Original Article

ARTIFICIAL INTELLIGENCE TECHNIQUES APPLIED TO PREDICT TEAMS POSITION OF THE BRAZILIAN FOOTBALL CHAMPIONSHIP

TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL APLICADAS PARA PREVISÃO DA POSIÇÃO DE TIMES DO CAMPEONATO BRASILEIRO DE FUTEBOL

Mariana Kleina[1], Mateus Noronha dos Santos[1], Tiago Noronha dos Santos[1], Marcos Augusto Mendes Marques[1] and **Wiliam de Assis Silva[1]**

[1]Federal University of Paraná, Curitiba-PR, Brasil.

RESUMO

Este trabalho apresenta uma previsão de classificação em grupos para as equipes do campeonato brasileiro de futebol tanto da série A quanto da série B a partir dos resultados do primeiro turno de cada campeonato. Com previsões assertivas do grupo onde um time irá finalizar o campeonato, pode-se realizar um planejamento estratégico no elenco tal como novas contratações, treinos específicos dos atletas e possíveis campeonatos que o time terá direito de participar de acordo com o grupo em que se classificar. Para encontrar as previsões, aplicou-se as técnicas rede neural artificial Multi Layer Perceptron (MLP) e Support Vector Machine (SVM). Resultados preliminares indicam que a metodologia proposta é bastante promissora, acertando em mais de 40% dos casos com a MLP e quase 50% com o SVM. Além disso, os resultados indicam que a metodologia também é capaz de realizar uma boa previsão errando em um grupo do verdadeiro grupo ao final do campeonato. A técnica SVM se mostrou um pouco superior à MLP. Um pós processamento nos resultados do SVM é aplicado aos dados do ano de 2018 da série A do campeonato brasileiro, resultando em 85% de acertos dos grupos.

Palavras-chave: Campeonato brasileiro de futebol. Previsão de grupos. Pós processamento dos resultados.

ABSTRACT

This study presents a classifier prediction in groups for the Brazilian Football Championship of both A and B leagues, from the results of the first half of each championship. With assertive predictions of the group where a team will end the championship, strategic planning can be performed in the squad, such as new hiring, specific training for athletes, and possible championships that the team will be entitled to participate in according to the group classification. In order to find the predictions, two techniques of artificial intelligence were applied: Multi-Layer Perceptron (MLP), which is a type of artificial neural network, and Support Vector Machine (SVM). Preliminary results show that the proposed methodology is very promising, with more than 40% successful cases with MLP and almost 50% with SVM. Moreover, results indicate that the methodology is able to make a reasonable prediction by missing one group of the true group at the end of the championship. The SVM technique was slightly better than MLP. A post-processing analysis of the SVM results was applied to the 2018 A league data from the Brazilian championship, resulting in 85% success indicator of groups.

Keywords: Brazilian football championship. Reasonable predictions. Post-processing results.

Introduction

Football is considered the most popular sport in the world[1] and it is the most valued in Brazil, which is the fifth largest country that has this sport, regarding financial capital (R$ 544 million), behind England (R$ 1.45 billion), Germany (R$ 1.05 billion), Italy (R$ 599 million) and France (R$ 591 million), according to Época[2]. These figures show that football moves a lot of capital, is highly valued, and in Brazil, it is a profession highly desired by millions of boys and girls, whom from a very early age have been preparing to become a professional football player. Such motivation starts in monitoring especially for the Brazilian Football Championship.

The Brazilian Football Championships (A and B leagues) are managed by The Brazilian Football Confederation[3] (CBF), and consist of 40 teams (20 teams in each one) from around the country. In each league, the teams compete among themselves twice, playing at the

opposing team's stadium and at their own stadium, aiming to gain the most points[4]. The winning team of a match gets 3 points. In the case of a tie, each team gets 1 point, and the losing team does not score. At the end of 38 rounds, the team with the highest score is the champion. The last 4 teams are demoted to the lower level league (this rule applies to the A and B leagues), and the first 4 teams from the B league are promoted to the A league. In addition to scoring, other factors are important for tiebreaker criteria, such as goals scored for the team, goals scored against the team, and the number of yellow and/or red cards.

