

Sistema de Recomendação Autoexplicativo de Jogadores de Futebol

Duarte Meneses

duartemeneses@student.dei.uc.pt

Patrícia Costa

patriciacosta@student.dei.uc.pt

Departamento de Engenharia Informática
Inteligência Artificial Centrada no Humano - MECD

Abstract

Na sociedade atual, a quantidade de dados disponível impõe que exista uma seriação da informação que chega ao consumidor final. Os sistemas de recomendação têm um papel importante nessa questão visto que apenas apresentam ao utilizador dados que lhe interessam (definidos por ele manualmente ou através de um perfil). No entanto, para os sistemas de recomendação serem verdadeiramente aceites pelos humanos, devem explicar as recomendações que fazem. Só assim um utilizador se sente esclarecido quanto às opções que o sistema lhe gerou. É com isso em vista que decidimos desenvolver um sistema de recomendação autoexplicativo de jogadores de futebol. Pretendemos que este simplifique a vida aos clubes de futebol, poupando-lhes tempo, dinheiro e recursos humanos, explicando sempre as vantagens e desvantagens de contratar determinado jogador.

Palavras-chave – futebol, sistema de recomendação, inteligência artificial, jogadores, explicabilidade

1 Introdução

Atualmente, o futebol gera e faz gerar milhares de milhões de euros. No entanto, as recentes regras do *fair play* financeiro, sobretudo na Europa (impostas pela UEFA – *Union of European Football Associations*), fazem com que os clubes tenham de ser bem geridos e saber exatamente onde e como gastar o seu dinheiro. Não são poucas as vezes em que um clube gasta milhões de euros em contratações e depois esses jogadores ou não se encaixam na equipa, ou não rendem o que era esperado. Com isto, os clubes perdem dinheiro já que não têm retorno financeiro nem desportivo com esses jogadores. Uma das razões para o falhanço nas contratações pode estar numa fraca rede de olheiros e analistas de dados. Se um clube tivesse à sua disposição um sistema de recomendação de jogadores que lhe sugeria futebolistas com as características pretendidas/semelhantes a outros atletas, talvez o erro nas contratações reduzisse.

Imagine-se que o Benfica planeia vender o Gonçalo Ramos na próxima janela de transferências e quer um substituto com as mesmas características para que a ideia de jogo da equipa não se altere. Habitualmente, um clube procura por alternativas vendendo milhares de jogos, utilizando para isso inúmeros olheiros espalhados pelo mundo, levando a um esforço gigantesco em termos de recursos humanos e um enorme gasto financeiro. No entanto, se o Benfica tiver um sistema de recomendação como o referido acima, pode apenas pedir que este lhe recomende jogadores com as características semelhantes às do Gonçalo. Deste modo, apenas tem de analisar os atletas que o sistema sugere (teoricamente, o filtro já está tão apertado que não irão perder tempo a analisar jogadores muito diferentes do pretendido).

Este é assim um processo mais rápido, eficaz e menos dispendioso financeiramente que o tradicional. Se, aliado a isto, o sistema conseguir explicar a razão para recomendar determinado jogador e aspetos negativos e positivos da contratação do mesmo, o clube teria ainda o trabalho mais facilitado. Apenas indicando as características que gostava que um jogador tivesse, o clube obteria uma lista restrita de atletas (os que mais se enquadrariam nas preferências). Aliado a isto, teria também uma explicação para serem estes os recomendados e uma lista de aspetos positivos e negativos de uma eventual contratação. Por exemplo, um jogador que tenha muitas lesões, embora até possa ter características semelhantes, não deve ser tão recomendado e o sistema deve explicar essa mesma razão. Deste modo, pretendemos que a nossa aplicação recomende 10 jogadores, ordenados por semelhança, e explique a razão para aquele jogador estar a ser recomendado, a par de algumas objeções para tal contratação. Nas próximas secções abordaremos o trabalho relacionado com o nosso projeto, apresentaremos os materiais por nós utilizados e analisaremos o dataset. Após isto, iremos explicar como decidimos abordar o problema, mostrar um pouco da fase de experimentação, apresentando de seguida o modo como validamos o sistema.

2 Trabalho relacionado

Existem inúmeros sistemas de recomendação *content-based* já implementados e com enorme sucesso. É o caso dos sistemas das lojas de roupa *online* que recomendam artigos semelhantes aos que o cliente viu ou adquiriu. Outro sistema muito conhecido é utilizado pelo *Spotify* que gera um perfil de utilizador através da música que este ouve, as classificações que dá, Tendo o perfil de utilizador traçado, fica mais fácil para o *Spotify* recomendar músicas que o cliente possa gostar de ouvir.

Segundo o artigo “*Content-based Recommendation Systems*” [9], o que um sistema de recomendação faz não é mais do que filtrar informação conforme as preferências do utilizador. Enquanto, nos exemplos explanados acima, o sistema vai construindo perfis de utilizadores segundo a sua atividade, no nosso sistema, a ideia passa por ser o utilizador a introduzir as suas preferências aquando de cada recomendação. De qualquer maneira, o sistema compara (correlaciona) as preferências do utilizador com os inúmeros objetos presentes na base de dados, recomendando assim aqueles que mais se assemelham.

