



**CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E
DESENVOLVIMENTO DE SISTEMAS**

SISTEMA DE ANÁLISE DE SCOUT NO FUTEBOL BRASILEIRO

**SABRINA DO ESPIRITO SANTO MENDONÇA
VINICIUS SILVA CRUZ**

**GUARULHOS
2025**

**SABRINA DO ESPIRITO SANTO MENDONÇA
VINICIUS SILVA CRUZ**

SISTEMA DE ANÁLISE DE SCOUT NO FUTEBOL BRASILEIRO

Trabalho de Conclusão de Curso apresentado ao Curso de Análise e Desenvolvimento de Sistemas como requisito parcial para obtenção do Título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador (a): Prof (ª) Esp. Jane Maria dos Santos Ebersson

GUARULHOS

2025

**SABRINA DO ESPIRITO SANTO MENDONÇA
VINICIUS SILVA CRUZ**

SISTEMA DE ANÁLISE DE SCOUT NO FUTEBOL BRASILEIRO

Trabalho de Conclusão de Curso/ Monografia/ Dissertação/Tese apresentado (a) ao Curso de Análise e Desenvolvimento de Sistemas como requisito parcial para obtenção do Título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Banca Examinadora

Orientador: _____

Prof (ª) Esp. Jane Maria dos Santos Eberson
Fatec Guarulhos

Banca: _____

Titulação + nome completo
Fatec Guarulhos

Banca: _____

Titulação + nome completo
Fatec Guarulhos

Guarulhos, data (28/11/2025)

RESUMO

CRUZ, Vinicius Silva; MENDONÇA, Sabrina do Espírito Santo. **Sistema de Análise de Scout no Futebol Brasileiro**. 2025. Trabalho de Conclusão de Curso – Faculdade de Tecnologia de Guarulhos, Guarulhos.

Este Trabalho de Conclusão de Curso propõe o desenvolvimento de uma ferramenta de análise de *scout* no futebol brasileiro. Utilizando dados estatísticos acessíveis da plataforma *FBref*, buscou-se coletar, organizar e analisar métricas de desempenho de jogadores com o objetivo de auxiliar os clubes no processo de identificação de talentos. A proposta surge como uma alternativa viável aos métodos tradicionais baseados em observação subjetiva, promovendo maior objetividade na análise e seleção de jogadores.

Palavras-chave: futebol; scout; análise de dados; identificação de talentos.

ABSTRACT

This undergraduate thesis proposes the development of a scouting analysis tool for Brazilian football. Using accessible statistical data from the FBref platform, the study aimed to collect, organize, and analyze player performance metrics in order to assist clubs in the talent identification process. The proposal emerges as a feasible alternative to traditional methods based on subjective observation, promoting greater objectivity in the analysis and selection of players.

Keywords: *football; scouting; data analysis; talent identification.*

SUMÁRIO

1. INTRODUÇÃO	7
2. FUNDAMENTAÇÃO TEÓRICA	8
2.1. O Scouting no futebol: conceito e aplicação	8
2.2. A evolução do scouting e o uso de dados no futebol	8
2.3. Scouting baseado em dados	9
2.4. O que são dados?	10
2.4.1. Importância dos dados	11
2.4.2. Qualidade dos dados	13
2.4.3. Ciclo de vida dos dados	14
3. METODOLOGIA	16
4. DESENVOLVIMENTO	18
4.1. Escopo do software	18
4.2. Objetivos do desenvolvimento	18
4.3. Requisitos do sistema	18
4.3.1 Requisitos funcionais	19
4.3.2. Requisitos não funcionais	19
4.4. Casos de uso	20
4.5. Diagramas de sequência	23
4.5.1. UC01 - Consultar lista de jogadores	24
4.5.2. UC02 - Visualizar ficha do jogador	25
4.5.3. UC03 - Visualizar estatística média	26
4.5.4. UC04 - Visualizar estatística geral	26

4.5.5. UC05 - Visualizar partida	27
4.5.6. UC06 - Analisar desempenho técnico	28
4.6. Projeto de interface	29
4.6.1. Tela inicial	29
4.6.2. Ficha do jogador - Estrutura geral	30
4.6.3. Média geral	31
4.6.3. Estatística geral	33
4.6.4. Partidas	35
4.6.5. Análise do jogador	35
4.7. Diagrama de classes	36
4.8. Diagrama de entidade e relacionamento	41
5. RESULTADOS E DISCUSSÕES	48
6. CONSIDERAÇÕES FINAIS	55
7. REFERÊNCIAS	57

1. INTRODUÇÃO

O futebol, ao longo das últimas décadas, tem passado por um processo de transformação impulsionado pelo avanço das tecnologias de informação e análise de dados. Clubes ao redor do mundo vêm investindo cada vez mais em departamentos de análise de desempenho, utilizando ferramentas digitais para extrair informações objetivas que possam subsidiar decisões estratégicas. Essas inovações têm impacto direto sobre o processo de *scouting*, tradicionalmente baseado na observação subjetiva de olheiros experientes.

Contudo, enquanto grandes clubes contam com equipes especializadas e acesso a plataformas robustas, muitos clubes de pequeno e médio porte enfrentam limitações significativas no que diz respeito à estrutura e à capacidade técnica para aplicar essas soluções. A realidade do futebol brasileiro, marcada por desigualdades estruturais, evidencia a necessidade de ferramentas acessíveis que democratizem o acesso à tecnologia no esporte (REIS, 2019).

Nesse cenário, este trabalho propõe o desenvolvimento de um sistema digital que visa coletar, organizar e interpretar dados estatísticos de jogadores de futebol, com o objetivo de apoiar o processo de identificação de talentos por meio de uma abordagem fundamentada em evidências. A proposta busca reduzir a subjetividade nas análises, ampliar as possibilidades de comparação entre atletas e fornecer uma base de informações confiável para auxiliar analistas, treinadores e gestores na tomada de decisões.

Além disso, o sistema permitirá o registro contínuo do desempenho dos jogadores observados, possibilitando o acompanhamento da evolução ao longo do tempo e a construção de um histórico individual útil em futuras avaliações.

Para sustentar e validar o desenvolvimento desta ferramenta, este trabalho adota uma abordagem metodológica compatível com a natureza exploratória da pesquisa, buscando familiaridade com o problema e a proposição de soluções viáveis. A metodologia, apresentada em detalhes no capítulo subsequente, orienta o processo de construção da solução e permite analisar de que maneira o uso estruturado de informações pode contribuir efetivamente para a identificação de talentos no futebol brasileiro.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. O Scouting no futebol: conceito e aplicação

O termo *scouting*, amplamente utilizado no meio esportivo, refere-se ao processo de observação, avaliação e identificação de jogadores com potencial de desenvolvimento. Segundo Bergkamp et al. (2021), a prática tradicionalmente se baseia em avaliações visuais realizadas por olheiros experientes, que observam aspectos como posicionamento, leitura de jogo, capacidade técnica, tomada de decisão e comportamento em campo. Essa metodologia, embora ainda amplamente utilizada, está sujeita a vieses subjetivas e limitações operacionais.