Since the beginning of the Brazilian Football Championship, several teams have participated in the competition to achieve leadership or positions that offer benefits, such as the opportunity to participate in major championships, in addition to financial rewards. This makes these speculations and projections increasingly studied and observed and with assertive projections, not only in the sports area but also decision making, which may be facilitated and objectives may be more easily achieved.

Predictions are made in several areas, such as medicine[5], meteorology[6], and supply chain sales[7]. Several studies have been conducted to predict team performance in competitions using statistical methods, such as studies by Añon et al.[8], Araújo et al.[9] and Santos[10]. Artificial intelligence techniques are also used to predict the results of football championships. Huang and Chang[11] applied Artificial Neural Network (ANN) to predict the 2006 FIFA World Cup matches. Duarte[12] used Support Vector Machine (SVM) and Random Forest to foresee results of the Portuguese League football matches. Bunker and Thabtah[13] and Langaroudi and Yamaghani[14] reviewed the research literature on using machine learning techniques to predict sports in general.

Performance prediction can help to facilitate possible changes in team strategies or to support sponsors in decision making on who to invest their money. In regards to strategies for a team, we can mention better physical preparation of the athletes, as this is fundamental to improve team's performance.

This paper aims to predict an interval position that a team will finish the A and B leagues of the Brazilian Football Championship, based on ranking, goal difference, wins, draws, and losses starting from the part of the competition, with data between 2006 and 2018. For predictions, we used ANN and SVM, both Machine Learning techniques, i.e., predictions were based on patterns that have occurred in the past through supervised learning. Many research studies address predictions through statistical techniques that model the number of goals scored by the teams in a match or the team that will win the match (or draw). Improving athletes' physical preparation or game strategies requires more time than just the interval between games (which normally varies between 3 and 7 days for the Brazilian Football Championship[3]). Our research uses unconventional techniques in football predictions to estimate results for approximately 100 days ahead, providing a final group that a team will finish the championship with, this time enough to make tactical modifications, to improve physical performance or to do hiring, if necessary.

Predicting the position of a team at the end of the championship is difficult and the scientific literature is scarce. The research by Tsakonas et al.[15] makes predictions for the Ukrainian Championship with 10 years of data using some techniques. This study presents the predictions regarding the position of 14 teams at the end of the championship.

The results from the Tsakonas et al.[15] study were compared to our results, since no research predicting final groups in football championships have been found.

In the literature, some studies that provide predictions for the second part of the Brazilian Football Championship were found. However, they have different methodologies and results, which are not suitable for comparison with the present study. Research from Saraiva et al.[16] provides the probability that a team will be either the champion, participating in the Libertadores Cup, or be demoted at the end of the 2015 championship, using a Poisson

regression model. Alves et al.[17] created two logit models to estimate the probability that a team will finish the championship in the top four positions and in the bottom four positions, applying data from the year 2008.

The present research has limited application to data from A and B leagues of the Brazilian Football Championship, as they have the same format, and have a history of results, sufficient to train the machine learning models. Another limitation of the study (for machine learning techniques in general) is that if a new variable needs to be added to the models, they need to be retrained.

Methods

In this study, we used two methods of artificial intelligence to predict the group in which a team will finish the championship. The variables used (ranking, goal difference, wins, draws, and losses in the first part of the competition) are the inputs for both methods and all variables initially have the same importance for the model. The idea of training a mathematical model is to adjust these weights so that the error in a set never seen by the model is as small as possible. Unfortunately, at the end of the adjustment of ANN and SVM models, it is impossible to verify which variables had the greatest influence on the results, as the weights cannot be expressed as coefficients of the variables, as in linear regression models[18], for example.

