Abstract

Sistema de Recomendação Autoexplicativo

de Jogadores de Futebol

**Duarte Meneses Patrícia Costa**

duartemeneses@student.dei.uc.pt patriciacosta@student.dei.uc.pt

Departamento de Engenharia Informática

Inteligência Artificial Centrada no Humano - MECD

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 os prós e os contras 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 pretende 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. Em vez de ter de procurar por uma alternativa vendo milhares de jogos de outros jogadores, utilizando para isso inúmeros olheiros espalhados pelo mundo (levando a um esforço gigantesco em termos de recursos humanos e um enorme gasto financeiro) como é comum, 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), tendo ainda a explicação para serem estes os recomendados e, em contrapartida, algumas razões que poderiam levar a que as contratações dos mesmos não fossem bem-sucedidas. Isto facilitaria em muito o trabalho de *scouting* e de analistas de dados dos clubes, reduzindo também o tempo despendido, o gasto financeiro e os recursos humanos necessários.

O grande objetivo deste projeto passa por conseguir oferecer aos clubes uma forma de conseguirem encontrar jogadores com as características que pretendem de forma rápida, eficaz e com menos impacte financeiro. Para tal, o nosso sistema de recomendação deve ser capaz de sugerir jogadores que tenham características semelhantes às pretendidas. Aliado a isto, deve ainda explicar a razão para recomendar determinado jogador, apresentando ainda alguns pontos menos positivos da contratação desse mesmo futebolista. Por exemplo, embora tenha características semelhantes, um jogador que tenha muitas lesões 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 alguns contras para tal contratação.

Nas próximas secções iremos abordar o trabalho relacionado com o nosso projeto, iremos apresentar os materiais por nós utilizados e analisar o dataset. Após isto, iremos explicar como decidimos abordar o problema, bem como explicar a fase de experimentação, apresentando os resultados e as conclusões de seguida.

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*” [10], o que um sistema de recomendação faz não é mais do que filtrar informação de acordo com as preferências do utilizador. Enquanto que, nos exemplos explanados acima, o sistema vai construindo perfis de utilizadores de acordo com a sua atividade, no nosso sistema, a ideia passa por ser o utilizador a introduzir as suas preferências aquando de cada recomendação. De qualquer das maneiras, o que o sistema faz é comparar (correlacionar) 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*” [8], a explicabilidade num sistema de recomendação leva a que este se torne mais transparente e confiável por parte dos utilizadores, o que, por conseguinte, aumenta o seu grau de satisfação.

Há duas grandes 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, apesar de ser mais livre e, assim, poder explicar algo melhor, pode também produzir conteúdo menos útil já que não é tão controlável [8].

Posto isto, decidimos utilizar os *templates* pré definidos no nosso projeto, já que são mais fáceis de controlar e de apresentarem a informação que se pretende.

Encontramos algumas implementações de sistemas de recomendação para este tipo de problema. No entanto, a maior parte dos programas já existentes ou apenas recomendam 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, pretendemos melhorar o que encontramos adicionando explicação textual para que as razões que levam o sistema a sugerir determinado jogador (tanto prós como contras) sejam mais compreensíveis. Pensamos que a explicação textual, trará benefícios a quem irá analisar as recomendações dos futebolistas.

4 Materiais (dados incluídos)

No que diz respeito à questão do sistema de recomendação, iremos utilizar como referência algumas implementações presentes na Internet [1, 4, 6].

Decidimos utilizar Python e algumas das suas bibliotecas para implementar o nosso sistema.

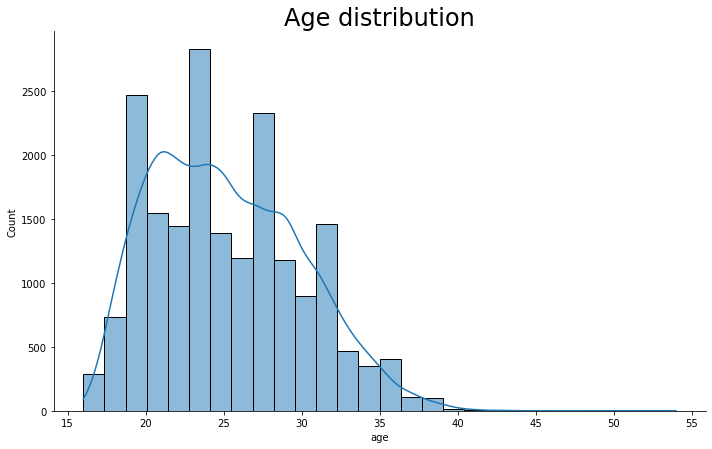
Quanto ao *dataset*, vamos utilizar os dados/estatísticas dos jogadores presentes no FIFA22 [2]. Este é um jogo de simulação de futebol baseado na realidade, pelo que os ratings dos jogadores estão de acordo com as suas qualidades e capacidades na vida real.