Com o tempo, o processo de *scouting* passou por uma transformação significativa. De uma prática predominantemente empírica e subjetiva, passou a incorporar métodos mais estruturados e objetivos. Hoje, o conceito moderno de *scouting* inclui tanto a análise qualitativa quanto a quantitativa do desempenho dos jogadores. Isso significa que, além da observação presencial, os clubes passaram a adotar métricas estatísticas como número de passes certos, desarmes, finalizações e posse de bola, entre outras (MACEDO; LEITE, 2009).

Essa mudança visa reduzir a margem de erro e ampliar o repertório de informações disponíveis para os tomadores de decisão. Além disso, o *scouting* também pode ser utilizado para analisar o desempenho de adversários e apoiar os treinadores na elaboração de estratégias de jogo. Dessa forma, a prática se consolida como um componente essencial da gestão esportiva atual, integrando dados objetivos com a experiência técnica dos profissionais envolvidos.

2.2. A evolução do scouting e o uso de dados no futebol

Historicamente, a identificação de talentos no futebol, principalmente em contextos de categorias de base, sempre foi fortemente baseada na percepção e intuição dos olheiros (BERGKAMP, 2021). A análise subjetiva, ainda muito utilizada, pode ser limitada por vieses individuais e por dificuldades práticas, como a escassez de tempo, recursos humanos e financeiros — especialmente em clubes de pequeno e médio porte.

No contexto do futebol brasileiro, essas limitações são ainda mais evidentes. Enquanto grandes clubes contam com departamentos estruturados e acesso a tecnologias avançadas, a realidade em muitas equipes menores é que existem muitos desafios básicos, como falta de infraestrutura, logística precária e ausência de profissionais especializados (REIS, 2019). Essas condições dificultam a implementação de processos sistemáticos de análise de desempenho e tornam os métodos tradicionais insuficientes para cobrir com qualidade a vasta quantidade de atletas em formação no país.

Diante dessas limitações, o avanço da tecnologia tem se mostrado uma solução promissora. Tecnologias como GPS, câmeras de rastreamento e softwares de análise vêm sendo amplamente utilizadas em clubes de elite para mensurar variáveis como carga de trabalho, distância percorrida, velocidade, intensidade e padrões táticos (CARLING, 2008). Além disso, o uso de algoritmos e ferramentas de *machine learning* permitem a identificação de padrões de desempenho, oferecendo *insights* baseados em dados históricos e contextuais (CHANDRA, 2024).

No entanto, mesmo essas soluções tecnológicas apresentam desafios, o alto custo dos sistemas comerciais, a ausência de validação científica em alguns casos e a necessidade de profissionais capacitados para operar e interpretar os dados são obstáculos significativos para a democratização dessas ferramentas (CARLING, 2008). Assim, surge a necessidade de desenvolver soluções mais acessíveis, adaptadas à realidade de clubes com estrutura limitada, mas que ainda assim desejam modernizar seus processos de captação e análise.

2.3. Scouting baseado em dados

A incorporação de dados objetivos ao processo de *scouting* representa uma mudança de paradigma na gestão esportiva. Em vez de se basear exclusivamente em impressões subjetivas, o novo modelo de análise busca identificar e comparar jogadores por meio de métricas técnicas (como taxa de acerto em passes e finalizações), físicas (como resistência e velocidade), táticas (como movimentação e posicionamento) e até comportamentais (como regularidade e disciplina) (BERGKAMP et al., 2021).

Plataformas de dados esportivos, como o *FBref*, desempenham papel essencial nesse novo cenário, ao disponibilizar informações estatísticas padronizadas e acessíveis. Tais ferramentas possibilitam que clubes de diferentes níveis tenham acesso a dados detalhados, mesmo sem infraestrutura complexa, permitindo a comparação entre atletas de diferentes regiões e contextos competitivos.

O uso de algoritmos de aprendizado de máquina, por sua vez, amplia ainda mais o escopo do *scouting* moderno. Por meio da análise de séries temporais e perfis históricos de jogadores, é possível identificar tendências de evolução semelhantes às de atletas consagrados, antecipando o potencial de jovens promessas antes mesmo que atinjam sua maturidade técnica (CHANDRA, 2024). Isso não apenas aumenta as chances de sucesso nas contratações, como também otimiza recursos e reduz riscos.

Apesar dessas vantagens, é fundamental reconhecer que o uso de dados deve atuar como ferramenta complementar — e não substitutiva — ao olhar técnico e sensível dos profissionais do futebol. Além disso, questões como padronização, qualidade e interpretação adequada dos dados precisam ser enfrentadas com responsabilidade. Métricas mal definidas ou aplicadas fora de contexto podem gerar diagnósticos equivocados.

Portanto, o *scouting* baseado em dados se apresenta como uma alternativa inovadora e promissora para tornar o processo de identificação de talentos mais eficiente, justo e acessível. Para clubes com recursos limitados, em especial, representa uma oportunidade de competir em um mercado cada vez mais profissionalizado, rompendo barreiras históricas de acesso à tecnologia e informação no esporte.

2.4. O que são dados?

Dado é a representação de fatos, conceitos ou instruções de maneira formalizada, adequada para comunicação, interpretação ou processamento para seres humanos ou até meios automáticos (ISO/IEC 2382:2015, 2015). No entanto, por si só, o dado possui caráter bruto, carecendo de significado até que seja devidamente contextualizado. Quando submetido a processos de organização,

análise e interpretação, transforma-se em informação, a qual assume papel central na geração de conhecimento (FÁVERO, 2017).

Os dados, quando estruturados e interpretados, passam a sustentar decisões, apoiar a resolução de problemas e gerar vantagens competitivas. Assim, enquanto o dado constitui a matéria-prima essencial, a informação resultante é aquilo que efetivamente orienta ações e possibilita a construção de inferências e proposições relevantes no contexto acadêmico, social ou empresarial (FÁVERO, 2017).

Dados podem assumir diversas naturezas e formatos, sendo normalmente classificados como estruturados, semi estruturados ou não estruturados. Dados estruturados apresentam organização rígida, geralmente em formato tabular, o que facilita o processamento por sistemas computacionais. Dados semi estruturados possuem algum nível de organização, como arquivos JSON ou XML, embora não sigam completamente um modelo fixo. Já os dados não estruturados, como textos, imagens, áudios e vídeos, não apresentam padrão definido e exigem técnicas especializadas para serem analisados. Além disso, os dados também podem ser qualitativos, quando descrevem características e percepções, ou quantitativos, quando representam valores mensuráveis e passíveis de tratamento estatístico (FÁVERO, 2017).

A forma como os dados são organizados influencia diretamente a capacidade de convertê-los em informação útil. Esse processo envolve atividades como classificação, estruturação, limpeza, preparação, normalização e armazenamento adequado, considerando o tipo e o volume dos dados. Também pode incluir a indexação, que permite localizar informações de maneira eficiente. Uma organização bem executada garante maior precisão, consistência e confiabilidade às análises subsequentes, fortalecendo processos decisórios e contribuindo para a produção de conhecimento fundamentado (FÁVERO, 2017).