Artificial Neural Network

Artificial Neural Networks simulate the behavior of biological neurons and can model complex problems[19]. An ANN receives information $(x_1, x_2, ..., x_m)$ in the input layer and it is multiplied by weights $(w_{k1}, w_{k2}, ..., w_{km})$. This weighting is summed, generating a signal u_k, according to Equation 1.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{1}$$

A term called bias b_k is added in Equation 1 in order to provide greater freedom in the weighting of input data, as well as greater network approachability. An activation function $\varphi(.)$ is applied to this signal to limit y_k, the output layer signal, avoiding progressive additions (Equation 2).

$$y_k = \varphi(u_k + b_k) \tag{2}$$

The activation functions more frequently used are linear, binary step, sigmoid/logistic, hyperbolic tangent and rectified linear unit.

A Multi-Layer Perceptron (MLP) is an ANN that has more layers (hidden) than input and output layers. Normally, a hidden layer can solve any continuous problem[19]. An MLP uses supervised learning and in order to update the weights, it uses the backpropagation algorithm which is based on error correction. Thus, the input variables are weighted in order to minimize the final error.

Support Vector Machine

Support Vector Machine is applied in two problem types: classification and regression. The classification technique called Support Vector Classification (SVC), is based on a separation of the d-dimensional data into two classes with a hyperplane building a margin with

the maximum geometrical distance of blank space between the two species[20]. The discriminant hyperplane is defined by Equation 3.

$$\langle w, x \rangle + b = 0 \tag{3}$$

In Equation 3, w is a weighted vector indicating the orientation of a hyperplane and b is a scalar that compensates the hyperplane from the origin. Considering the outputs $y = 1$ and $y = -1$ for two classes, an i-th point of the training dataset is correctly classified if $y_i(\langle w, x_i \rangle + b) \geq 1$. Then, to find the optimal hyperplane, a quadratic problem should be solved by Equations 4.

$$
\begin{aligned}
min \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i \\
s.t. \quad & y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \forall i = 1, \dots, n \\
& \xi_i \geq 0, \qquad\qquad \forall i = 1, \dots, n
\end{aligned}
\tag{4}
$$

The first part of objective function in Equations 4 aims to seek a smaller vector w (that is equivalent to seek the maximum margin) and the second part aims to penalize constraint violation, where C is a scalar which determines the cost of constraint violation, ξ_i is the slack variable and n is the number of points of dataset training. The dual problem of Equations 4 is obtained because the computational complexity becomes dependent only on the number of support vectors[21].

$$
\begin{aligned}
max \quad & \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}y_iy_j\alpha_i\alpha_j\langle x_i, x_j \rangle \\
s.t. \quad & \sum_{i=1}^{l}y_i\alpha_i = 0 \\
& 0 \leq \alpha_i \leq C \qquad \forall i = 1, \dots, n
\end{aligned}
\tag{5}
$$

The variable α is a Lagrange multiplier. Solving the problem in Equations 5, we found the support vectors ($\alpha \neq 0$) that were used to discover w and b in Equation 3, with the support of the KKT conditions.

To solve nonlinearly separable classifications, a nonlinear mapping of the input data X to a space of high dimensionality F, called the feature space, was performed. A common technique for performing this procedure is to change the data representation for Equation 6,

$$x = (x_1, x_2, \dots, x_n) \longmapsto \phi(x) = (\phi_1(x), \dots, \phi_N(x)) \tag{6}$$

where $F = \{\phi(x) \mid x \in X\}$ and $N \gg n$. However, acknowledging that $\langle \phi(x_i), \phi(x_j) \rangle$ directly in the feature space can become computationally infeasible due to its high dimensionality, a kernel function K was used to compute the inner product, according to Equation 7.

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \tag{7}$$

The most common kernel functions are linear, polynomial, sigmoid and RBF.

