Pesquisa Operacional (2022) 42: e243537 p.1-21
doi: 10.1590/0101-7438.2022.042.00243537

Printed version ISSN 0101-7438 / Online version ISSN 1678-5142
www.scielo.br/pope
ARTICLES

MULTIPLE CHOICE METHOD WITH GENETIC ALGORITHM FOR THE FORMATION OF SOCCER TEAMS

Sérgio Augusto Faria Salles[1], Henrique Rego Monteiro da Hora[2], Milton Erthal Júnior[3], André Soares Velasco[4] and Paulo Rossi Croce[*5]

Received September 3, 2020 / Accepted January 10, 2022

ABSTRACT. For soccer managers, player selection lineups are a key process for better performance, both on financial and sports matters. However, to determine a quality solution to this problem it becomes more complex, as the number of alternatives and criteria increases and the number of viable solutions grows exponentially. This paper proposes a multicriteria method with genetic algorithm for the evaluation of soccer teams based on Brazilian Championship data. A team complementation percentage was calculated considering a total of 322 athletes and 18 criteria. The results presented a 3-6-1 format as the ideal for this case study, obtaining a team complementation value of 43.04%. The method adaptability for the decision-maker is highlighted, showing it was possible to determinate the most complementary team according to the desired tactical formation and importance attributed for each criterion.

Keywords: evolutionary algorithm, multicriteria, operational research.

1 INTRODUCTION

During the 20th century, sport has been consolidating as a cultural phenomenon, causing great social, economic and political impacts. Lately, soccer is one of the most important means of expression in sports, as well being a business of high economic importance, growing constantly and forcing clubs to become more efficient in their business, seeking success on and out the field (Pyatunin et al., 2016; Zambom-Ferraresi et al., 2017).

*Corresponding author

[1]Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, RJ, Brazil - E-mail: safsalles@hotmail.com – http://orcid.org/0000-0003-3701-5743

[2]Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, RJ, Brazil - E-mail: dahora@gmail.com – http://orcid.org/0000-0001-7192-9245

[3]Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, RJ, Brazil - E-mail: miltonerthal@hotmail.com – http://orcid.org/0000-0002-9959-3568

[4]Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, RJ, Brazil - E-mail: asvelasco@iff.edu.br – http://orcid.org/0000-0001-9769-6232

[5]Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, RJ, Brazil - E-mail: paulorossicroce@gmail.com – http://orcid.org/0000-0001-8229-090X

The best players selection is a crucial issue for soccer club's managers considering the costs and performance of their team (Arabzad et al., 2014). In many cases, decision-making about transferring athletes is questionable since the low performance of them. It is often associated the way in which players are evaluated, disregarding the contribution of their attributes to the collective performance of the contracting team (Al-Shboul et al., 2017; Jarvandi et al., 2013; Tavana et al., 2013).

The soccer team formation is a process by which individuals are assigned to positions that were determined according to the characteristics of the team (Budak *et al.*, 2017). Approaching this problem, decision makers aim to form an ideal team, in which the best agents are selected for each function. The mathematical modelling for these problems, considering different variables and points of view, can be performed by decision analysis techniques (Costa, 2017, Sales *et al.*, 2019). This kind of problem, characterized by the high number of alternatives, variety of positions and evaluative criteria (e.g., price, correct passes, tackles, goals etc.), the number of feasible solutions increases exponentially and identifying the ideal solution for this type of problem demands an excessive effort. According to Kramer (2017) optimization algorithms, such as genetic algorithms, can be used for this purpose, being able to solve most of the optimization problems that occur in practice, identifying efficient solutions.

In this context, by using the Multiplex Electionis Methods (Geneticae Algorithm), a decision-making method that uses a genetic algorithm for soccer team formation, this study performs a study case based on the first division of the Brazilian Championship, season 2017.

2 SOCCER AND MULTICRITERIA DECISION MAKING

Many data analysis methods are used in a range of sports (Kröckel, 2017) and the book "Moneyball" (Lewis, 2004) was the main trigger for this adoption. Large clubs have used data analysis in order to support their decision, since this information means competitive advantage, most of their techniques and results remains a secret (Kröckel, 2017).

However, Wright (2009) reports that the lack of a more holistic view of managers and researchers can compromise the efficiency of these techniques. Many publishing papers have a research fragmentation problem, whereas authors no longer seek to delve deeper into this field. Despite the advantages of data analysis for the decision-making process, the culture of using these techniques is still not widespread in soccer clubs (Zhu et al., 2015).

Some publishing papers employ a metaheuristic method of Genetic Algorithms (GA) in soccer. Atan and Hüseyinoğlu (2017) show a GA application for the assignment of referees for Turkish soccer league games. The authors used integer programming to assign constraints on the referees' workload, with the GA being inserted to reduce the required computational effort for the problem. A Genetic Algorithm model was proposed by Cakmak *et al.* (2018), to measure the effectiveness of passes made between players during a soccer match. The proposed model in this work is able to differentiate common passes from so-called "key passes", that generate real contributions to the teams. Rotshtein *et al.* (2005) also used GA in a predictive model for match results aiming for a discrepancy minimization between predicted and actual results.

Techniques of Operational Research (OR) have been successfully used to study the formation of soccer teams (Sales *et al*., 2019). The Multicriteria decision Aid (MCDA) has been applied in soccer, also in club's management or athletes' evaluation. The Analytic Hierarch Process (AHP) was proposed by Mu *et al*. (2016) as a new way to evaluate the best players of the world in 2014 FIFA Ballon d'Or award. The method identifies that the chosen player, Lionel Messi, did not appear as the best evaluated in any of the model sensitivity analyses, which had named James Rodriguez as the best player of the season. Qader *et al*. (2017) propose an evaluation of the performance of players, but with the objective of assisting the selection for youth teams using the TOPSIS method. The method M-MACBETH was used to select an ideal midfielder, a key position a Brazilian team (Magalhaes *et al*., 2016). The soccer players' evaluations were also performed through the Data Envelopment Analysis (DEA) by Salles *et al.* (2018). The authors explore the efficiency of more than 3,000 athletes, comparing this result with their market values but they did not find a relationship between these variables.

Regarding club evaluation, Principe *et al.* (2017) analyzed the performances of the English national league participants, using the VIKOR method, with 23 criteria and two other techniques, to distinguish the best and worst teams from the league. The MCDA method PROMETHEE II was proposed by Galariotis *et al.* (2018), to rank the participating teams in the French league and to show the relationship between finances and sports performance of the clubs. They identified that clubs with higher income are also the ones with better sports performance. A similar work, using the PROMETHEE II method, was published by Chelmis *et al.* (2017), to ranking the Greek national league clubs between 2012 and 2014 years. In this paper, the authors did not identify significant correlations between financial and commercial dimensions.

The formation of a good team is vital to assure its own success, since a bad player selection can lead to defeat, both in finance and in performance matters. The team formation's topic is not properly studied in the soccer industry or in its literature (Al-Madi *et al.*, 2016). These statements corroborate for this research purpose: the application of a method that allows, through players performance complementation, to assist teams in their initial formation.