2.4.1. Importância dos dados

Nesse contexto, o mundo entrou em uma nova etapa de desenvolvimento, na qual “o recurso mais valioso não é mais o petróleo, mas os dados” (The Economist, 2017). Essa analogia ao petróleo evidencia que, assim como o combustível fóssil foi o motor da revolução industrial e da expansão econômica do século XX, os dados

desempenham papel equivalente no século XXI, impulsionando modelos de negócios digitais, algoritmos de inteligência artificial e sistemas de tomada de decisão baseados em evidências (The Economist, 2017).

Empresas líderes globais como Google, Amazon e Facebook exemplificam esse fenômeno: seu crescimento exponencial e domínio de mercado decorrem diretamente da habilidade de coletar, armazenar, processar e utilizar grandes volumes de dados. Tais organizações conseguem prever comportamentos de consumo, personalizar ofertas, otimizar cadeias logísticas e até mesmo desenvolver novos produtos a partir da análise preditiva. Dessa forma, os dados configuram-se não apenas como insumos informacionais, mas como ativos estratégicos de altíssimo valor (Provost; Fawcett, 2016).

Contudo, a valorização dos dados traz consigo desafios, questões relacionadas à privacidade, à segurança da informação e à concentração de poder em poucas corporações que controlam grandes volumes de dados levantam debates éticos e jurídicos (MAYER-SCHÖNBERGER, 2014). A regulamentação sobre proteção de dados pessoais, como a *General Data Protection Regulation* (GDPR) na União Europeia e a Lei Geral de Proteção de Dados (LGPD) no Brasil, surge como tentativa de equilibrar inovação e competitividade com os direitos fundamentais dos indivíduos.

Portanto, pode-se afirmar que os dados constituem o **novo petróleo** não apenas pelo valor econômico que representam, mas sobretudo pelo poder estratégico de transformação que carregam. Assim como o petróleo precisou ser extraído, refinado e distribuído para mover a sociedade industrial, os dados necessitam ser coletados, tratados e analisados para impulsionar a sociedade digital. Sua relevância transcende o aspecto tecnológico, alcançando dimensões políticas, sociais e culturais, tornando-os um dos pilares centrais da era digital (The Economist, 2017).

A partir dos dados devidamente coletados, organizados e tratados, torna-se possível gerar informações que orientam processos decisórios com maior precisão e embasamento. Dados brutos, quando submetidos a técnicas analíticas e modelos estatísticos, revelam padrões, tendências e relações que não seriam perceptíveis de forma intuitiva (Provost; Fawcett, 2016). Essas informações permitem que gestores identifiquem problemas, antecipem cenários e escolham alternativas mais eficientes

para alcançar seus objetivos. No ambiente organizacional, essa capacidade de transformar dados em insights estratégicos reduz incertezas, aumenta a competitividade e possibilita a implementação de ações mais assertivas. Assim, a análise de dados deixa de ser apenas um recurso operacional e passa a constituir um elemento central para a tomada de decisões fundamentadas em evidências, fortalecendo processos de planejamento, controle e inovação em diferentes setores da sociedade (Provost; Fawcett, 2016).

2.4.2. Qualidade dos dados

Sendo um fator crítico para garantir que a informação extraída seja confiável e que decisões advindas disso sejam eficazes, a qualidade de dados é um fator comprometedor uma vez que dados de baixa qualidade possam comprometer análises, gerar interpretações errôneas e decisões estrategicamente erradas, esse fenômeno é conhecido como *garbage in, garbage out* (MARR,2016).

Existem quatro fatores principais que podem avaliar a qualidade de um dado, são eles:

1. **Consistência:** Trata-se da uniformidade dos dados de diferentes fontes e registros, assegurando que informações iguais ou relacionadas apresentem o mesmo valor em toda a base de dados.
2. **Completeness:** Refere-se a necessidade de que os dados sejam os mais completos possíveis, evitando lacunas significativas que poderiam comprometer a análise.
3. **Precisão:** A relação entre a exatidão dos dados registrados e a realidade, de forma a refletir o contexto verdadeiro.
4. **Atualidade (ou temporalidade):** Diz respeito à atualização e relevância temporal dos dados, de modo que as informações desatualizadas não comprometam os resultados da análise (CORRAR; DIAS FILHO, 2021).

A manutenção da qualidade dos dados exige processos contínuos de validação, limpeza e padronização, além do estabelecimento de políticas organizacionais claras sobre coleta e gestão de informações. Organizações que negligenciam a qualidade dos dados frequentemente enfrentam dificuldades na tomada de decisão e na confiabilidade de sistemas automatizados, impactando

negativamente sua performance operacional e estratégica (DAVENPORT; PRUSAK, 1998).

Portanto, a qualidade dos dados não deve ser vista apenas como uma questão técnica, mas como um elemento estratégico, pois a confiabilidade das análises, a eficiência dos processos de decisão e a competitividade das organizações dependem diretamente dela.

2.4.3. Ciclo de vida dos dados

O ciclo de vida dos dados representa as etapas fundamentais que os dados percorrem desde sua coleta inicial até a utilização na tomada de decisão. Compreender e gerenciar adequadamente esse ciclo é essencial para transformar dados brutos em informações confiáveis e insights estratégicos, garantindo que o valor do recurso seja plenamente explorado (MARR, 2016).

A primeira etapa do ciclo é a coleta, na qual os dados são obtidos a partir de fontes internas ou externas à organização, como transações, sensores, redes sociais, sistemas corporativos e bancos de dados públicos. A coleta eficiente assegura que os dados relevantes sejam capturados com qualidade e em volume adequado. Em seguida, ocorre o armazenamento, que consiste na preservação segura e organizada dos dados, garantindo integridade, confidencialidade e disponibilidade. Tecnologias como bancos de dados relacionais, *data warehouses* e *data lakes* são utilizadas para armazenar dados de forma estruturada ou não estruturada, dependendo das necessidades analíticas da organização.

O processamento é a etapa em que os dados brutos são transformados em formatos que permitem análise e interpretação, incluindo limpeza, padronização, normalização e integração de diferentes fontes. Processos de ETL (*Extract, Transform, Load*) são frequentemente empregados para preparar os dados para análises avançadas. Posteriormente, na fase de análise, são explorados padrões, correlações e tendências, utilizando técnicas estatísticas, algoritmos de *machine learning* e mineração de dados, com o objetivo de extrair valor e gerar informações significativas (MARR, 2016).

A visualização é a etapa que transforma os resultados das análises em representações gráficas ou *dashboards* compreensíveis, facilitando a interpretação

dos dados pelos tomadores de decisão e aumentando a efetividade na comunicação dos *insights*. Por fim, a tomada de decisão é a fase em que os dados, agora transformados em informações e conhecimento, orientam decisões estratégicas, operacionais ou táticas, impactando diretamente o desempenho e a competitividade da organização (FÁVERO, 2017).

A gestão eficiente de cada etapa do ciclo de vida dos dados é fundamental, pois erros ou falhas em qualquer fase podem comprometer todo o processo analítico, diminuindo a confiabilidade dos *insights* e o valor estratégico dos dados. Assim, o ciclo de vida dos dados não deve ser entendido apenas como um modelo conceitual, mas como uma ferramenta prática que sustenta a governança de dados, a melhoria contínua e a capacidade organizacional de se tornar uma empresa orientada por dados (*data-driven*) (FÁVERO, 2019).

