



UNIVERSIDADE DE
COIMBRA

José Miguel Pereira do Carmo

**RELATÓRIO DE ESTÁGIO, DESENVOLVIDO NO SEVILLA
FÚTBOL CLUB, INSERIDO NO DEPARTAMENTO DE
ANÁLISE DESPORTIVA, DURANTE A ÉPOCA
DESPORTIVA 2022/2023**

**AS FUNÇÕES DO NOVO ANALISTA
QUOTIDIANO, EXIGÊNCIAS E OBJETIVOS DE UM CLUBE DE ELITE**

Relatório de Estágio inerente à Unidade Curricular “Estágio”, realizado sob a orientação do Professor Doutor Hugo Miguel Sarmento, apresentados à Faculdade de Ciências do Desporto e Educação Física da Universidade de Coimbra, para apreciação e prestação de provas públicas do Mestrado em Treino Desportivo.

Setembro de 2023



UNIVERSIDADE D
COIMBRA

**RELATÓRIO DE ESTÁGIO, DESENVOLVIDO NO SEVILLA
FÚTBOL CLUB, INSERIDO NO DEPARTAMENTO DE ANÁLISE
DESPORTIVA, DURANTE A ÉPOCA DESPORTIVA 2022/2023**

**AS FUNÇÕES DO NOVO ANALISTA
QUOTIDIANO, EXIGÊNCIAS E OBJETIVOS DE UM CLUBE DE ELITE**

Relatório de Estágio apresentado à Faculdade de Ciências do
Desporto e Educação Física da Universidade de Coimbra, com
vista à obtenção do grau de mestre em Treino Desportivo.

Orientador: Professor Doutor Hugo Miguel Sarmento

José Miguel Pereira do Carmo

Setembro de 2023

Carmo, JM. (2023). *Relatório de Estágio, desenvolvido no Sevilla Fútbol Club, inserido no departamento de Análise Desportiva, durante a época desportiva 2022/2023*. Relatório de Estágio, Faculdade de Ciências do Desporto e Educação Física da Universidade de Coimbra, Coimbra, Portugal.

Palavras-chave: Futebol, Análise de Dados, *Performance Analysis*, Observação e Análise, Estágio.

*“Un día, cuando nos paremos a pensar,
quizá nos demos cuenta de lo que se ha hecho.”*

Paco Cepeda (2020)

Agradecimentos

“Não sou nada.

Nunca serei nada.

Não posso querer ser nada.

À parte isso, tenho em mim todos os sonhos do mundo.”

Álvaro de Campos (1944)

Início este documento enaltecendo as pessoas que foram parte integrante desta curta e árdua caminhada, referenciando que são, uma das motivações para esta obra existir. Como na citação inicial de Álvaro de Campos (1994), são estas pessoas que continuam a alimentar os meus sonhos diariamente.

Professor Doutor Hugo Sarmento, obrigado pelo acompanhamento, profissionalismo e sentido crítico, que sem dúvida me tornou num melhor profissional. É com total certeza que afirmo, que é um motivo de orgulho para mim, ter o seu nome associado como orientador no presente documento.

Professor Doutor João Duarte e Professor Vítor Gouveia, agradecer simplesmente é pouco. Foi graças às vossas inquietações e experiências relativamente ao treino, que fui inspirado a explorar áreas e variáveis de rendimento, cada vez mais próximas das problemáticas diárias de um treinador.

Aos restantes professores do meu curto, mas insaciável percurso académico, o meu obrigado por todos os ensinamentos.

Para a minha família não tenho palavras. Foi um ano distante, fora de Portugal, sozinho noutro país, e que consigo perfeitamente identificar que foi graças aos esforços financeiros e apoio psicológico deles, que este estágio foi possível de levar a cabo.

À minha namorada Sara, agradecer pela pessoa incansável que é, bem como, por conseguir aguentar todas as dores de cabeça que lhe proporcionei durante a minha fase de adaptação. Como dissemos desde o primeiro dia, este seria apenas o início de um percurso risonho, no qual quero que estejas presente.

Aos meus colegas do *Sevilla Fútbol Club*, não tenho palavras para vocês. Aceitaram-me, atribuíram-me responsabilidades como um profissional de alto nível, tal como vocês. Tenho a certeza que não poderia ter realizado este estágio em outro clube, porque as aprendizagens associadas são infundáveis.

O meu obrigado a todos vocês, que me acompanham, nesta jornada que é a vida.

Resumo

Um relatório de estágio profissionalizante procura transmitir o quotidiano, as exigências, as reflexões e inquietações sentidas ao longo do período de inserção, neste caso durante uma época desportiva.

Nesta lógica de pensamento, o presente documento foi desenvolvido no seio da academia do Sevilla Fútbol Club, atual detentor de 7 *Europa Leagues*, com o objetivo de apresentar de forma sintetizada as vivências e saberes da experiência profissionalizante relativos à Unidade Curricular Estágio, com vista à obtenção do grau de Mestre em Treino Desportivo pela Universidade de Coimbra.

Simultaneamente, é apresentado um estudo científico, intitulado como “O algoritmo de identificação de perfis similares de jogadores de futebol”, cujo desenvolvimento foi realizado com dados da entidade de acolhimento, sendo neste caso, um dos projetos implementados no clube. O projeto recorre à utilização de um algoritmo de *clustering K-means*, juntamente com a técnica de análise de vetores, *cosine similarity*, de forma a obter-se, como objetivo primordial, a identificação de perfis similares de jogadores de futebol. A sua finalidade passa pelo auxílio no processo de tomada de decisão na construção dos plantéis das distintas equipas da organização.

Sucintamente, o documento é composto por três pilares onde, primeiramente, é realizada uma contextualização da instituição de acolhimento, tanto histórica como atual. Seguidamente abordam-se alguns dos projetos desenvolvidos ao longo do período de estágio profissionalizante, onde o estagiário é membro integrante e de sua autoria. Por último, aborda-se o projeto de investigação, com o objetivo de dar-se a conhecer uma nova forma de obtenção de informação adicional, aplicável às transferências de jogadores de futebol.

Palavras-chave: Futebol, Análise de Dados, *Performance Analysis*, Observação e Análise, Estágio.

Abstract

A professional internship report seeks to convey the daily life, demands, reflections and concerns felt throughout the period of insertion, in this case during a sporting season.

In this logic of thought, this document was developed within the academy of Sevilla Fútbol Club, current holder of 7 Europa Leagues, in order to present in a synthesized way the experiences and knowledge of the professional experience related to the Curricular Internship, in order to obtain the degree of Master in Sports Training from the University of Coimbra.

Simultaneously in the document, an integral member is a scientific study, entitled as "The algorithm for identifying similar profiles of football players", whose development was carried out with data from the host entity, one of the projects implemented in the club. The project uses a clustering algorithm K-Means together with the vector analysis technique, Cosine Similarity, in order to obtain, as a primary objective, the identification of similar profiles of soccer players. The purpose is to assist in the decision-making process in the construction of the different teams of the organization.

Briefly, the document is composed of three pillars, where first is performed a contextualization of the host institution, both historical and current. It is then addressed some of the projects developed over the period of professional internship, where the trainee is an integral member and his own. Finally, the research project is approached, with the objective of making known a new way of obtaining additional information, applicable to transfers of football players.

Keywords: Football, Data Analysis, Performance Analysis, Observation and Analysis, Internship.

Índice

Agradecimentos	vi
Resumo	viii
Abstract	ix
Índice de Figuras	xii
Índice de Tabelas	xiv
Capítulo I - Introdução- objetivos e expectativas iniciais.....	1
1.1 Introdução	2
1.2 Expectativas iniciais para o estágio	4
1.3 Objetivos do estágio	6
Capítulo II - Contextualização e caracterização do clube de acolhimento	8
2.1 História do Sevilla Fútbol Club	9
2.2 Recursos Espaciais e Materiais	10
2.3 Estrutura Organizacional do Clube	13
2.4 Projeto do Sevilla FC	15
2.4.1 Departamento de Análise Desportiva	20
Capítulo III - Planeamento e projetos do departamento de análise desportiva	23
3.1 Planeamento anual	24
3.2 Planeamento semanal	27
3.3 Tratamento, Visualização e Análise de dados	33
3.4 Projetos Autopropostos	38
Capítulo IV - Considerações finais	42
Capítulo V - O algoritmo de identificação de perfis similares de jogadores de futebol ..	45
Abstract	46
1. Introdução	46
2. Materiais e métodos	48

3. Resultados e Discussão	54
4. Conclusão	57
Referências Bibliográficas	58

Índice de Figuras

Figura 1- Figura ilustrativa do Estádio Ramón Sánchez Pizjuán após requalificação.....	10
Figura 2- Mapa de acesso ao Estadio Ramón Sánchez Pizjuán.	11
Figura 3- Representação ilustrativa da <i>Ciudad Deportiva Ramón Cisneros Palacios</i> do <i>Sevilla Fútbol Club</i>	12
Figura 4- Projeto de requalificação da cidade desportiva	13
Figura 5- Representação Esquemática do organigrama Administrativo do <i>Sevilla Fútbol Club</i>	14
Figura 6- Atletas da formação do SFC, Sérgio Ramos, Jesús Navas e David Prieto.....	15
Figura 7- Imagem elucidativa da residência da academia do SFC.	16
Figura 8- Representação Esquemática do SFC da Organização do Departamento de R+D+I (fonte Documentos Oficiais)	18
Figura 9- Trajeto do departamento de Análise Desportiva do SFC após criação.	21
Figura 10- 6 Taças Europa Leagues pertencentes ao museu do SFC.....	22
Figura 11- Esquematização temporal do planeamento anual do departamento de análise desportiva	24
Figura 12- Publicação do SFC nas páginas oficiais sobre a ferramenta desenvolvida internamente e comercializada à LaLiga, <i>Transfer Tracker</i>	26
Figura 13-Representação gráfica da janela de codificação de ações individuais e coletivas utilizadas no <i>Nacsports</i>	29
Figura 14- Representação ilustrativa de um ficheiro exportável de <i>event data</i>	30
Figura 15- Ilustração do equipamento de gravação <i>VEO</i> e respetiva <i>cloud</i>	32
Figura 16- <i>Dashboard</i> desenvolvida em <i>PowerBi</i> representativa da eficácia de atuação em determinado corredor (esquerdo, central, direito).	35

Figura 17- <i>Dashboard</i> desenvolvida em <i>PowerBi</i> com a finalidade de comparação de performance individual entre atletas.	36
Figura 18- Representação esquemática de jogadas que terminam em gol e esquematização de associação de passes em pontapé de baliza (mapas desenvolvidos de autoria própria).....	37
Figura 19- Esquematização da arquitetura do projeto de <i>Webscraping</i>	39
Figura 20- <i>Screenshot</i> exemplificativo do ficheiro extraído após <i>webscraping</i>	39
Figura 21- <i>Dashboard</i> desenvolvida em <i>PowerBi</i> com informações dos dados obtidos através do projeto <i>Webscraping</i>	40
Figura 22- Ilustração das bibliotecas necessárias para o projeto de normalização de dados. ...	40
Figura 23- Documento de avaliação oficial do estágio por parte da entidade de acolhimento, SFC.....	44
Figura 24- Gráfico de barras do número de atletas por equipa da temporada 2022-2023 a atuar na academia do <i>Sevilla Fútbol Club</i> e pertencentes à amostra do estudo.	48
Figura 25- Esquematização do número de atletas válidos e incluídos no estudo.	50
Figura 26- Esquematização da arquitetura do modelo desenvolvido desde importação até atualização.	53
Figura 27- Representação do método <i>Elbow</i> . Número de <i>Clusters</i> por <i>Within-Cluster Sum of Squares (WCSS)</i>	54
Figura 28- Representação do grupo de cluster desenvolvido em <i>PowerBi</i> para a posição Medio Centro.	54

Índice de Tabelas

Tabela 1- Enumeração dos títulos conquistados pelo <i>Sevilla Fútbol Club</i> a nível Sênior.....	22
Tabela 2- Correspondência de etapas de formação da modalidade de futebol de acordo com a idade cronológica.	27
Tabela 3- Representação do Horário Oficial Semanal de um analista do SFC.....	28
Tabela 4- Representação das variáveis utilizadas nos diferentes dataset (Ofensivo, Defensivo e Global).....	50
Tabela 5- Representação dos resultados obtidos do top 3 de similitude global para substituição do Jogador A.	55

Capítulo I

Introdução- objetivos e expectativas iniciais

1.1 Introdução

O Futebol é a modalidade desportiva de maior abrangência a nível global (Soriano, 2010).

Alicerçado à sua popularidade está a procura de melhoria constante, de forma a obter-se vantagem competitiva face ao rival. Esta melhoria é desenvolvida através, sobretudo, de estudos científicos. Sarmiento et al. (2014) expressam que um dos primeiros trabalhos desenvolvidos na análise de jogo remonta ao ano de 1910, apontando como autor Hugh Fullerton. Com isto, interprete-se a antiguidade e a necessidade de adaptação de muitas formas de operacionalização relacionadas com a modalidade.

Uma das áreas que mais tem evoluído nos últimos tempos é a área de *performance analysis*. Esta aposta está diretamente relacionada com a ideia transmitida por Bill James (citado por Anderson & Sally, 2013) que no desporto, o que é verdadeiro, é mais poderoso do que aquilo em que se acredita, pois o que é verdadeiro fornecerá uma vantagem. Sendo assim, a aposta em obtenção de análises avançadas constituiu um passo subsequente. São exemplos de métricas avançadas o *Expected Goals (xG)* e a avaliação de jogadores por posição (*Positional Player Rating*), entre outras que começam a auxiliar na tomada de decisão.

Anderson e Sally (2013), afirmam que qualquer clube de renome internacional está ciente do facto que o futebol está em mudança em relação às tradicionais formas de trabalho. Os autores constataam ainda que estes clubes integram equipas de análise, compostas por especialistas na recolha e interpretação de dados, que exploram integralmente todas as informações que podem obter, com o objetivo de auxiliar a planear sessões de treino, conceber sistemas de jogo mais apropriados e delinear transferências. Valores monetários substanciais e um vasto número de troféus estão em causa. Cada clube demonstra prontidão para adotar quaisquer medidas necessárias com o intuito de adquirir a mínima vantagem possível. Estes autores, expressam algo ainda tão atual, apesar de a afirmação remontar ao ano de 2013.

Contrariamente, nem tudo terá de ser alterado porque muitas coisas foram bem realizadas até ao momento, porque estamos perante uma das melhores eras de futebol a nível tático. Contudo, os clubes que não acompanharem a era da digitalização ficarão em desvantagem competitiva, e desvantagem competitiva, é dinheiro no final das contas.

Estes são alguns dos tópicos a serem abordados no presente documento, que procuram demonstrar o novo papel de um analista de dados inserido num clube de elite, que se assume como um dos clubes mais desenvolvidos tecnologicamente a nível mundial. Contudo, primeiramente conheça-se a estrutura do documento.

De forma inicial, o presente relatório expressa o quotidiano, as exigências, as reflexões e inquietações de um analista e “eterno estudante” inserido em um contexto de elite, relativo à modalidade de futebol. Contudo, enaltece-se que este percurso, que se finalizou na redação de este relatório, foi iniciado muito antes de este último ano, no que ao sonho e ambição se concerne.

Ademais, reflexiona-se a importância do Mestrado em Treino Desportivo (MTD) da Faculdade de Ciências do Desporto e Educação Física (FCDEF-UC) para a estruturação de este documento, onde o mesmo é inerente. Acrescenta-se ainda, que a frequência do MTD foi a “Pré-Época” para uma das fases mais desafiantes da vida de um analista (até ao momento), onde se apresentaram diferentes perspetivas e saberes, com o objetivo de estimular a curiosidade de cada estudante, que é um “passe-chave” na área do treino.

Para esta obra ser apresentada foi necessário inscrever o analista no “campeonato”, e nada melhor do que essa inserção ser no clube referencia no que à análise de dados no mundo do futebol se trata, sendo neste caso o *Sevilla Fútbol Club*.

Assim sendo, este relatório pretende apresentar, de forma sucinta, as vivências e saberes da experiência profissionalizante relativos à Unidade Curricular Estágio, com vista à obtenção do grau de Mestre em Treino Desportivo pela Universidade de Coimbra.

Estruturalmente, a obra está dividida por um enquadramento inicial, onde se realçam informações relevantes sobre a instituição de acolhimento e uma simplista caracterização das funções exercidas pelo analista. A obra culmina, por fim, na apresentação de um dos projetos de investigação desenvolvido no seio da estrutura desportiva, com o devido realce do acréscimo que o mesmo pode ter, com o atual estado da arte da modalidade.

De forma mais detalhada o documento apresenta-se com um primeiro capítulo intitulado: “Introdução - objetivos e expectativas iniciais”. O mesmo pode ser associado à temática futebolística como o início de um jogo de futebol, devido à sua natureza maioritariamente introdutória e de gestão de expectativas iniciais, no qual antes do seu início, por exemplo, não é de conhecimento como se apresentará a nível tático o adversário. Neste caso, é aplicável um raciocínio similar, uma vez que, é desconhecido o desfecho e aprendizagens associadas ao estágio curricular.

De seguida, é possível encontrar-se o capítulo número II (dois): “Contextualização e caracterização do clube de acolhimento”. Compreenda-se que no mesmo, é possível obter-se informações relativamente à história do clube e respetiva evolução, fazendo referência à sua estrutura (de nível macro a nível micro), associando-se aos diferentes objetivos a serem perseguidos ao longo da temporada desportiva 2022/2023.

Por sua vez, no capítulo III (três), “Planeamento e projetos do departamento de análise desportiva”, é possível ler-se o quotidiano do analista. É referenciado durante o desenrolar do mesmo, como o próprio nome indica, a organização anual, semanal e projetos associados ao departamento do qual é membro integrante o estagiário, com o objetivo, de apresentar-se as relações práticas retiradas e desenvolvidas durante a época desportiva. Este capítulo apresenta-se como uma exposição de diferentes variáveis de rendimento (diferentes projetos desenvolvidos por parte do estagiário), possibilitadora de uma futura avaliação da *performance* do próprio Analista.

Com lógica sequencial, apresenta-se o capítulo número IV (quatro), intitulado como “Considerações finais”. Como o próprio título indicia trate-se de um aglomerado de conclusões e reflexões em relação à experiência vivenciada. Por fim, o presente documento termina com a apresentação de um trabalho de investigação, nomeado como “Capítulo V - O algoritmo de identificação de perfis similares de jogadores de futebol”.