Para conseguir implementar o nosso sistema de recomendação, necessitamos de tratar o dataset de modo a que o conseguíssemos utilizar como pretendíamos. Como tal, começamos o nosso trabalho por analisar e tratar os dados.

Inicialmente, decidimos analisar o dataset sem qualquer tratamento. Com isso, percebemos que a maioria dos atributos eram do tipo object ou int64. Existem ainda dados float64 e datetime64. Isto era espectável já que a maioria dos nossos dados resumem-se a avaliações (números) nos diversos atributos.

Antes de passar para o tratamento, 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 já de seguida 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, trocamos 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 fazer 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.

Figura 3 - Distribuição da idade dos jogadores

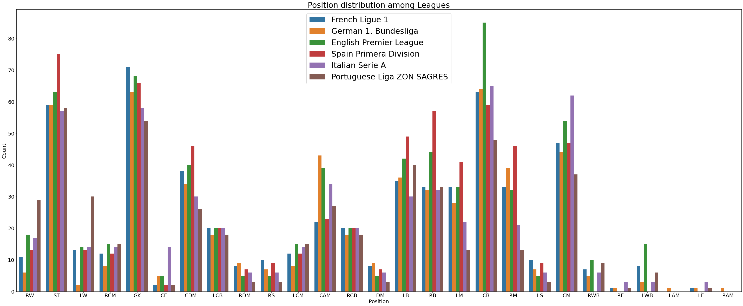
Consideramos importante analisar também a distribuição das posições e qualidade dos jogadores por liga, bem como a distribuição de idade dos mesmos. Nas análises por liga, como o nosso dataset tem imensos dados, optamos por apenas apresentar para as seis principais ligas (English Premier League, Spain 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

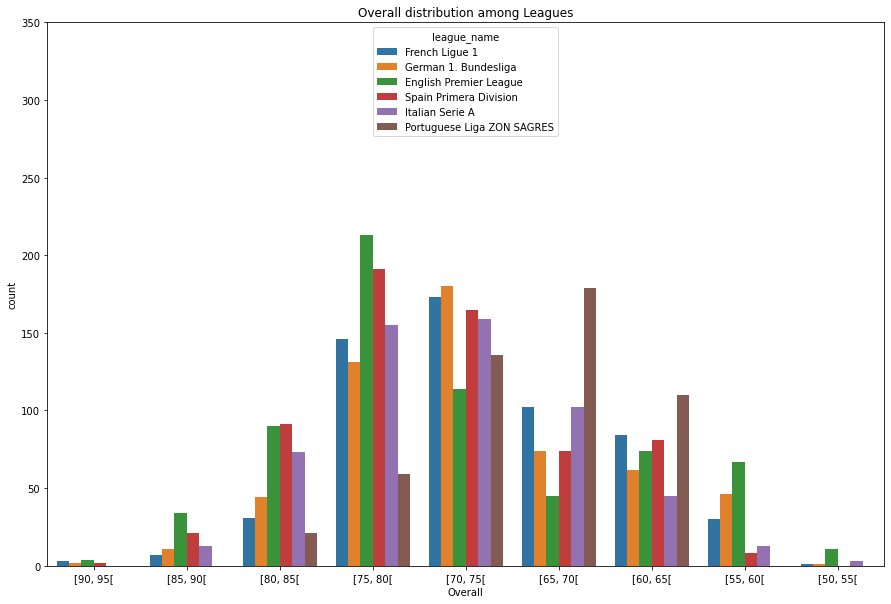


Figura 2 - Distribuição da qualidade dos jogadores por liga

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, por exemplo, ao Cristiano Ronaldo, muito dificilmente serão recomendados jogadores destas duas ligas acima mencionadas, 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 ser da liga espanhola.