Posto isto, fica evidente que sempre que existe maneira de categorizar informação (com *features*), fica relativamente fácil de comparar algo. Por conseguinte, definindo um objetivo, consegue-se fazer uma recomendação, visto que, inevitavelmente numa comparação, há sempre uma entidade que se sobrepõe a outra.

É com isso em vista que decidimos desenvolver um sistema de recomendação autoexplicativo de futebolistas.

Autoexplicativo porquê? De acordo com o artigo “*Towards Controllable Explanation Generation for Recommender Systems via Neural Template*” [7], a explicabilidade num sistema de recomendação leva a que este se torne mais transparente e confiável para os utilizadores, aumentando o seu grau de satisfação.

Encontramos algumas implementações de sistemas de recomendação para este tipo de problema. No entanto, a maioria dos programas já existentes ou apenas recomendava os futebolistas, não explicando a razão para tal, ou utilizam gráficos para a questão da explicabilidade. A nossa ideia assenta em utilizar como referência as implementações existentes para a questão da recomendação. Quanto à explicabilidade, aspiramos melhorar o que encontramos adicionando explicação textual para que as razões que levam o sistema a sugerir determinado jogador (tanto as vantagens e desvantagens de uma eventual contratação) sejam mais compreensíveis.

3 Materiais (dados incluídos)

No que diz respeito à questão do sistema de recomendação, utilizamos como referência algumas implementações presentes na Internet [1, 4, 5].

Para a implementação, decidimos utilizar Python e algumas das suas bibliotecas.

Quanto ao *dataset*, utilizamos os dados/estatísticas dos jogadores presentes no FIFA22 [2]. Este é um jogo de simulação de futebol baseado na realidade, pelo que os *ratings* dos jogadores correspondem às suas qualidades e capacidades na vida real.

Para conseguir implementar o nosso sistema de recomendação, necessitámos de tratar o *dataset* de modo utilizá-lo como pretendíamos.

Numa primeira instância, decidimos analisar o *dataset* sem qualquer tratamento. Com isso, percebemos que a maioria dos atributos são do tipo *object* ou *int64*. Existem ainda dados *float64* e *datetime64*. Isto era expectável já que a maioria dos nossos dados se resumem a avaliações (números) nos diversos atributos dos jogadores.

Antes de passar para o tratamento dos dados, consideramos também importante verificar a existência de valores nulos, uma vez que a ausência de valores pode impossibilitar o bom funcionamento do sistema, impossibilitando a correta comparação e posterior explicabilidade. Tal como esperávamos, alguns atributos continham vários *None*: os que dizem respeito ao clube do jogador, liga em que atua (no caso do jogador não ter contrato no momento), se está emprestado, número de camisola na seleção nacional, entre outros. Estes casos foram todos tratados devidamente, tal como iremos abordar mais adiante neste artigo.

Todas as colunas que não nos interessavam (quer contivessem valores *None* ou não), como, por exemplo, as que diziam respeito ao ID do jogador no FIFA, o URL correspondente, o número da camisola que utiliza na sua seleção, ..., foram eliminadas.

No entanto, continuávamos com colunas com valores nulos. Deste modo, trocámos todos os valores *None* por algo que nos fosse útil. Por exemplo, se um jogador não tiver clube, não tem valor na coluna correspondente. Para contornar isto, optamos por, nestes casos, colocar o valor “*Free Agent*”. Se um jogador não tem clube, também não tem liga. Deste modo, os valores na coluna correspondente são *None*, pelo que os trocamos por “*Free Agent*”. Por último, caso os jogadores não estejam emprestados, o valor que diz respeito a esse aspeto é também nulo, pelo que alteramos por “*No loan*”.

Após este tratamento, decidimos efetuar uma breve análise estatística dos dados. Analisamos a distribuição dos valores em cada atributo e percebemos, por exemplo, que o valor médio de qualidade dos jogadores presentes no *dataset* era 65.772182, sendo o máximo 93. No entanto, o maior potencial tem a nota de 95.

Consideramos também importante analisar a distribuição das posições e qualidade dos jogadores por liga, bem como a distribuição de idade dos mesmos. Nas análises por liga, como o nosso *dataset* tem imensos dados, optamos por apenas analisar as seis principais ligas (*English Premier League, Spain Primera Division, German 1. Bundesliga, Italian Serie A, French Ligue 1 e Portuguese Liga ZON SAGRES*). Os resultados foram os seguintes (para melhor visualização, enviamos as imagens em anexo):

