Multi-Criteria Decision Making with TOPSIS to Ranking Brazilian Championship Players

Igor Magno Silva[1], Rodrigo Colnago Contreras[2], Rodrigo Capobianco Guido[2]

[1]University of São Paulo (USP)
Ribeirão Preto - SP, 14040-905, Brazil

[2]São Paulo State University (UNESP)
São José do Rio Preto, SP, 15054-000, Brazil

Abstract. *The discussion of who is the best player reaches many football groups often being used more subjective standards than objectives to justify their choices, it happens even in professional football teams. Even in a world with increased access to data, there are still teams that base their choices on outdated methods. Therefore, this study was designed to develop a ranking of the Brazilian national championship players using the TOPSIS method, which uses benefit and cost criteria to find the suboptimal option. For this analysis data was collected from the website Football Stats and History, in the end a comparison between the Cartola FC rating and the TOPSIS rank was made.*
Keywords *Multi-Criteria Decision Making; Campeonato Brasileiro; Soccer; TOPSIS; Cartola FC.*

1. Introduction

Soccer is considered a national passion and moves high values in its transactions. The first division of the Brazilian Championship has great visibility, the final round of the championship was watched by nearly 11 million viewers. According to [Schaefer et al. 2019], Brazil is considered a great supplier of talented athletes, especially for European clubs, which usually invest the highest amounts in hiring players.

A valued market like soccer raises a question: who would be the best players in the Brazilian Championship? This issue afflicts the club's management, because they need to spend a budget efficiently to obtain a winning team, as explained by [Gerrard 2017] using the Moneyball: The Art of Winning an Unfair Game [Lewis 2004].

According to [Anderson e Sally 2013], the number of clubs using analysis teams for data collection and analytics grows, their work helps the management in decision-making, allowing them to find out if hiring a player will be a good investiment, after all, millions of dollars are spent and titles are in dispute.

As stated by [Anderson e Sally 2013], soccer is entering the age of the "numbers revolution", where more and more data are generated, but they have to use it. Baseball, basketball, and American football have been using data analytics for a long time, but in soccer, it is a new branch. The use of detailed data and statistics allows clubs to monitor the players performance [Gerrard 2016]. However, not all Brazilian clubs are used to using scientific elements to support their routine.

Thus, this work aimed to elaborate a classification of the ten best players, based on their statistics in Brazilian Championship. For this, the TOPSIS method was used, originally created by [Hwang et al. 1981], as this method allows players to be ranked based on their statistics. In the end, there was a comparison of the results obtained with the average score of the players in the fantasy game "Cartola FC" [Cartola 2020].

2. Material and Methods

2.1. Database

The research was executed from the statistics of soccer players obtained from the "FBref website" [FBref 2020]. This website makes its data available directly to users in CSV format, however, for each club in the championship there are 6 different tables for download, in this way a python script was created to read the tables directly from the website and transform them into data frames from the Pandas library, created by [pandas development team 2020] and published by [Wes McKinney 2010], as this would be the format used by Scikit-criteria, created by [Cabral et al. 2016], to perform the calculations for the TOPSIS method.

The target population of the study was the soccer players of the Brazilian Championship in the year 2020 and for this, only players who were on the field for at least 90 minutes, not necessarily consecutive, were selected. Thus, the database used required a pre-processing step, since its original version contains many players with less than 90 minutes of gameplay, and most of the statistics are zeroed, a situation that does not contribute to the research.

The first problem found in the data was the need for treatment of missing values of players with few pitch minutes, which in this case were all replaced by zero. Then, the correction with the separators was carried out, since the site uses both dots and commas in the decimal place separation. To finalize the treatment, the tables of each club were grouped into a single data frame with 608 lines and 24 collumns.

The data of the fantasy game Cartola FC were obtained through the API caRtola, created by [Gomide e Gualberto 2021]. Based on these data, a comparison of the ranking obtained in this study was made with the average score of players on Cartola FC.

2.2. Cartola FC

According to [Casagrande et al. 2021], a sport fantasy game is the type of game that uses in its base the performance of athletes in real competitions as the game's scoring system, existing for a long time in the United States in sports such as American football and baseball. In the fantasy game, the player chooses and manages his virtual team and his performance is according to the performance of the real athletes in the competition, thus allowing the player of the fantasy game a special interaction with the sport.

Cartola FC is the fantasy game created by Grupo Globo where its players, called cartoleiros, can select a virtual team during the 38 rounds using the athletes of the Brazilian Championship teams as a base, as reported by [Ribeiro et al. 2019]. During the Cartola FC competition, players face teams chosen by other players, based on statistical data, called scouts, which varies each time according to their income, as stated by [Junior 2014]. The scouts are scores, negative or positive, obtained by the athletes and the sum of all the scouts represent the athlete's score in the round, and they are based on the officials of the Brazilian Football Confederation (CBF).