Sucintamente trata-se de um resumo das aprendizagens adquiridas sobre o fenómeno desportivo vivenciado, com o objetivo de ser mais um “Período Preparatório”, tendo em vista o “Período Competitivo”, sendo neste caso, a futura prática profissional no mundo do treino desportivo.

1.2 Expectativas iniciais para o estágio

*“Quero, terei — Se não aqui,
Noutro lugar que inda não sei.
Nada perdi.
Tudo serei.”*

Fernando Pessoa (1990)

O estágio profissionalizante (EP) é a reta final do percurso de ensino do MTD, onde o estudante tenta encontrar respostas para as suas equitções originárias da explanação teórica e dos diferentes saberes absorvidos. O EP consiste também, no início da trajetória e preparação para uma futura inserção no competitivo mercado de trabalho, sendo caracterizado como o primeiro contacto com as exigências do fenómeno desportivo.

Em sùmula é a vivência diária do processo de transferência da teoria para a prática, podendo-se caracterizar esta transferência/adaptação, como um paradigma imprescindível no mundo do treino desportivo, devido à necessidade de atualização e adaptação constante a diversos cenários.

Por sua vez, as expectativas iniciais associadas à realização da prática profissionalizante vão ao encontro do excerto anteriormente mencionado de Fernando Pessoa (1990), na sua obra *Poesias Inéditas*, onde as mesmas se podem caracterizar como ambiciosas, e que, independentemente dos cenários futuros, a experiência vivenciada constará como uma aprendizagem e evolução, tanto profissional como pessoal, culminando certamente neste conceito-chave “(...) *Nada perdi. Tudo serei.*”.

Agregadamente, as expectativas também se podem adjetivar como desafiantes. Como se compreende, o contexto de realização do estágio curricular é considerado pela *Union of European Football Associations* (UEFA), na temporada transata (2021/2022) o 12º melhor clube europeu do Ranking de Coeficientes de Clubes, tendo sido membro integrante do top 10 na época 2019/2020, mais concretamente, o 8º lugar do *ranking* de clubes europeus.

Outro ponto, com conotação de importância, é o facto de o estagiário integrar a equipa de trabalho, com responsabilidades de ajuda na tomada de decisão, de forma indireta para com a direção desportiva do clube. Este ponto objetiva-se como procura da máxima absorção de conhecimento possível, em relação ao mecanismo de trabalho, daquele que é considerado atualmente, um dos melhores diretores desportivos, conhecido por “Monchi”.

De forma a justificar o ponto anteriormente mencionado, note-se que a empresa norte americana, do setor da informação de temáticas económico-financeiras, mais concretamente o *International Business Times* (citado por Daniel Pinilla, 2020), refere-se ao mesmo como “*El rey Midas del fútbol profesional*”, pelas diversas contratações e posteriores vendas valiosas de jogadores profissionais de futebol, oriundos de cenários competitivos de um nível inferior, com rumo para grandes ligas e clubes europeus, que acabaram por ser e são atualmente, jogadores de extremo impacto no futebol mundial, tais como, Jules Koundé, Jesus Navas, Sérgio Ramos (jogador da formação do clube), Ben Yedder, Daniel Alves, Ivan Rakitic, entre muitos outros. Assim sendo, espera-se aprender dos melhores e com os melhores, procurando-se de alguma forma aprofundar e aplicar os conhecimentos adquiridos sobre o fenómeno do treino desportivo e o mundo da análise de dados, no *Sevilla Fútbol Club*.

Desta forma, sumariamente, as expectativas para este estágio podem ser caracterizadas por procura constante de acréscimo de valor à estrutura desportiva nos diferentes processos de tomada de decisão, bem como aferir-se que no final da prática profissional o estagiário encontra-se com um leque de experiências e conhecimentos imersivos extremamente enriquecedores, evolucionando o mesmo, para outro nível em comparação ao perfil inicial, antes da realização do estágio curricular.

1.3 Objetivos do estágio

O primeiro objetivo, de cariz pessoal, considerado como o mais global, porém o pilar essencial para a concretização dos demais objetivos, é o estabelecimento de relações de confiança com os diferentes intervenientes do *Sevilla Fútbol Club*. Este ponto está assente no motivo da realização da experiência profissionalizante (EP), a quase 800 quilómetros de casa, fora das fronteiras de Portugal, onde o estudante apenas conta consigo e com os seus sonhos. Desta feita, tornar-se primordial uma rápida e eficiente adaptação ao novo contexto, criando o clima ótimo para a obtenção do máximo rendimento como estagiário em um dos maiores clubes do mundo.

O segundo objetivo, de cariz profissionalizante, passa por procurar acrescentar valor ao clube com os conhecimentos teóricos e técnicos, acreditando-se que Portugal tem muito a oferecer ao mundo do futebol.

O terceiro objetivo passa pela obtenção do grau de mestre em Treino Desportivo pela FCDEF-UC ao finalizar o EP, independentemente das circunstâncias.

O quarto objetivo, caracteriza-se pela procura constante de melhoria das suas capacidades técnico-práticas de análise de dados, onde se procura aprofundar os conhecimentos práticos relacionados com temáticas, tais como *Machine Learning* e tratamento de grandes bases de dados, comumente conhecido como *Big Data*. Associativamente a este objetivo, entenda-se, por sua vez, a melhoria de manuseamento de software de visualização dos dados tratados e analisados com as técnicas supramencionadas, como são o caso de ferramentas como *Power Bi* ou *Tableau*.

Em quinto lugar, ambiciona-se num futuro, a obtenção do grau de treinador. Este objetivo passa pela preocupação de comunicar no mesmo “idioma” que o treinador. Note-se desde já, que este é um dos pilares primordiais de um analista de dados, neste caso, aplicável na relação para com a equipa técnica, onde a comunicação e confiança são pilares basais. Para tal, é necessário a compreensão de terminologias técnicas, bem como refletir as necessidades e inquietações dos treinadores de futebol, culminando em respostas úteis para o processo de tomada de decisão do interveniente direto no rendimento desportivo, que será sempre o treinador da modalidade.

Por sua vez, um sexto objetivo, passa por ser um dos impulsionadores, num futuro próximo, do desenvolvimento da área de análise de dados aplicado ao futebol em Portugal. Este ponto reflete-se de extrema motivação, ambicionando-se criar maior competitividade no campeonato português de futebol e subseqüentemente, procurar esbater a diferença existente para com as demais ligas e clubes europeus.

Outro tipo de objetivo, não diretamente aplicável ao estágio curricular, mas merecedor de destaque, mais concretamente uma ambição que qualquer profissional da área desportiva sonha um dia lograr, qualificando-se uma vez mais no leque de objetivos futuros, independentemente da sua linha temporal de acontecimento (mais longínqua ou mais próxima), sendo neste caso, a participação nas principais e máximas competições do mundo futebolístico, tais como, *FIFA World Cup*, *UEFA European Championship*, *UEFA Champions League* e *UEFA Europa League*.

Todos os demais objetivos, na perspetiva do estagiário, não devem ser traçados numa fase inicial, mas sim, serem criados e reajustados em relação ao momento que se vivencia durante a época desportiva. Sucintamente, deve-se ter objetivos de direção, para possuir-se um rumo, respondendo à questão onde queremos chegar a longo prazo. Por sua vez, todos os demais objetivos, precisam ser alvo de reajuste conforme as exigências do quotidiano.

Capítulo II

Contextualização e caracterização do clube de acolhimento

“El club que aprendió a ganar”

Paco Cepeda (2020)

O *Sevilla Fútbol Club* será alvo de descrição no presente capítulo, assemelhando-se à apresentação de um calendário competitivo, onde, com o conhecimento do mesmo, é possível obter-se informações como dias de jogo, adversários, jogos sob a condição de local ou visitante, originando subsequentemente ajustes importantes, diretamente relacionados com o planeamento anual da temporada. Neste caso, a apresentação da entidade desportiva, possui o mesmo cariz de um calendário competitivo, onde iremos introduzir diversos detalhes da mesma, denotando-se da sua imprescindibilidade para o planeamento anual de tarefas do estagiário, subsequente organização e distribuição de projetos a desenvolver ao longo da temporada desportiva 2022/2023.

2.1 História do Sevilla Fútbol Club

Primeiramente deve-se dar a conhecer a história da instituição de acolhimento, que segundo Montes (1981), a 25 de janeiro de 1890 “El Club de Football de Sevilla” encontrava-se devidamente e legalmente constituído, com os respetivos cargos oficiais nomeados, tornando-se esta, a data de fundação oficial do clube.

Desde logo, realçar a sua dimensão a nível histórico, onde se pode compreender que se trata de um clube com 133 anos de muitas lutas, lágrimas e troféus. A mesma fonte bibliográfica detalha mais sobre a organização desportiva, referenciando três pessoas como fundadoras, sendo elas Edward F. Johnston (cargo de Presidente); Isaías White (cargo de Secretário) e por último, Hugo Maccoll (cargo de capitão). É então, no dia 8 de março do mesmo ano de fundação que o Sevilla Fútbol Club realiza o seu primeiro jogo, que haja registo, tendo como adversário o Real Club Recreativo de Huelva, iniciando aqui a sua histórica era de conquistas que passar-se-á a apresentar de seguida (Montes, 1981).

Um clube com 133 anos de história, à data de hoje, apresenta 29 presidências distintas, tendo como presidente atual José Castro Carmona, onde iniciou o seu mandato em 2013, cujo qual, vigora até ao momento.

Segundo o livro oficial do clube, “Historia del Sevilla Fútbol Club”, relativamente ao órgão máximo do clube, a pessoa que apresenta o maior número de anos sob o leme do clube é Ramón Sánchez Pizjuán, mais concretamente, de 1932 a 1941 (1º mandato) e de 1948 a 1956 (2º mandato), somando-se neste caso, 17 anos como presidente, de um dos históricos europeus.

Tal feito é merecedor de referência, como intuitivamente se interpreta, na homenagem à sua pessoa, foi realizado o “apadrinhamento” do mítico estádio do Sevilla FC, originando desta forma, o extremamente conhecido Estádio Ramón Sánchez Pizjuán.

2.2 Recursos Espaciais e Materiais

A infraestrutura de eventos desportivos do Sevilla Fútbol Club, localiza-se na Avenida de Eduardo Dato, bairro de Nervión, da cidade de Sevilha em Espanha.

Numa fase inicial os arquitetos António García de Castro e Antonio González Cordón desenvolveram uma obra com a capacidade para 30000 espectadores, tendo associada como data de inauguração, 7 de setembro de 1958, segundo a página oficial do *Sevilla Fútbol Club*. Pós as últimas intervenções, o estádio aumentou a sua lotação em 13 883 espectadores, culminando numa capacidade total de 43858 adeptos.

Observe-se na figura 1, o atual estado do palco da entidade desportiva de Nérvion.



Figura 1- Figura ilustrativa do Estádio Ramón Sánchez Pizjuán após requalificação.

A reestruturação do estádio foi realizada, não exclusivamente com o objetivo de ampliação do número de espectadores que poderiam assistir ao evento desportivo, mas também, possibilitou a aplicação da atual organização e distribuição dos aficionados ao longo do palco desportivo, com a finalidade de proporcionar uma melhor experiência ao adepto.

Esta reforma, foi diretamente associada ao desenvolvimento de um *software* próprio, considerado como um dos projetos mais importantes da entidade desportiva, neste caso intitulado como *Ticketing*. Com esta ferramenta, o clube consegue corrigir e ajustar o preço de cada jornada, através de algoritmos de previsão (quantos adeptos segundo dados históricos poderão comprar entrada para um determinado evento desportivo), e consequentemente, organizá-los e distribuí-los pelas distintas zonas do estádio, orientando os mesmos, para as diversas portas de entrada, evitando longas filas de espera.

De forma sucinta, visualize-se posteriormente a figura número 2, ilustrativa das zonas de acesso criadas com a nova reforma do palco desportivo.



Figura 2- Mapa de acesso ao Estádio Ramón Sánchez Pizjuán.

Dentro do tópico de infraestruturas que beneficia e possui a entidade desportiva, neste caso as instalações que possibilitam a preparação para a atuação no grande palco de Nervión, o Sevilla Fútbol Club dispõe de uma cidade desportiva intitulada de “Ciudad Deportiva José Ramón Cisneros Palacios”.

O nome anteriormente apresentado, associado à cidade desportiva da organização, deve-se ao órgão máximo da entidade, cuja figura apresenta-se como principal responsável pelo desenvolvimento do projeto, e aquisição dos terrenos para futura construção da mesma, sendo neste caso, o Presidente José Ramón Cisneros Palacios.

Durante a década de 60, o clube apresentava poucas infraestruturas, mais concretamente um estádio (anteriormente apresentado), destinado à equipa principal cuja utilidade era para treinos e jogos da mesma, e um campo de treinos, designado para a preparação dos jogos, neste caso utilizado pela equipa B “de la familia *roja y blanca*”.

Desta feita, José Palacios, procurava desenvolver um projeto que passasse pela construção de novas infraestruturas, com a finalidade de auxiliar os processos de treino das distintas equipas integrantes da instituição, procurando desta forma, a melhoria do seu projeto desportivo.

Tendo em conta as limitações financeiras da época, foi possível obter-se um recurso espacial de 150 mil metros quadrados, na atual localização do centro de treinos, mais concretamente *Carretera de Utrera, Km. 1, 41005 Sevilla, España*. Apesar de o primeiro passo de engrandecimento a nível de recursos espaciais ter sido alcançado, a equipa principal acaba por descer de divisão em 1972, sob o mandato do José Palacios, acabando com a destituição do mesmo, do cargo de presidente da administração do clube.

Assim sendo, a cidade desportiva é oficialmente inaugurada sob a ordem de Eugenio Montes Cabeza, mais concretamente, no dia 4 de fevereiro de 1974. Contudo, o antigo

presidente Eugenio Cabeza, não é associado como o idealizador da obra devido aos detalhes anteriormente apontados.

Desta forma, é em assembleia aprovado, em homenagem à sua pessoa, mais concretamente ao antigo presidente José Ramón Cisneros Palacios, a atribuição do seu nome ao atual centro de treinos da instituição (*Historia del Sevilla Fútbol Club*, Montes, 1981).

Com o objetivo de se identificar o estado atual da cidade desportiva, visualiza-se seguidamente a figura 3, composta por 3 imagens, ilustrativas da mesma.



Figura 3- Representação ilustrativa da Ciudad Deportiva Ramón Cisneros Palacios do Sevilla Fútbol Club.

Como se observa, atualmente o centro de treinos do Sevilla Fútbol Club, após a segunda fase de renovação das infraestruturas, apresenta:

- 1 Estádio, conhecido como Estádio Jesús Navas, com uma capacidade para 6000 espectadores, cujo nome da infraestrutura dispensa apresentações;
- 7 campos de relva natural, sendo 3 deles de apoio exclusivo à primeira equipa;
- 4 campos de relva artificial sintética;
- 2 ginásios (1 para a formação e outro para a primeira equipa);
- 1 restaurante/cafeteria aberta ao público diariamente;
- 1 refeitório de apoio à primeira equipa;
- Departamento médico com recurso a material especializado;
- 1 sala de conferência de imprensa;
- 1 balneário de apoio à equipa filial (*Sevilla Atlético Club*), onde os atletas têm à sua disposição diversos recursos materiais, como por exemplo material de recuperação desportiva (exemplo: crioterapia, jacuzzi, etc);
- 2 balneários destinados às equipas de etapas superiores da formação (*Sevilla C e Sevilla Juvenil División de Honor*);
- 10 balneários de apoio aos distintos escalões de formação do clube;
- 2 lavandarias;
- 5 armazéns (vinculados ao serviço de manutenção das infraestruturas e apoio às equipas da formação);

- 6 Escritórios e 2 espaços de *coworking* localizados no Estádio Jesus Navas;
- 5 salas de reunião, de acesso indiscriminável, com o objetivo de auxiliar todas as necessidades dos diferentes intervenientes e membros integrantes da entidade.

Com a nova fase de expansão da cidade desportiva, o clube oferecerá acesso a novas infraestruturas, neste caso à disposição da equipa principal masculina, que por sua vez, possuirá um espaço integralmente independente.

O novo edifício, consta com um espaço residencial composto por 40 dormitórios, 2 balneários, um restaurante, um ginásio, um centro médico, bem como um espaço de reabilitação e recuperação.

De forma a elucidar mais facilmente a dimensão atual do presente projeto do clube, veja-se a figura 4.



Figura 4- Projeto de requalificação da cidade desportiva

2.3 Estrutura Organizacional do Clube

Estas infraestruturas, anteriormente reportadas são as ferramentas essenciais para dar suporte e possibilitar o máximo rendimento dos recursos humanos da organização, mais concretamente aos 455 funcionários.

Nesta vertente de recursos humanos conhece-se, primeiramente, o organigrama da estrutura, na figura 5, proveniente de documentos oficiais da entidade desportiva. Como se observa, no presente ano de realização do EP, o Presidente da instituição é José Castro Carmona, cujo mandato vigora desde 2013.

Associado ao cargo de presidente, encontra-se os membros integrantes da direção desportiva, onde, se deve dar realce à posição de Diretor Geral Desportivo, cargo exercido durante a presente temporada por Ramón Rodríguez, comumente conhecido por Monchi.

Este último, merece especial destaque e plausível apresentação, não só por ser o órgão máximo associado ao trabalho desenvolvido por parte do estudante, bem como pelo mérito e

respeito alcançado pela sua pessoa no meio desportivo, por inúmeros sucessos, tanto em contratações de futebolistas profissionais, bem como em títulos alcançados no comando do clube.