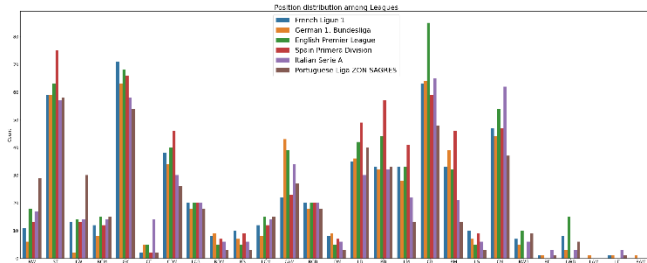


Figura 1 - Distribuição das posições dos jogadores por liga

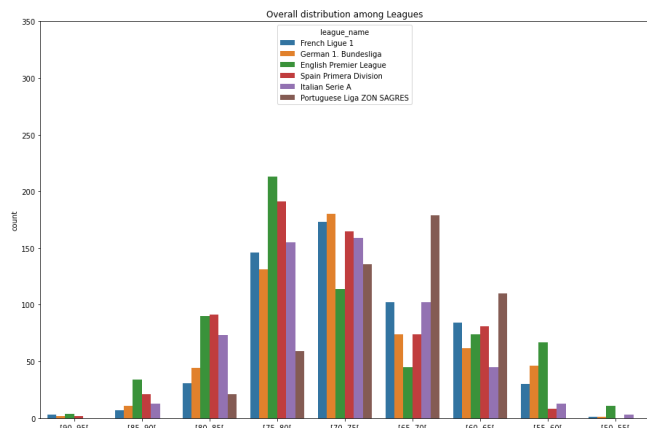


Figura 2 - Distribuição da qualidade dos jogadores por liga

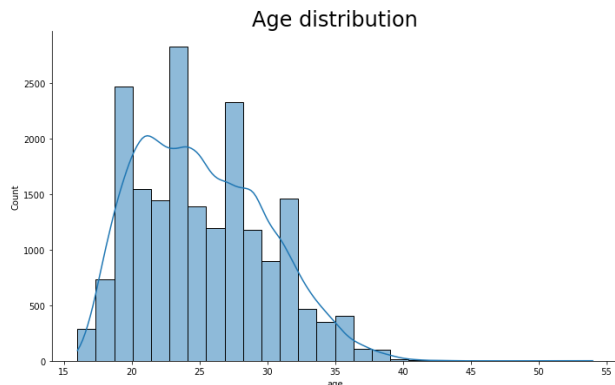


Figura 3 - Distribuição da idade dos jogadores

Pelos gráficos acima, percebemos que dentro das seis principais ligas de futebol, a qualidade dos jogadores varia, bem como as suas posições. Quanto à qualidade, percebemos que nas ligas portuguesa e italiana, não existem jogadores com qualidade superior a 89. Inclusive, na liga portuguesa, não

existem jogadores com qualidade superior a 84. Isto é relevante, uma vez que, se pretendermos uma recomendação de um jogador semelhante ao Cristiano Ronaldo, muito dificilmente serão recomendados jogadores destas duas ligas, uma vez que não têm jogadores com qualidade semelhante.

Quanto às posições, pela figura 1 percebemos que, por exemplo, a probabilidade de um defesa central recomendado jogar na liga inglesa é maior, visto que há mais quantidade de jogadores dessa posição nessa liga. Pelo contrário, se pretendermos um avançado centro, há mais probabilidade de este jogar na liga espanhola.

Quanto às idades dos jogadores, fica evidente, tal como era expectável, que a maioria se encontra na faixa entre os 20 e os 31. Isto é positivo, pois, em princípio, é a idade em que os jogadores evoluem ou estão na sua melhor forma, sendo mais atrativos para contratar. Isto era também detetável na análise da distribuição dos dados, uma vez que a idade média era 25.210822 anos, com um máximo de 54.

Antes de avançarmos para o desenvolvimento do sistema de recomendação, tivemos ainda de fazer mais um breve tratamento de dados. Existiam jogadores no nosso *dataset* que tinham posições no clube ou na seleção diferentes daquelas que eram anunciadas nas posições do jogador. Deste modo, caso essa posição não fosse suplente ou reserva, optamos por juntá-la às posições do jogador em si, uma vez que, se joga assim no clube ou na seleção, tem de ser contabilizada como posição do jogador.

Muitos dados numéricos que demonstram a qualidade do jogador em determinados atributos vinham com potencial, isto é, em vez de ter um valor fixo, vinham com uma expressão (84+3, por exemplo). Isto significa que atualmente o jogador tem qualidade de 84, mas pode, se evoluir, chegar a 87. Decidimos que no nosso sistema iríamos considerar a avaliação com a conta feita, ou seja, neste caso, consideraríamos uma nota de 87.

Para finalizar o tratamento dos dados, resolvemos a questão dos nomes repetidos. Sendo que é através dos nomes que indexamos os jogadores no *Dataframe*, convém que cada identificador seja único. No entanto, percebemos que existiam jogadores com o mesmo nome. Para resolver este problema, decidimos que a melhor opção seria o identificador ter o nome do jogador e a equipa em que joga, “Rui Patrício (Roma)”, por exemplo. Contudo, mesmo assim, descobrimos que existiam jogadores com o mesmo nome a jogar na mesma equipa. Com isto, o problema subsistia. Lembramo-nos que os números dos jogadores no seu clube são únicos. Com isto, definimos os identificadores com o nome e o clube e número entre parênteses, ou seja, “Rui Patrício (Roma, 1)”. Esta foi a forma por nós encontrada de indexar os jogadores no *Dataframe* com algo único e de fácil procura.