The regression technique called Support Vector Regression (SVR), is similar to SVC and was not explained in detail, but the main purpose was to find a function $f(x) = \langle w, x \rangle + b$ which has at most one deviation ε from the target values y_i. The dual problem to be solved was given by Equations 8,

$$
\begin{aligned}
max \quad &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \beta_i)(\alpha_j - \beta_j)K(x_i, x_j) - \varepsilon\sum_{i=1}^{n}(\alpha_i + \beta_i) + \sum_{i=1}^{n}y_i(\alpha_i - \beta_i) \\
s.t. \quad &\sum_{i=1}^{n}(\alpha_i - \beta_i) = 0 \\
&\alpha_i, \beta_i \in [0, C], \quad \forall i = 1, \dots, n
\end{aligned}
\tag{8}
$$

where α and β are the Lagrange multiplier.

For this method, the weights w give importance to the input variables, which help in the construction of the separating hyperplane. Specifically, the variables ranking, goals difference, wins, draws, and losses in the first part the competition was used to maximize the distance between the groups of teams.

Data Set

Data from the A and B leagues of the Brazilian Football Championship (20 teams each one), between 2006 and 2017 was used in this study to train and to validate the mathematical models. Indeed, 2003 was the first year in which championships began to be played with continuous scoring, but in 2003, 2004, and 2005 there were more than 20 teams in each championship. The year of 2018 was the full year before the completion of this research and it was used to exemplify the results, by comparing predictions with real groups. This data set was collected from CBF[3] and is constituted by ranking, score, goals difference, wins, draws, and losses in the first part of the championships. All variables used to express the team's performance are correlated, that is, one influences the other directly, as it can easily be observed with the variables position and score: teams with higher scores are in greater positions.

In addition, the ranking for each team at the end of the championships was obtained and used as response variable for the ANN and the SVM techniques. This variable was divided into five groups as follows: group 1 (1st to 4th ranking), group 2 (5th to 8th ranking), group 3 (9th to 12th ranking), group 4 (13th to 16th ranking) and group 5 (17th to 20th ranking). The formation of groups with four teams was chosen mainly due to the fact that the top four teams placed in the A league have direct access to the group stage of the Copa Libertadores da América, the top four teams placed in the B league are promoted to A league, and the bottom four teams from A and B leagues are downgraded to the lower level leagues[4].

Results and Discussion

The *scikit-learn* version 0.20.1 in Python 2.7 was used to apply the methodology presented[22]. For the ANN and SVM techniques, the dataset was randomly divided into two sets: training (70% of total) and testing (30% of total). To measure the effectiveness of ANN and SVM results, we used Root Mean Square Error (RMSE) and some success and error indicators: Success means that the method hit the group correctly; Error$\pm$1 means that the method made a mistake for one group difference; Error means that the method missed for more than one group difference.

In the ANN method the default *scikit-learn* module settings were used, and to choose the number of neurons in the middle layer, a variation was made (1 to 10, 15, 20, 30, 50 and 100 neurons). Table 1 shows the results of the ANN method for the test data.

Table 1. Results of the Artificial Neural Network for the test data

Neurons number	RMSE	*Success* (%)	*Error±1*(%)	*Error* (%)
1	1.14	45.76	36.72	17.51
2	1.12	40.68	43.50	15.82
3	1.05	49.72	35.59	14.69
4	1.09	44.07	40.11	15.82
5	1.07	46.33	39.55	14.12
6	1.04	44.07	42.37	13.56
7	0.96	42.94	47.46	9.60
8	1.05	43.50	34.60	21.90
9	1.24	38.98	39.55	21.47
10	1.07	44.07	40.11	15.82
15	1.13	42.94	37.85	19.21
20	1.07	41.81	45.20	12.99
30	1.18	33.33	44.63	22.03
50	1.04	40.11	45.76	14.12
100	1.12	44.63	39.55	15.82