3 METHODOLOGY

3.1 Data collection and technical procedures

As performance measure, the present research used statistics and information from players and clubs participating in the 1st division of the Brazilian Championship, the season of 2017. The data used to construct the model were extracted from three websites: WhoScored (2017), Squawka (2017) and Transfermarkt (2017).

For the team formation, the results were considered and analyzed in two situations: a) divergent as the functions performed by the players, i.e. in an analysis the result was not restricted to any predetermined position and; b) a restriction for exactly two players from the five positions (totaling 10 athletes), divided according to the functions of “Defenders”, “Fullbacks”, “Defensive Midfielders”, “Offensive Midfielders” and “Strikers”. The designation for each of these positions occurred according to the number of matches played in each one of them, with the athlete being assigned to the one with most games through the championship.

The criteria used for evaluation, as well as their acronyms and descriptions, are presented in Table 1. Eighteen criteria were used for the evaluation of the athletes’, chosen by their capacities and unique abilities. Previously to team selection, the athletes that did not appear in, at least, 25% of the championship games were removed from the analysis. This procedure pointed out a total of 322 players capable of being evaluated.

Table 1 – Description and acronym of eighteen criteria.

Acronym	Criterion	Description
C1	Disarms	Balls recovered from opposing team possession
C2	Interceptions	Balls intercepted from the opposing team
C3	Fouls suffered	Fouls suffered by the player
C4	Fouls committed	Infractions committed by the player
C5	Yellow cards	Infractions with yellow card received
C6	Red cards	Infractions with red card received
C7	Clarence	Clear the ball from the defense in opponents’ attack situations
C8	Shots on target	Shots on target, but not converted into goals
C9	Goals inside the area	Goals made inside the opponent’s great area
C10	Goals outside the area	Goals scored from outside the opponent’s great area
C11	Dribbling	Overtaking an opponent keeping possession of the ball
C12	Possession lost	Number of times the athlete lost the ball to the opponent
C13	Aerial duel won	Winning a header in a contest with the opponent
C14	Long passes	Certain passes made greater than 22 meters
C15	Short passes	Certain passes made less than 22 meters
C16	Long key passes	Long passes resulting in shots
C17	Short key passes	Short passes resulting in shots
C18	Assistence	Passes that resulted in goals

Source: Authors (2020)

3.2 Multiplex Electionis Methodus (MEM)

MEM is a method for solving problems that seek the best set of alternatives, aiming to obtain the combination of “n” individuals that are the most complementary, regarding the established criteria. Its calculation is divided into five stages, explained according to da Hora and da Costa (2015):

i. Elaborate the matrix $A_{i,j}$ of the payments of the “i” alternatives on the “j” criteria;

ii. Define the matrix $C_{j,j'}$ of complementarity between criteria;

iii. Establish the values of the weight for each evaluating criterion (vector W_j);

iv. Determine the value of tau (τ);

v. Calculate the matrix $B_{l,m}$, combining the criteria two by two, and combining the alternatives, taken 'n' to 'n', both without repetition.

The Matrix $A_{i,j}$ is given by the arrangement of alternatives (players), which are chosen in accordance of judgments of eighteen criteria and the position of the players on the soccer field. The complementarity matrix $C_{j,j'}$ was calculated to assign values between '0' and '1', related to the similarity between criteria, where the closer to '1' the more complementary and less related. The value determination was done through the application of questionnaires, where judgments of a total of 50 interviewees were analyzed. As the research deals with a subject known by a large part of the Brazilian population, such as soccer, the evaluators were enthusiasts, indicated by the authors through their ability to analyze and express opinions related to the research.

The criteria weighting was designed in a team formed by players with different characteristics, thus forming a most complementary team. However, the criteria weighting can be handled according the decision maker view, strategies or goals and forming a team more suited to their individual preferences.

According to da Hora and da Costa (2015), the tau is a measure that allows alternatives to have their performances considered "satisfactory" according to an established cut-off value, which decharacterizes it as a compensatory method. In this research, a cut-off value of tau was assigned for each criterion, chosen through its ability to drop alternatives that, after normalization of their performances, had values lower than the rate. Table 2 shows the values are chosen and the number of alternatives higher to them. The tau values chosen were the ones that could cut about 80% of alternatives, to select the group with closer performance in each criterion.

To find the most complementary alternative, the method MEM calculates the results for all combinations of alternatives. In this research, it is not possible to perform this procedure because the high number of possibilities, through combinations of 10 to 10 players, with 327 possibilities, that were judged by eighteen criteria and it would be necessary to evaluate a total of 3.35 x 10^{18} groups. Therefore, the genetic algorithm (Holland, 1992) was inserted into the method because of its capacity to assist in these types of problems, reducing significantly the required computational effort.

The metaheuristic methods could be applied to this kind of problem, where the satisfactory solution cannot be obtained in an acceptable time through conventional optimization techniques. The genetic algorithm (GA) is a metaheuristic that have been successfully applied to several practical optimization problems. The theoretical foundation of GA is based on the mechanisms of evolutionary biology, whose individuals in a population (set of possible solutions) are differentiated from one another by gene recombination, mutations and natural selection. GA looks for possible results in face of many possible solutions. The computational programming of GA uses the combination between the pull of solutions (chromosomes) and make small random modifications (mutation) to enlarge the number of possible results. In each generation, chromosomes (feasible

Table 2 – Criteria and tau values of them.

Acronym	Criterion	Tau	alternatives above tau %
C1	Disarms	0.35	20%
C2	Interceptions	0.41	19%
C3	Fouls suffered	0.34	20%
C4	Fouls committed	0.45	22%
C5	Yellow cards	0.51	17%
C6	Red cards	0.50	17%
C7	Clearances	0.25	20%
C8	Shots on target	0.25	22%
C9	Goals inside the area	0.20	20%
C10	Goals outside the area	0.15	24%
C11	Dribbling	0.25	22%
C12	Possession lost	0.40	20%
C13	Aerial duel won	0.30	19%
C14	Long passes	0.20	21%
C15	Short passes	0.48	20%
C16	Long key passes	0.19	18%
C17	Short key passes	0.40	20%
C18	Assistance	0.25	18%

Source: Authors (2020)

solutions) are created and tested, in accordance with an objective function, to select the better solutions and to eliminate the low performance chromosomes. After many generations the latest chromosome represents the best result and the solution of the problem (Beojone & Souza, 2020).

3.3 MEM-GA

According to Chambers (2001), one of the main purposes for Genetic Algorithms is the selection of parameters to optimize the performance of a system, which usually depends on the decision parameters chosen by the decision maker. The appropriate choice of these variables and decision parameters directly influences the functioning of the system, whether for better or worse, as measured by some goal. In these types of systems, the researcher must use appropriate search techniques, like GA, to optimize operating systems. Considering the problematic of this research – search an optimal combination of variables (players) for the formation of a team - it is confirmed the efficiency of this technique to reach this objective. The pseudocode representing the steps of the algorithm is described in Table 3.