Tendo tudo isto tratado, fizemos novamente uma pequena análise do *dataset*, já sem valores nulos. Verificamos que a quantidade de valores *None* em cada *feature* era agora nula e

que, nos atributos onde as contas das avaliações foram realizadas (nota atual com o potencial), os valores da média, desvio padrão, máximo, mínimo e quartis alteraram como era expetável, já que os valores das colunas foram também eles modificados.

4 Abordagem

Após o tratamento de dados explanado na secção anterior, estávamos finalmente aptos a desenvolver o nosso sistema de recomendação. Decidimos calcular as semelhanças entre jogadores através da distância de cosseno, uma vez que é uma ótima forma de calcular a semelhança entre dois vetores com dimensionalidade elevada (ao contrário da distância euclidiana, por exemplo). Deste modo, necessitamos de selecionar apenas os valores numéricos (as avaliações que têm nos diversos atributos), uma vez que nenhum outro tipo de dados pode ser comparado com esta técnica. Optámos por utilizar os atributos cujos valores são textuais na parte da explicabilidade, ajudando a justificar se determinado jogador é uma boa ou má contratação. Mais à frente iremos explicar de forma mais detalhada esse processo.

Com os dados numéricos selecionados, decidimos recorrer a PCA para melhorar o desempenho do nosso sistema de recomendação, diminuindo o tempo de execução (que é elevado, já que cada jogador tem de ser comparado com todos os outros e o nosso *dataset* tem quase 20 mil). Este é um algoritmo que reduz o número de atributos a considerar, descartando aqueles que estão fortemente correlacionados. Analisando o gráfico de cotovelo seguinte, percebemos que 50 *features* são suficientes para distinguir os jogadores entre si, sem perder informação.

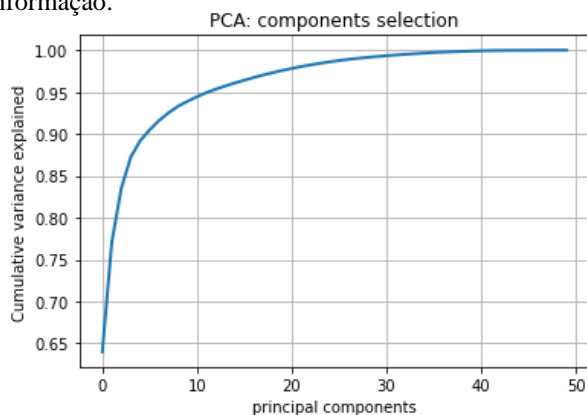


Figura 4 - Gráfico de cotovelo para seleção de features

Com isto, reduzimos a dimensionalidade do nosso problema, eliminando alguma redundância, ficando com os dados prontos para comparação.

Neste sentido, comparámos todos os jogadores do *dataset* com todos os outros, de modo que existissem métricas de semelhança entre todos os jogadores. Isto permite que qualquer jogador possa ser recomendado quando se procura por outro,

apenas através das suas qualidades. Embora seja improvável, um guarda-redes pode ser recomendado quando se procura por um avançado. Esta é também uma maneira de, através das qualidades nos diversos atributos, perceber se um jogador pode ser adaptado a outra posição.

Com todas as métricas de semelhança calculadas, basta pedir ao utilizador que introduza o jogador que quer que seja o modelo a seguir na recomendação. Como indexamos por nome, para termos a certeza que o utilizador procura exatamente da mesma forma que temos guardada na nossa base de dados, apresentamos todas as opções disponíveis. Primeiro apresentamos a liga e, após este parâmetro ser escolhido, apresentamos as equipas dessa liga. Escolhendo a equipa, apresentamos os nomes dos jogadores dessa equipa e respetivo número da camisola. Introduzindo o nome, o nosso sistema apresenta os jogadores que mais se assemelham ao escolhido, ordenados segundo o grau de semelhança. Se o jogador for “*Free agent*”, a parte da escolha da equipa é ignorada pelo sistema, considerando “*Free agent*” como uma liga disponível.

