Using the MLP Classifier Model with Markov Chains Observing the Bovespa Index

Bruno Oliveira Cardoso[1*], José Roberto de Araujo[2]
[1]University Bahia State, Alagoinhas, Bahia, Brazi

The application of resources in the financial market in the form of investment is primarily linked to variable income, which has a higher risk, requiring a more thorough assessment to increase its assertiveness. Successful past experiences influence decisions and can lead to correct choices. Mathematical and statistical models should assist in decision-making to optimize their success. This work involves creating a web service connector to capture data from the BOVESPA index, randomly used to choose between three websites: Infomoney, Investnews, and Valor Invest. This work used an Artificial Neural Network (ANN) Classification model, the MLPClassifier, and compared this model alone and about itself with the insertion of the modeling with Markov Chains. The MLPClassifier model proved satisfactory, given that performance metrics measure its classifications, and the speed will depend on the number of hidden layers used. By including Markov Chains in the MLPClassifier model, it was possible to refine the classification process of the model, which already had a very high level of assertiveness.
Keywords: Stock Exchange. Markov Chain. MLPClassifier. Bovespa Index.
Abbreviations: Format. Microsoft Word Template. Style. Insert. Template.

The application of resources in the financial market in the form of investment is linked mainly to variable income, which involves greater risk, and a more detailed assessment is necessary to increase its assertiveness. Successful past experiences influence decisions and can lead to correct choices. Mathematical and statistical models should assist in decision-making to optimize success.

In this sense, Nascimento and colleagues [7] states that the mind avoids information overload, causing individuals to be induced into mental shortcuts, and a significant part of this process is executed unconsciously and automatically. In other words, such shortcuts guide us to escape from danger zones. However, using a mobile equipment management application, Mota [11] used technology to persuade users regarding energy consumption. Therefore, technology is one of the tools to shield such triggers and place the investor in a strategic position.

Received on 17 September 2024; revised 18 October 2024.
Address for correspondence: Bruno Oliveira Cardoso. Rodovia Alagoinhas / Salvador, BR110, Km 03. Zipcode: 48.000-000. Alagoinhas, Bahia, Brazil. E-mail: cardosocontabil.bahia@gmail.com.

J Bioeng. Tech. Health 2024;7(Suppl 1):35-42

Therefore, when analyzing the financial market, the BOVESPA Index (IBOVESPA) [4] is the leading performance indicator of stock exchanges on the B3. It comprises approximately 80% of the trades and is reassessed every four months. In other words, it is nothing more than the result of a theoretical portfolio of assets, a thermometer of how the investment market is doing in Brazil. Therefore, this primary performance indicator is a reference for an intelligent assessment of the financial market.

For this work, the strategy used is technical analysis that according to Piazza [12], [...] *It is the science that seeks, through the study of multifaceted records associated with mathematical and statistical formulations, applied to past and current prices, volumes, and open contracts of different financial assets, to provide, by analyzing recurring patterns, conditions that allow us to project the future price movement within a logic of higher probabilities.* [...]

Therefore, technical analysis is a study that observes the movement of the stock exchange intending to obtain information. Therefore, one method used in this type of analysis is Markov Chains, which, according to Bolson and colleagues [3], is a stochastic process, that is, a set of random variables that describes the behavior of this process in a certain period and that the systematics for Markov Chains is that only the current experience can affect the following result. This work aims to use Markov Chains to observe the BOVESPA index.

Materials and Methods

The purpose of this work is to create a web service connector. For Daigneau [6], a web service provides the means to integrate different systems and expose reusable business functions via HTTP. They use HTTP as a simple method of transporting data, and in this sense, creating a web service connector to capture data from the BOVESPA index using a random selection of three websites: Infomoney, Investnews, and Valor Invest.

With the information obtained from IBOVESPA, modeling will be performed with Markov Chains since IBOVESPA is affected by stochastic values. In this sense, a Markov Chain is a set of random variables that describes the behavior of variables over a certain period [5].

This work is based on verifying the frequency of historical movement. Therefore, probability information helps in decision-making for investors who deal with these probabilities of the Stock Exchange. For this classification, descriptive scope with a quantitative approach was considered.

This work used an Artificial Neural Network (ANN) Classification model, the MLPClassifier, and compared this model alone and about itself with the insertion of modeling with Markov Chains: A study of the BOVESPA index, whose purpose was to create a reliable computational algorithm to assist in decision-making. Data preprocessing is necessary since Gorgens and colleagues [8] state that normalization is an approach that aims to homogenize the variables involved in the analysis. Thus, normalization was performed using Natural Logarithm, which, according to Zhang and You [19], is a local logarithm normalization method proposed to improve the accuracy of the local prediction of the ANN.