Table 3 presents each calculation step for the method, dividing it in the creation and evolution of the population. The chromosomes formation, create in a random way, is possible to form any combinations of alternatives, as long as 10 individuals were present, being made in a binary way, where '1' represents the selection of the individual for the composition of the chromosome and '0' representing the opposite.

Table 3 – Genetic algorithm pseudocode.

Beginning
Population
Import data
Define Population Size *(pop)*
While population size $<$ *pop*
Generate chromosome *(crom)*
Randomly select 10 individuals
pop = [pop; chrom];
End while
Evaluate population chromosomes (fit.pop);
best.set = bigger (fit.pop)
Evolution
Set amount of iterations for crossover (*cross*)
Define number of stagnation of best set for mutation (*n.estag*);
Define number of mutations (*n.mut*);
While iteration $<$ cross
Select randomly two chromosomes from the population (parents)
Generate combination (child) between parents
Evaluate child (*fit.f*)
If *fit.f* $>$ *minor(fit.pop)*
Delete *minor*(*fit.pop*);
*p*op = [pop; fit.f];
End if
Refresh *better.set*
If *greater(fit.pop)* $>$ *better.set*
best.set = *bigger(fit.pop)*
stag = 0;
Else
*s*tag = stag + 1;
End if
If *estag* = *n.estag*
*c*ont.mut = 0;
While *cont.mut* $<$ *n.mut*
Change randomly individuals from chromosome
Select individual from *best.set*
Replace selected individual
Create new chromosome (*n.crom*)
Evaluate new chromosome (*fit.ncrom*)
If *fit.ncrom* $>$ *best.set*
*b*est.set = fit.ncrom
End if
*c*ont.mut = cont.mut +1;
End while
End if
*i*teration = iteration + 1;
End while
End of code

Source: Authors (2020)

The ‘*best.set*’ variable (line 11) give the highest complementation value, obtained among the combination of the selected alternatives, therefore, this is a value that seeks to be maximized. From line 16 it is noticed that the selection by tournament was chosen as a selection operator. This type of selection chooses a random set of solutions and, within this subset, the best solutions are selected as new parents. This choice offers a positive probability for each solution to survive, even if it has worse values of capability than others, thus increasing the technique search location (Das & Das, 2018; Kramer, 2017).

The parameters chosen by the user were: (i) population size (line 4), referring to how many combinations of alternatives will be saved to perform the crossover; (ii) number of iterations for crossover (line 12) which deals with the number of times the algorithm will repeat the combinations of alternatives of the population and it calculation; (iii) number of stagnation for mutation (line 13) which refers to the limit of stagnation, acting as a trigger for mutation; (iv) number of mutation calculations (line 14) determines how many times mutations will be performed within the optimal alternative.

The method was developed and applied in MATLAB software R2015, the code is available in the Appendix of this work. Among other procedures performed within the method, also determined by the user's choice, are the determination of the number of individuals in the group (n), which, in this research, was equal to 10, and the choice of the team's tactical formation (4-4-2).

4 ANALYSIS AND RESULTS

Table 4 presents the complementation between criteria, based on the judgment of 50 experts. To help the visualization of the level of complementarity between criteria, these were marked in reddish, yellowish or greenish tones, indicating a low, medium or high level of complementarity, respectively. Twelve criteria present low complementarity, six of them are associated with defensive functions (C1, C2, C3, C4, C5 and C7) and others six (C13, C14, C15, C16, C17 and C18) are about of types of passes performed. The criteria C8 to C12, associate with more offensive features show high level of complementarity with the others criteria.

Before the definitive application of the method, ten tests carried out were performed with different parameters for checking the results in each measurement and determine the amounts that to be chosen (Table 5). The population size, stagnation, mutation and complementation were made. The calculation time for each of the iterations was approximately 20 seconds and the total calculation time between 6 (1st test) and 37 hours (10th test).

The formation for a team without restriction of minimum or maximum number of players by position were checked. Figure 1 illustrates the complementation evolution results among alternatives during the 10 performed tests. In all performed tests the main developments occurred in the early iterations and the stabilization was reach after thousandth of iteration. The best result for this formation (0.4304) reached in five of the tests performed, with the ninth test being the one who found it the earliest, during the 1355th iteration, remaining stagnant until the end of the 3,000 iterations established. Considering there was a convergence for a single and better result in

Table 4 – Criteria complementation values. C1: Disarms; C2: Interceptions; C3: Fouls suffered; C4: Fouls committed; C5: Yellow cards; C6: Red cards; C7: Clearance; C8: Shots on target; C9: Goals inside the area; C10: Goals outside the area; C11: Dribbling; C12: Possession lost; C13: Aerial duel won; C14: Long Passes; C15: Short Passes; C16: Long key passes; C17: Short key passes; C18: Assists.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
C1		0.15	0.19	0.32	0.35	0.4	0.24	0.89	0.90	0.85	0.97	0.87	0.34	0.84	0.87	0.90	0.91	0.94
C2	0.15		0.55	0.31	0.25	0.34	0.34	0.91	0.92	0.94	0.92	0.84	0.32	0.84	0.84	0.91	0.92	0.96
C3	0.19	0.55		0.44	0.54	0.55	0.34	0.51	0.49	0.62	0.72	0.72	0.88	0.85	0.90	0.95	0.90	0.90
C4	0.32	0.31	0.44		0.32	0.33	0.46	0.88	0.95	0.97	0.91	0.86	0.55	0.85	0.99	0.98	0.99	0.98
C5	0.35	0.25	0.54	0.32		0.15	0.97	0.99	0.96	0.97	0.99	0.98	0.89	0.97	0.99	0.99	0.98	1.00
C6	0.40	0.34	0.55	0.23	0.15		0.98	1.00	0.98	0.98	1.00	0.99	0.90	0.99	0.99	1.00	0.99	1.00
C7	0.24	0.34	0.34	0.46	0.97	0.98		0.88	0.90	0.87	0.87	0.90	0.44	0.95	0.86	0.95	0.96	1.00
C8	0.89	0.91	0.51	0.88	0.99	1.00	0.88		0.25	0.25	0.52	0.58	0.57	0.69	0.70	0.75	0.65	0.69
C9	0.90	0.92	0.49	0.95	0.96	0.98	0.90	0.25		0.36	0.61	0.46	0.68	0.79	0.80	0.82	0.83	0.82
C10	0.85	0.94	0.62	0.97	0.97	0.98	0.87	0.25	0.36		0.81	0.76	0.75	0.76	0.69	0.73	0.75	0.75
C11	0.97	0.92	0.72	0.91	0.99	1.00	0.87	0.52	0.61	0.81		0.25	0.85	0.90	0.92	0.92	0.94	0.91
C12	0.87	0.84	0.72	0.86	0.98	0.99	0.9	0.58	0.46	0.76	0.25		0.90	0.91	0.93	0.92	0.93	0.93
C13	0.34	0.32	0.88	0.55	0.89	0.90	0.44	0.57	0.68	0.75	0.85	0.90		0.82	0.81	0.81	0.76	0.82
C14	0.84	0.84	0.85	0.85	0.97	0.99	0.95	0.69	0.79	0.76	0.90	0.91	0.82		0.25	0.16	0.17	0.34
C15	0.87	0.84	0.9	0.99	0.99	0.99	0.86	0.70	0.80	0.69	0.92	0.93	0.81	0.25		0.23	0.16	0.32
C16	0.9	0.91	0.95	0.98	0.99	1.00	0.95	0.75	0.82	0.73	0.92	0.92	0.81	0.16	0.23		0.16	0.27
C17	0.91	0.92	0.90	0.99	0.98	0.99	0.96	0.65	0.83	0.75	0.94	0.93	0.76	0.17	0.16	0.16		0.15
C18	0.94	0.96	0.90	0.98	1.00	1.00	1.00	0.69	0.82	0.75	0.91	0.93	0.82	0.34	0.32	0.27	0.15	