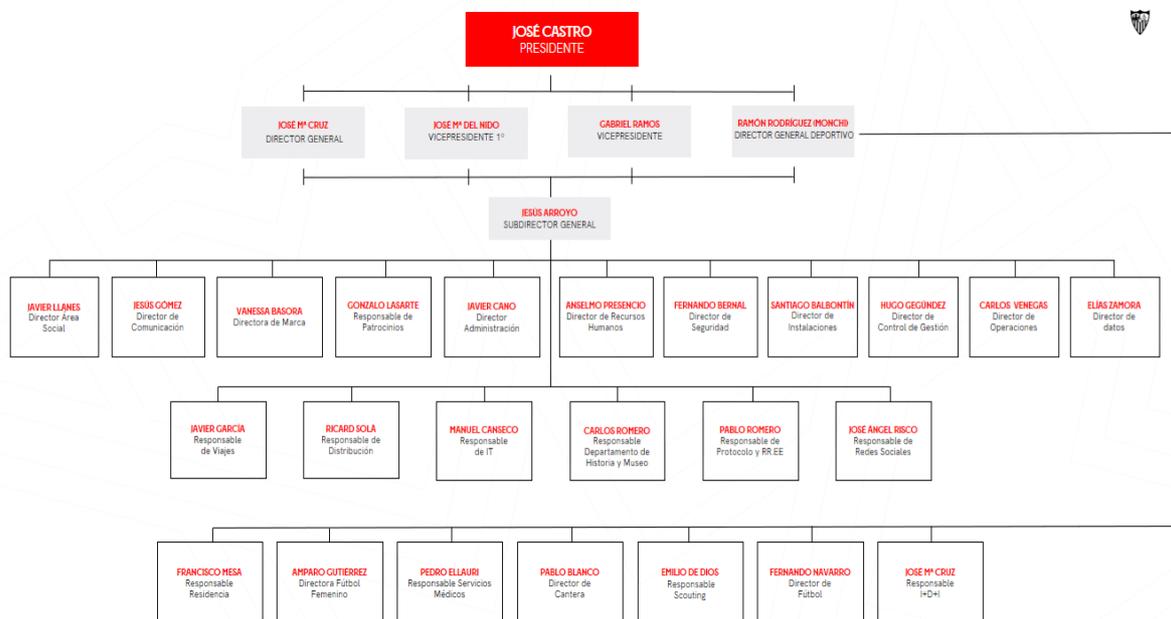


Figura 5- Representação Esquemática do organograma Administrativo do *Sevilla Fútbol Club*.

Segundo Pinilla (2020) “Cuando Monchi tomó las riendas de la dirección deportiva del Sevilla FC en el año 2000, no pensó que su objetivo era comprar jugadores baratos para venderlos caros, no ideó una metodología de trabajo para firmar estrellas, no organizó la cantera para ahorrar costes en fichajes.”. Como se pode interpretar, a figura máxima associada à gestão desportiva é a imagem mais conhecida do *Sevilla Fútbol Club*, onde o autor afirma, que Monchi, no seu primeiro mandato, com início nos anos 2000 até 2017, não seria fácil prever todo o sucesso que seria alcançado com a sua ideologia de gestão.

Deste modo, é durante o seu legado, 21 anos no total (contabilizando-se com o segundo mandato pós regresso em 2019, proveniente do AS Roma), que o Sevilla FC renasce para o futebol mundial, obtendo atualmente o número recorde de conquistas da competição *UEFA Europa League*, mais concretamente 7 conquistas.

Contudo, o seu sucesso como diretor desportivo, não deve ser exclusivamente mensurado através da obtenção de títulos (considerado para o adepto comum, um dos grandes medidores de sucesso desportivo), mas sim, pelo feito da aplicação de um projeto desportivo centrado no jogador da formação, com o objetivo de os formar para o futebol profissional.

Durante os anos de mandato, foi possível, em relação a este ponto, observar-se jogadores da academia do clube a alcançarem este “sucesso desportivo”, tais como Jesús Navas e Sérgio

Ramos. Observe-se seguidamente uma imagem icónica, onde é possível visualizar-se três jogadores da formação, que debutaram na equipa principal (Sérgio Ramos, Jesús Navas e David Prieto, da esquerda para a direita respetivamente).



Figura 6- Atletas da formação do SFC, Sérgio Ramos, Jesús Navas e David Prieto.

Os atletas na figura anteriormente apresentados, são exemplos sonantes de casos de sucesso, que possibilitam afirmar que o projeto desportivo implementado por Ramón Rodríguez é merecedor de destaque.

2.4 Projeto do Sevilla FC

Para se conhecer melhor o projeto desportivo do *Sevilla Fútbol Club*, deve-se saber primeiramente que o Diretor Desportivo, tem ao seu dispor inúmeros departamentos com objetivo de auxiliar as tomadas de decisão em diferentes áreas de atuação, sendo estas preponderantes para o correto crescimento e acompanhamento do futebolista, desde a formação até à primeira equipa, tais como:

- Departamento de Residência;
- Departamento de Futebol Feminino;
- Departamento da Formação;
- Departamento de *Scouting*;
- Departamento de Futebol;
- Departamento de Investigação, Desenvolvimento e Inovação.

Relativamente ao primeiro departamento apresentado, que é um dos pontos cruciais para a criação da mística e sentimento de pertença à família de Nervión, bem como, o assumir da função de suporte para os que se encontram longe da sua família é a residência do clube.

A residência da instituição, atualmente, encontra-se a 2,6 quilómetros da cidade desportiva, mais concretamente localizada na *Calle Historiador Juan Manzano, 2, 41089 Montequinto, Sevilla, España*. A última fase de reestruturação da cidade desportiva, apresentada anteriormente a figura elucidativa do projeto, tem um edifício única e exclusivamente destinado para a nova residência dos atletas da formação, procurando integrar todas as infraestruturas de suporte aos atletas no mesmo local. Veja-se seguidamente uma figura elucidativa da residência dos atletas da formação do clube.



Figura 7- Imagem elucidativa da residência da academia do SFC.

Em Espanha, o futebol feminino está a ser alvo de uma aposta muito elevada a nível financeiro, por parte da La Liga. Os diferentes projetos, financiados pela Liga, têm como objetivo a aquisição de recursos materiais para auxiliar o correto desenvolvimento da competição no país, procurando esbater a diferença existente de suporte para com o futebol masculino. Desta forma, o clube, apresenta um departamento de futebol feminino, detendo neste caso como órgão máximo, Ámparo Gutiérrez.

O Sevilla Fútbol Club feminino, possui uma forte aposta a nível de recursos à disposição da equipa, como por exemplo, um campo de treino próprio para auxílio à equipa principal, material de avaliações físicas de terreno e GPS (*Global Positioning System*) destinados à formação da jovem atleta, uma realidade que em Portugal, infelizmente apenas poucas equipas integrantes da Liga BPI têm acesso. Este trabalho, é mais um dos distintos casos de sucesso do projeto desportivo do Sevilla FC.

Por sua vez, o departamento de formação, que se pode considerar como a administração da formação, é constituída pelos distintos decisores de diversas matérias tais como, por exemplo, salários dos jogadores da “*cantera*”, requisitos de acesso à residência, despesas de suporte aos distintos escalões de formação da academia, etc. Para assumir este cargo, Monchi tem um nome sonante, que transporta com ele a mística da família de Sevilla, mais

concretamente Pablo Blanco, antigo jogador do clube, com 14 anos em representação do mesmo.

O objetivo de possuir alguém com a identidade do clube, nos cargos de administração da formação, passa pela transferência e manutenção de determinados valores para os novos/futuros atletas, mantendo-se desta forma, ao longo dos anos, a essência da organização. Com o objetivo de futura melhor interpretação, um clube que partilha o mesmo tipo de valores, que é expressamente visível em Portugal, nas suas equipas da formação é o Futebol Clube do Porto.

Note-se que até ao presente departamento, na perspetiva do estudante, o Sevilla FC não apresenta muitas disparidades organizacionais em comparação aos clubes a atuar em Portugal. Porém, é precisamente o próximo departamento a ser abordado que apresenta formas de trabalho completamente distintas às comumente e tradicionalmente conhecidas, neste caso o departamento de *Scouting*.

A forma de trabalho assenta na procura de combater a lacuna e dependência do atual panorama de deteção e identificação de talentos, que neste caso, segundo Reeves et al. (2018, citado por Bergkamp et al. 2021), os jogadores talentosos geralmente são identificados por *scouts* talentosos, por sua vez. Desde logo, pode-se apontar que uma instituição não deveria ser dependente de deter, os melhores e mais talentosos “olheiros”, para conseguir captar talento para os seus quadros de ativos.

Ao encontro da presente linha de raciocínio, a instituição procurou aportar informação a este departamento recorrendo a um software próprio, desenvolvido pelo departamento de dados do clube, intitulado como *AI Scouting*, com o objetivo de minimizar o erro de tomadas de decisão no que a contratação de novos atletas diz respeito, independentemente do “talento” do *scout*.

Contudo, como estamos perante um clube de elite, o Sevilla FC apresenta profissionais de deteção de talento extremamente competentes, os *scouts* ao serviço da formação apresentam o mesmo método de recrutamento que o apresentado por Reeve et al. (2019 citado por Bergkamp et al. 2022), mais detalhadamente, eles observam e avaliam principalmente jogadores que ainda não foram recrutados por academias de futebol de elite, em testes ou jogos de futebol. Neste caso, a instituição seleciona por Andaluzia (maioritariamente) um número elevado de atletas que estão sinalizados com potencial (relatórios semanais cotados com letras de rendimento A, B e C), realizando-se todas as semanas um jogo amigável entre os candidatos, nas instalações do clube, onde o mesmo é gravado e posteriormente analisado.

É sob este ponto, onde o mesmo se apresenta como merecedor de debate, que se deve enaltecer o primeiro pilar do quotidiano do analista de dados, mais concretamente, a capacidade de reconhecer que os dados são meramente uma complementaridade, tendo como principal finalidade, prestar auxílio no processo de tomada de decisão, com base nas informações fornecidas.

Intrinsecamente, com colaboração direta, associado ao departamento de *Scouting*, depara-se com um dos departamentos mais inovadores de Espanha, no que ao mundo do futebol diz respeito, mais concretamente o departamento de *R+D+I Football (Research + Development + Innovation)*.

É sob a alçada do mesmo, que se insere o gabinete de Análise, mais concretamente o local onde exerceu funções o estagiário da Universidade de Coimbra. Introdutoriamente, conheça-se a estrutura organizacional do departamento, visualizando-se o seguinte organigrama (figura 8).

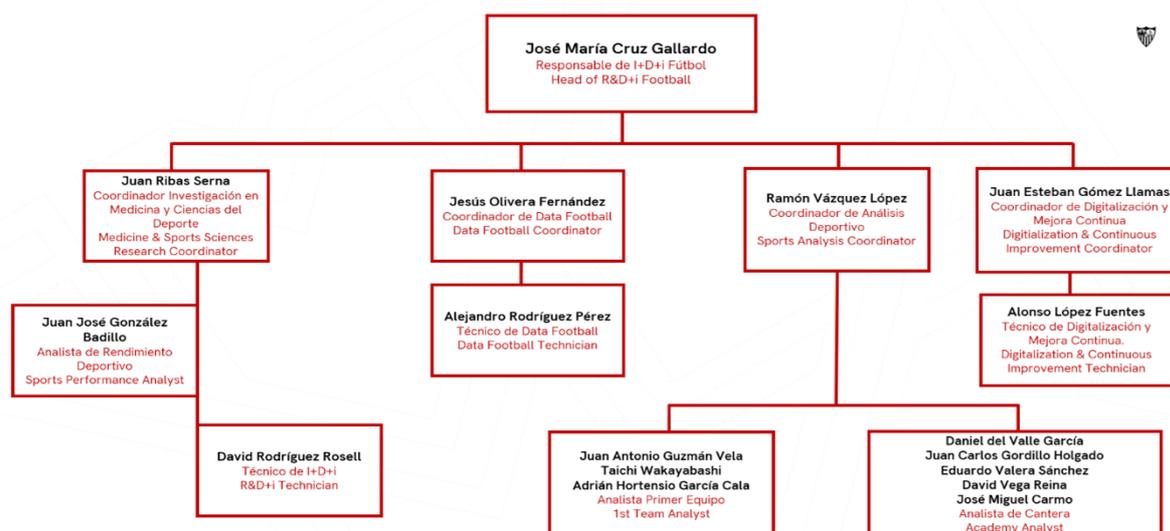


Figura 8- Representação Esquemática do SFC da Organização do Departamento de R+D+I (fonte Documentos Oficiais).

Como é visível no esquema, o máximo responsável da área tecnológica e inovadora do clube é o José María Cruz Gallardo, tutor do estagiário na instituição de acolhimento.

De forma a se conhecer melhor as áreas integrantes no ecossistema inovador do Sevilla FC, leia-se uma simplista explicação do “core” desenvolvido pelos diferentes membros integrantes do mesmo:

- Departamento de Investigação em Medicina e Ciências do Desporto: Este departamento é responsável por recolher, tratar e analisar dados relativos à área medicinal e de rendimento desportivo de todos os atletas da academia do clube, com o objetivo final de gerar um conhecimento científico do estado atual dos

atletas da formação (exemplo: estado maturacional). Este departamento possui ligação direta com área de rendimento desportivo (ginásio e preparadores físicos das equipas técnicas). Apresentam como responsabilidade a realização de projetos de investigação diretamente aplicados à “amostra *Sevilla FC*”, tendo como recurso humano o investigador Juan Badillo, renomeado na área do treino recorrendo ao *velocity based training*. Assim sendo, o trabalho de ginásio é controlado conforme os valores de referência obtidos pelo departamento de investigação em medicina e ciências do desporto.

- *Data Football*: Este departamento assume responsabilidades de análise de dados mais vinculadas ao departamento de *Scouting*, prestando apoio e assessoria direta à primeira equipa masculina. Recorrem a diferentes provedores de dados, bem como ao *software* próprio da La Liga, intitulado de *MediaCoach*, para dar resposta às diferentes inquietações no que à temática de contratações e vendas de jogadores diz respeito. Este departamento apresenta deste modo, uma ligação direta com o departamento de *scouting* (consumidor direto das análises desenvolvidas) e departamento de análise desportiva (auxiliadores do desenvolvimento de projetos devido ao conhecimento técnico da área aplicada ao futebol).
- Departamento de Análise Desportiva: Estamos perante o departamento e área de atuação do estagiário, onde as principais funções do mesmo passam pela procura quotidiana de identificar respostas possíveis a problemáticas técnico-táticas, e consequente avaliação da evolução de rendimento dos atletas da academia, bem como o papel dos treinadores de cada etapa. Este departamento possui ligação direta com a direção desportiva (maior figura administrativa da academia Pablo Blanco, diretor geral desportivo Ramón Rodriguez e as distintas equipas técnicas). O presente tópico será aprofundado de forma mais detalhada posteriormente, no atual capítulo.
- Departamento de Digitalização: O objetivo de este departamento passa pela implementação de uma nova era digital nas diferentes áreas adjacentes ao clube, procurando evolucionar o mesmo para um novo patamar, tornando-se mais tecnológico e subsequentemente, mais sustentável. O mesmo possui impacto direto com todos os demais departamentos anteriormente apresentados devido à sua responsabilidade de dotar os diferentes membros integrantes da entidade. No seguimento deste raciocínio, este departamento é responsável pela capacitação

e formação das equipas técnicas de utilização do *software* ISM (reportado no presente documento posteriormente), onde as unidades de treino serão guardadas na plataforma, com o objetivo de se digitalizar a informação e facilitar uma posteriori análise do trabalho realizado ao longo da temporada desportiva. Para além desta avaliação reportada, também é possibilitadora de se possuir um histórico do estado da academia em diferentes áreas, desde treino até relatórios de *scouting*. Assim sendo, o departamento de digitalização é um dos impulsionadores da inovação na entidade desportiva.

De forma articulada, os diferentes departamentos integrantes da grande área de inovação e investigação do SFC, apresenta dinâmicas cooperativas. A título de exemplo, o departamento de investigação em medicina e ciências do desporto, possui os dados recolhidos através de testes específicos de terreno, que apos a sua análise, é incluído no relatório final do departamento de análise desportiva. O objetivo dos diferentes departamentos é no final de cada trimestre, reportar aos órgãos de tomada de decisão o panorama atual do clube, e se o projeto caminha numa vertente de melhora ou não, como delineado.

Deste modo, as decisões serão tomadas informativamente, através de relatórios de evolução das distintas áreas de impacto direto na gestão do clube.

Simplistamente, os pontos anteriormente representados, bem como os membros integrantes de cada área de atuação, são o projeto do *Sevilla FC* no que à área de inovação diz respeito. Sendo assim, sucintamente, procura-se uma nova vantagem competitiva face a organizações com uma capacidade financeira superior, que disputam o mesmo mercado que o clube.

2.4.1 Departamento de Análise Desportiva

O departamento de análise desportiva pode considerar-se como uma evolução do departamento de análise de vídeo, comumente conhecido em Portugal. Observe-se o esquema (figura 9), apresentando-se com ordem cronológica associada à evolução do departamento, onde é possível, interpretar-se alguns tópicos relacionados com os enfoques essenciais de atuação do mesmo.

Historicamente, a secção iniciou a sua caminhada na temporada de 2015/2016, tendo como principal área de abrangência a tradicional e atual do cenário português, análise de jogo.

O tipo de relatório gerado pelo departamento era maioritariamente em vídeo, com diversas temáticas representadas, desde análise de rival, análise da própria equipa e análise de rendimento individual. Este tipo de forma de trabalho massificou-se rapidamente pelos principais clubes de elite da modalidade de futebol, onde por sua vez a organização procurou, com o regresso do diretor desportivo em 2019, Monchi, proveniente do AS Roma, inovar a sua forma de trabalho, procurando criar uma vantagem competitiva e inclusive, um projeto inovador, considerado atualmente a imagem de marca do clube.

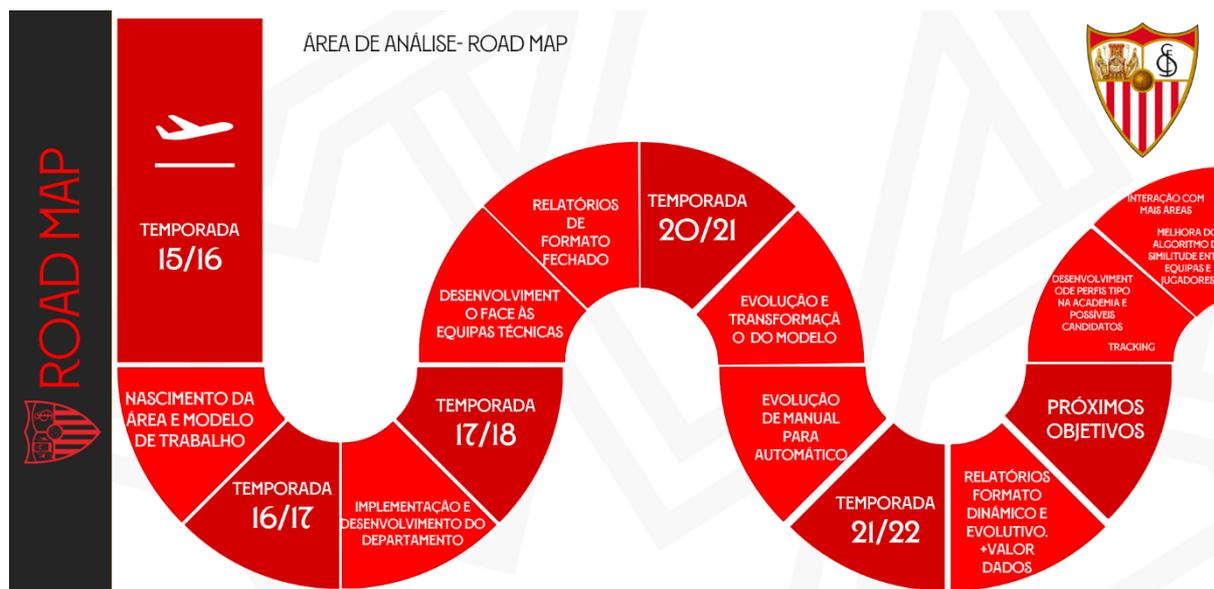


Figura 9- Trajeto do departamento de Análise Desportiva do SFC após criação.