Markov Chains

Markov Chain is a mathematics model whose author, Andreyevich Markov, initiated studies about Russian theory methodology; it is known as the Markov Chain [3]. Only experience could affect the future state, and this is known as the Markovian Property. The authors emphasize that even if past events are not considered relevant to the future state, they are not ignored because the model uses the information from the past as a basis for the present state of the process.

Cechin and Cordo [5] describe that Markov Chain is a set of random variables, that is, a stochastic process that describes the behavior of variables $\{Xtn\} = \{x1, x2, ..., xn\}$ within a certain period T and that such a process is a Markov chain if the probability of occurrence of a future state depends exclusive on the present state that is if it is independent of past events.

The sequence P $\{Xn, n>0\}$ is said to be a Markov chain if for all state values $i0, i1, i2, ..., in \in I$:

$$P\{X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_n = i\} = P\{X_{n+1} = j \mid X_n = i\}$$

Where, $i0$, $i1$, $i2$, . . ., in are the states in the state set I. This kind of probability is called Markov chain probability. So, this setup indicates that regardless of its history prior to time n, the probability that it will transition to another state j depends only on state i.

Therefore, Bhusal [2] affirms that the transition probability, defined by the Markov chain, is called transition or jump probability from state i to state j.

$$P\{X_{n+1} = j \mid X_n = i\} = P_{ij}$$

This is known as one-step transition probability. If the transition probabilities defined above are independent of time n, then such an assumption is called a stationary Markov chain. Thus,

$$P\{X_{n+1} = j \mid X_n = i\} = P\{X_1 = j \mid X_0 = i\}P_{ij}$$

The transition probabilities Pij can be written or arranged in a matrix form as:

$$P = [p_{ij}], i, j \in I$$

The matrix P is called the transition probability matrix or stochastic matrix. The values from matrix P must be non-negative elements with row sum unity equals 1. Hence, The matrix P insists on non-negative elements with row sum unity. Hence,

$$0 \leqslant p_{ij} \leqslant 1 \text{ e } \sum_{j=1}^{n} p_{ij} = 1, \forall i \in j$$

Cechin and Corso [5] show that the transition probability can be represented in a matrix.

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1j} \\ P_{21} & P_{22} & P_{23} & \dots & P_{2j} \\ \dots & \dots & \dots & \dots & \dots \\ P_{i1} & P_{i2} & P_{i3} & \dots & P_{ij} \end{pmatrix}$$

When a Markovian process can go from one state to another, any in n-time steps are classified as an ergodic matrix. The stable probability state πj to an ergodic Markov Chain is represented by the equation represented above.

Gathering Data

Data collection was carried out by searching for historical data from the BOVESPA index, as previously reported, in which each piece of information was randomly selected between March 11, 2024, and August 30, 2024, totaling 4,166 pieces of data. All of this data found was stored in a SQLite Database for easy access and standard integration with the Python programming language. This data was used to perform the Markovian process. The data from IbovespaReal, IbovespaAjustado, DataHoraAtual, DataAtual, Porcentagem, and delay comprise the information stored in the SQLite Database. Therefore:

- IbovespaReal - Data collected from the period collected from the chosen website and contains decimal places;
- IbovespaAjustado - Adjustment made to use only the whole number;
- DataHoraAtual - Date and time of the collection;
- DataHoraAjustado - Date and time with the adjustment about the delay;
- DataAtual - Use only the current date;
- Delay - interval time of each data.

Making a Transition Matrix

For this work, the database was divided into 8 classes to perform data discretization since the Bovespa index deals with continuous values. The choice for these 8 classes, whose division strategy is used by the works cited on Markov Chains in this research, used a percentage of 0.5% between each class since there was not much volatility within the scope of variation during the day. In this sense, when evaluating the IBOVESPA, as seen in this work, it is a thermometer of how the investment market is in Brazil.

The intervals with this separation of 0.5% are between -2.0% and 2% since, when evaluating the database, it was noticed that there were no values below or above these values. Markov Chains do not have an initial quantity or limit to perform. The criterion for this research is based on the understanding that the smaller these distances, the better the refinement of the results will be.