Legend: Shades of red indicate low complementarity; shades of yellow indicate medium complementarity; and shades of green indicate high complementarity.
Source: Authors (2020)

Table 5 – Parameters estimation used in MEM-GA.

Parameters	*Interactions*	*Population Size*	*Stagnation*	*Mutation*	*Complementation*
1º Test	1.000	50	20	5	0.4061
2º Test	1.000	100	25	8	0.4023
3º Test	1.500	75	15	5	0.4093
4º Test	1.500	100	20	8	0.4061
5º Test	3.000	80	25	5	0.4104
6º Test	2.500	60	14	7	0.4304
7º Test	2.500	55	12	7	0.4304
8º Test	2.500	50	10	5	0.4304
9º Test	3.000	60	12	6	0.4304
10º Test	4.500	50	12	6	0.4304

Source: Authors (2020)

50% of the tests, with the same remaining unchanged for more than 2 thousand iterations (10th test) it is possible to consider the value of complementation percentage '0.4304' as a solution sufficiently satisfactory for this case study.

In order to evaluate the performance of the chosen team, through the normalized performance of their athletes, Figure 2 was elaborated. We can see how the athletes' performances behave in relation to the population, with the selected group showing the best performance in five of the 18 examined criteria. Most criteria achieved performances higher them 0.6.

The chosen players and their positions, according to their original functions are presented on Figure 3. Considering the tactical disposition used and its representation, it is possible to say that the ideal lineup suggested was 3-6-1. It is noticed that the method prioritized midfielders and fullbacks', having only one striker and no defenders, being a team with an offensive char-

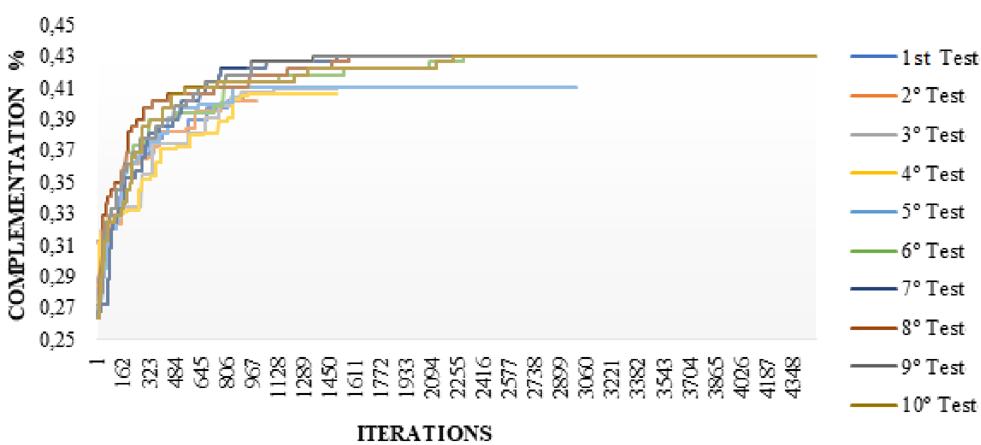

Figure 1 – Evolution of complementation values during tests.

Source: Authors (2020).

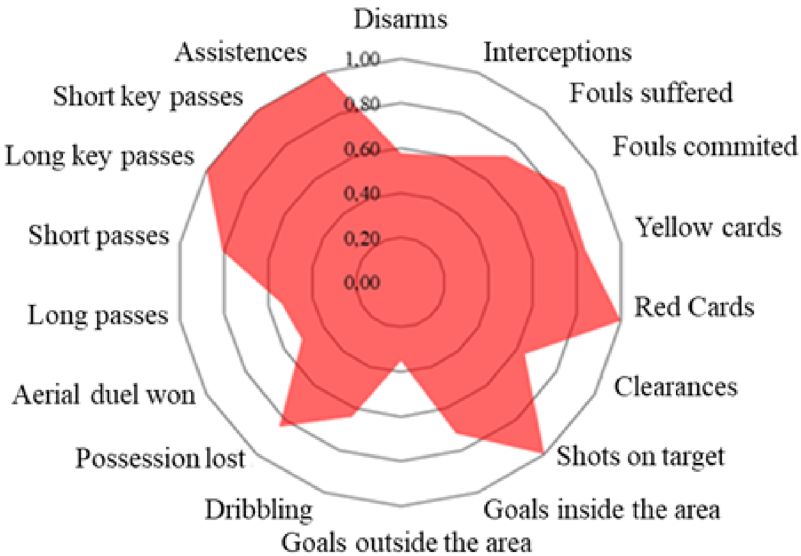

Figure 2 – Performance representation for the unrestricted team.

Source: Authors (2019).

acteristic. The proposed team should be composed by three fullbacks' (Fagner, Fábio Santos e Reinaldo), Juninho as a defensive midfielder, five offensive midfielders (Gustavo Scarpa, Bruno Silva, Hernandes, João Paulo and Thiago Neves) and Luan as striker. The absence of defenders can be explained by the characteristics of the function, very concentrated in one role, with little action in offensive functions, such as midfielders and fullbacks' that, in many cases, need to cover both functions.

The same calculations were made for a team with restrictions of two players in each of every position. The parameters used in the calculations were the same included in the 9th test, previously performed (Table 4), because they were more efficient in their search. The results evolution according to the number of iterations are illustrated in Figure 4. Regarding the evolution of values during iterations, the black line of the graphic has a similar pattern in relationship the lines of Figure 1. In both cases, there was a greater initial evolution that stabilized during the intermediate stage, peaking at the 1616th iteration, with a final value of '0.4021', a difference of '0.0283' (2.83%) to the previously obtained result.

Figure 5 shows the performance results for the athletes selected in the 4-4-2 formation, normalized according to the population performance. The team's performance reaches the greatest value in five of the 18 criteria. Some similarities are found with the performances of the earlier formation, given by the presence of some players, scaled in both formations.