3. METODOLOGIA

Este trabalho se caracteriza como uma pesquisa exploratória de abordagem qualitativa, voltada à investigação do uso de dados estatísticos como suporte à identificação de novos talentos no futebol brasileiro. O caráter exploratório justifica-se pela necessidade de compreender um cenário ainda pouco estruturado em clubes de menor porte, buscando uma familiaridade com o problema e a proposição de soluções iniciais viáveis. A abordagem qualitativa é adequada, pois os dados, embora quantitativos em sua origem, são interpretados com foco na extração de significados e padrões que subsidiem o processo de tomada de decisão no futebol.

A base de dados utilizada nesta pesquisa provém exclusivamente da plataforma *FBref*, escolhida por sua disponibilidade pública, ampla cobertura de campeonatos e acessibilidade. Entretanto, a limitação dessa fonte é um aspecto relevante: os dados oferecidos são estatísticos e descritivos, não abrangendo informações visuais detalhadas como mapas de calor, posicionamento tático, ritmo de jogo ou carga de trabalho físico dos atletas.

Além disso devido à dificuldade de manter uma base atualizada, por conta de custos, decidimos fazer um corte dos dados que se estendem do dia 05/04/2024 até 13/09/2025, acreditamos que tal corte ainda seja relevante, entregando métricas relativamente atuais até a presente entrega do Trabalho de graduação.

Diante dessas limitações, optou-se por focar em métricas fundamentais como número de passes, finalizações, desarmes, posse de bola, cruzamentos e outras ações relevantes para cada posição, acessadas por meio de técnicas de *scraping*¹. Mesmo com essas restrições, acredita-se que a sistematização desses dados já representa um avanço para clubes que, atualmente, não dispõem de qualquer ferramenta estruturada de análise.

O desenvolvimento técnico do sistema foi integralmente concluído, contemplando as etapas de levantamento de requisitos, modelagem, implementação das telas e funcionalidades, integração com a base de dados e validação das saídas estatísticas. O sistema foi desenvolvido majoritariamente em C# e Python, utilizando

¹ Scraping vem de raspagem, é o processo automatizado de coleta de dados de sites, geralmente por meio de scripts ou ferramentas específicas.

bibliotecas como *Pandas*, *Numpy* e *Seaborn* para tratamento de análise de dados, enquanto as interfaces foram construídas em *Typescript* com a biblioteca *React*.

4. DESENVOLVIMENTO

4.1. Escopo do software

O escopo do sistema abrange o desenvolvimento de uma aplicação web voltada à análise de desempenho de jogadores de futebol, utilizando dados estatísticos obtidos da plataforma *FBref*. O sistema foi projetado para oferecer consultas, filtros, e comparações entre atletas, além de uma interpretação automatizada das métricas coletadas, possibilitando uma análise rápida e objetiva.

4.2. Objetivos do desenvolvimento

O desenvolvimento visa implementar a estrutura funcional e visual do sistema proposto, garantindo que as principais etapas do processo de *scouting* sejam apoiadas digitalmente.

Os objetivos dessa fase são:

- Implementar módulos de coleta, processamento e exibição de dados;
- Desenvolver telas interativas para busca e visualização de informações;
- Criar relatórios interpretativos que apresentem o desempenho dos jogadores;
- Assegurar boa usabilidade e desempenho, mesmo com grande volume de dados;

4.3. Requisitos do sistema

O levantamento dos requisitos foi realizado com base nos objetivos definidos e nas necessidades identificadas durante a pesquisa. Eles definem as funcionalidades obrigatórias e as características da qualidade que o sistema deve apresentar.

4.3.1 Requisitos funcionais

Os requisitos funcionais especificam as ações que o sistema executa, ou seja, as operações diretamente ligadas ao comportamento e às respostas esperadas do software.

Tabela 1 - Requisitos Funcionais do Software

Requisito	Descrição
RF01	O sistema deve coletar automaticamente dados de jogadores da plataforma <i>FBRef</i> .
RF02	O sistema deve exibir uma lista de jogadores com informações básicas como: nome, idade, clube e posição.
RF03	O sistema deve permitir filtrar jogadores por posição.
RF04	Ao selecionar um jogador, o sistema deve exibir uma ficha individual com informações detalhadas.
RF05	A ficha do jogador deve possuir navegação em abas: “Média Geral”, “Estatística Geral”, “Partidas” e “Análise do Jogador”.
RF06	A aba “Média Geral” deve apresentar as médias agregadas de desempenho do jogador.
RF07	A aba “Estatística Geral” deve exibir métricas detalhadas divididas por categoria (ofensivas, defensivas, posse de bola e passes).
RF08	A aba “Partidas” deve exibir as estatísticas por jogo do jogador, listando desempenho individual em cada partida.
RF09	A aba “Análise do Jogador” deve apresentar uma interpretação visual e textual dos dados, destacando pontos fortes e fracos do jogador.

Fonte: Elaborado pelos autores

4.3.2. Requisitos não funcionais

Os requisitos não funcionais dizem respeito aos atributos de qualidade do sistema, como desempenho, usabilidade, confiabilidade e compatibilidade, assegurando que funcione adequadamente em diferentes contextos de uso.

Tabela 2 - Requisitos Não Funcionais do Software

Requisito	Propriedade	Medida
RNF01	Usabilidade	A interface deve ser simples e intuitiva, com navegação fluida entre filtros, visão geral e análise detalhada.
RNF02	Desempenho	A resposta aos filtros e carregamento de dados devem ocorrer em até x segundos para até y jogadores.
RNF03	Escalabilidade	O sistema deve ser capaz de lidar com grandes volumes de dados, acompanhando o crescimento da base alimentada pelo <i>FBRef</i> .
RNF04	Confiabilidade	Os dados devem refletir fielmente as informações da <i>FBRef</i> .
RNF05	Portabilidade	O sistema deve estar funcionando em navegadores modernos.

Fonte: Elaborado pelos autores

4.4. Casos de uso

Os casos de uso são representações da forma como o usuário interage com o sistema, descrevendo os principais fluxos de funcionamento a partir dos requisitos identificados. Eles são utilizados para modelar o comportamento do sistema sob o ponto de vista do usuário, permitindo compreender cada funcionalidade que se conecta às necessidades práticas do projeto.

No contexto deste trabalho, o ator principal é o técnico ou analista esportivo, responsável por consultar os jogadores e interpretar os dados disponíveis. O Diagrama de Caso de Uso a seguir (Figura 1) ilustra essas interações, evidenciando as ações possíveis e suas relações com o ator principal.

Figura 1 - Casos de Uso do Software



Fonte: Elaborado pelos Autores

UC01 - Consultar Lista de Jogadores

- **Atores:** Treinador
- **Descrição:** Permite visualizar lista de jogadores importados da *FBRef* e aplicar filtros
- **Pré-condição:** Dados carregados com sucesso da *FBRef*
- **Fluxo principal:**
 - O usuário acessa a tela inicial.
 - O sistema exibe a lista de jogadores com nome, posição e clube.
 - O usuário pode aplicar filtros por posição.
 - O sistema atualiza a listagem.