Making a Webservice Connector

To search for the Bovespa indexes of the day, a data collector was used directly from three reliable websites: Infomoney, Investnews, and Valor Invest. Since these websites update in approximately 15 minutes, care must be taken to scrape the data without blocking the computer's Internet Protocol (IP) to continue.

In this sense, this web scraping-type connector was created with the request to randomly select among these websites. This schedule was created using the Schedule library and programmed to run every business day between 10:00 a.m. and 6:00 p.m. with an interval of between 10 and 15 minutes. The time chosen was because the counter opens and closes during this period, and the delay varies within this time range.

Metrics

For the process of measuring the classification results of a model, Prati and colleagues [13] state that: "*To induce a classifier, a supervised algorithm uses a sample of cases for which the proper classification is known. [...] Each case is labeled with a unique attribute called a class to distinguish cases among the possible classifications*."

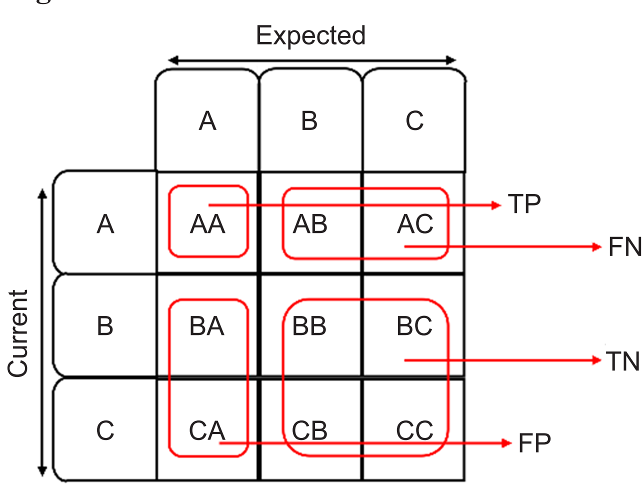
That said, to calculate Precision, Accuracy, Specificity, Recall, and F1-score, the Confusion Matrix is used (Figure 1) [17]. Wabang and colleagues [17] highlight this discussion as a binary classification problem. There are only two classes: positive and negative. What is the confusion matrix for? Ramos [14] explains that: "*The confusion matrix of a model provides a straightforward way to assess whether the model accurately predicts genuine and fraudulent transactions. [...] It displays the frequency of matches between observed and predicted classifications based on a given threshold*."

The model performance metrics [1] are:

Accuracy: The percentage of entire instances rightly predicted by the model.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

Figure 1. Confusion matrix.



TP is True Positive, FP is False Positive, TN is True Negative, and FP is False Negative.

F1-score: This is a harmonic mean of precision and recall.

$$F1 - score = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

Specificity: This shows a model's ability to classify actual negative instances as negative [1]. Therefore, this proportion of negative instances rightly predicted divided the classifier out of the total instances that are actually negative.

$$Specificity = \frac{tn}{tn + fp}$$

Precision: According to Wabang and colleagues [17], the precision is the true positives divided by the total prediction.

$$Specificity = \frac{tn}{tn + fp}$$

Recall: It measures the fraction of true positives that were correctly classified.

$$Recall = \frac{tp}{tp + fn}$$

ROC Curve

According to Prati and colleagues [13], receiver operating characteristic is a method used for prediction evaluation. It was introduced for machine learning to evaluate classification models. The author explains that the most efficient way is to order all test cases according to the continuous value to generate the curve.

According to Santos and colleagues [15], the ROC Curve is a great tool for evaluating a balance between sensitivity and specificity across all possible cutoff points for prediction. The authors continue that the overall performance of a classifier can be evaluated by the area under the ROC curve (AUC): The higher the AUC (closer to 1), the better the model's performance. Therefore, AUC ROC is an excellent way to compare two or more models.

Results and Discussion

As shown in the previous session, the database was divided into 8 classes with a 0.5% interval between each class. There was no further discretization since data below -2% and above 2% were not counted. Then, the transition count for the Markov Chain was performed, as shown in Table 1.

After carrying the transition counting in Table 1, the transition matrix from the Markov Chain is demonstrated in Table 2.

When performing the transition matrix Markov Chain in Table 2, it has to calculate what means the initial vector is the probability of zero state from these 8 classes. The probability of occurrence can be obtained by counting from 4166 from Bovespa index data. The initial vector is indicated by π (0) and given by π (0) = (π (1), π (2), π (3), · · · , π (8)).