Source: The authors

Figure 1 shows the *Success*, *Error±1* and *Error* indicators for the test data, varying the number of neurons in the middle layer.
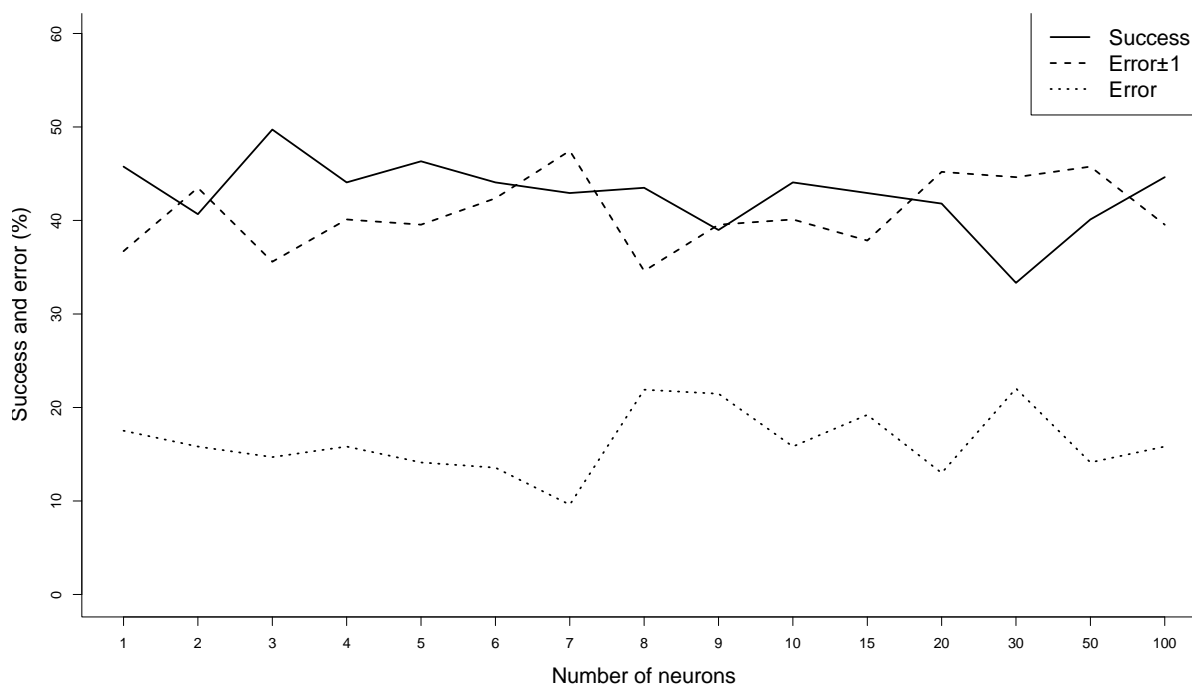

Figure 1. Results of the Artificial Neural Network for the test data varying the number of neurons in the middle layer
Source: The authors

Analyzing Table 1 and Figure 1, it is concluded that ANN achieved satisfactory results, with about 40% success indicator. There was also a low error indicator.

In the SVM method, using the same input variables, we varied the kernel function: linear, radial base function (RBF), sigmoidal and quadratic. Table 2 shows the same indicators previously defined for the test data.

Table 2. Results of the Support Vector Machine for the test data

Kernel function	RMSE	*Success* (%)	*Error±1*(%)	*Error* (%)
Linear	1.16	43.50	36.15	20.35
RBF	1.13	44.65	39.54	15.81
Sigmoidal	1.39	18.08	44.63	37.29
Quadratic	0.95	49.71	37.85	12.44

Source: The authors

The SVM technique also showed satisfactory results, except when using sigmoidal kernel function, with less than 20% success indicator. It should be noted that *scikit-learn* as default was used for SVM technique parameters.

In order to compare the techniques used, the best configurations of each one were chosen, that is, the ANN MLP with 7 neurons in the middle layer and the Quadratic kernel function in the SVM. The techniques were applied to the 2018 A series data (that was not used during training step), and post-processing was proposed in order to improve the results.

