Programação Linear - Trabalho Final

Buscando o Elenco Ótimo para o Campeonato Brasileiro de Futebol

Equipe 7

Felipe Melo, Igor Pacheco e Matheus Vargas

1. Introdução

O futebol é sem dúvidas o esporte mais popular do mundo. Ele movimenta bilhões de euros a cada temporada em transferências de jogadores, vendas de produtos e com a receita gerada em partidas. Recentemente, foi introduzido um modelo especial de constituição de empresas voltado para o futebol, que são as chamadas Sociedades Anônimas de Futebol (SAFs).

Com a popularização das SAFs no futebol brasileiro, pensamos no cenário hipotético de alguma empresa ou mesmo pessoa adquirir as ações de um clube nacional. Nesse contexto, deseja-se encontrar o melhor elenco possível formado por jogadores que atuam no Brasil, tendo como entrada um valor total disponível para transferências.

Para isso, utilizamos dados da plataforma SofaScore, que reúne informações sobre placares em tempo real, calendários de jogos e estatísticas de jogadores, bem como suas avaliações e mapas de calor. Ademais, usamos também dados da plataforma Transfermarkt, que possui, entre outras coisas, referências para o valor de mercado dos jogadores de futebol. Por fim, os dados utilizados são referentes à temporada de 2021, uma vez que a de 2022 ainda está em andamento.

2. Os Dados

2.1 Refinando a Seleção dos Dados

Nessa etapa, restringimos a coleta dos dados, buscando uma abordagem mais próxima à realidade. Nesse sentido, foram considerados jogadores que disputaram mais de 10 partidas no Campeonato Brasileiro de 2021, para que jogadores que tiveram poucas participações fossem descartados, já que amostras pequenas podem gerar anomalias estatísticas. Como nosso objetivo era selecionar os melhores jogadores possíveis, sujeito a um limite de preço específico, estávamos principalmente interessados em quão bem avaliado é o jogador e em seu valor de mercado, que está disponível em milhões de euros.

Outros dados utilizados foram a posição do campo em que o jogador atua (goleiro e centroavante, por exemplo) e o seu time. Foi importante coletar o clube de cada jogador, pois isso facilitou a junção das tabelas do Transfermarkt e do SofaScore.

#	Time	Nome	Nota SofaScore
			▼
1		Giorgian De Arrascaeta	7.49
2		Gustavo Scarpa	7.34
3	8	Artur	7.33
4	*	Nikão	7.32
5	3	João Ricardo	7.32
6	Ÿ	Lucas Crispim	7.28
7	8	Léo Ortiz	7.26
8	(3)	Edenilson	7.25
9	Ψ	Hulk	7.24
10		Raphael Veiga	7.23
11	-	Filipe Luís	7.23
12	=	Gabriel Barbosa	7.21
13	*	Mailson	7.20
14	*	David Terans	7.19
15	•	Danilo Barcelos	7.18
16	ð	Vina	7.18
17	•	Bruno Henrique	7.18
18	(F)	Ademir	7.17
19	0	Daniel	7.17
20	The state of	Robert Arboleda	7.16
< Ant	terior		<u>1</u> 2 24 Próximo >

Figura 1: Tabela de jogadores do SofaScore

Existem ainda jogadores sem um valor de mercado associado. Nesse caso, optamos por descartar esses atletas ao invés de atribuir o valor zero a eles.

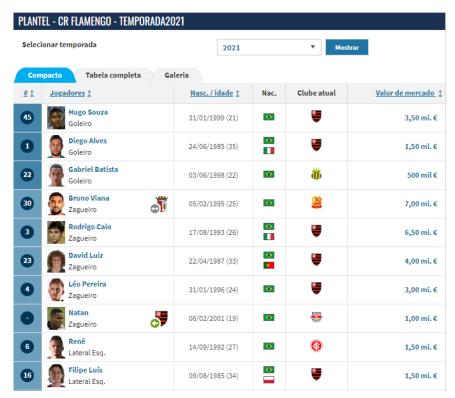


Figura 2: Tabela de jogadores por time do Transfermarkt

2.2 Obtendo e Tratando os Dados

A primeira etapa foi a coleta de dados das duas plataformas. Para a retirada de dados do Transfermarkt, utilizamos a técnica de Web Scraping por meio do módulo BeautifulSoup, disponível na linguagem de programação Python. O SofaScore, entretanto, é mais restritivo quanto a essa prática. Por isso, tivemos de copiar manualmente os dados disponíveis nas tabelas de avaliações de jogadores.