Quanto às idades dos jogadores, fica evidente, tal como era espectável, 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 estão a evoluir ou na sua melhor forma, sendo mais atrativos para contratações. Já tínhamos visto isso 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 que 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 conta (84+3, por exemplo). Isto significa que atualmente o jogador tem qualidade em 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 passar a ter o nome do jogador e a equipa em que joga (“Rui Patricio (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 Patricio (Roma, 1)”. Esta foi a forma por nós encontrada de indexar os jogadores no Dataframe com algo que se percebesse e única.

Tendo tudo isto tratado, fizemos novamente uma pequena análise do dataset, agora já sem valores nulos. Verificamos que na análise da quantidade de valores None em cada feature era agora nula. Nas colunas em que efetuamos a conta da avaliação atual com o potencial, os valores da média, desvio padrão, máximo, mínimo e quartis alteraram como era espetável, já que os valores das colunas foram também eles alterados.

Fizemos também uma análise bivariada entre as features que consideramos mais relevantes (overall, velocidade, remate, passe, drible, potencial). Os resultados apresentam-se de seguida:

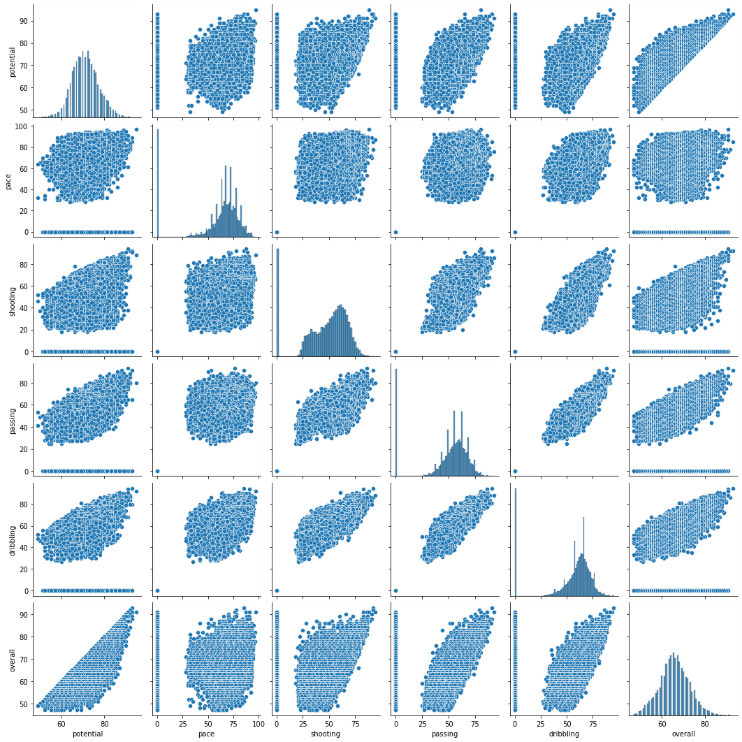


Figura 4 - Análise bivariada de alguns atributos

5 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 de que forma estes dados são utilizados.

Com os dados numéricos selecionados, decidimos utilizar PCA para melhorar o desempenho do nosso sistema de recomendação, diminuindo o tempo de execução (que é elevado, já que cada jogador tem que 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.

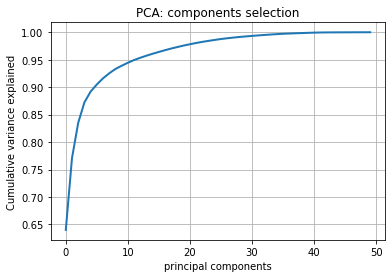


Figura 5 - Gráfico de cotovelo para seleção de features

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 a 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 diversas 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 no nosso dataset, apresentamos todas as opções disponíveis. Primeiro apresentamos a liga e, após este parâmetro ser escolhido, apresentamos as equipas dessa liga. Escolhendo a equipa, os nomes dos jogadores dessa equipa e respetivo número da camisola são apresentados ao utilizador. Introduzindo o nome, o nosso sistema apresenta os jogadores que mais se assemelham ao escolhido, ordenados segundo o grau de semelhança.

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 respetivamente e a liga em que o jogador atua. Isto são parâmetros editáveis no código antes de chamar a função de recomendação.