Post-processing

It was observed in both techniques that some groups did not contain four teams, exactly, which is ideal. Therefore post-processing of the results was done as follows: the position of each team was divided by the highest position (called $\overline{position}$) and the score of each team was divided by the highest score (called $\overline{score}$). Thus, a new indicator (*NI*) was created and calculated by Equation 9.

$$NI = \overline{score} - \overline{position} \qquad\qquad (9)$$

For example, a team with a score of 29 (where the best team had a score of 38) and it was 5th in the middle of the championship will have $NI = 29/38 - 5/20 \approx 0.51$. This indicator was calculated for all teams of group 1, and if group 1 had five teams, for example, the team with smaller *NI* was allocated to group 2. Thus, the same analysis was conducted for group 2 and etc. Similarly, we analyzed when the number of classified teams in the group was less than four.

Box 3 shows a comparison between MLP and SVM with better performance during the application of the test set. The data used in Box 3 is from the 2018 A league championship. Results are also presented after the post-processing step.

Box3. Comparison between MLP (7 neurons in the middle layer) and SVM (Quadratic kernel function) techniques applied to the 2018 A league data

Team	Real group	MLP forecast	Post-processing MLP forecast	SVM forecast	Post-processing SVM forecast
Palmeiras	1	1	1	1	1
Flamengo	1	1	1	1	1
Internacional	1	1	1	1	1
Grêmio	1	1	1	1	1
São Paulo	2	2	2	2	2
Atlético-MG	2	1	2	2	2
Atlético-PR	2	4	4	2	2
Cruzeiro	2	5	5	5	2
Botafogo	3	4	3	4	4
Santos	3	3	3	5	3
Bahia	3	3	2	2	3
Fluminense	3	4	4	4	3
Corinthians	4	5	5	4	4
Chapecoense	4	3	3	5	5
Ceará	4	1	2	4	4
Vasco	4	4	3	4	4
Sport	5	4	4	5	5
América	5	5	5	5	3
Vitória	5	4	4	5	5
Paraná	5	5	5	4	5
Success	-	**10**	**10**	**13**	**17**

Note: Errors are in gray
Source: The authors

The first column of Table 3 shows the real group of the Brazilian Football Championship 2018 A league for each team; the second one is the MLP forecast with 10 success indicators; the third one shows the results after post-processing MLP forecast with 10 success indicators, i.e., it was not possible to improve the number of success indicators; the fourth one shows SVM forecast results with 13 success indicators; and the last one shows the results after post-processing SVM forecast with 17 success indicators, with a 20% improvement in the results and equalling 85% success.

The results of Tsakonas et al.[15] was adapted for comparison with our results. We divided the 14 teams of the Ukrainian Championship into 4 groups: group 1 (1st to 4th ranking), group 2 (5th to 8th ranking), group 3 (9th to 12th ranking), and group 4 (13th to 14th ranking). As the research of Tsakonas et al.[15] predicted the final position for a given season, the results were easily converted into predictions for the final group. Both the Fuzzy model and the Neural Network of Tsakonas et al.[15] had 10 success indicators and 4 error indicators, i.e., 71% success, approximately. This result shows that the model proposed in this study is promising since it achieved a better performance than that of Tsakonas et al.[15] (85% success versus 71% success).

Using the 2018 results from the SVM post-processing predicion, the following analysis can be made in relation to squad (signings), tactical scheme, athletes' physical preparation, among other action plans:

- Palmeiras, Flamengo, Internacional and Grêmio could maintain their performance until the middle of the championship, because according to the model, these teams would finish the championship in the first group;

- São Paulo, Atlético-MG, Atlético-PR and Cruzeiro should improve their performance to access the group stage of the Libertadores Cup or to obtain the national championship title;
- Chapecoense, Sport, Vitória and Paraná, according to the model, would need to improve performance to remain in the A series. As the model is not 100% successful, errors happened, as is the case of the incorrect classification of Chapecoense in the fifth group, and the non-classification of America in this group;
- Groups 3 and 4 would have a similar analysis if they wanted to have a better position at the end of the championship.