Depois, utilizamos o módulo Pandas para juntar as tabelas. As duas tabelas foram unidas com base no time do jogador e em seu nome. Apenas um conflito ocorreu, com dois jogadores do Cuiabá: o lateral João Lucas e o volante João Lucas. Como o volante não tinha rating SofaScore, ele foi excluído. Além disso, outros problemas com relação aos nomes dos jogadores e times também tiveram que ser solucionados.

	Time	Jogador	Posição	Preço	Rating	Colocação
0	Flamengo	Giorgian De Arrascaeta	Meia Ofensivo	18.000	7.49	1
1	Palmeiras	Gustavo Scarpa	Meia Ofensivo	3.500	7.34	2
2	Red Bull Bragantino	Artur	Ponta Direita	5.000	7.33	3
3	Athletico	Nikao	Meia Ofensivo	1.400	7.32	4
4	Ceará	Joao Ricardo	Goleiro	0.400	7.32	5
			•••			
447	Red Bull Bragantino	Weverson	Lateral Esq.	0.500	6.49	460
448	Bahia	Oscar Ruiz	Meia Esquerda	0.500	6.47	461
449	Fortaleza	Edinho	Meia Ofensivo	0.475	6.47	462
450	Athletico	Carlos Eduardo	Ponta Direita	3.000	6.46	463
451	Bahia	Lucas Araujo	Volante	0.050	6.43	464

Figura 3: Junção dos dados das duas plataformas

Ao juntarmos as tabelas, notamos posições nas quais havia poucos jogadores:

- Meia-esquerda (4 jogadores)
- Meia-direita (1 jogador)
- Segundo atacante (7 jogadores)

	Time	Jogador	Rating	Colocação	Preço
Posição					
Meia Esquerda	Fortaleza	Lucas Crispim	7.28	6	0.30
Meia Esquerda	Internacional	Patrick	6.74	320	4.50
Meia Esquerda	Flamengo	Kenedy	6.57	439	10.00
Meia Esquerda	Bahia	Oscar Ruiz	6.47	461	0.50
Meia Direita	Fortaleza	Yago Pikachu	7.01	83	1.40
Seg. Atacante	América Mineiro	Mauro Zarate	6.94	131	1.00
Seg. Atacante	São Paulo	Eder	6.68	364	0.80
Seg. Atacante	São Paulo	Vitor Bueno	6.66	378	1.60
Seg. Atacante	Atlético Mineiro	Eduardo Vargas	6.91	158	3.00
Seg. Atacante	Chapecoense	Fernandinho	6.65	397	0.40
Seg. Atacante	Juventude	Roberson	6.54	456	0.20
Seg. Atacante	Fortaleza	Robson	6.90	164	0.85

Figura 4: Jogadores de posições escassas

Para evitar situações nas quais se tornaria inviável a montagem do time devido a falta de jogadores nestas posições, decidimos eliminá-las e adicionar estes jogadores em posições com funções similares. Desse modo, jogadores cujas funções originais eram Meia-esquerda e Meia-direita viraram Alas, ao passo que os atletas que atuam como Segundo Atacante viraram Centroavantes.

Por fim, precisávamos de dados para restringir os jogadores pela sua posição. Para isso, coletamos manualmente 14 formações da Wikipedia (da seção Formações Contemporâneas) e adicionamos 5 outras com pequenas variações das iniciais, totalizando 19 formações.





Figura 5: Formações 4-4-2 e 4-3-3, retiradas da Wikipedia

2.3 Analisando os Dados

Vamos agora explorar os dados para obter informações relevantes à etapa de modelagem.

- 452 jogadores
- Rating:
 - o Média de 6,84
 - Mediana de 6.84
 - Maior valor: 7,49 Giorgian De Arrascaeta (Flamengo)
 - Menor valor: 6,43 Lucas Araújo (Bahia)

- Valor de mercado (em milhões de euros):
 - o Média de 1,99
 - Mediana de 1
 - Maior valor: 26 Gabriel Barbosa (Flamengo)
 - Menor valor: 0,025 Igor Torres (Fortaleza)
- Posições:
 - o 5 alas
 - 60 centroavantes
 - o 28 goleiros
 - o 35 laterais direitos
 - o 37 laterais esquerdos
 - o 43 meias centrais
 - 50 meias ofensivos
 - 39 pontas direitas
 - 37 pontas esquerdas
 - o 49 volantes
 - o 69 zagueiros

Observe que a diferença entre o maior rating e o menor rating é bastante bem pequena (7,49-6,43=1,06), o que indica que as diferenças de ratings entre times selecionados também não será grande, indicando maior relevância aos números decimais.