According to Figure 6, the 4-4-2 team has some similarities with the unrestricted team, sharing six of its 10 athletes, demonstrating that these individuals have a high complementation rate among themselves. The defense of proposed team should be composed by two centerbacks

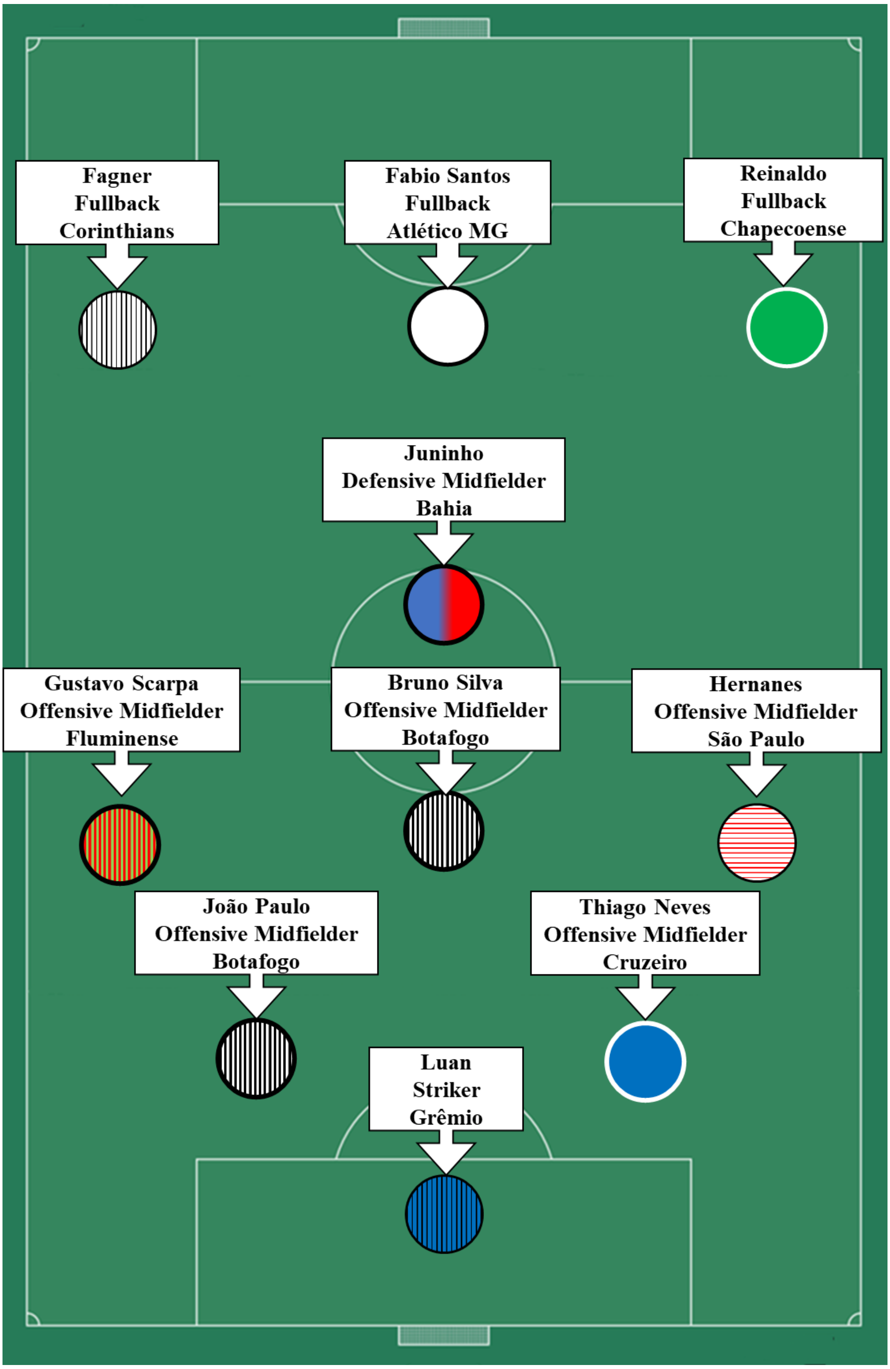

Figure 3 – Unrestricted team line-up.

Source: Authors (2019).

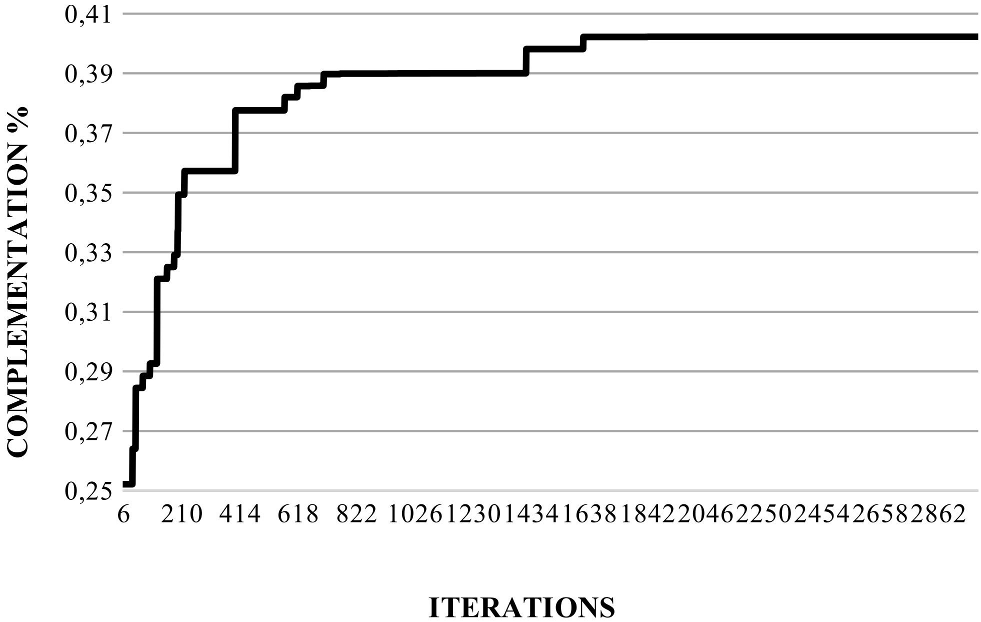

Figure 4 – Complementation between alternatives in a team with restrictions.

Source: Authors (2019).