- O usuário seleciona um jogador para abrir sua ficha.
- **Pós-condição:** Lista visível e pronta para aplicação de filtros por posição.

UC02 - Visualizar Ficha do Jogador

- **Atores:** Treinador
- **Descrição:** Permite visualizar informações básicas e médias gerais do jogador.
- **Pré-condição:** Jogador selecionado.
- **Fluxo principal:**
 - O usuário seleciona um jogador.
 - O sistema exibe dados como nome, idade, posição, clube e estatísticas principais.
 - O sistema apresenta a ficha com as abas “Média”, “Gerais”, “Partidas” e “Análise”.
- **Pós-condição:** Ficha do jogador aberta com navegação disponível entre abas.

UC03 - Visualizar Estatísticas Médias

- **Atores:** Técnico
- **Descrição:** Exibe um resumo das médias de desempenho do jogador.
- **Pré-condição:** Ficha do jogador aberta.
- **Fluxo principal:**
 - O usuário acessa a aba “Média Geral”.
 - O sistema busca as médias do jogador em diferentes métricas.
 - As informações são apresentadas de forma resumida e visual.
- **Pós-condição:** Médias exibidas corretamente.

UC04 - Visualizar Estatísticas Gerais

- **Atores:** Técnico
- **Descrição:** Exibe as métricas detalhadas divididas por categoria.
- **Pré-condição:** Jogador já selecionado.
- **Fluxo principal:**
 - O usuário clica na aba “Estatística Geral”.

- O sistema consulta as métricas completas e as organiza em categorias: ofensivas, defensivas, posse e passes.
- O sistema exibe os dados em *cards*.
- **Pós-condição:** Estatísticas detalhadas exibidas com sucesso.

UC05 - Visualizar Partidas

Atores: Técnico

- **Descrição:** Exibe o desempenho do jogador em cada partida disputada.
- **Pré-condição:** Jogador já selecionado.
- **Fluxo principal:**
 - O usuário acessa a aba “Partidas”.
 - O sistema lista todas as partidas registradas do jogador.
 - São exibidas as estatísticas individuais por jogo (como gols, passes, finalizações, desarmes, etc.).
- **Pós-condição:** Lista de partidas com estatísticas exibidas.

UC06 - Analisar Desempenho Técnico

- **Atores:** Técnico
- **Descrição:** Exibe uma análise interpretativa com base nas métricas do jogador.
- **Pré-condição:** Jogador já selecionado.
- **Fluxo principal:**
 - O usuário acessa a aba “Análise do Jogador”.
 - O sistema apresenta uma visão geral interpretando os dados, destacando pontos fortes e fracos.
 - Indicadores visuais podem ilustrar o desempenho geral.
- **Pós-condição:** Análise exibida com sucesso.

4.5. Diagramas de sequência

Os diagramas de sequência apresentados a seguir descrevem o fluxo de interação entre os componentes do sistema durante a execução de cada caso de uso.

Em cada cenário, são representados o ator principal (técnico), a interface do sistema, o controlador (camada de aplicação) e, quando necessário, o banco de dados ou serviços auxiliares.

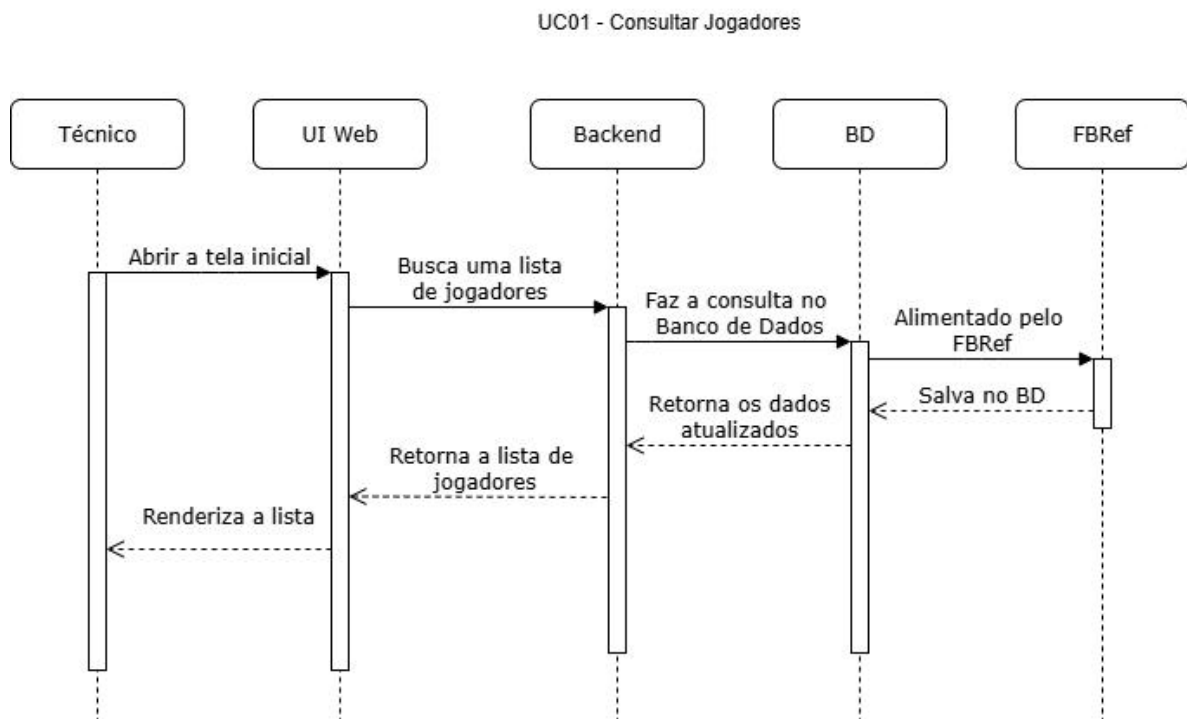
O objetivo é demonstrar a troca de mensagens, a ordem dos eventos e o comportamento dinâmico da aplicação.

4.5.1. UC01 - Consultar lista de jogadores

O diagrama de sequência do caso de uso “Consultar lista de jogadores” (Figura 2) demonstra o processo de carregamento e exibição dos dados coletados da plataforma *FBRef*. Primeiro, o usuário acessa a tela inicial, acionando a interface para requisitar a lista de jogadores. A interface envia a solicitação ao sistema, que consulta o banco de dados em busca das informações disponíveis.

Os dados são retornados à interface, que exibe a lista completa de jogadores, permitindo ao usuário aplicar filtros ou realizar uma nova busca.