The analysis is important since the actions mentioned above may be taken at the beginning of the second part of the championship, and there is enough time for the action plans to take effect until the end of the tournament.

Conclusion

The purpose of this paper was to predict the group in where a football team will end in the Brazilian Championship only considering statistical information up to the first part of the competition. Artificial Neural Network and Support Vector Machine techniques were used to make the predictions. Both techniques have parameters that were chosen empirically. The best configuration for ANN was with 7 neurons in the middle layer and for SVM was with a quadratic kernel function.

The best results for ANN shows 42.94% success, 47.46% error for one difference group, and 9.6% error for more than one group. The best results for SVM shows 49.71% success, 37.85% error for one difference group, and 12.44% error for more than one group. Considering that a team has 20% probability of being in a group at the end of the championship, the results of the techniques were very good.

In order to improve the results of the techniques, post-processing was proposed. It consisted of balancing the groups with exactly the same number of teams in each one. This technique was applied in the 2018 A league data. For ANN the post-processing did not improve results, however, for SVM there was an improvement of about 30.77% in the results, totaling 17 success out of 20. Thus, in this study, the SVM was better than the ANN technique, both unprocessed and post-processed.

To know a priori the group that a team will end up during the championship can lead to strategies being taken by the team leader, to conduct a plan for the remainder of the year or to plan future championships. For example, if an A league team has as a result from the prediction model that will finish the championship in the first group, this team will be a contestant for the title of the competition and will be entitled to participate in the Libertadores Cup. Having access to this information in the first part of the championship, allows the team to maintain the squad, tactical scheme, and physical preparation of the athletes, and to repeat the performance obtained during the first part of the championship. However, if a team is expected to finish the championship in the last group, obviously measures need to be taken so that prediction does not occur.

The main contribution of this research is precisely to anticipate important decision-making, so that teams have time (e.g., approximately 100 days) to improve their performance in the championship, in the case that the predicted group is not the desired one. The application of this research was in a football competition, however, it is emphasized the fact that these models can be trained and applied for team sports, contributing to the diagnosis of athletes` professional training.