3. Modelagem

3.1 As Variáveis

Fizemos um esforço para condensar os dados do problema em variáveis de simples entendimento, com isto ficamos com o seguinte conjunto:

- R_i → Rating associado ao jogador j.
- T_i → Seletor associado ao jogador j, para o time titular.
- B_i → Seletor associado ao jogador j, para o time reserva.
- P_{ii} → Variável que informa se o jogador j é capaz de atuar na posição i.
- C_i → Custo associado ao jogador j.
- I → Investimento disponível.

3.2 As Equações

O objetivo é então maximizar a qualidade do time, representada pelo rating SofaScore, restringindo pelo quanto se quer gastar em compras de jogadores e pela posição de tais jogadores em campo. Com isto em mente, chegamos a equação abaixo:

$$Z = \sum_{j} R_j T_j + R_j B_j$$

Esta equação representa a soma total dos ratings de cada jogador selecionado, tanto para o time titular como para o time reserva. Visto que T_j e B_j são variáveis binárias e mutuamente excludentes, explicado abaixo, no máximo uma parcela da soma será considerada no somatório para um dado jogador.

Como já mencionado, existem algumas restrições que nos impedem de selecionar jogadores livremente. Essas restrições levam em conta critérios financeiros e também devem respeitar a quantidade necessária de jogadores em cada posição específica para a formação escolhida. Chegamos às seguintes restrições:

$$\sum_{j} T_{j} = 11$$

$$\sum_{i} B_{i} = 11$$

Essas equações limitam o número de jogadores selecionados para o time titular e reserva, respectivamente, que deve respeitar a quantidade estipulada pelas regras do esporte, ou seja onze atletas em campo para cada time.

$$T_i + B_i \le 1, \ \forall j$$

Logicamente, um jogador não pode estar simultaneamente listado como titular e reserva, a equação anterior faz justamente este papel de garantir que um dado atleta não seja selecionado para os dois papéis.

$$\sum_{i} \sum_{j} P_{ij} T_{j} = N_{i}$$

$$\sum_{i} \sum_{j} P_{ij} B_{j} = N_{i}$$

Um dos critérios de seleção de jogadores é a posição na qual ele se sente mais confortável para jogar, já que a formação escolhida determina uma quantidade fixa de jogadores necessários em cada posição. As duas equações acima cuidam desta necessidade, tanto para o time titular quanto para o time de reservas.

$$\sum_{j} C_j T_j + C_j B_j \le I$$

Por fim, a seleção de jogadores deve respeitar um teto financeiro, nunca devendo ultrapassar o valor do investimento inicial recebido como entrada.

4. Transformando em Código

Foi utilizada API em Python do software de otimização da Google, o OR-Tools, com o solver de programação inteira mista SCIP.

Na modelagem da seção anterior, foram utilizadas duas variáveis de decisão, T (que representa os titulares) e B (que representa os jogadores do banco). Ao transformar isso em código, no entanto, utilizamos apenas uma variável, S, com os jogadores duplicados, para simplificar o programa.

Um dos problemas que tivemos foi que o algoritmo não diferenciava os jogadores titulares dos reservas. Portanto, era possível que os reservas fossem todos melhores do que os titulares. Para que o algoritmo priorizasse os titulares, foi criada uma nova coluna na tabela: Rating Reserva, com valor igual a 50% da coluna Rating. Isso força o algoritmo a considerar os melhores jogadores de um time selecionado como titulares, como desejado. Depois, é possível recuperar os valores originais de rating do banco, portanto isso não é um problema.

4.1 Função Objetivo

A função objetivo do problema foi assim definida:

```
dados['coeficientes objetivo'] = list(juncao['Rating']) + list(juncao['Rating Reserva'])
```

Que é a concatenação das listas de Rating dos titulares e reservas.

4.2 Restrições

Os coeficientes da restrição de custo:

```
list(juncao['Preço']) + list(juncao['Preço'])
```

Que tem como limite inferior 0 e limite superior o parâmetro "investimento". Lembrando que todos os coeficientes são duplicados para representar os reservas.