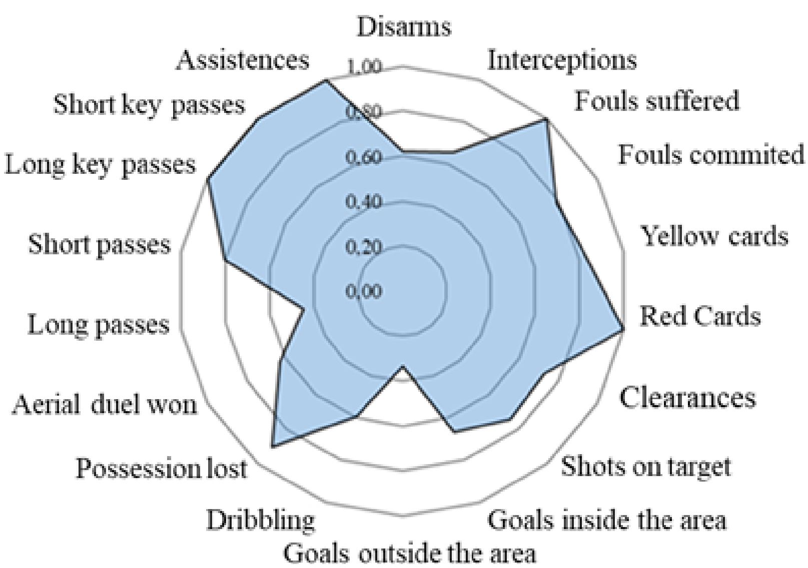

Figure 5 – Performance representation of 4-4-2 lineup.

Source: Authors (2019).

(Balbuena and Rever) and two fullbacks' (Fagner and Fábio Santos). Four midfielders (Rodrigo Lindoso, Juninho, Gustavo Scarpa and Hernanes) and Lucca and Luan as strikers. The winner of the championship of 2017, the Corinthias club, have two selected players: Fagner (also selected in the previous lineup) and Balbuena.

5 DISCUSSION

The model used in this research could be compared with other studies also faced to the formation of soccer teams. Salles et al. (2018) propose a similar case, also applying the MEM for team lineup, but in a segmented manner, dividing it for each of the 5 functions and analyzing the complementation only in groups of two athletes. In the present research we analyses 10 athletes selected at a single time aiming the complementation of the team as a whole. The two practical applications of the study, with and without restrictions imposed on the model, show seven athletes in common: Balbuena, Rever, Reinaldo, Rodrigo Lindoso, Gustavo Scarpa, Hernanes and Lucca.

The under-21 team from Indonesia was assigned with the AHP method (Purwanto *et al*., 2018). The weights and criteria used were determined according to the coach and team managers. Their results presented a lineup in the 4-3-3 tactical formation, different from the one preferred in this work. Ozceylan (2016) uses a two-stage approach to the best players selection of Fernerbahçe, a club of the Turkish 1st division. In the first phase, it was carried out an identification of the main attributes for each position using the AHP method. In the second phase, the problem was modelled using integer programming to determine the best set of players, as shown in Figure 6. The work of Ozceylan (2016) are resemblance of the team format (4-4-2) in relationship the present study. However, the case study on Fernerbahçe is restricted to just one club, less than 10% of the amount investigated in the present research.

Tavana *et al*. (2013), propose a methodology for team formation, exploring the interaction of players with their teammates, being a method capable of improving the club's teamwork. The work, applied on Parsan Soccer Club from the Iranian national league, proposed a fuzzy inference system, where fuzzy sets were used to transform linguistic variables, used in the evaluation of performance, into discrete values. The method is divided in two stages, where the first one was responsible for ranking and evaluating team members according to their performance, in a set of criteria, and within their respective importance for each position. In their second stage, the number of matches shared through individuals was considered, as for the number of times they passed the ball to each other, thus analyzing the players level of teamwork. The result of the model is the combined percentage for the best sets of players in each position (defense, midfield and attack), dividing the team in the same scheme presented in this work (4-4-2). The works of Ozceylan (2016) and Tavana *et al*. (2013) considered just a fill number of players.

Boon and Sierksma (2003) proposed a model capable of helping managers to determine optimum team formation. Initially, all the relevant qualities for a team were decided by the AHP method. Next, the importance of each criterion (by position), reflecting the game system and team tactics. By mathematical programming, the possible choices for 11 positions and 26 individuals are

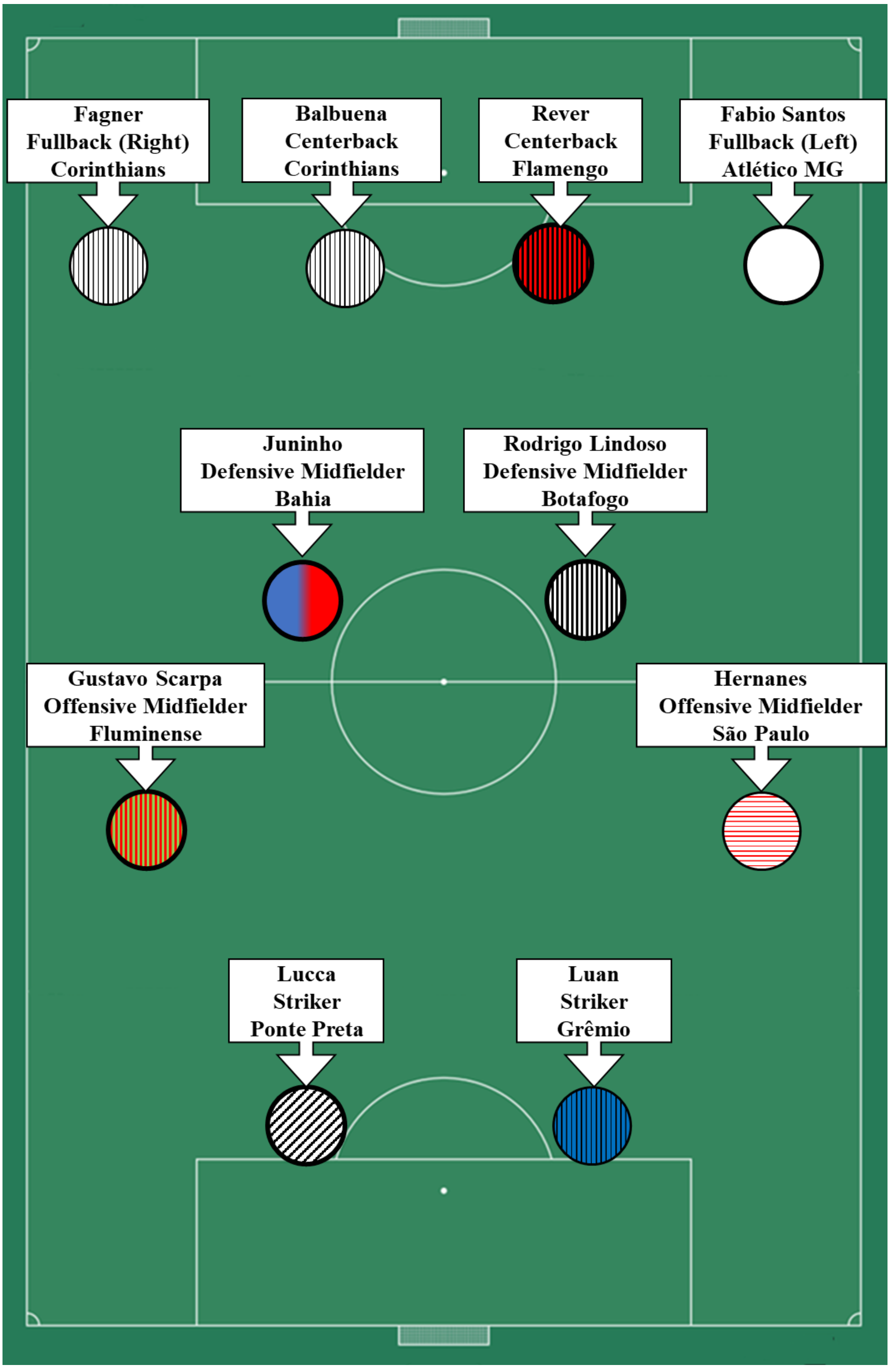

Figure 6 – Team lineup in 4-4-2 formation.