Para filtrar melhor as características consoante o que os clubes necessitam, quisemos que fosse possível selecionar características básicas dos jogadores, tais como o pé preferencial, a idade, a intensidade no ataque e na defesa e a liga onde o jogador atua. Isto são parâmetros editáveis no código antes de chamar a função de recomendação. No caso do pé preferencial e liga em que atuam, o sistema retorna exatamente aquilo que se definiu nos parâmetros. Já na idade, é possível definir um intervalo em que se quer que um jogador esteja inserido. Quanto à intensidade no ataque e na defesa, o sistema recomenda sempre jogadores de intensidades iguais ou superiores às especificadas. O valor por defeito de todos estes parâmetros é “*All*”. Tendo isto, o sistema apresenta os 10 jogadores mais semelhantes ao modelo, com os filtros pedidos. Com a parte da recomendação implementada, fica apenas a faltar a explicabilidade. Há duas formas de efetuar a explicação: através de *templates* pré-definidos ou gerando o próprio texto. Enquanto os *templates* são mais limitativos visto terem uma estrutura não moldável, a geração de texto é mais livre e, assim, pode explicar algo melhor. No entanto, pode também produzir conteúdo menos útil já que não é tão controlável [8]. Posto isto, decidimos utilizar os *templates* pré-definidos em que variáveis servem para tornar o texto coerente e explicativo, já que são mais fáceis de controlar e de apresentarem a informação pretendida.

Deste modo, para cada jogador recomendado, começamos por compará-lo com o que serviu de modelo para a comparação, definindo limiares. Para cada atributo, se este distar menos do que 5 unidades da nota do jogador comparado, assumimos que os jogadores são semelhantes nessas características. No entanto, se estes distarem mais do que 10 unidades entre si num atributo, consideramos que os jogadores diferem nesse aspeto. Com isto, apresentamos ao utilizador os atributos em que os jogadores mais coincidem e mais diferem.

Na parte da explicabilidade, tal como já falamos anteriormente, pretendíamos também expor ao utilizador os motivos que fariam com que fosse benéfico ao clube contratar determinado jogador, bem como aqueles que seriam prejudiciais. Para tal, definimos entre nós o que seria positivo ou negativo, tendo sempre em mente os clubes como entidade principal a ser beneficiada. Com isto, o nosso sistema apresenta o que é benéfico como pontos positivos e o que é prejudicial como pontos negativos. É aqui que os atributos que não foram utilizados na comparação através da distância de cosseno entram, não só textuais como numéricos. Por exemplo, as *features* que dizem respeito ao salário, idade, cláusula de rescisão, data do fim do contrato, valor e potencial são numéricas, mas não fazia sentido compará-las com a distância de cosseno. Entendemos isto uma vez que a nossa intenção passa por recomendar jogadores em que as características de jogo se assemelham e não este tipo de variáveis que muitas vezes dependem do clube que representam e da sua nacionalidade. Para cada atributo, definimos métricas que indicam se é um aspeto positivo ou negativo. Se for positivo ao acontecer, se não acontecer, consideramos como aspeto negativo e, de igual modo, o seu contrário. Conseguimos perceber melhor este conceito quando um jogador é atreito a lesão. Se o for, consideramos um aspeto negativo. Se não o for, consideramos como algo benéfico.

Em suma, definimos os seguintes critérios para o atributo no jogador ser positivo ou negativo:

Feature	Positivo	Negativo
Posição	Se for reserva ou suplente, provavelmente está descontente e será mais fácil contratá-lo.	Se for reserva ou suplente, provavelmente não está em boa forma.
		Se não for reserva ou suplente, é sinal que está a jogar e provavelmente está feliz, pelo que será mais difícil de o contratar.
Idade	Se tiver menos que 24 anos, irá evoluir e pode render a nível desportivo e financeiro.	Se tiver menos que 24 anos, terá de ganhar alguma experiência.
	Se tiver entre 24 e 32 anos, está na sua melhor forma.	Se tiver mais que 32 anos, provavelmente irá diminuir o seu rendimento desportivo em breve e não irá render financeiramente.

Salário	Se receber menos que o jogador que estamos a comparar.	Se receber mais que o jogador que estamos a comparar.
Cláusula de rescisão	Se a diferença entre a cláusula de rescisão e o valor do jogador for menor que metade da cláusula de rescisão, é um valor aceitável.	Se a diferença entre a cláusula de rescisão e o valor do jogador for maior ou igual a metade da cláusula de rescisão, é um valor exagerado.
Término do contrato	Se o contrato terminar nos próximos 2 anos, poderá ser mais fácil e mais barato contratar um jogador, visto que o seu clube tem menos poder negocial.	Se o contrato não terminar nos próximos 2 anos, o clube que detém o seu passe terá mais margem negocial, pelo que será mais complicado contratá-lo.
	Se o contrato terminar este ano, poderemos contratar o jogador a custo 0.	
Nível de liga	Se o jogador jogar numa liga de nível inferior, provavelmente ser-lhe-á aliciante jogar numa melhor.	Se o jogador jogar numa liga de nível superior, provavelmente não lhe será aliciante jogar numa pior.
Clube	Se o jogador não tiver clube, podemos contratá-lo a custo zero.	-
Potencial	Se o seu potencial for superior em 2 unidades à sua atual qualidade, irá evoluir e pode render a nível desportivo e financeiro.	Se o seu potencial não for superior em 2 unidades à sua atual qualidade, não irá evoluir muito, pelo que não valerá muito mais financeiramente.
Posições	Se jogar nas mesmas posições que o jogador modelo, a adaptação será mais fácil.	Se não jogar nas mesmas posições que o jogador modelo, a adaptação será mais difícil e poderá necessitar de mais tempo.
Tags	Todas as <i>tags</i> em que o jogador recomendado coincide	Todas as <i>tags</i> que o jogador modelo tem e o recomendado não.

	com o modelo ou que tem a mais.	
Características	Todas em que o jogador recomendado coincide com o modelo ou que tem a mais, à exceção de ser atreito a lesões e jogador de um clube apenas.	Todas em que o jogador modelo tem e o recomendado não, à exceção de ser atreito a lesões e jogador de um clube apenas.
		Se o jogador recomendado for atreito a lesões.