π (1) = 16/4166 = 0.43%
π (2) = 162/4166 = 3.89%
π (3) = 495/4166 = 11.88%
π (4) = 1001/4166 = 24.03%
π (5) = 1612/4166 = 38.69%
π (6) = 612/4166 = 14.69%
π (7) = 222/4166 = 5.33%
π (8)=44/4166 = 1.06%

Table 1. Transitions counting to calculate the transition matrix.

	[-2.0~-1.5]	[-1.5~-1.0]	[-1.0~-0.5]	[-0.5~0.0]	[0.0~0.5]	[0.5~ 1.0]	[1.0~ 1.5]	[1.5~ 2.0]
[-2.0~-1.5]	7	7	2	1	1	0	0	0
[-1.5~-1.0]	7	115	25	1	14	0	0	0
[-1.0~-0.5]	2	29	363	60	40	1	0	0
[-0.5~0.0]	0	0	61	759	167	10	1	3
[0.0~0.5]	2	11	43	166	1249	122	19	0
[0.5~ 1.0]	0	0	1	9	122	461	19	0
[1.0~ 1.5]	0	0	0	2	16	17	177	10
[1.5~ 2.0]	0	0	0	4	2	1	6	31

Table 2. Markov chain transition matrix.

	[-2.0~-1.5]	[-1.5~-1.0]	[-1.0~-0.5]	[-0.5~0.0]	[0.0~0.5]	[0.5~ 1.0]	[1.0~ 1.5]	[1.5~ 2.0]
[-2.0~-1.5]	38.89%	38.89%	11.11%	5.56%	5.56%	0.00%	0.00%	0.00%
[-1.5~-1.0]	4.32%	70.00%	15.43%	0.62%	8.64%	0.00%	0.00%	0.00%
[-1.0~-0.5]	0.40%	5.86%	73.33%	12.12%	8.08%	0.20%	0.00%	0.00%
[-0.5~0.0]	0.00%	0.00%	6.09%	75.82%	16.68%	1.00%	0.10%	0.00%
[0.0~0.5]	0.12%	0.68%	2.67%	10.30%	77.48%	7.57%	1.18%	0.00%
[0.5~ 1.0]	0.00%	0.00%	0.16%	1.47%	19.93%	75.33%	3.10%	0.00%
[1.0~ 1.5]	0.00%	0.00%	0.00%	0.90%	7.21%	7.66%	29.23%	4.50%
[1.5~ 2.0]	0.00%	0.00%	0.00%	9.08%	4.55%	2.27%	13.64%	70.45%

The Markov Chain model suggests that the probability of states by many periods can be obtained by multiplying the probabilities transition matrix and the initial state vector π (i + 1) = π (i) · P. Where π (i) is the vector state given by equation 8 to the state i and P is the probability of matrix transition. When the Markovian process can go from one state to another, it reaches the state of the stationary condition and is described as an ergodic matrix. So to aid the stationary state in this paper was multiplied the transition matrix many times with the auxiliary computational and the Python language programming and obtained the stationary state with vector π = (0,43%, 3,89%, 11,89%, 24,1%, 38,63%, 14,67%, 5,32%, 1,06%).

MLPClassifier

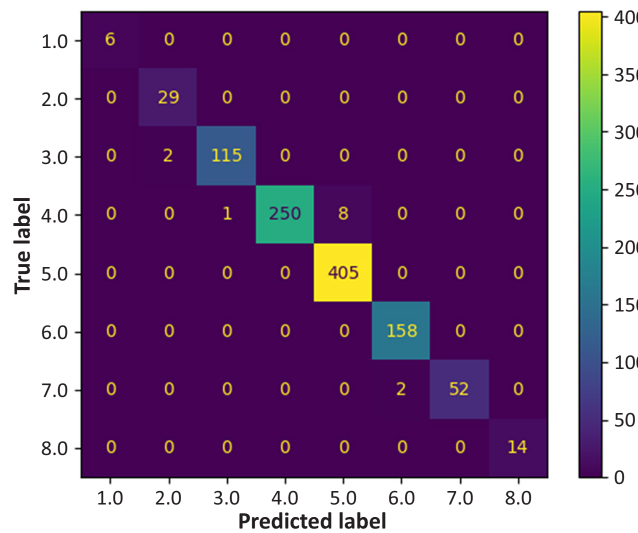
To perform the MLPCLassifier classification calculation, it is necessary to divide it into training and testing for this work. Ladi'c and Mandeki'c (2021) indicated the division using the language's library in the train_teste_split function at 75% and 25%, respectively, using the random_state = 20 parameters, as it was noticed that there was an improvement in not placing it. This library originates from open-source machine learning.