Para se compreender melhor a ideia anteriormente apresentada, simplistamente, o departamento, após o seu regresso, inicia durante a época de 2020, a desenhar um projeto distinto, onde existiu a criação e implementação, por exemplo de um departamento de dados, associado à estrutura do clube, comandado por Elías Zamora (*Chief Data Officer Sevilla FC*). Esta mudança, é visível na figura 9, mais concretamente na temporada 20/21, onde o departamento passa por uma transformação do seu modelo de trabalho, e evolucionando, por sua vez, para um formato automatizado.

Seguindo-se a linha temporal, a época 2021/2022 é caracterizada pela implementação oficial e valorização dos dados por parte dos distintos departamentos da organização. Posto isto, o departamento de análise desportiva começa a integrar profissionais com um novo perfil, orientados para o tratamento e análise de dados, representando funções como o desenvolvimento de relatórios automatizados, recorrendo-se a *software* especializados (*NacSport, PowerBi, Visual Studio Code*, entre outros) para essa finalidade (temática abordada no seguinte capítulo).

A aposta por parte da organização, na procura de obtenção de respostas, não única e exclusivamente dependentes da subjetividade do ser humano, obrigou ao aumento considerável de recursos humanos com este perfil tecnológico, sendo possível observar-se duas equipas de atuação distintas:

- Primeira Equipa: Ramón Vázquez, Juan Antonio Guzmán, Taichi Wakabayashi, Adrián García.
- Formação: Daniel Del Valle, Juan Carlos Gordillo, Eduardo Valera Sánchez, David Vega e o Estagiário da Universidade de Coimbra.

Como último ponto, merecedor de destaque, catalogado como o produto da obra, neste caso o projeto da organização, o *Sevilla Fútbol Club*, logrou os seguintes títulos, fruto de todos os fatores detalhados anteriormente do contexto, desde acesso a recursos espaciais de excelência, até recursos humanos qualificadíssimos, visíveis na tabela número 1 e figura 10.

Tabela 1- Enumeração dos títulos conquistados pelo Sevilla Fútbol Club a nível Sénior.

Competição	Títulos
<i>Copa del Rey</i>	1935, 1939, 1948, 2007, 2010
<i>Supercopa de España</i>	2007
UEFA Europa League	2006, 2007, 2014, 2015, 2016, 2020, 2023
UEFA Supercup	2006
Taça Euro-Americana	2013
Campeonato Espanhol	1945-46
<i>Copa del Sol</i>	2015
<i>Copa Antonio Puerta</i>	2007, 2009, 2011, 2017



Figura 10- 6 Taças Europa Leagues pertencentes ao museu do SFC.

Sucintamente, este é o contexto integrante, possibilitador da realização de um estágio curricular, aplicado à análise de dados na modalidade de futebol.

Capítulo III

Planeamento e projetos do departamento de análise desportiva

O presente capítulo é possibilitador de conhecimento de detalhes do departamento anteriormente apresentado, do qual é elemento integrante o estudante da Universidade de Coimbra. Será aprofundado informações, tais como projetos desenvolvidos ao longo da temporada desportiva 2022/2023 por parte do mesmo, bem como quotidiano e ferramentas de trabalho essenciais para o desenvolvimento pleno da área num contexto de elite.

3.1 Planeamento anual

Numa perspetiva inicial, de forma a se compreender as tarefas associadas ao longo da temporada desportiva, posteriormente detalhadas no presente capítulo, deve-se abordar o planeamento anual de trabalhos do departamento de análise do clube.

Associativamente, observe-se a seguinte imagem, denotada de cariz elucidativo do anteriormente mencionado.

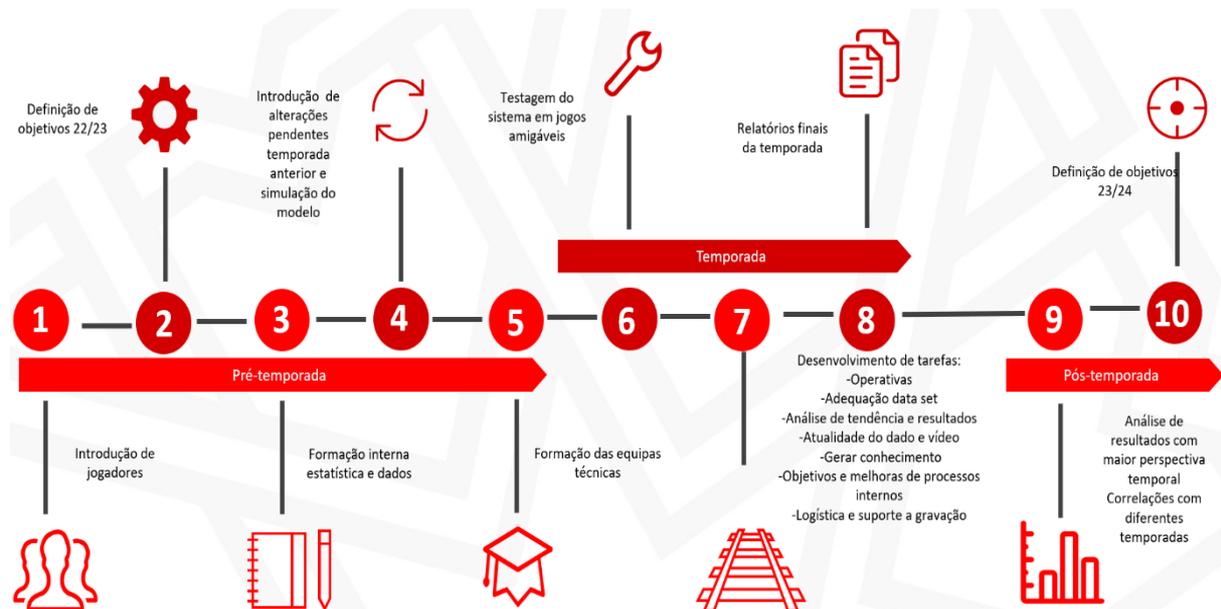


Figura 11- Esquematização temporal do planeamento anual do departamento de análise desportiva.

Como é visível e interpretável na figura 11 é durante a pré-temporada que se encara a maior parte dos objetivos, mais concretamente 5 pontos essenciais e passíveis de referência.

Cronologicamente, inicia-se com um ponto primordial para um dos trabalhos quotidianos dos analistas do clube, com influência direta na análise individual de jogadores, sendo mais concretamente a inserção de novos jogadores na base de dados, com alguns pontos cruciais a serem respeitados a nível de arquitetura de dados, para o seu correto funcionamento:

- Nome;
- Idade;

- País de nascimento;
- *Url* de foto oficial;
- Número de internacionalizações;
- Equipa época transata;
- Equipa época atual;
- Clube de empréstimo;
- ID da plataforma *Football ISM*.

Este primeiro ponto, como se pode observar, funciona como uma tabela de referência para futuras *requests*, conforme as *dashboards* desenvolvidas, como por exemplo, associação ao ID, equipa atual e nome de jogador, todo o *eventing* (definido posteriormente no presente capítulo) que o jogador realize durante os diferentes jogos da presente temporada. Deste modo apenas é necessário a pesquisa por ID de determinado jogador para se obter dados de rendimento em campo.

Como se pode compreender, possuir uma base de dados com toda esta informação de atletas que são ou foram pertencentes aos quadros da formação do *Sevilla Fútbol Club*, tem interação direta com um dos *software* próprios desenvolvidos pelo departamento de dados em parceria com o departamento jurídico, intitulado de *Transfer Tracker*.

Esta ferramenta é um serviço tecnológico, cuja finalidade é de consultoria jurídica, onde é possível rastrear-se transferências de futebolistas e solicitar-se o retorno de pagamentos de compensação que não foram efetuados de determinado jogador do qual o clube exerce de seu direito.

Segundo a página oficial do *Sevilla FC*, o *transfer tracker* é a única solução disponível atualmente no mercado, que possui determinadas características, como por exemplo verificar e reivindicar pagamentos retroativos antes que eles expirem.

Um dos seus algoritmos de previsão, estima que existam mais de mil milhões de euros em compensações não reclamadas, que os clubes com esta ferramenta podem reivindicar. Observe-se a seguinte imagem (figura 12) do excerto da publicação do clube sobre o desenvolvimento do *software*.

Como é interpretável na figura, através do *transfer tracker*, o *Sevilla FC* identificou mais de 700 movimentações de jogadores que pertenceram aos quadros do clube, desde formação até primeira equipa, correspondentes a 53 competições diferentes.

Este rastreamento, originou por parte do *Sevilla FC*, a recuperação de mais de 1.000.000 de euros em pagamentos associados ao mecanismo de solidariedade.

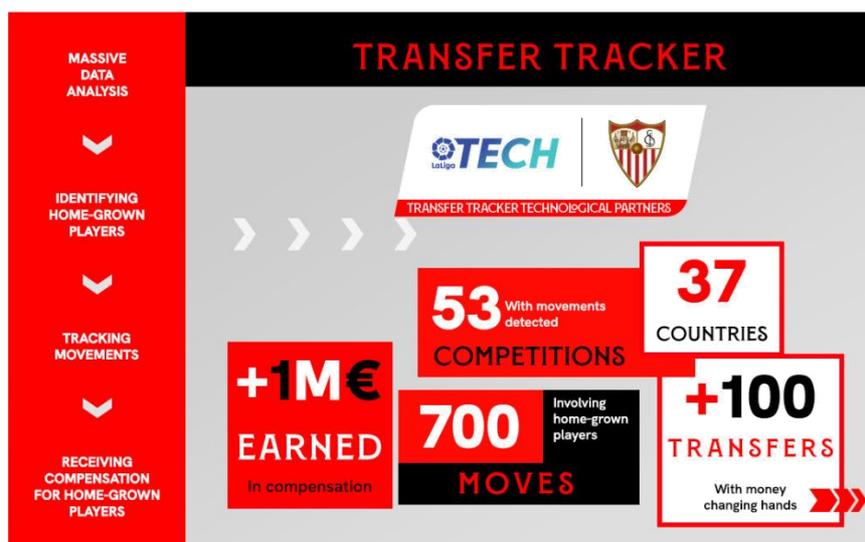


Figura 12- Publicação do SFC nas páginas oficiais sobre a ferramenta desenvolvida internamente e comercializada à LaLiga, *Transfer Tracker*.

Desta forma, e interligando-se com a base de dados de jogadores da formação, consegue-se detalhar as percentagens a receber de determinada transferência de jogadores até ao final da sua carreira como jogador, conforme o seu percurso na academia no clube.

Como segundo ponto do cronograma de planeamento anual do departamento, apresenta-se a definição de objetivos. Associativamente, é neste momento que se realiza uma reflexão da temporada anterior e se idealizam projetos prioritários a desenvolver durante a presente temporada, como por exemplo, um *software* associado ao departamento de recursos humanos. O objetivo da plataforma é realizar a monitorização das cláusulas associadas aos contratos dos jogadores, servindo de exemplo, minutos de jogo, número de convocatórias, número de jogos a titular, entre outros, facilitando os respetivos pagamentos de cada atleta, com o objetivo de diminuir a margem de erro e respetivo impacto financeiro na entidade desportiva.

Na perspetiva do estagiário, são os pontos 3 e 5 os mais importantes do funcionamento do departamento, onde ambos estão diretamente associados à capacitação dos profissionais do clube.

Numa primeira fase, existe uma capacitação dos recursos humanos, considerados como intervenientes diretos na área de análise de dados, através de conhecimentos mais estatísticos e novas ferramentas de trabalho. Consequentemente, e de forma sequencial, procura-se instruir os corpos técnicos, por exemplo, elucidando as novas métricas de rendimento coletivo desenvolvidas durante a temporada passada e implementadas na nova temporada, como foi o caso da iniciativa de jogo.

Esta procura de transferência de conhecimento é uma das formas de trabalho mais eficazes da organização, onde possibilita que os diferentes intervenientes tenham formação interna e consigam num futuro, usufruir de todas as ferramentas disponíveis, objetivando-se subsequentemente a colocação dos seus conhecimentos a prol da “empresa”.

No período apresentado no cronograma como “Temporada”, os pontos intrinsecamente associados são de cariz de desenvolvimento dos projetos previamente definidos, bem como a entrega de relatórios trimestrais com a avaliação de *performance* tanto individual como coletiva de cada equipa de futebol de 11, da organização e, inclusive, rendimento do próprio treinador tendo em conta a organização e estrutura da etapa do atleta.

Se realizarmos uma associação direta à organização das competições por escalão etário, obter-se-ia os dados apresentados na tabela número 2.

Sucintamente, estes são os escalões de formação que possuem acesso à informação que será apresentado no presente documento, fruto do projeto desportivo do clube, que procura ter o máximo de informação possível de todos os atletas, desde o primeiro escalão de futebol de 11.

Tabela 2- Correspondência de etapas de formação da modalidade de futebol de acordo com a idade cronológica.

<i>Infantis</i>	Iniciados
<i>Cadetes</i>	Juvenis
<i>Juveniles</i>	Juniores
<i>Equipo C</i>	Sub-23
<i>Equipo B</i>	Equipa B

Como é possível interpretar-se, a fase final da temporada (pontos 9 e 10) revela-se por uma fase de reflexão sobre os projetos desenvolvidos e objetivos alcançados, denotando que o cronograma se apresenta como cíclico, iniciando desta feita, uma nova temporada com uma cronologia semelhante a nível estrutural.

3.2 Planeamento semanal

Seguidamente após apresentação da estrutura organizacional do departamento a nível macro, com o objetivo de seguir-se uma linha de raciocínio, conhece-se a realidade do fluxo semanal de trabalho de um analista do *Sevilla FC*, expressa na seguinte tabela, onde se pormenoriza uma semana comum, com jogo de sábado a sábado (cenário hipotético).

Tabela 3- Representação do Horário Oficial Semanal de um analista do SFC.

SEVILLA FC Horário Oficial							
	segunda-feira	terça-feira	quarta-feira	quinta-feira	sexta-feira	sábado	domingo
Manhã	<i>Tags dos Events</i> Individuais durante o jogo.	<i>Tags</i> do modelo descritivo de jogo.	Análise de dados e de vídeo de ações individuais e coletivas das duas equipas.	Reuniões com a equipa técnica das equipas profissionais.	Desenvolvimento das novas métricas; novos livros de <i>Power BI</i> ; novos projetos de outros departamentos; etc.	Gravar o jogo e cortar ações estratégicas.	
Tarde	<i>Tags dos Events</i> Individuais durante o jogo.	<i>Tags</i> do modelo descritivo de jogo.	Criação de um relatório final com os outputs dessa jornada.	Reuniões com a equipa técnica de escalões de formação mais inferiores.		Resumo do jogo em vídeo.	

Interpretando-se a anterior tabela, segunda-feira e terça-feira são dias categorizados pela criação de *eventig* próprio. É importante perceber-se o que se entende por *event* e *tag*, visto que este fenómeno é o possibilitador do conhecido KPI (*Key Performance Indicator*), onde vários autores têm apontado a sua utilidade no meio futebolístico, quando combinados com análise de vídeo, como uma excelente forma de obter-se novas informações, cujo objetivo passa por auxiliar as tomadas de decisão dos diferentes intervenientes, onde a análise pode ser focada numa perspetiva mais coletiva ou individual (Ruiz-Ruiz et al., 2013; Król et al., 2017 citados por Herold, et al., 2021).

Desta forma, o *Sevilla Fútbol Club*, numa das suas formações internas, expressa que “*Eventing*” é uma terminologia proveniente da língua inglesa, que faz referência a todos os eventos/acometimentos que ocorrem num determinado desporto. Note-se que são casos exemplificativos de *eventing* simples, a título de ilustração, os passes, remates ou golos.

Por sua vez, o fenómeno de classificar cada ação realizada, por exemplo a nível individual é, comumente, intitulado por *Tagging*.

Este ponto deve-se expressar de extrema importância, devido ao processo de “codificação” estar interligado com a imaginação do analista, desde que as métricas tenham aplicação direta no contexto de inserção. Sendo assim, este processo encontra-se diretamente associado aos primeiros dois dias da semana.

De forma mais concreta, o processo de “Tag” anteriormente explicitado é possível recorrendo-se a um software comumente/ tradicionalmente utilizado para análise de vídeo, mais especificamente *Nacsport Video Analysis Software* (Espanha, 2008), versão Elite.

Intuitivamente, a empresa desenvolveu em parceria com os analistas do *Sevilla FC*, uma janela de codificação de ações específicas, com o objetivo de possibilitar o acesso à informação visível no vídeo de cada jogo das equipas da formação (*event data*), devido, à inexistência de empresas que realizam este tipo de trabalho especificamente em escalões inferiores.

Deste modo, a sua tradicional utilização para corte de ações individuais ou coletivas em vídeo, passou a integrar *data*, sendo a título de exemplo, as coordenadas de um passe, ou de um remate.

Com o objetivo de elucidar a janela de código, observe-se uma janela hipotética (como se deve compreender estamos perante uma ferramenta única e exclusiva) aproximada à do quotidiano dos analistas da organização, seguidamente.

É visível na figura 13, que existem duas janelas de codificação de ações, mais detalhadamente, à esquerda uma janela descritiva do modelo de jogo apresentado pela equipa, e à direita uma janela de análise individual.