Source: Authors (2019).

Figure 7 – Fernerbahçe initial eleven.

Source: Ozceylan (2016).

calculated, seeking to acquire a team with maximum weights per player position. As seen in Ozceylan (2016), they used a limited population of players for the lineup of FC Groningen, a club that participates in the Dutch league.

Among the works of Salles et al. (2018), Purwanto et al (2018), Ozceylan (2016), Tavana et al. (2013), and Boon and Sireksma (2003), only the first one takes into consideration the complementation between the criteria, where the team players were evaluated, seeking the most complete team, such as the one performed in this research. The others perform evaluations about the individuals' performance, combining their attributes in order to maximize them, without taking into account that several of these criteria are directly related, which tends to generate a repeated evaluation on similar criteria, thus reducing the actual level of team complementation.

The tactic proposed in this research (3-6-1) considered the complementation of all players, without restrictions. Some research is found in the literature addressing aspects that reinforce the utility of this type of formation, as in Miyamoto and Kaneki (2018, p. 201), where they assess the importance of disarms and interceptions made in the opposite field, for improving the club's performance. Their results showed that the tactics used to accelerate changes between attack and defense increases possessions and are considered more effective. In other words, tactics where the distance between players is shortened, approaching the distance between the defensive and offensive line, rather than separating them between attack and defense.

Although it is not possible to assure the efficiency of the tactic proposed in this research, it is conceivable that it meets the characteristics listed by Miyamoto and Kaneki (2018), where it is stated that most of the athletes should be concentrated in the central area of the field. In this scenario, ignoring the defenders, the rest of the team are forced to supply the essentially defensive characteristics of the position and, consequently, to recover the ball in the defense field of the opposing team. It is also possible to relate the statement made by Pep Guardiola about his team the way of play, where players are required to participate in all game actions, playing offensively, and always closer to the opponent's defense field (Braun, 2013).

6 CONCLUSION

This research fulfilled its objectives by presenting two soccer teams' lineups, through the application of the MEM-GA method in the 1st Division Brazilian Soccer Championship, 2017. The method allow to carry out a team formation, while considering a superior population to other correlated works, and also avoiding the problem of segmentation in several parts, thus pointing to a formation considered more the most complementary, presenting the formation 3-6-1 (three defenders, six midfielders and a striker) as satisfactory for this case study.

The adaptability of the method to the decision maker is emphasized, where it is possible to determine, through the insertion of constraints, the one preferred tactical formation. This result was shown by presenting a team in the 4-4-2 format.

For future works, we recommend the analysis of other tactical formations, as well as the insertion of the goalkeeper position, excluded in this research. The complementation model analysis of ex-champions teams, in order to identify the relationship between team performance and complementation value.

References

AL-MADI F, AL-TARAWNEH K & ALSHAMMARI MA. 2016. HR practices in the soccer industry: Promising research arena. *International Review of Management and Marketing*, **6**(4): 641–653. Scopus.

AL-SHBOUL R, SYED T, MEMON J & KHAN F. 2017. Automated Player Selection for Sports Team using Competitive Neural Networks. *International Journal of Advanced Computer Science and Applications*, **8**(8), Article 8. Available at: http://www.academia.edu/34895719/Automated_Player_Selection_for_Sports_Team_using_Competitive_Neural_Networks

ARABZAD SM, GHORBANI M & SHIROUYEHZAD H. 2014. A new hybrid method for seed determination in sport competitions: The case of European Football Championship 2012. *International Journal of Industrial and Systems Engineering*, **17**(3): 259–274. Available at: https://doi.org/10.1504/IJISE.2014.062537

ATAN T & HÜSEYINOĞLU OP. 2017. Simultaneous scheduling of football games and referees using Turkish league data. *International Transactions in Operational Research*, **24**(3): 465–484. Scopus. Available at: https://doi.org/10.1111/itor.12201

BEOJONE CV & SOUZA RM. 2020. Improving the shift-scheduling problem using non-stationary queueing models with local heuristic and genetic algorithm. *Pesquisa Operacional*, 40: e220764, 1-22. Available at: https://doi.org/10.1590/0101-7438.2020.040.00220764

BOON BH & SIERKSMA G. 2003. Team formation: Matching quality supply and quality demand. *European Journal of Operational Research*, **148**(2): 277–292. Available at: https://doi.org/10.1016/S0377-2217(02)00684-7

BRAUN H-J. (2013). Soccer Tactics as Science? On "Scotch Professors", a Ukrainian Soccer Buddha, and a Catalonian Who Tries to Learn German. *Icon*, **19**, 216–243.

BUDAK G, KARA İ, İÇ YT & KASIMBEYLI R. 2017. New mathematical models for team formation of sports clubs before the match. *Central European Journal of Operations Research*, 1–17. Available at: https://doi.org/10.1007/s10100-017-0491-x

CAKMAK A, UZUN A & DELIBAS E. 2018. Computational modeling of pass effectiveness in soccer. *Advances in Complex Systems*, **21**(3–4), Article 3–4. Scopus. Available at: https://doi.org/10.1142/S0219525918500108

CHAMBERS LD. 2001. *The practical handbook of genetic algorithms: Applications: Vol. 1* (2nd ed.). Chapman & Hall/CRC. Available at: http://gen.lib.rus.ec/book/index.php?md5=2169269B29A04F6CAC9548E27A7CA208

CHELMIS E, NIKLIS D, BAOURAKIS G & ZOPOUNIDIS C. 2017. Multiciteria evaluation of football clubs: The Greek Superleague. *Operational Research*, 1–30. Available at: https://doi.org/10.1007/s12351-017-0300-2

COSTA HG. 2017. AHP-De Borda: A hybrid multicriteria ranking method. *Brazilian Journal of Operations & Production Management*, **14**(3): 281–287.