Características ambíguas: Ser um jogador de um clube apenas, tanto pode ser benéfico como prejudicial ao clube que o quer contratar. Se algum clube o conseguir, este irá estar muitos anos na nova equipa, esforçando-se sempre. Por outro lado, será muito complicado convencê-lo a trocar de clube, uma vez que tem a mentalidade de ter um só.

5 Experimentação

Para experimentar o nosso sistema, tal como referimos anteriormente, basta indicar a liga, clube e nome do jogador em questão (ou apenas que não tem liga e o nome no caso de ser “Free agent”). Com isto, para os primeiros 5772 jogadores do *dataset* (30% dos dados), ao selecionar a Liga Francesa, o *Paris Saint-Germain* e L. Messi, o resultado é o seguinte:

	short name	Similarity	Position	overall	potential	age	league name	preferred foot	weak foot	work rate	body type	player tags
1	E. Hazard (Real Madrid CF, 7)	0.9884	DM LF	85	85	30	Spain Primera Division	Right	4	Medium/Medium	Unique	#Dribbler
2	P. Uyleke (Juventus, 10)	0.987	CF CAM	87	88	27	Italian Serie A	Left	3	Medium/Medium	Unique	#Dribbler, #Playmaker, #FK Specialist, #Acrobat...
3	Ronaldo Nazari (Girona, 7)	0.9855	CAM RM RW	83	83	29	Campeonato Brasileiro Serie A	Right	4	Medium/Low	Normal (75-185)	#Acrobat
4	Neymar Jr (Paris Saint-Germain, 10)	0.9835	DM CAM	91	91	23	French Ligue 1	Right	5	High/Medium	Unique	#Playmaker, #Dribbler, #Playmaker, #FK Specialist...
5	S. Grubny (FC Bayern Munchen, 7)	0.9828	RM LM RW	85	86	25	German 1. Bundesliga	Right	4	High/Medium	Unique	#Dribbler, #Acrobat, #Clinical Finisher, #Comple...
6	Marcos Asensio (Real Madrid CF, 11)	0.9826	RW LW	83	86	25	Spain Primera Division	Left	3	High/Medium	Normal (70-185)	0
7	N. Iker (Real Betis, Baleares, 8)	0.9823	CAM RM S	84	84	27	Spain Primera Division	Left	4	High/Medium	Slender (70-180)	#Dribbler
8	Gonzalo Guedes (Valencia CF, 7)	0.980	LM ST CF LS	81	86	24	Spain Primera Division	Right	4	High/Low	Normal (70-185)	0
9	M. Salah (Liverpool, 11)	0.9817	RW	89	89	23	English Premier League	Left	3	High/Medium	Unique	#Playmaker, #Dribbler, #Acrobat, #Clinical Fin...
10	M. Reus (Borussia Dortmund, 11)	0.9817	CAM CF	85	85	32	German 1. Bundesliga	Right	4	High/Medium	Unique	0

Figura 5 - Recomendações tendo como modelo L. Messi

De seguida, apresentamos a explicação do sistema para recomendar o E. Hazard, bem como os pontos positivos e negativos de uma eventual contratação deste mesmo jogador.

Como pontos positivos e negativos à sua contratação, o nosso sistema apresenta:

He is being recommended because L. Messi has similar grades at: weight_kg, weak_foot, skill_moves, pace, defending, physic, movement_sprint_speed, movement_agility, power_shot_power, mentality_interceptions, defending_sliding_tackle, goalkeeping_handling, goalkeeping_reflexes, goalkeeping_speed, lwb, ldm, cdm, rdm, rwb, lb, lcb, cb, rcb, rb, gk

However, E. Hazard has great differences at: shooting, attacking_finishing, skill_fk_accuracy, movement_reactions, power_jumping, power_long_shots, mentality_vision, mentality_penalties, defending_marking_awareness

❖ Pontos positivos:

- E. Hazard tem 30 anos, por isso está na sua melhor forma;

- E. Hazard ganha menos salário (240000.0) do que L. Messi (320000.0). Talvez possamos comprá-lo e poupar algum dinheiro em salários;
- O contrato de Hazard expirará em 2024. Será mais barato comprá-lo porque o seu clube não o quer perder a custo zero;
- E. Hazard é #Dribbler, assim como L. Messi;
- E. Hazard tem as seguintes características, bem como L. Messi: *Finesse Shot*, *Playmaker (AI)*, *Outside Foot Shot*, *Technical Dribbler (AI)*. Também tem *Flair*, que L. Messi não tem;