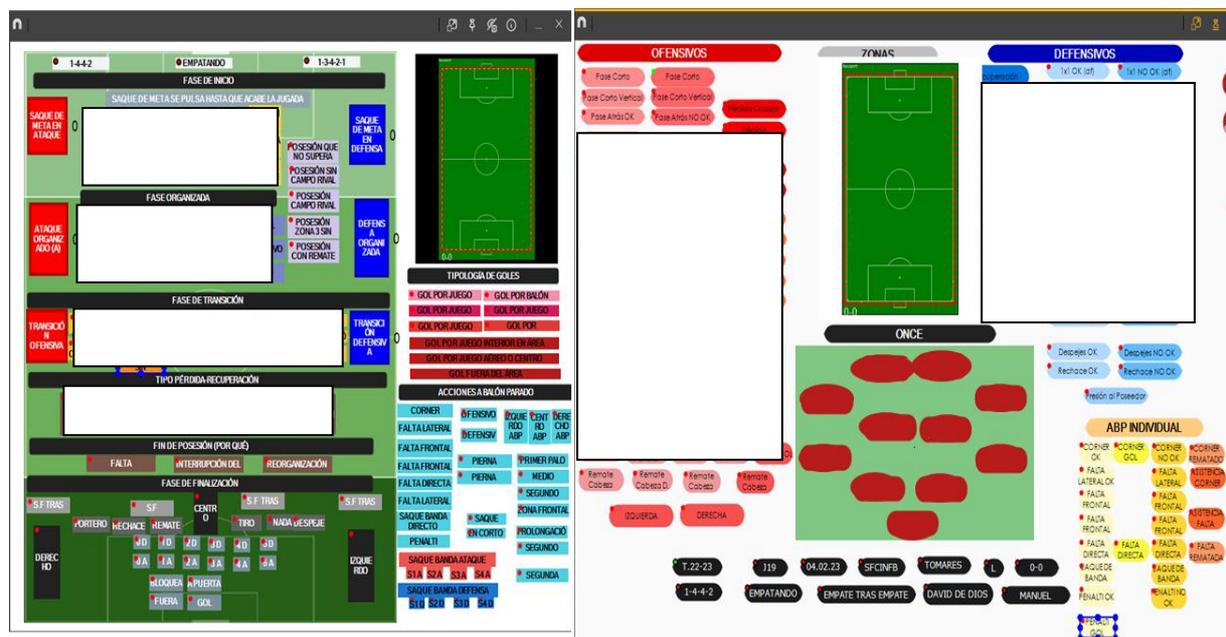


Figura 13-Representação gráfica da janela de codificação de ações individuais e coletivas utilizadas no Nacsports.

A título de demonstração, irá explicitar-se o funcionamento da janela de código à direita (funcionalidade de análise individual), onde é possível visualizar-se os nomes dos jogadores (a vermelho), no painel intitulado como “Once”.

Posteriormente é visível o campo, com coordenadas cartesianas de dois eixos associadas, apresentando X uma variação de 0 a 68, e por sua vez, a coordenada Y de 0 a 100. Em último, é representado o *Event* (anteriormente explicitado), sendo neste caso, as ações realizadas no terreno de jogo, variando de ações ofensivas a vermelho e ações defensivas a azul.

A especificidade e o detalhe que poderá ser apresentado no documento final com o *event data* de um jogo de futebol é diretamente associada às necessidades do contexto e público-alvo. A título de ilustração, o mecanismo de *Tag* seria realizado pela seguinte ordem cronológica:

1. Selecionar o jogador que realizou a ação;

2. Seleccionar a zona do campo onde se efetuou a mesma;
3. Categorizar a ação (por exemplo, remate intercetado);
4. Especificação da ação em caso de aplicação (por exemplo, o remate foi efetuado com o pé direito).

Por outro lado, a janela de codificação, visível à esquerda é utilizada com a finalidade de descrever, como anteriormente mencionado, o modelo de jogo do treinador/equipa. Esta, apresenta alguns campos, como por exemplo:

- “Saque de meta” (pontapé de baliza), onde é possível caracterizar-se a equipa oponente consoante a sua pressão e a forma que ultrapassamos a pressão;
- “Ataque organizado” (Organização Ofensiva), com a possibilidade de se categorizar a zona onde se iniciou, o corredor e como terminou (por exemplo, em finalização, e posterior detalhe da mesma);
- “Transición Ofensiva” (Transição Ofensiva), onde possui *events* categorizadores, como pressão pós perdida, e se a equipa realizou um contra-ataque ou se optou por manter a posse de bola, após a recuperação da mesma.
- “Fase de Finalización” (Fase de Finalização), com o objetivo de se categorizar a finalização existente, apresentando-se detalhes como corredor, se foi através de um *key pass*, ou um centro, bem como, categorização do destino do remate.

Intuitivamente, as “categorias” de codificação da descrição das ações coletivas que são visíveis no jogo, demonstram-se como adaptáveis ao cenário de utilização futura. Como a janela de codificação individual, as métricas a obter-se no ficheiro final, sendo neste caso, exportável em formato Excel, visível na figura 14, são dependentes da imaginação e necessidades do analista e restantes membros integrantes do fenómeno desportivo.

NR	Categori	Inicio	Click	Fin	XY	Des 1	Des 2	Des 3	Des 4	Des 5	Des 6	Des 7	Des 8	Des 9	Des 10	Des 11	Des 12
1	MARCO	00:39	00:43	00:48	48:24	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:24	Duelo Aéreo Frontal	OK		ALTA
1	CRISTO	00:41	00:45	00:50	52:31	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:29	Pase Largo Vertical	NO OK		ALTA
2	MARCO	01:08	01:12	01:17	48:25	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:24	Duelo Aéreo Frontal	OK		ALTA
1	PABLO	01:28	01:32	01:37	9:50	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:26	Pase Corto Vertical	NO OK		ALTA
2	PABLO	01:29	01:33	01:38	24:58	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:212	Pérdida			ALTA
1	ANGEL	01:34	01:38	01:43	19:42	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:27	Anticipación	NO OK		ALTA
1	ISMAEL	01:42	01:46	01:51	47:21	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:24	1x1 NO OK (df)			ALTA
2	ANGEL	01:44	01:48	01:53	35:19	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:23	Bloqueo centro-tiro	OK		ALTA
1	MARCO	02:04	02:08	02:13	25:7	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:22	Pase Corto Horizontal	OK		ALTA
1	MATIAS	02:07	02:11	02:16	35:8	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:23	Pase Largo Vertical	NO OK		ALTA
3	MARCO	02:13	02:17	02:22	51:38	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:29	Falta Realizada			ALTA
2	MARCO	03:12	03:16	03:21	24:7	J22	30:01:22	SFCO	POZOBLA 1-1	V	14-4-2	T:21-22	EMPATAN:22	Pase Corto Horizontal	OK		ALTA

Figura 14- Representação ilustrativa de um ficheiro exportável de *event data*.

Como é interpretável, na figura anteriormente apresentada, possuímos dois momentos integrantes; (1) o primeiro à esquerda, que apresenta as opções de formato de exportação do *Nacsport Video Analysis Software*, e; (2) à direita, o ficheiro exportado de um jogo da equipa C, do *Sevilla FC*.

Analisando-se o ficheiro de forma detalhada, é visível a coluna intitulada como “Categoria” que corresponde neste caso, ao nome do jogador que efetuou a ação. Subsequentemente, todas as colunas, integrantes da mesma fila do documento, são a caracterização da ação que realizou o jogador, como por exemplo as coordenadas (coluna “XY”), a ação “Duelo Aéreo Frontal Ok” na coluna “Des 11”, associadas a variáveis de contexto, como por exemplo estrutura tática nesse momento aquando da ação, ou o tipo de dificuldade do jogo, para futura análise de performance coletiva e individual, agrupados por nível de dificuldade.

O algoritmo de dificuldade é um dos projetos desenvolvidos pelos analistas do Sevilla FC, que possui informações, como por exemplo se o jogo é como local ou visitante, jornada da temporada (jogos no final da temporada normalmente são de maior dificuldade) entre outros pontos que apresentem relação direta e indireta com a dificuldade do encontro. Como se deve compreender, não é possível entrar-se em maior detalhe sobre o algoritmo, devido a ser matéria confidencial do clube.

Um aspeto relevante a ser enfatizado refere-se ao registo em vídeo do jogo. É expectável que para se obter o *Event Data* do jogo, seja necessário possuir-se o elemento audiovisual do mesmo. Desta forma, a organização possui câmaras com inteligência artificial, cujo algoritmo de *machine learning* aplicado a imagens e objetos como é o caso de *Convolutional Neural Network* (CNNs).

A sua definição técnica, segundo Yu An et al. (2020), as recentes variações e evoluções do algoritmo têm proporcionado *performances* extraordinárias em várias tarefas de visão computacional. De forma geral, os modelos CNN recebem *pixels* de imagem como entrada e utilizam múltiplas camadas “convolucionais” e camadas de *pooling* para capturar características da imagem.

Simplistamente, o objetivo do algoritmo integrante no *software* da câmara passa pela identificação da posição da bola em cada *frame* de vídeo, que em seguida, estima a sua posição nos *frames* subsequentes, permitindo que a câmara seja ajustada automaticamente para manter a bola centralizada na filmagem.

O acesso a este tipo de ferramenta possibilita a liberação de um recurso humano, neste caso o analista responsável pela gravação tanto de treino e jogo, para a realização de outras

tarefas de maior preponderância e impacto na organização. Detalhadamente, o equipamento de gravação é da empresa *VEO Sports Camera (Copenhagen, Denmark, 2015)*, umas das mais populares, apresentando como grande parte do seu público alvo a transmissão de jogos recreativos de diferentes modalidades desportivas.

Observe-se de seguida uma figura ilustrativa da câmara e respetivo *software* (figura 15). Como é interpretável a própria empresa disponibiliza uma *cloud* para o armazenamento dos ficheiros de vídeo, que no caso do *Sevilla FC*, é utilizado uma pasta para cada equipa, inserindo-se na mesma, a gravação de todos os treinos e jogos de cada equipa. O acesso à *cloud* como é intuitivamente expectável, é apenas a contas eletrónicas profissionais pertencentes à organização.

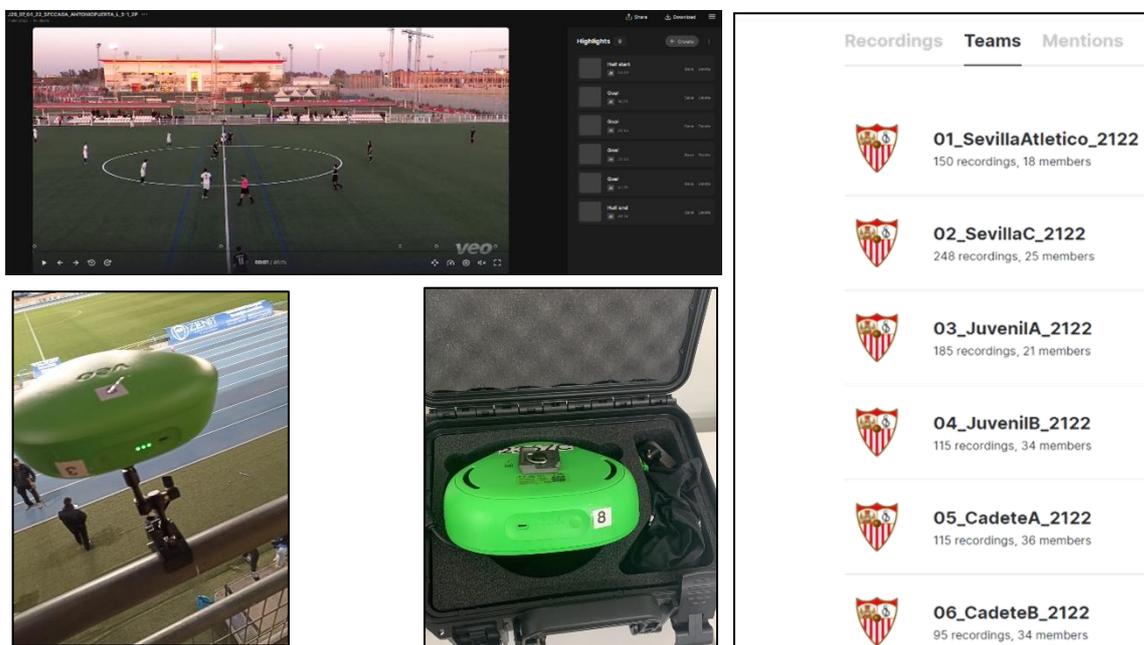


Figura 15- Ilustração do equipamento de gravação VEO e respetiva *cloud*.

Desta forma, o treinador ou pessoas que possuem acesso à respetiva pasta, poderão ter acesso ao vídeo de jogo ou treino após inserção do respetivo equipamento na sua caixa, após o *upload* do mesmo na *cloud* da empresa.

Globalmente, todo este processo, desde o registo da gravação do jogo até codificação das ações coletivas e individuais, são categorizados, figurativamente, como a bola num jogo de futebol. Neste caso, sem acesso ao elemento audiovisual da partida, não conseguimos o *event data*, e por sua vez, sem *event data* não se consegue realizar a última fase do processo, mais concretamente a análise de dados.

Cronologicamente, este processo de análise de dados através da visualização de *dashboards* desenvolvidas em *PowerBi* (grupo *Microsoft*, Washington, United States, 2015), correspondem ao terceiro dia semanal das atividades do analista.

A estratégia empresarial do clube, em relação ao departamento, é a realização da análise de dados do jogo ser realizada pela pessoa que o codificou, devido ao conhecimento prévio do mesmo, possibilitando um futuro enquadramento aos dados. Serve de exemplo, um lateral que apresente um maior número de centros em comparação à sua própria média, mas que por sua vez, está diretamente relacionado com a mudança de estrutura para 3-5-2 e o mesmo, apresenta uma maior profundidade de jogo. Este tipo de contexto apenas é possível através do conhecimento prévio da estratégia de jogo bem como a visualização e análise do jogo.

3.3 Tratamento, Visualização e Análise de dados

O maior impacto do novo recurso humano junto dos clubes de futebol, diretamente ou indiretamente ligado à equipa técnica é, especialmente, a vertente de análise de dados e respetivas conclusões aplicáveis às inquietações do quotidiano.

Para tal, o analista de dados deve possuir capacidades técnicas e teóricas de tratamento, interpretação e visualização de dados. Por outro lado, como se trata uma área inserida no contexto desportivo, sendo neste caso no futebol, é preferencial também um conhecimento teórico e prático da modalidade, objetivando-se que com este, seja possível a comunicação no mesmo idioma entre os diferentes intervenientes da área.

Em título de ilustração, o antigo *Head Data Scientist* do Futebol Clube Barcelona, Raúl Peláez, no Sports Data Fórum 2023, organizado pelo Sevilla FC em colaboração com a *Sports Data Campus* (empresa especializada na formação de temáticas tecnológicas), abordou um tema como o anteriormente referenciado, mais concretamente a falta de conhecimento da modalidade no perfil de analista/cientista de dados. O mesmo, aprofundou, que o treinador do FC Barcelona na época, procurava obter dados das vigilâncias da sua linha defensiva em organização e transição defensiva. Simultaneamente a equipa de analistas de dados acabava de aumentar o seu número de recursos humanos, recaindo sobre o novo elemento este novo projeto.

Inequivocamente, o perfil do profissional contratado apenas possuía conhecimentos técnicos e teóricos da área tecnológica, o que originou por sua vez, uma questão por parte do profissional para com o seu chefe de departamento, mais concretamente o que são “vigilâncias defensivas”.

A solução encontrada pelo *Head* do clube foi que o mesmo teria de integrar os treinos da equipa B, desde presença em palestra a exposição de balneário em dia de preparação de jogos, com o objetivo de conhecer as dinâmicas e inquietações da realidade da modalidade.

Sumariamente, a título de exemplo, uma métrica avançada popularmente utilizada por muitas empresas de consultadoria de dados é *Expected goals (xG)*. A métrica recorre a um algoritmo de previsão, que tem em conta variáveis como o ângulo de remate, pé de remate (dominante ou não), distância de remate, entre outras.

Sam Green (2012), apresentou um artigo pela sua empresa OPTA, uma referência a nível mundial na área estatística aplicada ao futebol, onde analisava a eficácia de finalização de Luís Suarez, na sua temporada de 2011-2012 com a camisola do Liverpool FC. Para tal abordou a métrica avançada de xG para avaliar o seu rendimento, dizendo que o jogador teve oportunidade de fazer o dobro dos golos, se interpretássemos o xG, apenas a sua conversão não tinha sido eficaz.

Correlacionando a notícia com o anteriormente apresentado pelo *Head* do FC Barcelona, o analista com poucos conhecimentos da modalidade utilizaria esta métrica para caracterizar a eficácia de finalização dos avançados da equipa, originando um relatório com este tipo de informação.

Um dos grandes pilares da atuação do analista de dados é responder a perguntas (não a perguntas para consumo próprio), dos intervenientes diretos, neste caso o treinador. Sendo assim, que impacto tem no treino o xG? O treinador irá alterar os conteúdos de uma unidade de treino de finalização pelo seu jogador ter apresentado baixo rendimento na variável golos em comparação ao esperado? A resposta é definitivamente não. É necessário conhecer-se o contexto, esta métrica por exemplo, terá muito mais influencia para diferenciar jogadores da posição avançado no departamento de *Scouting*, ajudando a tomada de decisão.

Contudo, não se denota com extrema importância ou impacto no quotidiano do treinador este tipo de variável que tanto observamos consultoras ou jornais desportivos a referenciar sobre o resultado do jogo, como a título de exemplo, a partida terminou 3-0, mas a variável xG demonstra que deveria ser empate 1-1. No final do dia, a equipa técnica continua com o mesmo problema, mais concretamente, a perda de 3 pontos no último jogo.

Desta feita, simplistamente, as métricas devem ser respostas às inquietações do cenário do qual o relatório será meio integrante.

Dentro deste ponto, as variáveis do *Sevilla FC*, apresentadas nas visualizações que serão detalhadas posteriormente no presente capítulo, são fruto de debate e interrogações das equipas técnicas face ao seu quotidiano como treinadores.

A título de ilustração, uma inquietação de um treinador da equipa da etapa cadete do clube era a procura de percepção de rendimento quando um jogador se encontrava a jogar pelo corredor lateral esquerdo ou direito, devido ao pé dominante.