2.3. TOPSIS Method

The TOPSIS method belongs to the Multi-Criteria Decision Making (MCDM) methodology, which consists of a grouping of decision-support techniques and procedures for evaluating alternatives based on multiple conflicting criteria, as reported by [Gomes e Gomes 2019]. According to [Lima Junior e Carpinetti 2015], this method serving managers as a support tool in uncertain, complex scenarios or with conflicting objectives.

However, even in evaluations using MCDM decision methods, there will hardly be an optimal or perfect solution, which would be the one with the maximum score in each criterion. In this case, as recommended by [Alves e Souza 2017], a suboptimal solution must be found in certain criteria instead of a non-existent solution that maximizes all criteria.

In addition, as [Lima Junior e Carpinetti 2015] said, the classification in the TOPSIS method is generated by calculating the distance of each alternative with the positive and negative ideal solutions, being the positive ideal solution the one that has the best values achieved between all alternatives, and similarly, the negative ideal solution is the one with the worst values.

3. Results and Discussion

3.1. TOPSIS ranking analysis

Following all the steps in the TOPSIS method, it was obtained the Closeness Coefficient and generate the ranking of the ten best athletes, as shown in Table 1.

Tabela 1. TOPSIS rank.

Player	Matches	Closeness
Claudinho - Bragantino	31.5	0.635802817
Giorgian De Arrascaeta - Flamengo	24.5	0.626234907
Marinho - Santos	23.8	0.605743729
Everton Ribeiro - Flamengo	29.3	0.602289322
Guilherme Arana - Atlético Mineiro	35.5	0.594674915
Patrick - Internacional	28.9	0.593040622
Luciano - São Paulo	27.7	0.592848566
Vinícius - Ceará	28.8	0.589746877
Pepê - Grêmio	26.9	0.589111514
Jefferson Savarino - Atlético Mineiro	26.7	0.579460366

After the classification, the next step was to understand the data distribution from an analysis of the boxplots of each criterion. Of the 24 criteria used for the players, some presented certain peculiarities, in some cases even outliers, as highlighted in Figure 1.

The top 10 players in the analysis were the best in one or more criteria, however Claudinho of Bragantino does not stand out from the others players in any criterion, and even so, he presented the higher Closeness Coefficient as showned in Table 1, it was because he was among the best in each criterion, but not the first in it.

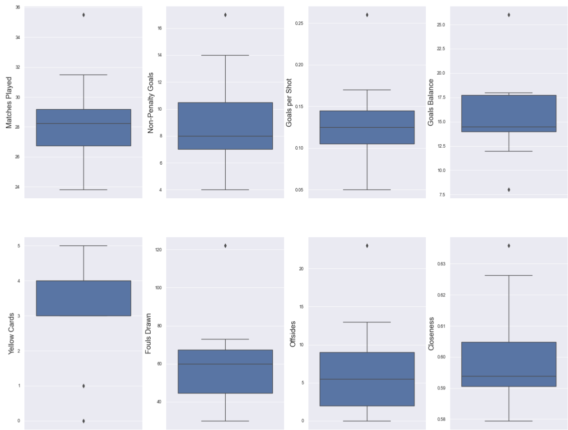

Figura 1. Boxplot of the relevant criteria for the top 10 players.

3.2. Comparison with Cartola FC

The best players in Cartola FC are on the Table 2.It is possible to observe the existence of players who participated in a few matches and were among the top 10 in the Cartola FC. It was because the use of the average score in matches to obtain the final grade, and thus, ends up being subject to cases where an athlete has a great performance in a few games and ends up standing out.

Tabela 2. Cartola FC Rank.

Player	Matches	Mean
Paolo Guerrero - Internacional	3	11.2
Marinho - Santos	26	10.25
Saravia - Internacional	8	7.96
Claudinho - Bragantino	34	6.92
Thiago Galhardo - Internacional	29	6.8
Alyson - Goiás	1	6.6
Keno - Atlético Mineiro	29	6.55
Vinícius - Ceará	30	6.5
Jemerson - Corinthians	4	6.45
Giorgian De Arrascaeta - Flamengo	26	6.45

However, cases like this are not ideal to verify the player's classification, since there is no guarantee that he would continue with this performance if he played more games during the championship.

To verify if the difference between the classification generated by the TOPSIS method and the Cartola FC is statistically significant, a t-test was performed with the averages of the games played by the athletes. As stated by [Verzani 2004], the t-test can be used for two samples as long as the population is normally distributed.

For the statistical test in this study, a significance of 5% was used. Therefore, first, a normality test was carried out for the mean of matches played in the year 2020, which proved to be statistically significant with a p-value < 0.001, the graphic distribution was shown in the Figure 2 histogram.

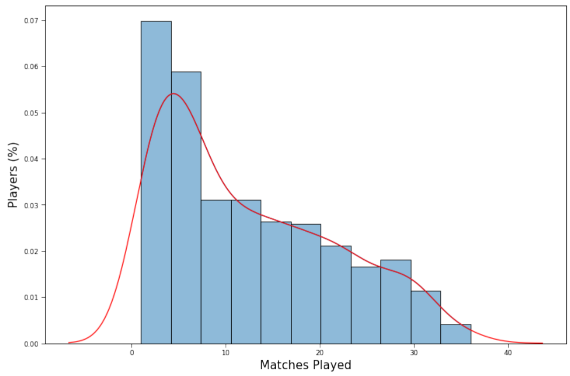

Figura 2. Histogram of matches played.