Restrições de posição:

```
# restringe pela posição do jogador. Com coluna = 0, restringe o goleiro, por exemplo
tabela_pos = cria_tabela_pos(matriz_pos)
for coluna in tabela_pos:
    dados['coeficientes_restrição'].append(list(tabela_pos[coluna]))
```

Que tem que ser exatamente igual a lista "formacao", tanto para os titulares quanto para os reservas.

Restrições de não-repetição de um mesmo jogador:

```
# evita que jogadores duplicados possam aparecer no time. Uma restrição por jogador
for i in range(len(juncao)):
    jogadores_duplicados = [0] * len(juncao)
    jogadores_duplicados[i] = 1

dados['coeficientes_restrição'].append(jogadores_duplicados+jogadores_duplicados)
```

Que tem que ser maior que 0 e menor que 1, para que se selecione a "versão" do jogador titular ou a reserva ou nenhum dos dois.

• E aqui estão os limites superiores e inferiores mencionados acima:

A primeira coluna é com relação ao custo, a segunda é a de posicionamento e a última a não-repetição do mesmo jogador.

4.3 Código Fonte

O programa completo está disponível no GitHub: https://github.com/theuvargas/programacao-linear

5. Exemplos

Para ilustrar o problema proposto, testamos nosso programa para dois cenários distintos de investimento:

- 1. Dinheiro "infinito"
- 2. Investimento de 20 milhões de euros

É esperado que no cenário 1 o algoritmo nos retorne o melhor elenco de 22 jogadores possível, sem levar em conta o seu preço. No cenário 2, ele deve retornar um elenco ótimo com valor máximo de 20 milhões de euros.

5.1 O Time Ideal

A seguir está o input necessário para encontrar o time ideal e o output do programa:

Figura 6: Resultado do time ideal

A função melhor_time foi chamada com argumento de investimento = infinito, e um time com rating de 158,71 foi encontrado. O preço desse time é de 98,5 milhões de euros, e a formação ótima é o 4-3-3 ofensivo. As linhas 2 e 3 do output do programa representam os índices dos titulares e reservas na tabela "juncao", respectivamente.

	Time	Jogador	Posição	Preço	Rating	Colocação
0	Flamengo	Giorgian De Arrascaeta	Meia Ofensivo	18.0	7.49	1
1	Palmeiras	Gustavo Scarpa	Meia Ofensivo	3.5	7.34	2
2	Red Bull Bragantino	Artur	Ponta Direita	5.0	7.33	3
4	Ceará	Joao Ricardo	Goleiro	0.4	7.32	5
6	Red Bull Bragantino	Leo Ortiz	Zagueiro	2.0	7.26	7
8	Atlético Mineiro	Hulk	Centroavante	2.5	7.24	9
10	Flamengo	Filipe Luis	Lateral Esq.	1.5	7.23	11
16	Flamengo	Bruno Henrique	Ponta Esquerda	4.5	7.18	17
19	São Paulo	Robert Arboleda	Zagueiro	1.6	7.16	20
24	Palmeiras	Marcos Rocha	Lateral Dir.	1.5	7.14	25
26	Internacional	Rodrigo Dourado	Volante	4.0	7.13	27

Figura 7: Titulares do time ideal

	Time	Jogador	Posição	Preço	Rating	Colocação
3	Athletico	Nikao	Meia Ofensivo	1.40	7.32	4
9	Palmeiras	Raphael Veiga	Meia Ofensivo	4.50	7.23	10
11	Flamengo	Gabriel Barbosa	Centroavante	26.00	7.21	12
12	Sport Recife	Mailson	Goleiro	0.50	7.20	13
14	Fluminense	Danilo Barcelos	Lateral Esq.	0.85	7.18	15
17	América Mineiro	Ademir	Ponta Direita	0.50	7.17	18
23	Corinthians	Fagner	Lateral Dir.	3.00	7.14	24
27	Corinthians	Roger Guedes	Ponta Esquerda	9.00	7.13	28
29	Palmeiras	Gustavo Gomez	Zagueiro	6.00	7.12	30
30	Athletico	Thiago Heleno	Zagueiro	1.60	7.11	31
42	Sport Recife	Ze Welison	Volante	0.65	7.08	43