Figura 2 - Diagrama de Sequência para o Caso de Uso Consultar Jogadores

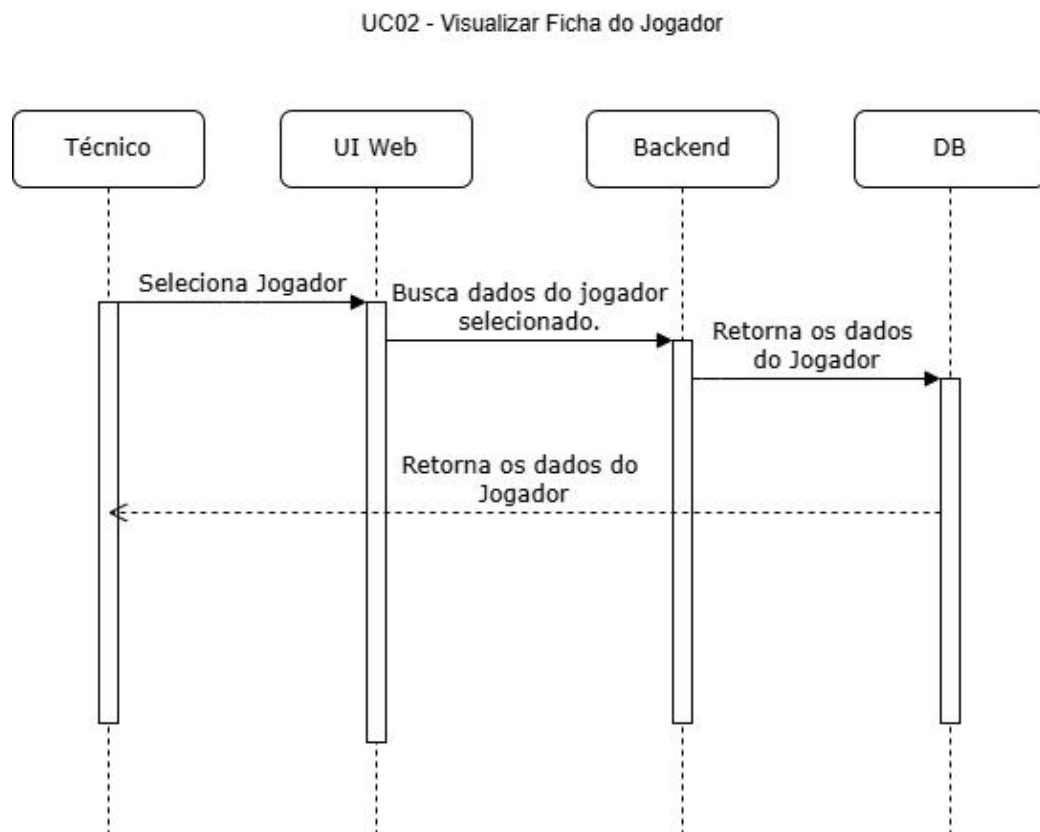


Fonte: Elaborado pelos autores

4.5.2. UC02 - Visualizar ficha do jogador

Este diagrama (Figura 3) mostra o comportamento do sistema quando o usuário seleciona um jogador específico. A interface envia ao sistema a requisição de detalhes do jogador selecionado. O sistema busca as informações no banco de dados e retorna o conjunto de dados para a interface. Por fim, a ficha do jogador é exibida, apresentando o cabeçalho com dados básicos (nome, idade, posição e clube) e as abas disponíveis (“Média Geral”, “Estatística Geral”, “Partidas” e “Análise do Jogador”).

Figura 3 - Diagrama de Sequência para o Caso de Uso Visualizar Ficha do Jogador



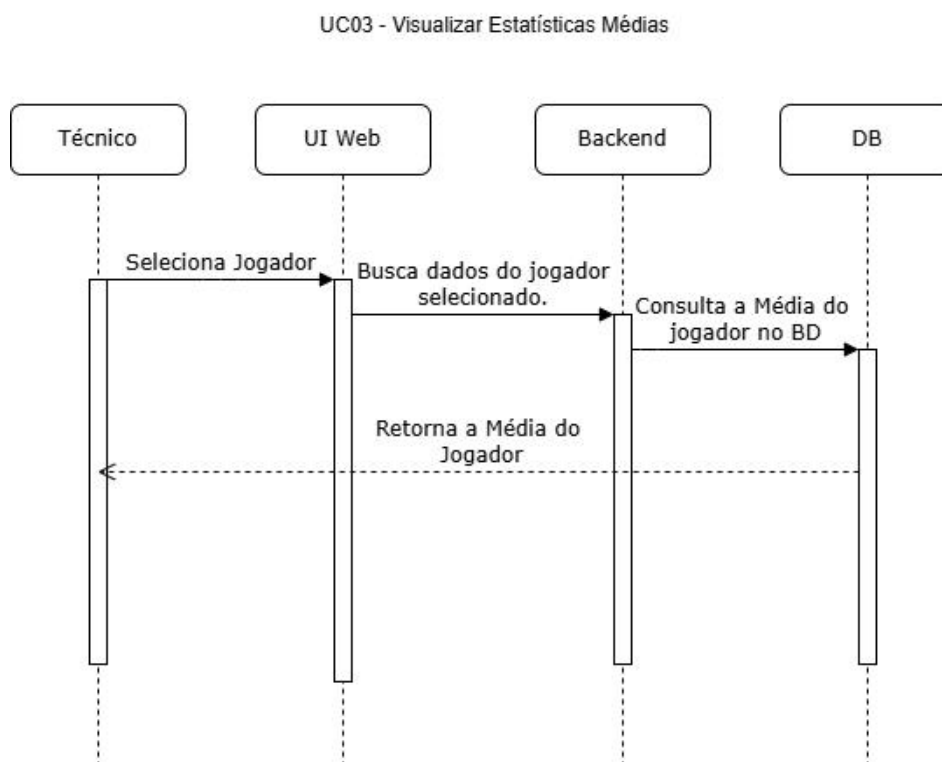
Fonte: Elaborado pelos autores

4.5.3. UC03 - Visualizar estatística média

O diagrama de sequência da aba “Média Geral” (Figura 4) representa o processo de renderizar as estatísticas médias do jogador. Ao selecionar essa aba, a interface solicita ao sistema as médias consolidadas registradas no banco de dados.

Os dados são então retornados e exibidos na tela em formato de *cards*, possibilitando ao analista uma visão geral do desempenho do atleta.

Figura 4 - Diagrama de Sequência para o Caso de Uso Consultar Estatística Média do Jogador