Uma das *dashboards* que foram desenvolvidas ao longo da época procurava debater este ponto (figura 16), onde foi possível retirar-se elações como a eficácia de cruzamentos em zona 3 (último terço do campo em momento ofensivo), um jogador que preferencial e subjetivamente encontrava-se mais confortável pela esquerda, apresentava melhores números de êxito pela direita. Com esta ferramenta foi possível a equipa técnica ajustar a estrutura da equipa para se conseguir responder a esta nova necessidade sendo que, conseqüentemente, a equipa começou a expressar um número superior de cruzamentos e cruzamentos finalizados por jogo.



Figura 16- Dashboard desenvolvida em PowerBi representativa da eficácia de atuação em determinado corredor (esquerdo, central, direito).

Outra janela de visualização, passível de referência é a *dashboard* de comparativa de rendimento. Este “livro” de *PowerBi* procura responder a questões como por exemplo, qual o rendimento de determinado jogador neste jogo, em variáveis aplicáveis ao perfil de cada posição em comparação à média da sua posição. Basicamente, estamos perante um documento de avaliação da performance individual do jogador em comparação a um valor de referência de cada variável.

O seu público-alvo poderá ser de um espectro desde equipa técnica até direção desportiva, tendo utilidade no caso deste último, a título de ilustração, auxiliar a tomada de decisão de renovação de contratos. Veja-se a figura 17, de cariz elucidativo das visualizações criadas a presente temporada pela entidade desportiva:

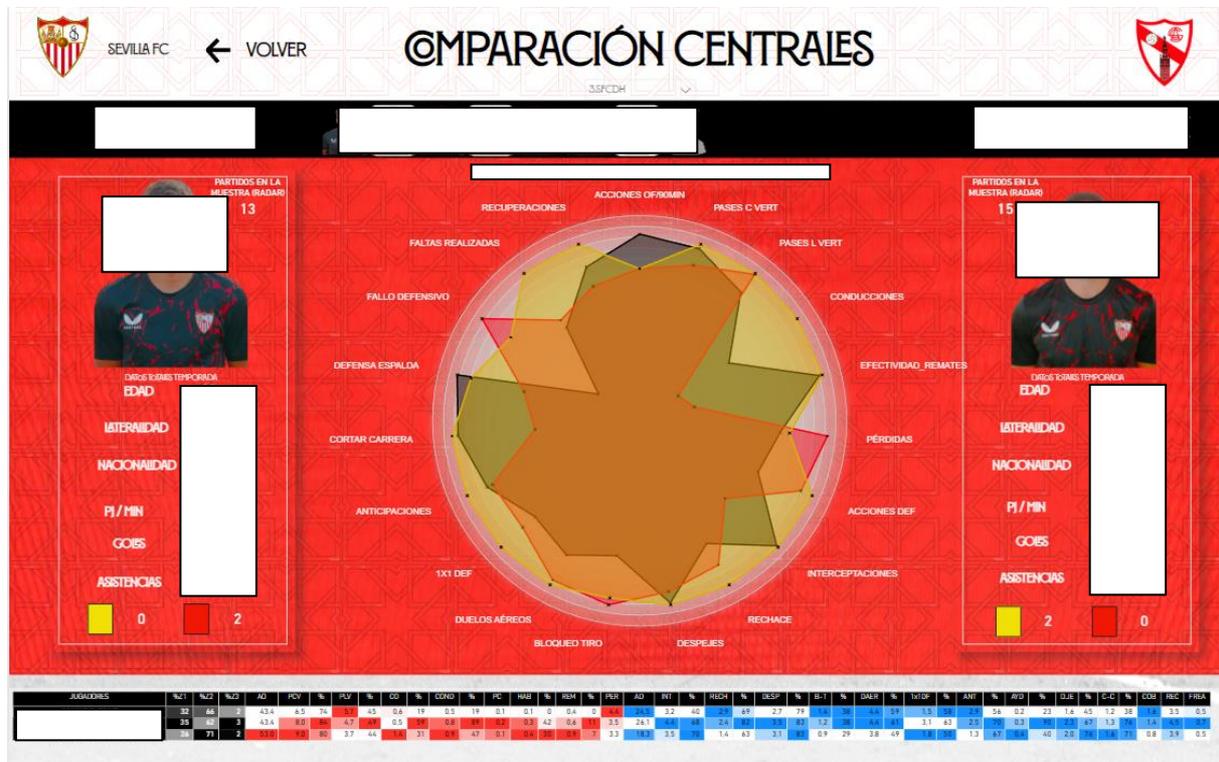


Figura 17- Dashboard desenvolvida em PowerBi com a finalidade de comparação de performance individual entre atletas.

Com esta ferramenta, consegue-se de forma intuitiva e rápida aceder a informações como é visível no gráfico de radar, que o jogador A, em comparação ao B e à média da sua posição é forte na variável condução de bola.

Um dos últimos projetos a ser desenvolvido pelo estagiário durante a sua inserção como membro do clube foi a realização de um mapa de identificação de padrões conforme os diferentes momentos de jogo, interpretável na figura 18.

A forma de criação e alimentação das visualizações serão abordados posteriormente no ponto 3.4 do presente capítulo, onde será detalhado a arquitetura de alguns projetos desenvolvidos pelo analista da Universidade de Coimbra.

Relativamente ao projeto de identificação de padrões de jogo, o mesmo foi desenvolvido através de um script de python e integrado na ferramenta de visualização de PowerBi.

De forma a elucidar o script desenvolvido de visualizações de dados em python, saiba-se que as bibliotecas utilizadas para o seu funcionamento foram:

1. *Scipy*, foi utilizada para o processamento da imagem para suavizar a mesma, bem como reduzir o ruído e melhorar a qualidade a sua qualidade;
2. *MplSoccer*, cuja utilidade foi a representação do campo, que por sua vez foi personalizado em código, desde cores até espessura das linhas;
3. *Matplotlib*, recorreu-se para a criação do gráfico.

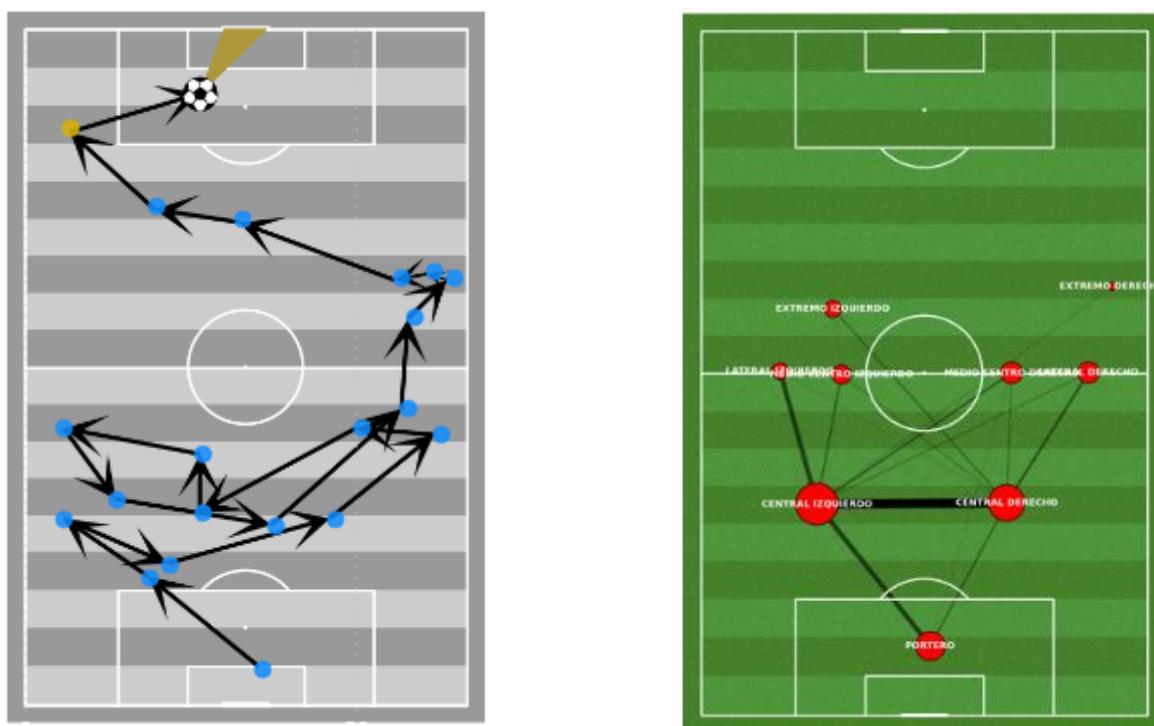


Figura 18- Representação esquemática de jogadas que terminam em gol e esquematização de associação de passes em pontapé de baliza (mapas desenvolvidos de autoria própria).

Posteriormente, após importação das bibliotecas, foram criados os vetores, neste caso as coordenadas do jogador que realiza o passe e as coordenadas do jogador que recebe o passe.

De forma de ilustração, a imagem da esquerda foi utilizada para caracterizar padrões de finalizações, como por exemplo uma equipa que apresente um estilo de jogo combinativo, que associações eram visíveis, bem como se estávamos perante padrões para se obter finalizações.

Por outro lado, o esquema da direita, na perspectiva do estudante, denota-se de informação de maior utilidade, sendo mais concretamente, a representação das associações dos diferentes jogadores de campo, dependentes de variáveis de influência tais como, estrutura da equipa, momento de jogo, jornada, entre outros fatores.

Assim sendo, a imagem da direita, que se refere a dados da equipa X, apresenta variáveis condicionais como:

- Momento de jogo: pontapé de baliza;

- Ligação entre jogadores deve ser superior a 4 passes;
- Estrutura: 4-4-2.

Tendo em conta estas variáveis é expressamente visível, uma tendência de utilização de construção pelo corredor lateral esquerdo.

Detalhadamente, a ligação entre o guarda-redes para com o central esquerdo (CE) em comparação com o central direito expressa-se com maior frequência de recurso, que por sua vez, o CE demonstra uma maior utilização de saída pelo lado esquerdo, como é o caso da linha de ligação entre central e lateral do mesmo corredor.

Este tipo de gráfico poderá ter um grande impacto no que à análise de rivais diz respeito, com o objetivo por exemplo, de orientar a pressão em bloco alto aquando do pontapé de baliza, preferencialmente para o corredor lateral direito do rival. Esta orientação de pressão objetiva a recuperação da posse de bola em locais adiantados do terreno, devido à menor utilização do corredor direito para este momento de jogo.

3.4 Projetos Autopropostos

Com a digitalização do clube ser uma das metas do departamento de inovação, os trabalhos autopropostos pelo estudante da Universidade de Coimbra, tiveram especial inserção em automatização de tarefas diárias e desenvolvimento de projetos recorrendo a métricas avançadas com recurso a algoritmos de *machine learning*.

Como estamos perante matéria confidencial serão dados a conhecer dois dos projetos, mais concretamente, *Webscraping* de contexto de cada equipa de futebol de 11 (classificação, número de jogos, número de golos marcados, entre outras variáveis), bem como o algoritmo de identificação de perfis de jogadores, que será por sua vez detalhado na vertente investigacional do presente documento.

O objetivo primordial do projeto de *Webscraping* passa pela automatização de consultas e atualização da informação inserida na base de dados do clube recorrendo a *Webscraping*.

Primeiramente, deve-se entender o conceito de *Webscraping*, que para Ryan Mitchell (2018) é a prática de recolha de dados através de qualquer meio que não envolva um programa, interagindo com uma API por exemplo, (ou por meio humano recorrendo a um website). O tipo de código é comumente integrado por um programa automatizado que faz consultas a um servidor web, que por sua vez solicita os dados, onde por fim, analisa-os e extrai as informações necessárias.

No caso do projeto desenvolvido para o *Sevilla FC*, a sua arquitetura de funcionamento caracteriza-se pelo exposto na figura 19.

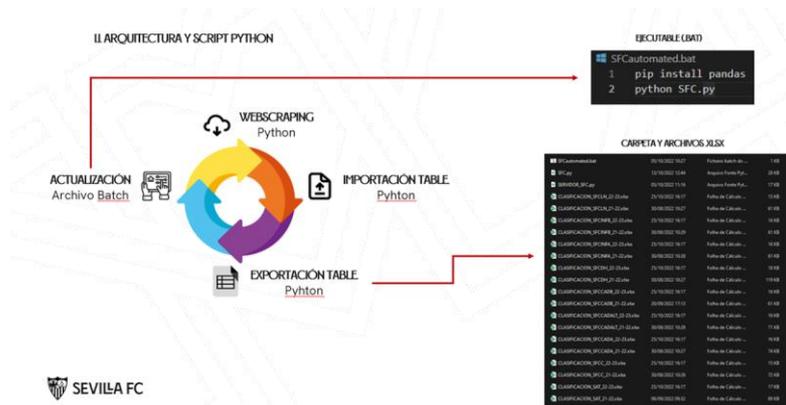


Figura 19- Esquemática da arquitetura do projeto de *Web scraping*.

Interpretando-se a figura anteriormente apresentada, é possível observar-se que se trata de uma arquitetura cíclica, sendo que a forma utilizada para *web scraping*, para importar a tabela anterior e comparar se os dados da consulta são os mesmos integrantes da última versão do ficheiro, que por sua vez, em caso negativo exporta um novo ficheiro em formato .xlsx (Excel), é através de um script em linguagem de programação *Python* (versão 3.11).

Por sua vez, a automatização e ativação do script com uma frequência diária de uma vez, é através de um ficheiro de formato *batch*, um executável para o sistema operacional *Windows*.

O documento exportável, apresenta a seguinte organização, visível na figura 20.

Posición	Equipo	Puntos	PJ	PG	PE	PP	GF	GC	DG	PUNTOS/PJ	GF/PJ	GC/PJ	COMPETICIÓN	TEMPORADA	LOGO LIGA	SITUACIÓN	JORNADA
1	REALBETIS	3	1	1	0	0	3	0	3	3	3	0	3. DH JUVENIL GRUP T.22-23	https://uploac	CAMPEÓN	1	
2	VILLARRE	3	1	1	0	0	3	0	3	3	3	0	3. DH JUVENIL GRUP T.22-23	https://uploac	CHAMPIONS LE	1	
3	ATCOMAF	3	1	1	0	0	3	0	3	3	3	0	3. DH JUVENIL GRUP T.22-23	https://uploac	CHAMPIONS LE	1	
4	REALMAD	3	1	1	0	0	2	1	1	3	2	1	3. DH JUVENIL GRUP T.22-23	https://uploac	CHAMPIONS LE	1	
5	OSASUNA	3	1	1	0	0	2	1	1	3	2	1	3. DH JUVENIL GRUP T.22-23	https://uploac	EUROPA LEAGU	1	
6	REALSOCIA	3	1	1	0	0	1	0	1	3	1	0	3. DH JUVENIL GRUP T.22-23	https://uploac	EUROPA LEAGU	1	
7	VALENCIA	3	1	1	0	0	1	0	1	3	1	0	3. DH JUVENIL GRUP T.22-23	https://uploac	CONFERENCE LI	1	
8	ESPANYOI	1	1	0	1	0	2	2	0	1	2	2	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
9	CELTA	1	1	0	1	0	2	2	0	1	2	2	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
10	MALLORC	1	1	0	1	0	0	0	0	1	0	0	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
11	RAYOVALI	1	1	0	1	0	0	0	0	1	0	0	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
12	ATHLETIC	1	1	0	1	0	0	0	0	1	0	0	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
13	BARCELON	1	1	0	1	0	0	0	0	1	0	0	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
14	PRIMER E	0	1	0	0	1	1	2	-1	0	1	2	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
15	ALMERIA	0	1	0	0	1	1	2	-1	0	1	2	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
16	CADIZ	0	1	0	0	1	0	1	-1	0	0	1	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
17	GIRONA	0	1	0	0	1	0	1	-1	0	0	1	3. DH JUVENIL GRUP T.22-23	https://uploac	PERMANENCIA	1	
18	GETAFE	0	1	0	0	1	0	3	-3	0	0	3	3. DH JUVENIL GRUP T.22-23	https://uploac	DESCENSO	1	
19	ELCHE	0	1	0	0	1	0	3	-3	0	0	3	3. DH JUVENIL GRUP T.22-23	https://uploac	DESCENSO	1	
20	VALLADOI	0	1	0	0	1	0	3	-3	0	0	3	3. DH JUVENIL GRUP T.22-23	https://uploac	DESCENSO	1	

Figura 20- Screenshot exemplificativo do ficheiro extraído após *web scraping*.

Em forma de interpretação, a imagem anterior refere-se à equipa A de juvenis, possuindo informações extraídas de páginas web, como por exemplo: (1) a jornada que se encontra a informação; (2) jogos ganhos, perdidos e empatados; (3) golos marcados e sofridos; (4) entre outras variáveis. Como anteriormente, a visualização de dados é realizada em *Powerbi*. Visualize-se a *dashboard* de consulta correspondente ao mesmo (figura 21).



Figura 21- Dashboard desenvolvida em PowerBi com informações dos dados obtidos através do projeto Webscraping.

Relativamente ao projeto de comparação de atletas (figura 21), o estagiário desenvolveu um passo importante antes de ser possível realizar a comparação entre atletas, comumente intitulado de normalização.

Simplistamente, normalização em *python* é o processo de ajustar os valores de um conjunto de dados para a mesma escala. A título de exemplo, no caso do futebol, o jogador A fez 1,2 remates por 90 minutos e fez 10 passes por 90 minutos. Estamos claramente perante escalas entre variáveis totalmente distintas, o que pode afetar alguns modelos *machine learning* que são sensíveis a escalas. Desta forma, objetiva-se garantir com este processo que os dados não sejam enviesados por diferentes unidades de medida ou escalas. Interprete-se seguidamente a figura elucidativa do passo inicial do processo de normalização (figura 22).