DA HORA HRM & COSTA HG. 2015. Proposta de um método multicritério para escolha múltipla. *Production*, **25**(2): 441–453. Available at: https://doi.org/10.1590/0103-6513.084812

DAS AK & DAS S. 2018. A Comparative Study on Different Versions of Multi-Objective Genetic Algorithm for Simultaneous Gene Selection and Sample Categorization. In: JK Mandal, S Mukhopadhyay & P Dutta (Eds.), *Multi-Objective Optimization: Evolutionary to Hybrid Framework* (pp. 243–267). Springer Singapore. Available at: https://doi.org/10.1007/978-981-13-1471-1_11

GALARIOTIS E, GERMAIN C & ZOPOUNIDIS C. 2018. A combined methodology for the concurrent evaluation of the business, financial and sports performance of football clubs: The case of France. *Annals of Operations Research*, **266**(1-2): 589–612. Scopus. Available at: https://doi.org/10.1007/s10479-017-2631-z

HOLLAND JH. 1992. Genetic Algorithms. *Scientific American*, **267**(1): 66–73.

JARVANDI A, SARKANI S & MAZZUCHI T. 2013. Modeling team compatibility factors using a semi-Markov decision process: A data-driven approach to player selection in soccer. *Journal of Quantitative Analysis in Sports*, **9**(4): 347–366. Scopus. Available at: https://doi.org/10.1515/jqas-2012-0054

KRAMER O. 2017. *Genetic Algorithm Essentials*. Springer International Publishing. Available at: http://www.springer.com/gp/book/9783319521558

KRÖCKEL P. 2017. Decision support enhancement for player substitution in football: A design science approach. *Lecture Notes in Business Information Processing*, **263**, 357–366. Scopus. Available at: https://doi.org/10.1007/978-3-319-52464-1_33

LEWIS M. 2004. *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company.

MAGALHAES LB, CASTRONEVES T, CHAVES MCC, SIMÕES CFG & PEREIRA ER. 2016. Estudo de apoio à Decisão: A escolha do camisa 10 ideal baseado no método MACBETH. *Revista Brasileira de Futsal e Futebol*, 8, 113-128.

MIYAMOTO M & KANEKI Y. 2018. Measuring Tactics of Taking the Ball Away from Defenders in the Japanese Football League. In: RHM Goossens (Ed.), *Advances in Social and Occupational Ergonomics* (pp. 336-348). Springer International Publishing.

MU E. 2016. Who really won the FIFA 2014 Golden Ball Award?: What sports can learn from multi-criteria decision analysis. *International Journal of Sport Management and Marketing*, **16**(3): 239–258. Available at: https://doi.org/10.1504/IJSMM.2016.077933

OZCEYLAN E. 2016. A mathematical model using ahp priorities for soccer player selection: A case study. *South African Journal of Industrial Engineering*, **27**(2): 190–205. Scopus. Available at: https://doi.org/10.7166/27-2-1265

PRINCIPE V, GAVIÃO LO, HENRIQUES R, LOBO V, LIMA GBA & SANT'ANNA AP. 2017. Multicriteria analysis of football match performances: Composition of probabilistic preferences applied to the English premier league 2015/2016. *Pesquisa Operacional*, **37**(2): 333–363. Scopus. Available at: https://doi.org/10.1590/0101-7438.2017.037.02.0333

PURWANTO IN, WIDODO A & HANDOYO S. 2018. System for selection starting lineup of a football players by using analytical hierarchy process (AHP). *Journal of Theoretical and Applied Information Technology*, **96**(1): 19–31. Scopus.

PYATUNIN AV, VISHNYAKOVA AB, SHERSTNEVA NL, MIRONOVA SP, DNEPROV SA & GRABOZDIN YP. 2016. The Economic Efficiency of European Football Clubs–Data Envelopment Analysis (DEA) Approach. *International Journal of Environmental and Science Education*, **11**(15): 7515–7534.

QADER MA, ZAIDAN BB, ZAIDAN AA, ALI SK, KAMALUDDIN MA & RADZI WB. 2017. A methodology for football players selection problem based on multi-measurements criteria analysis. *Measurement: Journal of the International Measurement Confederation*, **111**, 38–50. Scopus. Available at: https://doi.org/10.1016/j.measurement.2017.07.024

ROTSHTEIN AP, POSNER M & RAKITYANSKAYA AB. 2005. Football Predictions Based on a Fuzzy Model with Genetic and Neural Tuning. *Cybernetics and Systems Analysis*, **41**(4): 619–630. Available at: https://doi.org/10.1007/s10559-005-0098-4

SALLES SAF, ALMEIDA L DA C, HORA HRM & ERTHAL JR. M. 2018. Análise da eficiência de jogadores dos principais campeonatos europeus de futebol através de DEA. *Pesquisa Operacional Para o Desenvolvimento*, **10**(1): 9–26. Available at: https://doi.org/10.4322/PODes.2018.002

SALLES SAF, SARLO AMC, ALMEIDA L DA C, DA HORA HRM & ERTHAL M. 2018. Formação de equipe baseado em desempenho e interação: Aplicação ao Campeonato Brasileiro de 2017. *Apoio à Decisão Multicritério*. L Simpósio Brasileiro de Pesquisa Operacional (SBPO), Rio de Janeiro (RJ).

SALLES SAF, DA HORA HRM, ERTHAL JR. M, SANTOS ACSG & SHIMOYA A. 2019. Operations research contributions for football teams formation: a systematic review. *Pesquisa Operacional*, 39(2): 277-293. doi: 10.1590/0101-7438.2019.039.02.0277

SQUAWKA. 2017. *The home of Soccer, Transfer and Club news*. Available at: http://www.squawka.com/news/

TAVANA M, AZIZI F, AZIZI F & BEHZADIAN M. 2013. A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review*, **16**(1): 97–110. Scopus. Available at: https://doi.org/10.1016/j.smr.2012.06.002

TRANSFERMARKT. 2017. *Campeonato Brasileiro Série A - Receitas e despesas de transferências*. Available at: http://www.transfermarkt.pt/campeonato-brasileiro-serie-a/einnahmenausgaben/wettbewerb/BRA1

WHOSCORED. 2017. *Soccer Statistics | Soccer Live Scores | WhoScored.com*. Available at: https://www.whoscored.com

WRIGHT MB. 2009. 50 years of OR in sport. *Journal of the Operational Research Society*, **60**(S1), S161–S168. Available at: https://doi.org/10.1057/jors.2008.170

ZAMBOM-FERRARESI F, GARCÍA-CEBRIÁN LI, LERA-LÓPEZ F & IRÁIZOZ B. 2017. Performance evaluation in the UEFA Champions League. *Journal of Sports Economics*, **18**(5): 448–470.

ZHU F, LAKHANI KR, SCHMIDT SL & HERMAN K. 2015. *TSG Hoffenheim: Football in the Age of Analytics*. Available at: https://www.hbs.edu/faculty/Pages/item.aspx?num=49569

How to cite

SALLES SAF, DA HORA HRM, ERTHAL JÚNIOR M, VELASCO AS & CROCE PR. 2022. Multiple choice method with genetic algorithm for the formation of soccer teams. *Pesquisa Operacional*, **42**: e243537. doi: 10.1590/0101-7438.2022.042.00243537.