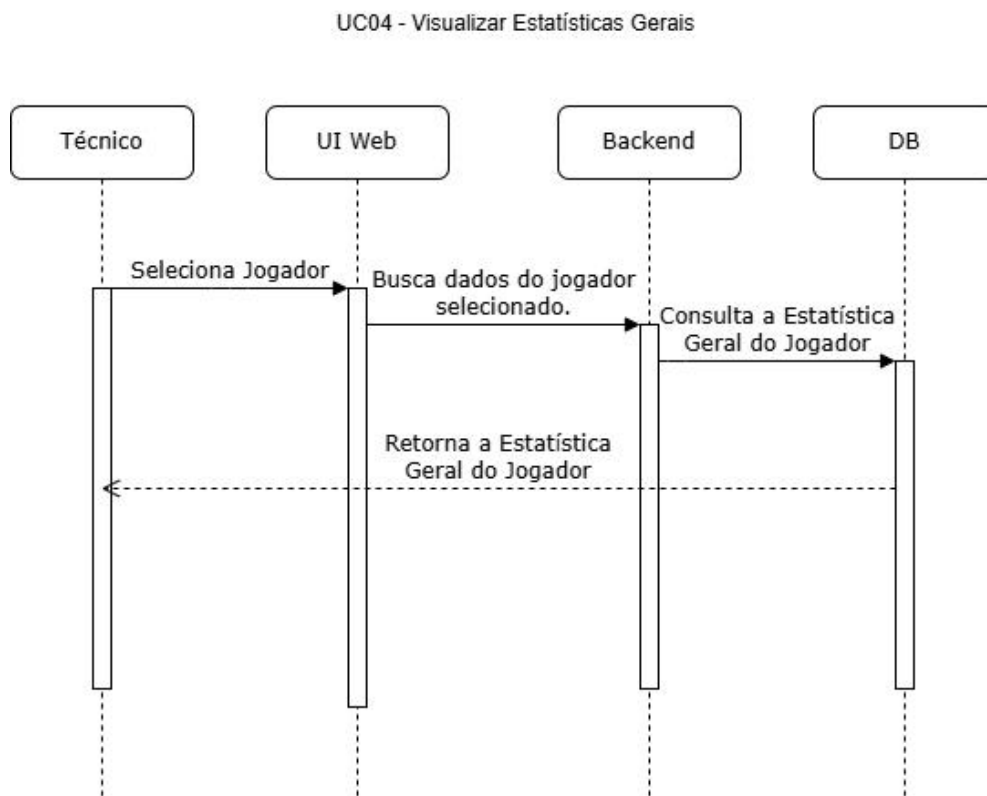
Fonte: Elaborado pelos autores

4.5.4. UC04 - Visualizar estatística geral

Neste diagrama (Figura 5) é descrita a sequência de interações ao acessar a aba “Estatística Geral”. A interface solicita ao controlador todas as métricas detalhadas do jogador, organizadas por categoria (ofensivas, defensivas, posse e passes). Após a recuperação dos dados no banco, a interface apresenta os

resultados em sessões categorizadas, permitindo uma análise detalhada de cada aspecto do desempenho do atleta.

Figura 5 - Diagrama de Sequência para Caso de Uso Visualizar Partidas

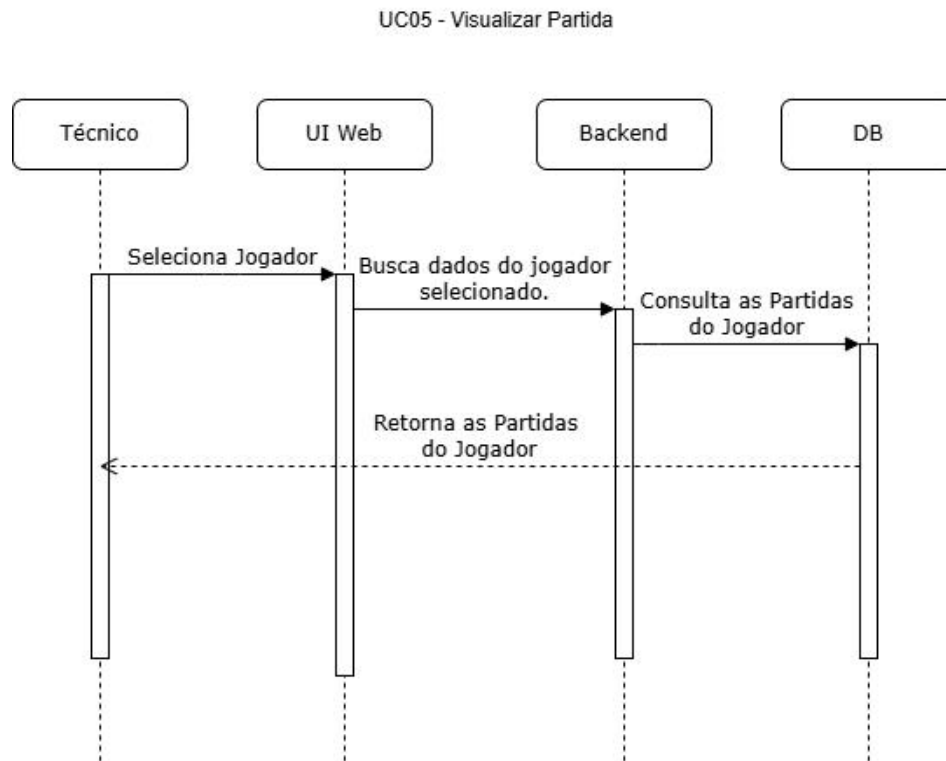


Fonte: Elaborado pelos autores

4.5.5. UC05 - Visualizar partida

O diagrama do caso de uso “Consultar aba Partidas” (Figura 6) mostra a consulta e exibição das estatísticas de cada jogo do jogador. Quando o usuário seleciona a aba, a interface envia uma requisição ao sistema para obter o histórico de partidas do atleta. O sistema acessa o banco de dados, recupera as informações e retorna a lista à interface, que exibe os dados da data, adversário, competição e indicadores de desempenho.

Figura 6 - Diagrama de Sequência para Caso de Uso Visualizar Partida



Fonte: Elaborado pelos autores

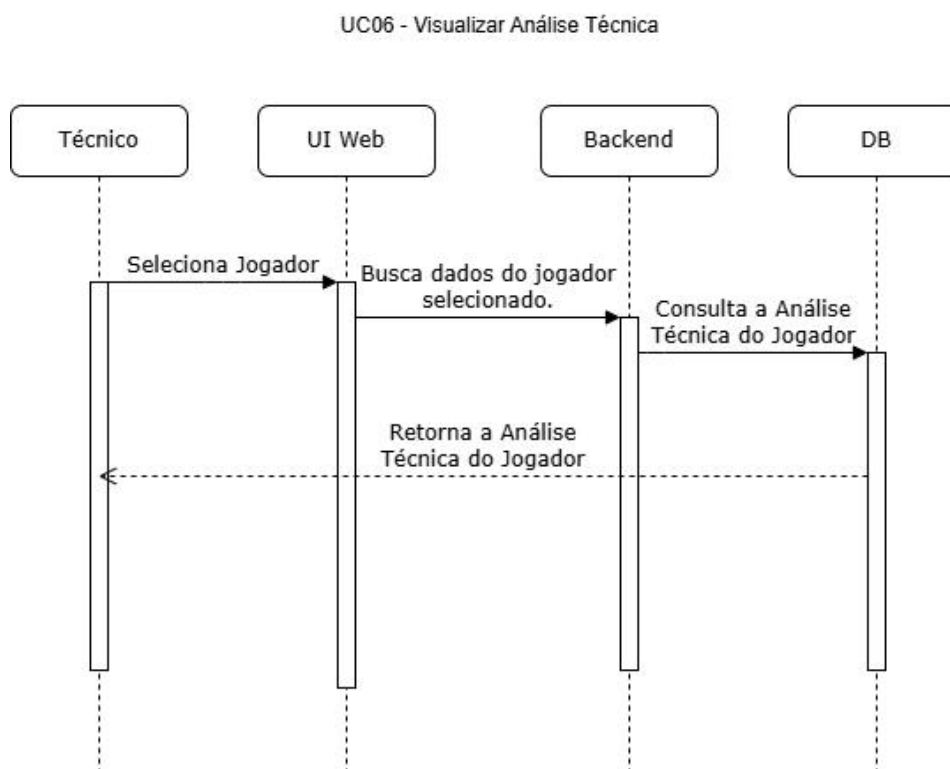
4.5.6. UC06 - Analisar desempenho técnico

O diagrama de sequência da aba “Análise do Jogador” (Figura 7) ilustra o processo de geração e exibição da análise interpretativa do jogador. Ao selecionar a aba, a interface requisita ao sistema o conjunto de dados consolidados de desempenho.