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 100))
```

Figura 22- Ilustração das bibliotecas necessárias para o projeto de normalização de dados.

Como é visível, primeiramente, importamos as bibliotecas que serão essenciais ao longo do *script*. Neste caso, como estamos a tratar de documentos em formato *Comma Separated Values (csv)* e *Xlsx* (extensão de arquivo de *Microsoft Excel*) basta recorrer à biblioteca *pandas* para a sua leitura e tratamento. Seguidamente, o processo relativo à normalização é efetuado através da biblioteca *sklearn*, mais especificamente utilizando o “recurso” *MinMaxScaler*, que neste caso em concreto, foi utilizado uma variação de escala entre 0 e 100.

Este processo, desde seleção de colunas adequadas para a realização da normalização, bem como o valor de variação da escala, foram obtidas através de conclusões grupais de debates internos, tendo em conta a finalidade de utilização do projeto.

Para se compreender melhor, este *script* de normalização, foi a base para a criação do algoritmo de similitude e o *rating* de jogadores.

Intuitivamente, para se realizar um *rating* de jogadores é essencial que todas as variáveis apresentem a mesma escala, para que, num cenário hipotético o jogador X, que possui determinada característica superior, não seja o jogador com melhor *rating*. Deste modo, para se obter o melhor *rating*, o atleta deve apresentar a melhor *performance* (dentro do percentil 90) em todas as variáveis que sejam incluídas para o mesmo, conforme a sua posição em campo. Com este ponto, é perceptível que a normalização foi realizada por posição e por escalão, com o objetivo de esbater futuros enviesamentos dos dados, devido à atuação em diferentes competições, por exemplo.

É de extrema relevância realçar que todas as ferramentas de visualização apresentadas no presente documento são propriedade do *Sevilla FC*, sendo que o seu uso para ilustrações futuras deve ser requerido ao mesmo.

Capítulo IV
Considerações finais

Este capítulo intitula-se, na perspectiva do estudante, como o processo de maior crescimento a nível individual. O mesmo acredita que é através de uma autorreflexão que existe a evolução. Posto isto passam-se a enumerar as ilações retiradas:

- Definição de objetivos de direção e de momento onde deve existir, intrinsecamente conectados, a capacidade de adaptação constante ao longo da época desportiva. Esta foi uma das maiores aprendizagens obtida no contexto profissional que o estudante esteve inserido. O futebol é momento e deve-se planear sobre ele. Assim sendo, muitos objetivos válidos na presente realidade, podem não ser aplicáveis quando se trate da segunda volta do campeonato, por exemplo;
- Sem trabalho de equipa nenhum “barco vai a bom porto”. O estágio foi realizado perante um cenário extremamente anormal, onde o clube apenas ultrapassou as adversidades através da resiliência e espírito de grupo de todos os recursos humanos do mesmo;
- Momentos de clareza são fulcrais. Este ponto é sem dúvida interligado com a autorreflexão. É sobre este, que mesmo em alturas de desespero ou de euforia é possível concluir-se aspetos realmente importantes e retirar as melhores elações dos acontecimentos;
- Os dados são meramente um complemento à tomada de decisão. Esta ideia era assimilada por parte do estagiário, mas como possui um grau elevado de importância é merecedora de destaque. Desta forma, não se deve, em nenhum momento questionar as tomadas de decisão de outros setores/ intervenientes, por ser contra o que os dados demonstram. Conjuntamente, os dados sem contexto não têm utilidade, e relativamente ao contexto, ninguém o conhece melhor que o ser humano que está diretamente associado a ele.
- Consistência no trabalho desenvolvido. A área do futebol é uma área de trabalho árduo de 24 horas sobre 7 dias semanais. É necessário apresentar-se o melhor relatório de dados possível tanto a um domingo como a uma quarta-feira. Este tipo de vivência, da realidade futebolística de um clube de elite foi sem dúvida outra ilação retirada. O quotidiano por sua vez, expressou contextos de pressão constantes, como por exemplo, o cumprimento das deadlines definidas semanalmente para os distintos projetos. Este contacto com a prática profissional foi estritamente recompensador.

Por último, melhor que o estudante realizar uma autorreflexão sobre o seu estágio, é uma reflexão direta, desenhada e atribuída pelo órgão máximo do departamento de inovação, José María Cruz Gallardo, visível na seguinte figura (figura 23).

INTERNSHIP EVALUATION FORM

This form corresponds to two semesters internship mobility.

CONTACT INFORMATION OF THE RECEIVING ORGANISATION

Organisation Sevilla Fútbol Club Department R+D+I

Supervisor Name José María Cruz Gallardo

Supervisor E-mail

STUDENT CONTACT INFORMATION

Name José Miguel Pereira do Carmo Start Date

Phone Number Completion date

E-mail Department Sports Analytics (R+D+I)

QUESTIONS

1. Please evaluate this student internship on the following items

	Excellent	Very Good	Satisfactory	Unsatisfactory	Not Applicable
Arrived to work on-time	X				
Oral communication skills	X				
Written communication skills	X				
Computer skills (software's used)	X				
Ability to work with others	X				
Decision-making, setting priorities	X				
Quality of work	X				
Making and meeting deadlines	X				
Presented work developed autonomously	X				
Demonstrated critical thinking and problem-solving skills	X				

2. Describe the ways in which the intern's performance benefited your organization?
The student has brought scientific rigor to the organization. He has managed to transmit enthusiasm for the work, for improving and has contributed to the discipline in our club

3. What development have you observed in the student's skills, knowledge, personal and/or professional performance?
I think he has improved his ability to interact personally with a professional environment. In addition, he has been able to apply theoretical approaches to reality.

4. Overall, how do you rate your experience with this intern (Excellent, Good, Average, Poor)?
Excellent.

Figura 23- Documento de avaliação oficial do estágio por parte da entidade de acolhimento, SFC.

Capítulo V

O algoritmo de identificação de perfis similares de jogadores de futebol

Abstract

O futebol, como modalidade com maior impacto a nível global foi obrigado a evoluir conjuntamente com a evolução tecnológica, de forma a possibilitar a tomada de decisões mais informadas.

Desta forma, o presente estudo procura fornecer um novo modelo de identificação de perfis de jogadores similares, recorrendo a técnicas de *machine learning*, aplicado à academia de elite do *Sevilla Fútbol Club*, com o objetivo de auxiliar e fornecer mais informação na construção dos plantéis, e evitar assim, os elevados custos associados na contratação de novos atletas.

Relativamente à amostra, foi utilizado um *dataset* com 199 atletas da academia do Sevilla Fútbol Club, distribuídos pelos escalões de iniciados até equipa B, apresentando uma média de idades de $16,73 \pm 2,81$ anos. Apenas foram incluídos para o projeto de investigação os atletas que cumprem os requisitos de terem participado em pelo menos um jogo de dificuldade média ou alta, onde deveriam possuir um mínimo de 50% dos minutos possíveis do respetivo encontro realizados. Foi excluída a posição de guarda-redes para o presente estudo.

No que concerne à metodologia associada ao algoritmo de *machine learning*, não supervisionado de *clustering*, intitulado por *K-means*, realizou-se técnicas de limpeza de dados, normalização das 50 variáveis utilizadas na mesma escala. De forma a encontrar-se o número de clusters ideal para a amostra disponível, recorreu-se à técnica de Elbow Method. Posteriormente, recorreu-se à técnica *cosine similarity* para identificar direção de vetores, visionando a identificação de perfis de atletas, pertencentes ao mesmo cluster.

Os resultados obtidos relativamente à posição médio centro é de 3 cluster. Com este resultado é possível afirmar-se que, para esta amostra, estamos perante 3 perfis de medio centros. Por sua vez, obteve-se que em relação à identificação de similitude entre 3 atletas da mesma posição, Jogador A, B e C, o jogador mais similar ao jogador A é o B, apresentando uma similitude de 82,20%.

Sucintamente os resultados obtidos demarcam-se auxiliares na tomada de decisão, apontando e discutindo um caso prático da academia de elite do sul de Espanha.

Keywords: Clustering, Machine Learning, Scouting, Football, Sports Analytics.

1. Introdução

O futebol, como modalidade desportiva, é o desporto mais aclamado a nível global (Behravan e Razavi, 2020), onde é visível uma transcendência de fronteiras culturais e

geográficas. Esta popularidade manifesta-se através das plataformas digitais, como são exemplo as transmissões televisivas, interação nas redes sociais e eventos ao vivo. Sendo assim, uma das novas preocupações dos clubes, dentro da problemática da digitalização, é a conexão entre adeptos e a organização desportiva, onde as estratégias de envolvimento se torna cruciais para manter uma base de adeptos/ sócios ativa e comprometida.

Alicerçado ao último tópico, existe a procura de sustentabilidade por parte das entidades desportivas. Cwiklinski et al. (2021), afirmam que as ligas e os clubes transformaram-se em indústrias, auferindo milhares de milhões de dólares de várias áreas de influência (de acordo com a dimensão clubística): patrocínios, receitas de bilhética, transferências de jogadores, eventos nas infraestruturas da entidade, direitos de transmissão, *merchandise*, entre outros.

Deste modo, como anteriormente referido, outro ponto de preponderância direta neste fenómeno, são as janelas de transferência de jogadores. Estas janelas não representam apenas um impacto direto na *performance* da equipa em diferentes momentos, fulcrais ao longo da temporada, mas também estimulam discussões apaixonadas e especulações fervorosas entre os adeptos, fenómeno intitulado de *fan engagement*.

Sucintamente, o mercado de transferências não se limita à mera movimentação de ativos e procura de estruturação de um plantel apto para os desafios da nova temporada, como também, possui um impacto direto em muitas outras áreas da organização. Assim sendo, a área de *Scouting* no futebol, foi obrigada a evolucionar juntamente com o progresso tecnológico, procurando desta forma, retirar a subjetividade das contratações, devido ao seu elevado impacto na entidade.

Segundo Calvin (2013, citado por Barron et al., 2018) os clubes de elite expandiram o número de países de “monotorização” global, através da utilização de novos recursos tecnológicos, como por exemplo *software* de análise de vídeo, criação de bases de dados, com o objetivo final de deter mais informação relativamente a diversos jogadores.

Por sua vez, as entidades desportivas, apresentam mais uma fonte de informação, para adicionar às análises subjetivas, possibilitando o aprimorar das suas políticas de recrutamento. Barron et al. (2018), justificam que esta aposta na evolução das práticas de prospeção e identificação precoce de jogadores talentosos, deve-se ao alto impacto que a mesma demonstra na estrutura do clube, a nível profissional.

Intrinsecamente associado ao ponto anterior, no seio de um projeto de clube de futebol sustentável, reside a valorização e promoção dos jogadores formados internamente. Este tipo de projeto centrado na formação do clube não fomenta apenas a perseverança dos ideais da organização, como também estabelece alicerces a longo prazo, mitigando a excessiva

dependência de contratações insustentáveis. Em sentido contrário às contratações, o mesmo possibilita futuras vendas de atletas formados na academia, ambicionando-se o equilíbrio financeiro do clube.

Desse modo, a utilização de recursos tecnológicos mais concretamente a *Artificial Intelligence* (IA) e Machine Learning (ML) tem sido cada vez mais notável na modalidade de futebol, com impacto em distintas áreas.

A título de exemplo, uma das áreas de utilização de algoritmos de ML no futebol são as transferências de jogador. O mesmo apresenta a finalidade de avaliar uma multiplicidade de fatores para determinar valores de mercado mais precisos, reduzindo assim a subjetividade frequentemente inerente a essas negociações. É exemplo de este tipo de trabalho o estudo de Behravan e Razavi (2020), onde apontam que obtiveram e desenvolveram, um novo e inovador método baseado em agrupamento automático e regressão de vetores de suporte otimizada.

Esse algoritmo de ML, desenvolvido pelos autores é capacitado para detetar automaticamente o fator importante para diferentes grupos de jogadores e determinar o seu valor de mercado de acordo com determinadas características.

Sucintamente, o presente artigo, procura ir ao encontro do trabalho anteriormente desenvolvido, onde por sua vez, o objetivo final não passa pela predição de valor de mercado, mas sim, identificação e agrupamento de perfis de jogadores de futebol de forma automática.

2. Materiais e métodos

2.1 Amostra

A amostra foi composta por 199 atletas da modalidade de futebol da academia de elite do *Sevilla Fútbol Club*.

A idade cronológica média foi de $16,73 \pm 2,81$ anos. Na figura 1 é detalhado o número de atletas por equipa, presentes no quadro final da temporada 2022-2023.



Figura 24- Número de atletas por equipa da temporada 2022-2023 a atuar na academia do Sevilla Fútbol Club e pertencentes à amostra do estudo.

2.2 Dataset- Aquisição e Pré-processamento de dados

Recolha de dados

O *dataset* (com dados de *event*) utilizado para a alimentação do algoritmo de *machine learning* é obtido através do software *Nacsport Video Analysis Software* (Espanha, 2008).

O processo de classificação de cada ação realizada com bola (*event*), por parte de cada atleta é realizada de forma manual recorrendo a uma janela de codificação desenhada especificamente para o estudo.

De realçar, que o *event* gerado é executado por 5 analistas experientes (mais de 100 jogos analisados com a ferramenta) e possuintes do grau de nível de treinador da modalidade. Deste modo, a aquisição de dados é obtida de forma autónoma e propriedade do *Sevilla Fútbol Club*.

Seleção de dados e limpeza de dados

O pré-processamento solicitado e aplicado ao *dataset* que contem os dados de todos os jogadores, regeu-se pelos seguintes passos:

1. Aplicação de filtros (atletas que possuam pelo menos 50% dos minutos possíveis da partida jogados; exclusão da posição de guarda-redes devido à falta de dados referentes à posição; inclusão apenas de jogos de dificuldade média e alta);
2. Agrupamento de dados (criação de uma fila única de valores médios por jogador e por métrica);
3. Seleção de métricas (colunas do *dataset*) a utilizar para o projeto (visível na tabela 1);
4. Limpeza de dados (a título de exemplo, substituição de valores vazios pelo valor 0, tendo em conta que, a não existência de uma ação por parte de um jogador é equivalente a 0 realizações).

Deste modo, o número de atletas obtidos após aplicação do filtro, ajustado ao número de minutos jogados conforme a etapa, exclusão da posição que não contem métricas analisadas e seleção da dificuldade das partidas, obteve-se o seguinte número de atletas para o modelo de ML visível na figura 25.

Outro ponto crucial a destacar-se é a seleção de colunas e sua respetiva divisão. Como estamos perante diferentes momentos do jogo que, respetivamente, correspondem a “sub-perfis” do jogador, neste caso a sua performance em momento defensivo e em momento ofensivo, realizou-se uma seleção de colunas e criação de 3 *datasets*, sendo eles:

- *Dataset* Global (sem diferenciação de métricas defensivas e ofensivas);

- *Dataset* Defensivo (seleção de métricas defensivas dentro das colunas disponíveis);
- *Dataset* Ofensivo (seleção de métricas ofensivas dentro das colunas disponíveis).

Com este mecanismo, obteve-se os seguintes conjuntos de dados visível na tabela 4.



Figura 25- Esquematização do número de atletas válidos e incluídos no estudo.

Tabela 4- Representação das variáveis utilizadas nos diferentes *dataset* (Ofensivo, Defensivo e Global).

	Número de colunas	Dataset
Colunas selecionadas		
% de êxito de conduções de bola		
% de êxito de passes largos verticais		
% de êxito de passes largos horizontais		
% de êxito de passes curtos verticais		
% de êxito de passes curtos horizontais		
% de êxito de assistências para remate		
% de êxito de mudança de orientação		
% de êxito de cruzamentos		
% de êxito de disputa aérea ofensiva	26	Ofensivo
% de êxito de disputas ofensivas		
% de êxito de 1 contra 1 ofensivo		
% de êxito de assistências		
% de êxito de remates		
*todas as colunas apresentadas em percentagem, foram selecionadas por sua vez, o seu respetivo valor relativo por 90 minutos.		
Colunas selecionadas		
% de êxito de 1 contra 1 defensivo		
% de êxito de antecipações defensivas		
% de êxito de contenção defensiva		
% de duelos defensivos		
% de êxito de alívios		
% de êxito de cortes defensivos		
% de êxito de bloqueios defensivos		
% de êxito de segunda jogada		
% de êxito de bloqueio de cruzamentos	24	Defensivo
% de êxito de bloqueio de remates		
% de êxito de ajuda defensiva		
% de êxito de disputa aérea defensiva		
*todas as colunas apresentadas em percentagem, foram selecionadas por sua vez, o seu respetivo valor relativo por 90 minutos.		
Colunas selecionadas		