When applying training and testing with the MLPCLassifier model with just a few modifications to the standard:

- max_iter=500—Determines the number of epochs (how many times each data point will be used). By default, the maximum number is 200; however, the classifications were refined, and no improvements were made above this number.
- hidden_layer_sizes=150—The number of neurons in hidden layer 14. By default, the number is 100; however, the classifications were refined, and no improvements were made above this number.

This research used the confusion matrix, which, as explained in this paper, is a table that allows visualization of the classification algorithm's performance (Figure 2).

Figure 2. Confusion matrix MLPClassifier.



In this segment, its performance metrics were calculated according to Tables 3 and 4.

Analyzing Tables 3 and 4, classes 4 and 5 obtained the most errors in their classifications. Regarding false positives, class 5 obtained a value of 8, and class 4 obtained a 9. In any case, considering only these isolated numbers may lead to the understanding that there is little data, which is why the Classifications calculated from these results are interesting.

In this sense, analyzing, for example, the recall of class 2 with a result of 1.00 indicates that the model correctly classifies the true negatives. Therefore, these metrics depend on analyzing the type of financial investor and what information will be helpful to them.

The Roc curve corroborates all the other information, indicating that the closer the curve is to 1, the better the model's performance. Therefore, the lowest value found is in class 7. It is interesting to note that in this class, the model incorrectly classifies its true negatives, which corroborates the recall found, which is also the lowest among the others.

As this work continued, the ANN from the MLPClassifier library was inserted into the Markov Chains, and all the metrics performed in this session were compared to see if there were any improvements (Figure 3).

Table 3. Metrics from MLPClassifier model.

Classes	VP	FP	FN	VN
[–2.0, –1.5, 1]	6	6	6	1036
[– 1.5, –1.0, 2]	29	2	0	1011
[–1.0, –0.5, 3]	115	1	2	924
[–0.5, 0.0, 4]	250	0	9	783
[0.0, 0.5, 5]	405	8	0	629
[0.5, 1.0, 6]	158	2	0	882
[1.0, 1.5, 7]	52	0	2	988
[1.5, 2.0, 8]	14	0	0	1028

Table 4. Classification report from MLPClassifier model.

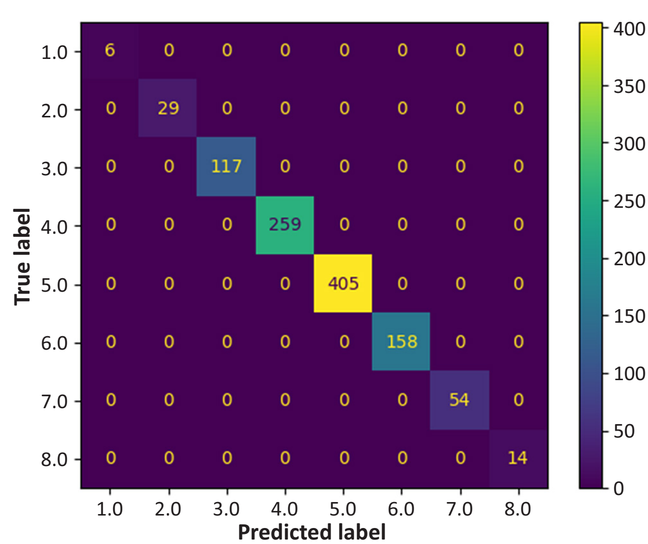
Classes	Precision	Recall	Accuracy	Specificity	F1	ROC
[–2.0, –1.5, 1]	1.0	1.0	1.0	1.0	1.0	1.0
[– 1.5, –1.0, 2]	1.0	1.0	1.0	1.0	1.0	1.0
[–1.0, –0.5, 3]	1.0	1.0	1.0	1.0	1.0	1.0
[–0.5, 0.0, 4]	1.0	1.0	1.0	1.0	1.0	1.0
[0.0, 0.5, 5]	1.0	1.0	1.0	1.0	1.0	1.0
[0.5, 1.0, 6]	1.0	1.0	1.0	1.0	1.0	1.0
[1.0, 1.5, 7]	1.0	1.0	1.0	1.0	1.0	1.0
[1.5, 2.0, 8]	1.0	1.0	1.0	1.0	1.0	1.0

The small classification errors that existed were eliminated with the insertion of Markov chains. Although small when analyzing Table 3, when calculating the existing classifications in Table 3, we check points for improvement. When inserting Markov Chains, there was a complete positive classification, thus eliminating the remnants of using only the MLPClassifier model.