Com isto, a parte que diz respeito ao sistema de recomendação fica feita. Fica apenas a faltar a explicabilidade que pretendemos que o nosso sistema também tenha. A ideia passou por utilizar templates pré definidos em que variáveis servem para tornar o texto coerente e explicativo.

Nesse aspeto, 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 X unidades da nota do jogador comparado, assumimos que os jogadores são semelhantes nessas características. No entanto, se estes distarem mais do que Y 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, não só textuais como numéricos, entram. Apesar de, por exemplo, as *features* que dizem respeito ao salário, idade, cláusula de rescisão, data do fim do contrato, valor e potencial sejam numéricas, não fazia sentido compará-las com a distância de cosseno, uma vez que a nossa intenção passa por recomendar jogadores em que as características de jogo se assemelhem e não este tipo de variáveis que muitas vezes dependem do clube que representam e da sua nacionalidade.

Deste modo, utilizamos todos os atributos que não foram utilizados antes para atribuir valor negativo ou positivo a uma eventual contratação.

Para cada atributo, definimos métricas que indicam se é um aspeto positivo ou negativo. Se for positivo ao acontecer, se não acontecer, consideramos um aspeto negativo. De igual modo, se for positivo não acontecendo, se acontecer consideramos negativo. 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.

De resto, definimos os seguintes limiares 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 que 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 tags em que o jogador recomendado coincide com o modelo ou que tem a mais. | Todas as tags que o jogador modelo tem e o recomendado não. |
| 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 alguém o conseguir contratar, este irá estar muitos anos na nossa equipa, esforçando-se sempre. Por outro lado, será muito complicado convencê-lo a trocar de clube, uma vez que tem a característica de ter um clube só.

6 Experimentação

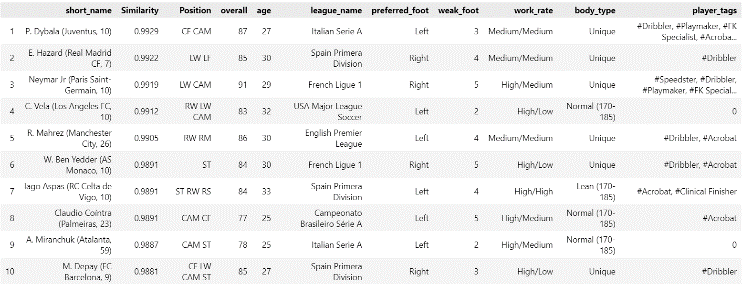
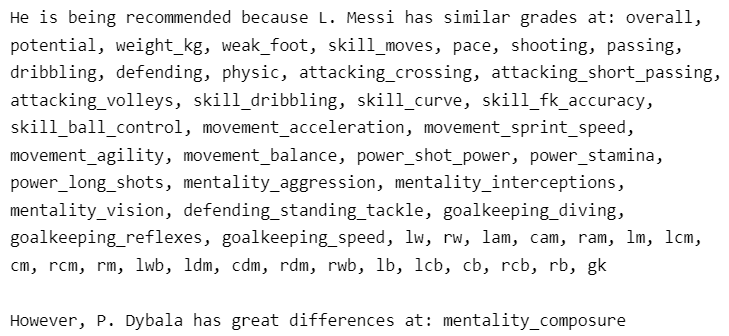
Para experimentar o nosso sistema, tal como referimos anteriormente, basta indicar a liga, clube e nome do jogador em questão (ou apenas que não tem liga e o nome no caso de ser “Free agent”). Com isto, para os primeiros 5772 jogadores do dataset (30% dos dados), ao selecionar a Liga Francesa, o Paris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

Figura 6 - Recomendações tendo como modelo L. Messi

De seguida, apresentamos a explicação do sistema para recomendar o P. Dybala, 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:

* **Pontos positivos:**
  + Tem 27 anos de idade, por isso está na sua melhor forma;
  + Ganha menos salário (160000.0) do que L. Messi (320000.0). Talvez possamos comprá-lo e poupar algum dinheiro em salários;
  + Tem uma cláusula de rescisão de 160400000.0 euros. É um valor aceitável em relação ao valor real do jogador (93000000.0);
  + O contrato de P. Dybala expirará este ano. Podemos contratá-lo de graça;
  + P. Dybala é #Dribbler, #FK Especialista, #Acrobat, bem como L. Messi. Ele é também #Playmaker, #Médio Completo que L. Messi não é;
  + P. Dybala tem as seguintes características, bem como L. Messi: Finesse Shot, Long Shot Taker (AI), Playmaker (AI), Outside Foot Shot, Technical Dribbler (AI). Também tem Flair, que L. Messi não tem.
* **Pontos negativos:**
  + Está atualmente a jogar no seu clube. Talvez ele esteja lá feliz, por isso seria difícil comprá-lo;
  + P. Dybala não vai evoluir muito. Ele não vai valer muito mais;
  + P. Dybala não joga em posições semelhantes às de L. Messi. Talvez seja necessário algum tempo para o adaptar;
  + P. Dybala não é #Atirador à Distância, #Afinador Clínico, #Atacante Completo como L. Messi;
  + P. Dybala não tem os seguintes traços que L. Messi tem: FlairChip Shot (AI).

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

7 Validação

Falta abordar a forma como avaliamos todas as componentes do sistema.

Na questão da recomendação, como esta é feita através de features exclusivamente numéricas, decidimos definir limiares. Por exemplo, no exemplo da secção anterior, procurando por um jogador semelhante ao Messi, o sistema recomenda o Dybala. Para avaliar a qualidade desta recomendação, comparamos os atributos de ambos. Se, no mesmo atributo, as notas distarem entre si menos que X unidades, consideramos que este atributo é semelhante. Por outro lado, se distarem mais do que Y unidades, consideramos que diferem nessa característica. Tendo o número total de atributos semelhantes e não idênticos, ao dividir ambos os valores pelo número total de features, obtemos uma métrica. Quanto maior a métrica proveniente dos atributos semelhantes, melhor a recomendação. Quanto menor a métrica relativa aos atributos não idênticos, pior a recomendação. Com isto, percebemos mais facilmente o grau de semelhança entre os jogadores, ajudando estas métricas também a explicar a razão da recomendação. Analisando a evolução das métricas ao descermos no ranking da recomendação, percebemos que, efetivamente, os jogadores vão deixando de ser tão parecidos:

|  |  |  |
| --- | --- | --- |
| Jogador | Métrica semelhança | Métrica diferença |
| P. Dybala | 71.6216 | 1.3513 |
| E. Hazard | 44.5946 | 2.7027 |
| Neymar Jr | 81.0810 | 4.0541 |
| C. Vela | 31.0811 | 2.7027 |
| R. Mahrez |  |  |
| Yedder |  |  |
| Iago Aspas |  |  |
| Coíntra |  |  |
| Miranchuk |  |  |
| Depay |  |  |

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 dos limiares que explicamos no parágrafo acima. Para avaliar se estas estão corretas, basta analisar o dataset e analisar a diferença das notas nesses atributos.

Quanto ao texto gerado pelo nosso sistema 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. 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.

Por outro lado, tendo algum conhecimento de futebol, podemos avaliar todas as componentes do nosso sistema de forma subjetiva, uma vez que dá para analisar se um jogador é parecido a outro e se as explicações fazem sentido, através de conhecimento prévio dos jogadores.

8 Conclusões

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 acedido a 15 de outubro 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 acedido a 15 de outubro de 2022

[3] Author's Kit | IJCAI. (s.d.). Welcome to IJCAI | IJCAI. https://www.ijcai.org/authors\_kit acedido 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 acedido a 15 de outubro de 2022

[5] [GitHub - Francesco-Sovrano, (s.d.)], *DoXpy:* *For Computing the Degree of Explainability*, GitHub, https://github.com/Francesco-Sovrano/DoXpy acedido a 15 de outubro de 2022

[6] [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 acedido a 15 de outubro de 2022

[7] [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/ acedido a 15 de outubro de 2022

[8] [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-tQ7xMAAAAA:Hn3PaqwKo1mKLrQIOlikzV2kVYt4MrszDyHoFijVbBrv0fy1WWU2MVBVXX86HTUT\_w39qEzBOk0 acedido a 15 de outubro de 2022

[9] [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 acedido a 15 de outubro de 2022

[10] [Zisopoulos C., Karagiannidis S., Demirtsoglou G., & Antaris S. (2008, novembro)], *Content-Based Recommendation Systems*, ResearchGate, https://www.researchgate.net/publication/236895069\_Content-Based\_Recommendation\_Systems acedido a 15 de outubro de 2022

[11] [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 acedido a 15 de outubro de 2022