❖ Pontos negativos:

- Joga atualmente no seu clube. Talvez ele esteja lá feliz, por isso seria difícil comprá-lo;
- E. Hazard tem uma cláusula de rescisão de 106600000.0 euros. É demasiado elevada em comparação com o valor real do jogador (52000000.0);
- E. Hazard não evoluirá muito. Ele não valerá muito mais;
- E. Hazard não joga em posições semelhantes às de L. Messi. Talvez seja necessário algum tempo para o adaptar;
- E. Hazard não é #Distance Shooter, #FK Specialist, #Acrobat, #Clinical Finisher, #Complete Forward como L. Messi;
- E. Hazard é propenso a lesões, portanto talvez não seja a melhor opção para contratar;
- E. Hazard não tem as seguintes características que L. Messi tem: *FlairLong Shot Taker (AI)*, *Chip Shot (AI)*.

O sistema apresenta este tipo de explicação para todos os jogadores recomendados.

6 Validação

Falta abordar como avaliamos todas as componentes do sistema.

No que diz respeito à explicabilidade para o jogador ser recomendado, optamos por apresentar as características em que estes são mais parecidos e mais diferentes, através de limiares que já explicamos anteriormente. Para cada atributo, se este distar menos do que 5 unidades da nota do jogador comparado, assumimos que os jogadores são semelhantes nessas características. No entanto, se estes distarem mais do que 10 unidades entre si num atributo, consideramos que os jogadores diferem nesse aspeto. Para avaliar se as explicações estão corretas, basta analisar o *dataset* e ver a diferença das notas nesses atributos.

Quanto ao texto gerado que expõe pontos negativos e positivos de uma eventual aquisição do jogador, avaliamos do ponto de vista gramatical e do conteúdo. No caso do conteúdo, avaliamos tendo em conta o *dataset*. Se a explicação disser que um jogador tem o *handicap* à sua contratação de se

lesionar frequentemente, significa que essa informação está no *dataset*. Deste modo, podemos avaliar essa informação consultando os dados. Do ponto de vista gramatical, teoricamente, como as explicações advêm de *templates* pré-definidos, pensamos que os resultados da avaliação serão sempre positivos, uma vez que, o programa apenas terá de introduzir no *template* as características para cada caso, o que é bastante controlado.

No entanto, tendo algum conhecimento em futebol, sobretudo conhecimento em jogadores, podemos avaliar todas as componentes do nosso sistema subjetivamente, uma vez que dá para analisar se um jogador é parecido a outro e se as explicações fazem sentido.

Deste modo, decidimos avaliar o desempenho do nosso sistema pedindo a opinião a várias pessoas. Seleccionamos algumas que sabíamos à partida que tinham conhecimento em futebol. Assim, cada uma delas interagiu com o nosso sistema, analisando os resultados. Após isto, preencheu um formulário com várias questões sobre as recomendações, as explicações e a análise dos pontos positivos e negativos de uma eventual contratação. Para garantirmos uma avaliação correta, apenas consideramos as respostas de utilizadores que consideraram o seu conhecimento em futebol e em jogadores acima de 4 (de 1 até 5). Tivemos 3 respostas no total. Das diversas perguntas presentes no formulário, seleccionamos as mais relevantes para apresentar neste artigo:

- Tem algum conhecimento sobre futebol? (de 1 a 5)

Média	Desvio padrão	Máximo	Mínimo
4,6667	0,5774	5	4

- Tem algum conhecimento sobre jogadores de futebol? (de 1 a 5)

Média	Desvio padrão	Máximo	Mínimo
4,6667	0,5774	5	4

- Considera que os jogadores recomendados são semelhantes ao jogador escolhido? (Sim/Não)

Sim	3
Não	0

- Considera correta a ordenação dos jogadores recomendados? (Sim/Não)

Sim	3
Não	0

- Concorda com os atributos semelhantes e diferentes entre o jogador recomendado e o modelo? (de 1 a 5)

Média	Desvio padrão	Máximo	Mínimo
4,3333	0,5774	5	4

- Concorda com os pontos positivos e negativos de uma eventual contratação de cada jogador? (de 1 a 5)

Média	Desvio padrão	Máximo	Mínimo
4,6667	0,5774	5	4

- Quão correto está o texto gramaticalmente? (de 1 a 5)

Média	Desvio padrão	Máximo	Mínimo
5	0	5	5

- O quão útil considera que seja este sistema? (de 1 a 10)

Média	Desvio padrão	Máximo	Mínimo
9,6667	0,5774	10	9

Uma vez que não obtivemos notas abaixo de 4 na avaliação do nosso sistema e que, inclusive, a recomendação foi avaliada como correta pela unanimidade, consideramos que o nosso sistema recomenda corretamente e explica igualmente bem as suas escolhas, de forma gramaticalmente aceite.

