Comput.* **2017**, *7*, 118–122. <https://doi.org/10.18178/ijmlc.2017.7.5.632>.
53. Stefenon, S.F.; Kasburg, C.; Nied, A.; Klaar, A.C.R.; Ferreira, F.C.S.; Branco, N.W. Hybrid deep learning for power generation forecasting in active solar trackers. *IET Gener. Transm. Distrib.* **2020**, *14*, 5667–5674. <https://doi.org/10.1049/iet-gtd.2020.0814>.
54. Kasburg, C.; Stefenon, S.F. Deep Learning for Photovoltaic Generation Forecast in Active Solar Trackers. *IEEE Lat. Am. Trans.* **2019**, *17*, 2013–2019. <https://doi.org/10.1109/TLA.2019.9011546>.
55. Stefenon, S.F.; Freire, R.Z.; Meyer, L.H.; Corso, M.P.; Sartori, A.; Nied, A.; Klaar, A.C.R.; Yow, K.C. Fault detection in insulators based on ultrasonic signal processing using a hybrid deep learning technique. *IET Sci. Meas. Technol.* **2020**, *14*, 953–961. <https://doi.org/10.1049/iet-smt.2020.0083>.

56. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.S.; Leithardt, V.R.Q.; Silva, L.A.; Seman, L.O. Hybrid Wavelet Stacking Ensemble Model for Insulators Contamination Forecasting. *IEEE Access* **2021**, *9*, 66387–66397. <https://doi.org/10.1109/ACCESS.2021.3076410>.
57. Ribeiro, M.H.D.M.; Stefenon, S.F.; de Lima, J.D.; Nied, A.; Mariani, V.C.; Coelho, L.S. Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning. *Energies* **2020**, *13*, 5190. <https://doi.org/10.3390/en13195190>.
58. Stefenon, S.F.; Bruns, R.; Sartori, A.; Meyer, L.H.; Ovejero, R.G.; Leithardt, V.R.Q. Analysis of the Ultrasonic Signal in Polymeric Contaminated Insulators Through Ensemble Learning Methods. *IEEE Access* **2022**, *10*, 33980–33991. <https://doi.org/10.1109/ACCESS.2022.3161506>.
59. Stefenon, S.F.; Freire, R.Z.; Coelho, L.S.; Meyer, L.H.; Grebogi, R.B.; Buratto, W.G.; Nied, A. Electrical Insulator Fault Forecasting Based on a Wavelet Neuro-Fuzzy System. *Energies* **2020**, *13*, 484. <https://doi.org/10.3390/en13020484>.
60. Stefenon, S.F.; Kasburg, C.; Freire, R.Z.; Silva Ferreira, F.C.; Bertol, D.W.; Nied, A. Photovoltaic power forecasting using wavelet neuro-Fuzzy for active solar trackers. *J. Intell. Fuzzy Syst.* **2021**, *40*, 1083–1096. <https://doi.org/10.3233/JIFS-201279>.
61. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.S.; da Rocha, D.F.M.; Grebogi, R.B.; Ruano, A.E.B. Wavelet group method of data handling for fault prediction in electrical power insulators. *Int. J. Electr. Power Energy Syst.* **2020**, *123*, 106269. <https://doi.org/10.1016/j.ijepes.2020.106269>.
62. Oceans14. Financial Information. <https://www.oceans14.com.br/> (accessed on 20 June 2021).
63. Yahoo. Finance: Stock Market. <https://finance.yahoo.com/> (accessed on 28 June 2021).
64. Araujo, F.B. *Análise Fundamentalista de Ações: Seleção das Melhores Ações do Mercado Acionário Brasileiro no Período de 2014 a 2019*; Universidade Federal de São Paulo: São Paulo, Brazil, 2021; pp. 1–42.
65. Stefenon, S.F.; Seman, L.O.; Schutel Furtado Neto, C.; Nied, A.; Seganfredo, D.M.; Garcia da Luz, F.; Sabino, P.H.; Torreblanca González, J.; Quietinho Leithardt, V.R. Electric Field Evaluation Using the Finite Element Method and Proxy Models for the Design of Stator Slots in a Permanent Magnet Synchronous Motor. *Electronics* **2020**, *9*, 1975. <https://doi.org/10.3390/electronics9111975>.
66. Dabiri, H.; Rahimzadeh, K.; Kheyroddin, A. A comparison of machine learning-and regression-based models for predicting ductility ratio of RC beam-column joints. In *Structures*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 37, pp. 69–81. <https://doi.org/10.1016/j.istruc.2021.12.083>.
67. Marquardt, D.W.; Snee, R.D. Ridge regression in practice. *Am. Stat.* **1975**, *29*, 3–20.
68. Huang, H.H.; Hsiao, C.K.; Huang, S.Y.; Peterson, P.; Baker, E.; McGaw, B. Nonlinear regression analysis. In: *International. Encyclopedia of Education*; Elsevier: Oxford, UK, 2010; pp. 339–346.