Then, the t-test was performed to verify if the difference between the ranking generated by the TOPSIS method and the Cartola FC was statistically significant, with the null hypothesis (H_0) being that the means are equal and that the alternative hypothesis (H_1) that the means are different.

Thus, with p-value < 0.05 obtained in the t-tests performed, the null hypothesis (H_0) was rejected, indicating that there was a statistically significant difference in the mean matches played between the classification by the TOPSIS method and the classification according to Cartola FC.

4. Final Considerations

The TOPSIS method proved to be a robust tool for the elaboration of the desired ranking in the study, it is believed that this method can be used also by other sports, to carry out performance analysis of players. However, the TOPSIS ranking of presented a statistically significant difference from the Cartola FC ranking, which, however, does not mean a failure or error in the classification system, since the ranking obtained in the study proved to be less exposed to cases of athletes who participated in one or a few matches and scored high in this small sample.

In addition, the TOPSIS can be compared in future studies with another MCDM methods, like AHP or VIKOR, or in conjunction with machine learning algorithms since its result generates a new variable for analysis. It can be used in conjunction with other variables, such as the purchase value of a player, thus verifying whether such an acquisition would be advantageous, or even in the search for talented players.

Another possible use for the continuation of this study would be the segregation of the players' on-field functions, dividing them by positions: goalkeeper, defender, winger, midfielder, and striker, because among the various criteria used in the research, some are more important for certain positions than others.

However, the MCDM should serve as a decision support tool and not as a substitute for human decision, there must also be other analyzes that must be considered when hiring a player, such as injury history, in addition to other behavioral, psychological and off-field factors.

5. References

Referências

Alves, M. A. e Souza, L. T. (2017). Método multicritério topsis aplicado à satisfação de usuários de smartphones com os sistemas operacionais android, ios e windows phone. *Rev. Sist. Informação FSMA*, 20:2–9.

Anderson, C. e Sally, D. (2013). *Os números do jogo: por que tudo o que você sabe sobre futebol está errado*, volume 1. São Paulo: Paralela.

Cabral, J. B., Luczywo, N. A., e Zanazzi, J. L. (2016). Scikit-criteria: colección de métodos de análisis multi-criterio integrado al stack científico de python. In *XIV Simposio Argentino de Investigación Operativa (SIO 2016)-JAIIO 45 (Tres de Febrero, 2016).*

Cartola (2020). Cartola fc. Available at: http://cartolafc.globo.com/.

Casagrande, M. C., de Lima Lavarda, S., e Silveira, A. C. M. (2021). Fantasy game e reconfiguração da cultura desportiva: um estudo de caso do cartola fc. *Intexto*, pages 95133–95133.

FBref (2020). Football stats and history statistics, scores and history for 100+ men's and women's club and national team competitions. Available at: https://fbref.com/en.

Gerrard, B. (2016). Analytics, technology and high-performance sport. In *Critical issues in global sport management*, pages 227–240. Routledge.

Gerrard, B. (2017). The role of analytics in assessing playing talent. In *Routledge handbook of talent identification and development in sport*, pages 422–431. Routledge.

Gomes, L. F. A. M. e Gomes, C. F. S. (2019). *Princípios e métodos para a tomada de decisão: Enfoque multicritério*, volume 6. São Paulo: Atlas.

Gomide, H. e Gualberto, A. (2021). *CaRtola: Extração de dados da API do CartolaFC, análise exploratória dos dados e modelos preditivos em R e Python*.

Hwang, C.-L., Yoon, K., Hwang, C.-L., e Yoon, K. (1981). Methods for multiple attribute decision making. *Multiple attribute decision making: methods and applications a state-of-the-art survey*, pages 58–191.

Junior, G. C. (2014). Burlando o círculo mágico: o esporte no bojo da gamificação. *Movimento*, pages 941–963.

Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.

Lima Junior, F. R. e Carpinetti, L. C. R. (2015). Uma comparação entre os métodos topsis e fuzzy-topsis no apoio à tomada de decisão multicritério para seleção de fornecedores. *Gestão & Produção*, 22:17–34.

pandas development team, T. (2020). pandas-dev/pandas: Pandas. https://doi.org/10.5281/zenodo.3509134.

Ribeiro, L. E. d. S. et al. (2019). Predição de escalações para o jogo cartolafc utilizando aprendizado de máquina e otimização.

Schaefer, J. L., Fagundes, B. J., Moraes, J., Nara, E. O. B., e Kothe, J. V. (2019). Aplicação de métodos multicritérios para ordenação e comparação da eficiência financeira dos clubes de futebol do campeonato brasileiro de futebol da série a. *RBFF-Revista Brasileira de Futsal e Futebol*, 11(42):31–43.

Verzani, J. (2004). *Using R for introductory statistics*. Chapman and Hall/CRC.

Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt e Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.