*todas as colunas anteriormente apresentadas do <i>dataset</i> defensivo e ofensivo	50	Global

Algoritmo K-means de machine learning e Cosine Similarity

O *event data*, que são neste caso os dados pertencentes aos 3 *datasets* de alimento ao algoritmo, tem sido a fonte padrão para quantificar e avaliar o desempenho individual e coletivo nas últimas décadas (Lago Peñas & Dellal, 2010; Sarmiento et al., 2014).

De forma global, Lago Peñas e Dellal (2010) apontam que estes dados de *event* são medidas de resultados como frequências, proporções e outros indicadores de desempenho acumulados, relacionados com as ações que ocorrem ao longo de um jogo. Agregadamente, estes estudos têm tido o intuito de identificar padrões táticos num jogo através de modelos de *machine learning* não supervisionada, ou prever o sucesso individual ou coletivo através de abordagens supervisionadas.

Na linha de raciocínio da literatura atual, o algoritmo utilizado para este artigo foi um algoritmo não supervisionado, mais concretamente K-means, com a associação posterior, de uma técnica de identificação de similitude entre dois vetores, intitulado por *Cosine Similarity*.

O algoritmo K-means é categorizado como um algoritmo de agrupamento (*clustering*) não supervisionado (modelo de *machine learning* sem supervisão) que divide um conjunto de dados em K clusters, onde K é definido previamente pelo *data analyst*.

O funcionamento mais preciso do algoritmo, é atribuir automaticamente os pontos de dados aos clusters, com base na distância em relação aos centroides, que são ajustados iterativamente para minimizar a variação interna. Sucintamente, o produto final é um conjunto de pontos de dados similares dentro do mesmo cluster.

Desta forma, para se trabalhar com o algoritmo de *clustering* K-means, é necessário um pré-processamento específico do conjunto de dados, tendo sido efetuado mais precisamente os seguintes pontos:

1. Sklearn -MinMaxScaler

Este ponto procura normalizar os dados na mesma escala de medida com o objetivo de que valores maiores não dominem/condicionem a formação dos grupos/*clusters*, devido à sua maior variação.

No estudo concreto, o número de passes curtos horizontais por 90 minutos pode variar até um valor máximo de 30 (a título de exemplo), enquanto o número de remates é 5. Assim sendo, realizou-se a aplicação da biblioteca *Sklearn*, a função *MinMaxScaler*, com a escala padrão de 0 a 1.

2. PCA

A técnica PCA (Análise de Componentes Principais) é uma técnica de redução de dimensionalidade que pode ajudar a lidar com a alta dimensionalidade de dados, mantendo as informações mais importantes e reduzindo o risco de problemas de dimensionamento no algoritmo K-Means. Simplistamente, o PCA é utilizado para reduzir a dimensionalidade de um *dataset* com muitas colunas/ métricas antes de aplicar o K-Means, como é a título de exemplo, o *dataset* global do presente estudo, com 50 colunas.

Este ponto de pré-processamento permite melhorar a qualidade dos clusters e a eficiência computacional. Este método de otimização de modelos foi também utilizado no trabalho de Gadekallu et al., (2020) com o mesmo propósito.

3. Elbow Method

Este método é associado como um dos mais utilizados para se encontrar o número ótimo de K a atribuir ao algoritmo de *clustering*. O mesmo, foi utilizado como referência ao longo dos anos através do estudo de Thorndike (1953), onde procurou agrupar por cluster 12 postos da força área tendo por base 19 variáveis. Dessa forma, para obter o valor adequado utilizou o Elbow Method, onde apresenta como justificção que o K ótimo como o ponto no qual o número de *cluster* corresponde exclusivamente à configuração de pontos, uma vez que há relativamente pouco ganho com um aumento adicional no número de *clusters*.

Por outras palavras, o ponto de flexão na curva, que é observado neste método, similar a um “cotovelo” (origem do nome), é considerado uma indicação do número ótimo de clusters, pois ele representa um compromisso entre a coesão interna dos grupos e a minimização da redundância. Deste modo, demarca-se nesse ponto que a adição de mais clusters não proporciona uma melhoria significativa na explicação dos dados.

4. Cosine Similarity

Segundo Krotha e Dabiru (2012), o *Cosine Similarity* é uma medida de similaridade entre dois vetores, calculada a partir do cosseno do ângulo formado entre eles. Assim, o cosseno do ângulo entre dois vetores, determina se esses vetores apresentam aproximadamente a mesma direção.

Sucintamente, a escala de medida varia entre -1 e 1, sendo que o cosseno de 0 é 1, o que significa que são totalmente similares no que à sua direção diz respeito, e -1 é interpretavelmente o oposto.

Na aplicação concreta do estudo, aplicou-se a similaridade do cosseno juntamente com o algoritmo de *clustering k-means*. O objetivo primordial passa pela procura de similitude entre jogadores do mesmo *cluster*.

A escolha de utilização deste método em relação ao *Euclidean Distance* (distância entre pontos), deve-se à procura de perfis de jogadores dentro de determinada posição, independentemente do número absoluto obtido em cada variável analisada, tendo em conta que se trata de uma amostra de jovens atletas.

Arquitetura

O processamento e tratamentos de dados, bem como aplicação do modelo de *machine learning*, anteriormente apresentado, foi realizado através de um script em *Python* (versão 3.11), desenvolvido no *Visual Studio Code* (produto Microsoft, Redmond, Washington, EUA).

Associativamente, o *script* em linguagem de programação *Python*, é ativado de forma automática através do gestor de tarefas, com uma regularidade semanal, recorrendo à interface de linha de comando operacional *Windows* (máquina local, com sistema operativo *Windows 11*) para realizar os distintos processos necessários para o projeto de *machine learning (ML)*.

Este processo revela-se autónomo desde a leitura dos diferentes *datasets* de cada jogo, realização da seleção, limpeza e tratamentos de dados, onde por fim, aplica o modelo de *machine learning* e exporta 3 ficheiros, em formato *Excel*, com os respetivos valores obtidos pelo modelo. Seguidamente, a sua visualização e interpretação de dados é realizada numa *dashboard* desenvolvida em *Microsoft PowerBI* (sistema operativo Microsoft *Windows*, 2015).

De forma a facilitar a interpretação da arquitetura do método desenvolvido analise-se a figura 3.

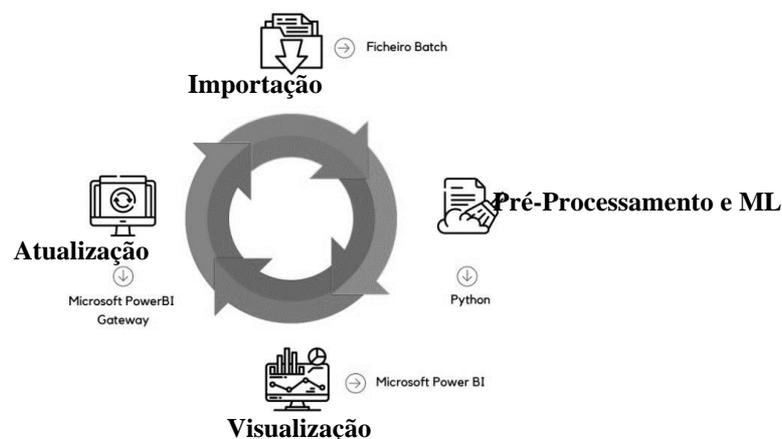


Figura 26- Esquematização da arquitetura do modelo desenvolvido desde importação até atualização.

3. Resultados e Discussão

Como se pode interpretar na figura 4, de cariz elucidativo da técnica *Elbow Method* aplicada ao *dataset*, foi visível que o número de cluster (K) ótimo para os dados do modelo são 3 grupos.

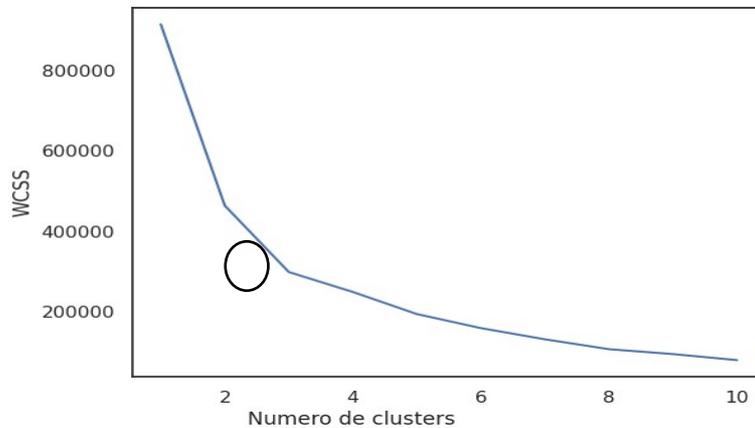


Figura 27- Representação do método Elbow. Número de Clusters por Within-Cluster Sum of Squares (WCSS)

A seleção de um número pequeno recaiu devido ao método, e em simultâneo, pela presença de um conjunto de dados, possuente de um pequeno número de amostras (174 jogadores da academia).

Com este número de cluster, o algoritmo K-means, agrupou os jogadores da posição médio centro em 3 grupos, que são visíveis a sua respetiva distribuição na figura 5.

Note-se que a ocultação dos nomes dos atletas do gráfico de representação dos resultados obtidos pelo algoritmo deve-se à clausula de confidencialidade de dados associada ao projeto desenvolvido.

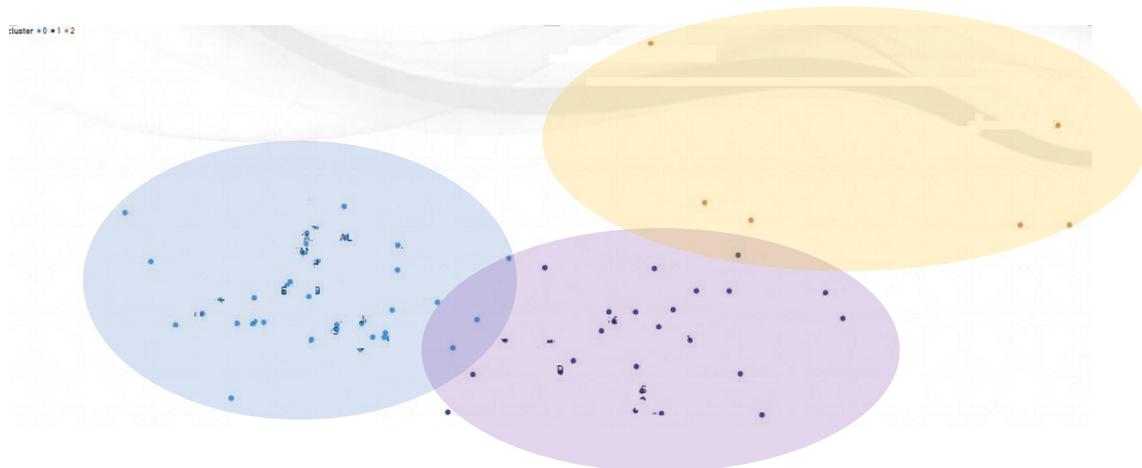


Figura 28- Representação do grupo de cluster desenvolvido em *PowerBi* para a posição Médio Centro.

Como é interpretável, possuímos 2 grandes grupos, de dimensões semelhantes, e um terceiro, que poderiam ser interpretados como limitações do estudo, sendo neste caso, atletas

que apenas possuíam em determinada posição um número reduzido de jogos por exemplo (2 jogos).

Por outro lado, subjetivamente, são atletas que atuaram na posição, mas que apresentam características de outra posição, como um avançado, atuar como um médio centro criativo / falso 9, o que, irá originar este tipo de cluster, sem identificação direta de perfil de médio centro. Esta limitação pode ser alterada e melhorada com a aplicação de um novo parâmetro no processo de processamento e tratamento de dados, definindo-se um número mínimo de jogos que o atleta deve apresentar em determinada posição, para no futuro, quando se procure a sua representação gráfica, não se obtenha um grupo “*Outlier*”.

O número de cluster obtidos para a posição de médio centro vai ao encontro do estudo recente de Li et al., (2022), onde os investigadores obtiveram 3 grupos, mais concretamente *Playmaker*, *Defensive Midfielders (Left and Right)*, e *Wide Midfielder*. O estudo procurava caracterizar o estilo de jogo de jogadores de futebol, recorrendo por sua vez, à técnica NMF (*Nonnegative Matrix Factorization*) aplicada a vetores. Desta forma, a utilização de outra técnica, mais concretamente a *cosine similarity* presente nesta investigação, obteve os mesmos resultados para a posição de médio centro.

A representatividade final, ou objetivo final do estudo passa pelo auxílio na tomada de decisão de promoção de jogadores dentro do clube, após ausência ou venda de determinado atleta. Esta aplicação pode ser efetuada tanto em contexto de equipa B para equipa A, bem como dentro de escalões da formação.

A tabela número 2, é exemplo de resultados obtidos com o método abordado no presente estudo, onde se procura auxiliar na tomada de decisão da substituição do jogador A, que atua como central na equipa de juniores de sub-19. Dentro dos dados disponíveis no *dataset*, e aplicando-se o *cosine similarity*, obteve-se 3 valores de similitude dentro das métricas selecionadas para cada *dataframe* (global, ofensivo e defensivo), apresentando-se apenas o top 3.

Tabela 5- Representação dos resultados obtidos do top 3 de similitude global para substituição do Jogador A.

Jogador referência	Jogador similar	% Similitude Global	% Similitude Ofensiva	% Similitude Defensiva
Jogador A (sub-19)	Jogador B (sub-17)	82,40	84,91	78,69
Jogador A (sub-19)	Jogador C (sub-18)	75,88	75,40	78,98
Jogador A (sub-19)	Jogador D (sub-19)	74,29	70,14	84,32

Como é interpretável, o jogador B a atuar na equipa de sub-17 na atual época, que por sua vez, no próximo ano apresenta idade de sub-18 (etapa júnior da modalidade), é o atleta que

apresenta maior valor de similitude em comparação ao Jogador A. Contudo, é o atleta que apresenta menor similitude a nível defensivo em comparação ao atleta referência.

O Jogador C, demarca-se como o jogador mais preparado a nível de idade cronológica. É um atleta que irá promover de escalão devido à idade, e atuou o último ano com a equipa B da categoria. Contudo a diferença de valores em comparação ao Jogador B apenas apresenta valores superiores de similitude em variáveis correspondentes ao momento de jogo defensivo.

A inclusão por sua vez do jogador D, deve-se ao mesmo ter idade para atuação de um ano mais no escalão, porém apresentou poucos minutos na época passada devido ao seu rival direto Jogador A. O valor mais significativo em comparação ao atleta A é a diferença na similitude defensiva. Desta forma, o parecer deste valor de forma isolada, é que estamos perante atletas que atuam de forma distinta no momento de jogo ofensivo.

Este tipo de análise deve ser acompanhada com mais dados, procurando elucidar-se que tipo de perfil é o atleta A no momento da construção (entre outros fatores), de forma a conseguir ser o mais detalhista possível no momento de aplicação do contexto ao número obtido do algoritmo de similitude.

Sucintamente, o jogador recomendado exclusivamente pelo algoritmo de similitude de direção de perfis, para substituir o Jogador A é o Jogador B.

O recurso à técnica de análise de vetores *cosine similarity* não se apresenta como uma técnica comum de avaliação de similitude entre atletas de futebol. O comumente utilizado para identificação de similitude, aplicado à modalidade de futebol é a *Euclidean Distance*, como é de exemplo o estudo de Cho et al., (2022), onde procurou encontrar a similitude nos diferentes tipos da variável passe, entre jogadores de futebol. Esta técnica é caracterizada, de forma simplista, por calcular a distância entre dois pontos.

Decroos et al., (2019), apontam que a técnica *Euclidean Distance* tende a penalizar injustamente quando existem grandes diferenças em algumas das variáveis do *dataset*. Este ponto é perceptível de interpretação, se imaginarmos 2 atletas, onde o atleta A expressa números elevados nas variáveis golos e assistências, e por sua vez, um número menor em drible. Por outro lado, o atleta B, apresenta um número menor nas mesmas variáveis (valores absolutos), contudo, apresenta a mesma tendência (superior em golos e assistências, e inferior na dimensão drible).

Recorrendo à distância euclidiana estaremos a afirmar que estamos perante jogadores não similares, porque estes dois atletas possuem valores absolutos dispares. Contudo, note-se que os jogadores podem não possuir os mesmos números absolutos por diversos fatores, mas, serem pertencentes ao mesmo grupo relativamente ao estilo de jogo.

Desta forma, o objetivo da utilização do *cosine similarity* passa pela identificação de atletas que possuam as mesmas variáveis como variáveis expressamente superiores às restantes, contrariamente à distância euclidiana, que se compara o número máximo obtidos entre elas.

4. Conclusão

O presente estudo obteve valores fiáveis através de um novo método automático de similitude entre perfis de jogadores, sendo neste caso obtido pela direção de vetores. A amostra, apresenta-se como uma debilidade, onde se deveria testar num futuro próximo, o mesmo modelo, porém com uma amostra com um número mais considerável, de forma a melhorar o algoritmo.

Outra desvantagem do estudo passa pela utilização de dados internos da entidade desportiva, o que impossibilita a reprodutibilidade do estudo para melhoramento do algoritmo.

De sucinta forma, o trabalho possibilita o auxílio na tomada de decisão dos diferentes intervenientes da entidade desportiva no que na temática de construção de equipas diz respeito, retirando alguma subjetividade ao fenómeno de identificação de talentos.

Referências Bibliográficas

- An, Yu., Ye, Q., Guo, J. & Dong, R. (2020). *Overlap Training to Mitigate Inconsistencies Caused by Image Tiling in CNNs*. Springer. Artificial Intelligence XXXVII. 35–48.
- Anderson, C. & Sally, D. (2013). *The numbers game: Why Everything You Know About Football is Wrong 1^a Edition*. Penguin Books.
- Barron, D., Ball, G., Robins, M. & Sunderland, C. (2018). *Artificial neural networks and player recruitment in professional soccer*. Journal PLoS ONE, 13(10). 1-11.
- Behravan, I. & Razavi, S. (2020). *A novel machine learning method for estimating football players' value in the transfer market*. Springer-Verlag GmbH Germany. 25(3). 2499-2511.
- Bergkamp, T., Frencken, W., Niessen, A., Meijer, R. & Hartigh, R. (2022). *How soccer scouts identify talented players*. European Journal of Sport Science. 22(7). 994-1004.
- Cepeda, P. (2020). *Nunca se rinde El club que aprendió a ganar 1^a Edición*. Samarcanda.
- Cho, H., Ryu, H. & Song, M. (2022). *Pass2vec: Analyzing soccer players' passing style using deep learning*. International Journal of Sports Science & Coaching. 17(2). 355–365.
- Cwiklinski, B., Giełczyk, A. & Choras, M. (2021). *Who Will Score? A Machine Learning Approach to Supporting Football Team Building and Transfers*. Entropy, MDPI. 23 (1). 1-12.
- Decroos, T. & Davis, J. (2019). *Player Vectors: Characterizing soccer players' playing style from match event streams*. Springer. Machine Learning and Knowledge Discovery in Databases. 569-584.
- Gadekallu, T., Rajput, D., Reddy, M., Lakshmana, K., Bhattacharya, S., Singh, S., Jolfaei, A. & Alazab, M. (2020). *A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU*. Journal of Real-Time Image Processing. 18(4). 1383-1396.
- Herold, M., Kempe, Matthias., Bauer, P. & Meyer, T. (2021). *Attacking Key Performance Indicators in Soccer: Current Practice and Perceptions from the Elite to Youth Academy Level*. Journal of Sports Science & Medicine. 20(1). 158–169.
- Krotha, R. & Dabbiru, M. (2012). *A Novel Approach for Compact Document Clustering*. International Journal on Computer Science and Engineering. 4(7). 1348-1353.
- Montes, E. (1981). *Historia del Sevilla Fútbol Club*. Biblioteca de Ediciones Analuzas, S.A.

Lago-Peñas, C. & Dellal, A. (2010). *Ball Possession Strategies in Elite Soccer According to the Evolution of the Match-Score: the Influence of Situational Variables*. *Journal of Human Kinetics*. 25(1). 93-100.

Li, Y., Zong, S., Shen, Y., Pu, Z., Gómez, M. & Cui, Y. (2022). *Characterizing player's playing styles based on player vectors for each playing position in the Chinese Football Super League*. *Journal of Sports Sciences*. 40 (14). 1629–1640.

Pinilla, D. (2020). *El método Monchi Las claves del sistema de trabajo del rey midas del fútbol mundial 3ª Edición*. Samarcanda.

Sarmiento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., & Leitão, J. C. (2014). *Match analysis in football: a systematic review*. *Journal of Sports Sciences*. 37(20). 2494-2507.

Soriano, F. (2010). *A Bola não entra por acaso – o que o Futebol tem a ensinar à Gestão*. Gestão Plus.

Thronthike, R. (1953). *Who Belongs in the family?*. *Psychometrika* 18. 267-276.