7 Conclusões

Ao início, propusemo-nos a construir um sistema de recomendação autoexplicativo que ajudasse os clubes a encontrar alternativas aos seus jogadores de forma mais automática e não tão dispendiosa a níveis financeiro e de recursos humanos que a habitual.

Como o *dataset* utilizado possui quase 20 mil jogadores, o tempo de execução na parte da comparação era, inicialmente, incomportável. Com isto, tivemos de encontrar uma solução para este problema (PCA). Esta foi a maior adversidade que encontramos ao longo da implementação do nosso programa, uma vez que, demorando imenso tempo a correr, o sistema tornar-se-ia inútil.


No entanto, como trabalho futuro, para reduzir ainda mais o tempo de execução do nosso sistema, temos pensado dividir inicialmente o nosso *dataset* em clusters para depois comparar os jogadores apenas com aqueles que estão no mesmo cluster que estes.

Na questão da validação, consideramos que apenas 3 respostas são escassas e gostávamos de avaliar melhor o nosso sistema, tendo mais respostas no futuro. No entanto, as métricas que obtivemos foram todas positivas, o que nos permite concluir que o sistema realmente funciona corretamente.

Gostávamos também que o nosso sistema oferecesse a possibilidade de recomendar jogadores que correspondem a mais características escolhidas através de filtros de atributos (definir, por exemplo, uma nota mínima de velocidade).

Em suma, pelos resultados que obtivemos, pensamos ter desenvolvido de forma correta o nosso sistema. Ele executa tudo aquilo que tínhamos em mente: recomenda, explica a recomendação e indica pontos positivos e negativos de uma eventual contratação. Isto torna o sistema mais transparente e confiável para os utilizadores, ajudando assim os clubes de futebol.

Referências

- [1] [Saini A. S., 2021, 30 de junho] *Building a Player Recommender Tool*, Medium, <https://medium.com/analytics-vidhya/building-a-player-recommender-tool-666b5892336f> acessado a 11 de dezembro de 2022
- [2] FIFA 22 complete player dataset. (s.d.), Kaggle: *Your Machine Learning and Data Science Community*. <https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset/code> acessado a 15 de outubro de 2022
- [3] Author's Kit | IJCAI. (s.d.). Welcome to IJCAI | IJCAI. https://www.ijcai.org/authors_kit acessado a 15 de outubro de 2022
- [4] [mbnb8317, 2020, 5 de junho],  *FIFA - Recommender System - if you are manager*, Kaggle: Your Machine Learning and Data Science Community, <https://www.kaggle.com/code/mbnb8317/fifa-recommender-system-if-you-are-manager/notebook> acessado a 15 de outubro de 2022
- [5] [kushojha12, (2022, 22 de maio)], *FIFA22 EDA and Feature Analysis using XGBoost*, Kaggle: Your Machine Learning and Data Science Community, <https://www.kaggle.com/code/kushojha12/fifa22-eda-and-feature-analysis-using-xgboost> acessado a 15 de outubro de 2022
- [6] [Das S., (2022, 11 de Agosto)], *Beginners Guide to learn about Content Based Recommender Engine*, Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/> acessado a 15 de outubro de 2022
- [7] [Li L., Chen L., & Zhang Y. (2020, abril)], *Towards Controllable Explanation Generation for Recommender Systems* via Neural Template | Companion Proceedings of the Web Conference 2020. ACM Conferences. https://dl.acm.org/doi/abs/10.1145/3366424.3383540?casa_token=L3Ida-tQ7xMA-AAAA:Hn3PaqwKo1mKLrQIO-likzV2kVYt4MrszDyHoFijVbBrv0fy1WWU2MVBVX X86HTUT_w39qEzBOk0 acessado a 15 de outubro de 2022
- [8] [Afchar D., Melchiorre A., Schedl M., Hennequin R., Epure E., & Moussallam M. (2008, novembro)], *Content-Based Recommendation Systems*, ResearchGate, https://www.researchgate.net/publication/236895069_Content-Based_Recommendation_Systems acessado a 15 de outubro de 2022
- [9] [Zisopoulos C., Karagiannidis S., Demirtoglou G., & Antaris S. (2008, novembro)], *Content-Based Recommendation Systems*, ResearchGate, https://www.researchgate.net/publication/236895069_Content-Based_Recommendation_Systems acessado a 15 de outubro de 2022
- [10] [Maanijou R. & Mirroshandel S. (2019, 25 de janeiro)], *Introducing an expert system for prediction of soccer player ranking using ensemble learning* - Neural Computing and Applications, SpringerLink. <https://link.springer.com/article/10.1007/s00521-019-04036-9> acessado a 15 de outubro de 2022
- [11] [Suriyadeepan, R. (2020, 30 de dezembro)], Exploratory Data Analysis of Titanic Survival Problem, Medium. <https://medium.com/analytics-vidhya/exploratory-data-analysis-of-titanic-survival-problem-e3af0fb1f276> acessado a 9 de dezembro de 2022

Link do repositório

- <https://github.com/patii01/IACH-project>