Conclusion

The creation of the web service connector to collect the movement of the BOVESPA index was effective since it can work in parallel with the construction of this work and others since the model created in Python collects the data directly from the database

Figure 3. Confusion Matrix MLPClassifier with Markov chain.



and thus has no connection with this connector. The Markov Chain model to analyze the behavior of the BOVESPA index shows that it is entirely affected by stochastic factors. In this sense, the oscillation of the IBOVESPA occurred gradually, as seen in the transition matrix in the previous session.

This Markov Chain study is applied to predict the behavior of the IBOVESPA. The results are expressed in terms of the probability of a specific state in the future. The model does not provide absolute state results. The initial state vector and the transition matrices estimate the next steps in different states.

The MLPClassifier model proved satisfactory, given its classifications measured by performance metrics. The speed will depend on the number of hidden layers used. Including Markov Chains in the MLPClassifier model showed a refinement in the classification process of the model that already had a very high level of assertiveness.

References

1. Ampomah EK, Nyame G, Qin Z, Addo PC, Gyamfi EO, Gyan M. Stock market prediction with Gaussian naïve Bayes machine learning algorithm. Informatica 2021;45(2).
2. Bhusal MK. Application of Markov chain model in the stock market trend analysis of Nepal. International Journal of Scientific & Engineering Research 2017;8(10):1733-1745.
3. Bolson M, Borella L, Tomé F, Orlandin B, Corso L. Aplicação de cadeias de markov para análise de variação do dólar americano. XXXIX Encontro Nacional de Engenharia de Produção 2019;Abepro: 2019:1-13.
4. Brasil BAB Bolsa. Índice Bovespa, 2023. Retrieved 14.09.2023, from https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm
5. Cechin RB, Corso LL. High-order multivariate 46 markov chain applied in dow jones and ibovespa indexes. 47 Pesquisa Operacional 2019;39:205–223.
6. Daigneau R. Service design patterns: fundamental design solutions for soap/wsdl and restful web services. AddisonWesley 2012.
7. do Nascimento AM, Taissun AS, Lopes ÍMSF. Atalhos mentais na confluência entre a psicologia cognitiva e o serviço social: Alerta nas tomadas de decisões de assistentes sociais. Revista Científica do UniRios 2020:307-321.
8. Gorgens EB, Leite HG, Santos HN, Gleriani JM. Estimação do volume de árvores utilizando redes neurais artificiais. Revista Árvore 2009;33:1141–1147. doi: https://doi.org/10.1590/S0100-67622009000600016.
9. Ladic T, Mandekic A. Face mask classification using mlp classifier. Ri-STEM-2021 2021;10(68):77.
10. Madeeh OD, Abdullah HS. An efficient prediction model based on machine learning techniques for prediction of the stock market. In Journal of Physics: Conference Series 2021;1804:012008. doi: 10.1088/1742-6596/1804/ 65 1/012008
11. Mota FP. Modelo computacional para análise da motivação e envio de gatilhos em sistemas persuasivos: estudo de caso em processos de aprendizagem ubíqua. (Unpublished master's thesis) Universidade Federal do Rio Grande 2018.
12. Piazza MC. Técnicas de investimento para o mercado de ações. Novo Conceito 2008.
13. Prati RC, Batista G, Monard MC et al. Curvas ROC para avaliação de classificadores. Revista IEEE América Latina 2008;6(2):215–222.
14. Ramos JE Redes bayesianas aplicadas a modelagem de fraudes em cartão de crédito (Unpublished doctoral dissertation) 2015.
15. Santos HG, Nascimento CF, Izbicki R, Duarte YAO, Chiavegatto Filho P, Dias A. Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de são paulo, brasil. Cadernos de Saúde Pública, 2019;35:e00050818.
16. Taha HA. Operations research: an introduction. 8th. ed. New Jersey: Pearson Prentice Hall 2007.
17. Wabang K, Nurhayati OD et al. Application of the naïve Bayes classifier algorithm to classify community complaints. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 2022;6(5):872–876.
18. Zhang D, Zhang X. Study on forecasting the stock market trend based on stochastic analysis method. International Journal of Business and Management 2009;4(6):163–170.
19. Zhang T, You X. Improvement of the training and normalization method of artificial neural network in the prediction of indoor environment. Procedia Engineering 2015;121:1245–12.