References

1. Worldatlas [Internet]. The Most Popular Sports In The World, 2018. [cited on Jul 18 2019] Available from: https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html
2. Época [Internet]. A CBF fatura alto com a seleção brasileira - eis de onde vem e para onde vai o dinheiro, 2018. [cited on Aug 05 2019]. Available from: https://epoca.globo.com/esporte/epoca-esporte-clube/noticia/2018/06/cbf-fatura-alto-com-selecao-brasileira-eis-de-onde-vem-e-para-onde-vai-o-dinheiro.html
3. Confederação Brasileira de Futebol [Internet]. Available from: https://www.cbf.com.br
4. MRV [Internet]. Entenda como funciona o comapeonato brasileiro de futebol, 2020. [cited on Aug 12 2020]. Available from: https://mrvnoesporte.com.br/entenda-como-funciona-o-campeonato-brasileiro-de-futebol/
5. Saritas I. Prediction of Breast Cancer Using Artificial Neural Networks. J Med Syst 2012;32:2901-2907. DOI: 10.1007/s10916-011-9768-0
6. Kleina M, Matioli LC, Leite EA. Identificação, monitoramento e previsão de tempestades elétricas utilizando métodos numéricos. Bol Cienc Geod 2016;22:589-612. Doi: https://doi.org/10.1590/s1982-21702016000400034
7. Boone T, Ganeshan R, Jain A, Sanders NR. Forecasting sales in the supply chain: Consumer analytics in the big data era. Int J Forecast 2019;35:170-80. Doi: https://doi.org/10.1016/j.ijforecast.2018.09.003
8. Añon IC, Yamanaka GK, Machado JC, Scaglia A. Performance da equipe da Espanha e seus adversários nos jogos da Copa do Mundo FIFA 2010. [cited on May 12 2019]. Rev Bras Fut 2013;06(1):33-44. Available from: https://rbf.ufv.br/index.php/RBFutebol/article/view/114
9. Araújo CTP, Tavares L, Alvares LG, Neto FL, Suzuki AK. Modelagem estatística para a previsão de jogos de futebol: uma aplicação no campeonato brasileiro de futebol 2014. Rev Estat UFOP 2015[cited on Jun 24 2019]; 4(2): 12-20. Available from: https://periodicos.ufop.br:8082/pp/index.php/rest/article/view/3337
10. Santos JMA. Previsões de resultados em partidas do campeonato brasileiro de futebol. [Dissertação de Mestrado em Matemática Aplicada]. Rio de Janeiro: Fundação Getúlio Vargas; 2019.
11. Huang KY, Chang WL. A neural network method for prediction of 2006 world cup football game. IEEE 2010;20:1-8. Doi: https://doi.org/10.1109/IJCNN.2010.5596458
12. Duarte LMS. 1X2 - Previsão de resultados de jogos de futebol. [Dissertação de Mestrado em Engenharia Eletrotécnica e de Computadores]. Porto: Faculdade de Engenharia da Universidade do Porto; 2015.
13. Bunker RP, Thabtah F. A machine learning framework for sport result prediction. Appl Comput Inform 2019;15(1) 27- 33. Doi: https://doi.org/10.1016/j.aci.2017.09.005
14. Langaroudi MK, Yamaghani MR. Sports result prediction based on machine learning and computational intelligence approaches: A Survey. J Adv Comp Eng Technol 2019. [cited on Oct 16 2019];5(1):27-36. Available from: https://jacet.srbiau.ac.ir/article_13599.html
15. Tsakonas A, Dounias G, Shtovba S, Vivdyuk V. Soft Computing-Based Result Prediction of Football Games. Proc of the First International Conference on Inductive Modelling 2002; 3: 15-21.
16. Saraiva EF, Suzuki AK, Filho CAO, Louzaba F. Predicting football scores via Poisson regression model: applications to the National Football League. Commun Stat Appl Methods 2016;23:297-319. Doi: https://doi.org/10.5351/CSAM.2016.23.4.297
17. Alves AM, Mello JCCBS, Ramos TG, Sant'Anna AP. Logit models for the probability of winning football games. Pesq Oper 2011;31(3):459-65. Doi: https://doi.org/10.1590/S0101-74382011000300003
18. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 5. ed. New Jersey: John Wiley & Sons; 2012.
19. Haykin S. Redes Neurais: princípios e prática. Porto Alegre: Bookman; 2001.
20. Yao Y, Liu Y, Yu Y, Xu H, Lv W, Li Z, et al. K-SVM: An effective SVM algorithm based on K-means clustering. J Comput 2013;8(10):2632–639. Doi: https://doi.org/10.4304/jcp.8.10.2632-2639
21. Birzhandi P, Kim KT, Lee B, Hee Youn YH. Reduction of training data using parallel hyperplane for support vector machine. Appl Artif Intell 2019;33(6):497-516. Doi: https://doi.org/10.1080/08839514.2019.1583449
22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. J Mach Learn Res 2011;12:2825- 830

ORCID number:

Mariana Kleina: https://orcid.org/0000-0001-8108-8793
Mateus Noronha dos Santos: https://orcid.org/0000-0002-3257-942X
Tiago Noronha dos Santos: https://orcid.org/0000-0002-6662-1825
Marcos Augusto Mendes Marques: https://orcid.org/0000-0002-7012-4362
Wiliam de Assis Silva: https://orcid.org/0000-0001-6495-9585

Correspondence address: Mariana Kleina. Rua Francisco Heráclito dos Santos, 210, Centro Politécnico da UFPR, Jardim das Américas, Curitiba, PR, CEP 81531-980. E-mail: marianakleina@ufpr.br