Figura 8: Reservas do time ideal

5.2 O Time de 20 Milhões de Euros

A seguir o input e output relacionados ao melhor time de 20 milhões de euros:

```
In [19]: melhor_time(20)

Out[19]: (157.18,
      [3, 4, 6, 7, 10, 17, 19, 24, 31, 42, 53],
      [13, 14, 18, 20, 36, 37, 44, 49, 60, 63, 92],
      '4-3-3 equilibrado',
      19.94999999999999)
```

Figura 9: Resultado do time de 20 milhões de euros

A função melhor_time agora foi chamada com argumento investimento = 20, e um time com rating 157,18 foi encontrado. O preço desse time é de 19,95 milhões de euros (bem próximo do teto de investimento), e a formação ótima é o 4-3-3 equilibrado. As linhas 2 e 3 do output do programa representam os índices dos titulares e reservas na tabela "juncao", respectivamente.

	Time	Jogador	Posição	Preço	Rating	Colocação
3	Athletico	Nikao	Meia Ofensivo	1.40	7.32	4
4	Ceará	Joao Ricardo	Goleiro	0.40	7.32	5
6	Red Bull Bragantino	Leo Ortiz	Zagueiro	2.00	7.26	7
7	Internacional	Edenilson	Meia Central	2.40	7.25	8
10	Flamengo	Filipe Luis	Lateral Esq.	1.50	7.23	11
17	América Mineiro	Ademir	Ponta Direita	0.50	7.17	18
19	São Paulo	Robert Arboleda	Zagueiro	1.60	7.16	20
24	Palmeiras	Marcos Rocha	Lateral Dir.	1.50	7.14	25
31	Red Bull Bragantino	Ytalo	Centroavante	0.90	7.10	32
42	Sport Recife	Ze Welison	Volante	0.65	7.08	43
53	Bahia	Rai Nascimento	Ponta Esquerda	0.20	7.06	54

Figura 10: Titulares do time com 20 milhões de euros de investimento

	Time	Jogador	Posição	Preço	Rating	Colocação
13	Athletico	David Terans	Meia Ofensivo	1.40	7.19	14
14	Fluminense	Danilo Barcelos	Lateral Esq.	0.85	7.18	15
18	Internacional	Daniel	Goleiro	0.05	7.17	19
20	América Mineiro	Juninho Valoura	Meia Central	0.50	7.16	21
36	Juventude	Matheus Peixoto	Centroavante	0.70	7.08	37
37	Cuiabá	Joao Lucas	Lateral Dir.	0.30	7.08	38
44	Ceará	Luiz Otavio	Zagueiro	0.60	7.08	45
49	Atlético Mineiro	Nathan Silva	Zagueiro	0.60	7.07	50
60	Palmeiras	Felipe Melo	Volante	0.60	7.05	61
63	Red Bull Bragantino	Tomas Cuello	Ponta Esquerda	0.80	7.04	64
92	Grêmio	Jhonata Robert	Ponta Direita	0.50	6.99	93

Figura 11: Reservas do time com 20 milhões de euros de investimento

6. Conclusões

Em suma, podemos tirar algumas conclusões desse cenário hipotético. A primeira delas é que cada centésimo do rating é bastante relevante, uma vez que a diferença de rating do jogador mais bem avaliado para o menos bem avaliado é de 1,06 ponto. Isso é refletido também na seleção dos jogadores que vão compor os times, visto que nos dois exemplos apresentados a diferença de rating total dos times é de 1,53, enquanto que a diferença de preços é de 78,55 milhões de euros.

A maioria das restrições existe para evitar que os jogadores não apareçam mais de uma vez no time. São 452 restrições desse tipo (uma por jogador). Apesar do número elevado de restrições, o programa não é lento, e leva menos de um segundo para encontrar o time ideal para uma formação. Medimos o tempo da função melhor_time, e ela leva em média 4,81 segundos. Considerando que são 19 formações, a média de tempo por formação é de 0,25 segundos.

Por fim, vale ressaltar que a qualidade dos times gerados pelo programa é totalmente dependente da qualidade dos dados coletados. Alguém que não concorde com alguma avaliação de preço do Transfermarkt ou alguma nota do SofaScore com certeza não concordará com a inclusão de certos jogadores.

7. Bibliografia

- https://www.sofascore.com/
- https://www.transfermarkt.com.br/
- https://www.ftm.eu/articles/transfermarkt-volunteers-european-football
- https://www.nytimes.com/2021/08/12/sports/soccer/soccer-football-transferma rkt.html
- https://pt.wikipedia.org/wiki/Esquemas t%C3%A1ticos do futebol