O sistema processa essas informações e as encaminha para a interface, que apresenta ao usuário uma interpretação visual e textual, acompanhada de gráficos comparativos (como o gráfico de radar).

Essa tela sintetiza o desempenho geral do atleta, destacando seus pontos fortes e fracos de forma visualmente acessível.

Figura 7 - Diagrama de Sequência do Caso de Uso Analisar Desempenho Técnico



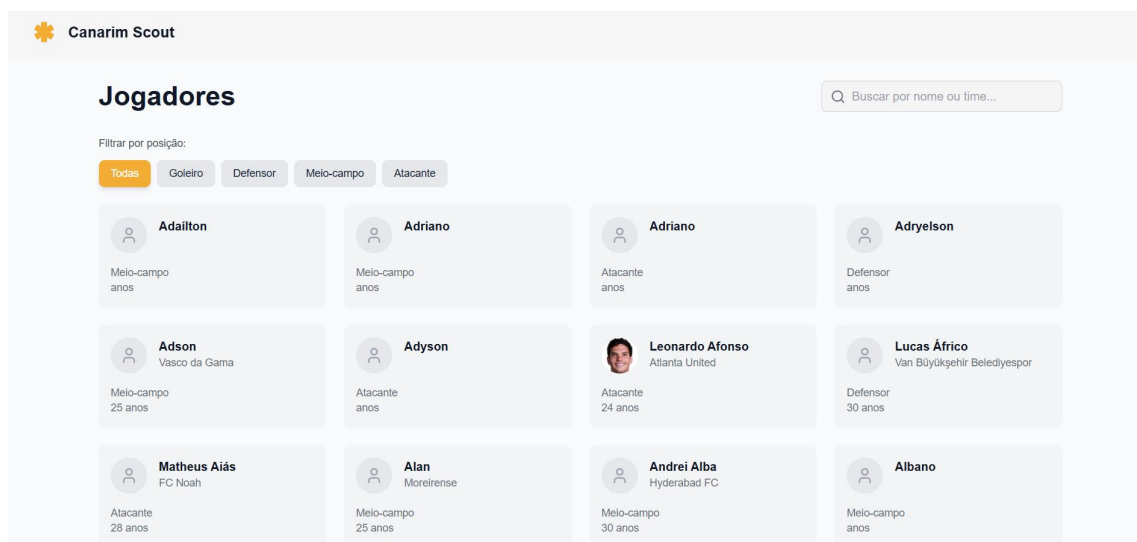
Fonte: Elaborado pelos autores

4.6. Projeto de interface

A interface do sistema foi projetada com foco na simplicidade e usabilidade, visando facilitar a análise de desempenho dos jogadores por parte de analistas e olheiros. Foram desenvolvidas telas que organizam os dados coletados em diferentes níveis de detalhe, proporcionando uma navegação fluida e acesso rápido às informações mais relevantes.

4.6.1. Tela inicial

A primeira tela (Figura 8) exibe um painel geral com um buscador de jogadores e filtros pelo nome do jogador e a sua posição. O objetivo é permitir uma visualização rápida e intuitiva do banco de dados disponível, servindo como ponto de partida para a análise. Os jogadores são exibidos com base no algoritmo definido na análise de dados, os que tiverem a melhor média serão exibidos primeiro.

Figura 8 - Tela inicial do sistema de *scouting*

Fonte: Elaborado pelos autores

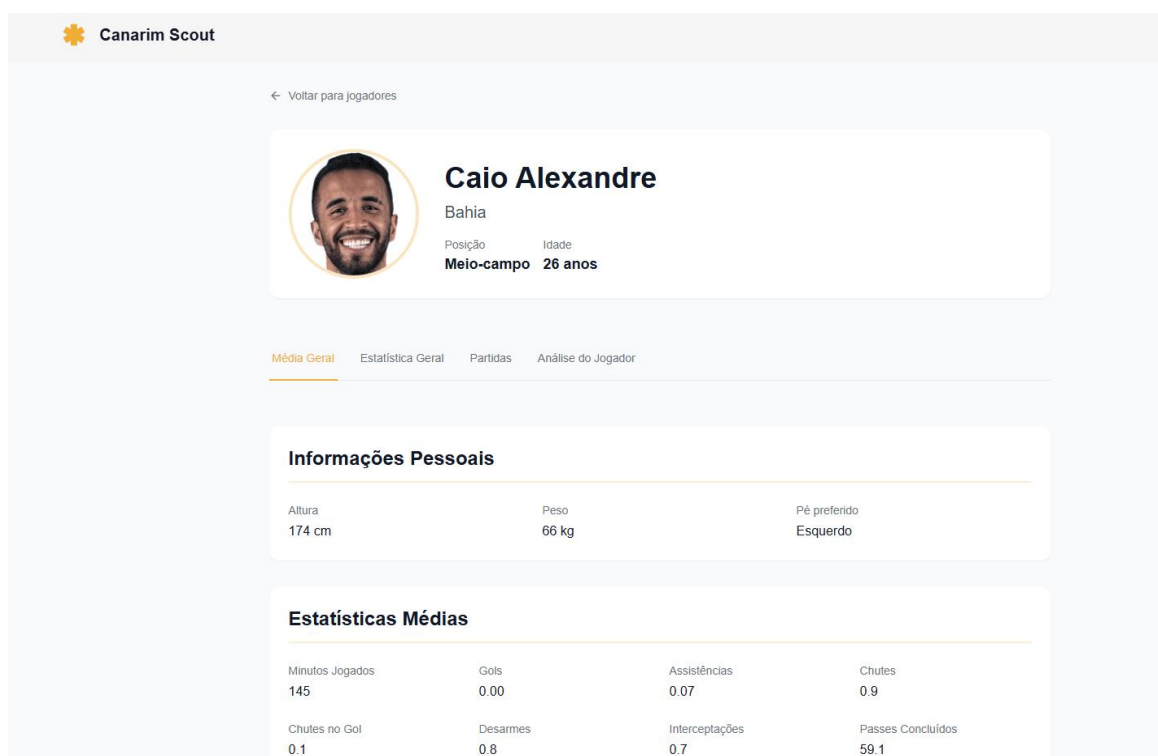
4.6.2. Ficha do jogador - Estrutura geral

A ficha do jogador (Figura 9) reúne suas principais estatísticas e é dividida em quatro abas de navegação:

- Média Geral
- Estatística Geral
- Partidas
- Análise do Jogador

Essas abas organizam as informações de forma hierárquica: o usuário pode começar pela visão geral e, se quiser, avançar para dados mais detalhados. O cabeçalho da ficha mantém o nome do jogador, idade, posição e clube, servindo como referência constante durante a navegação.

Figura 9 - Visualização da ficha do jogador selecionado



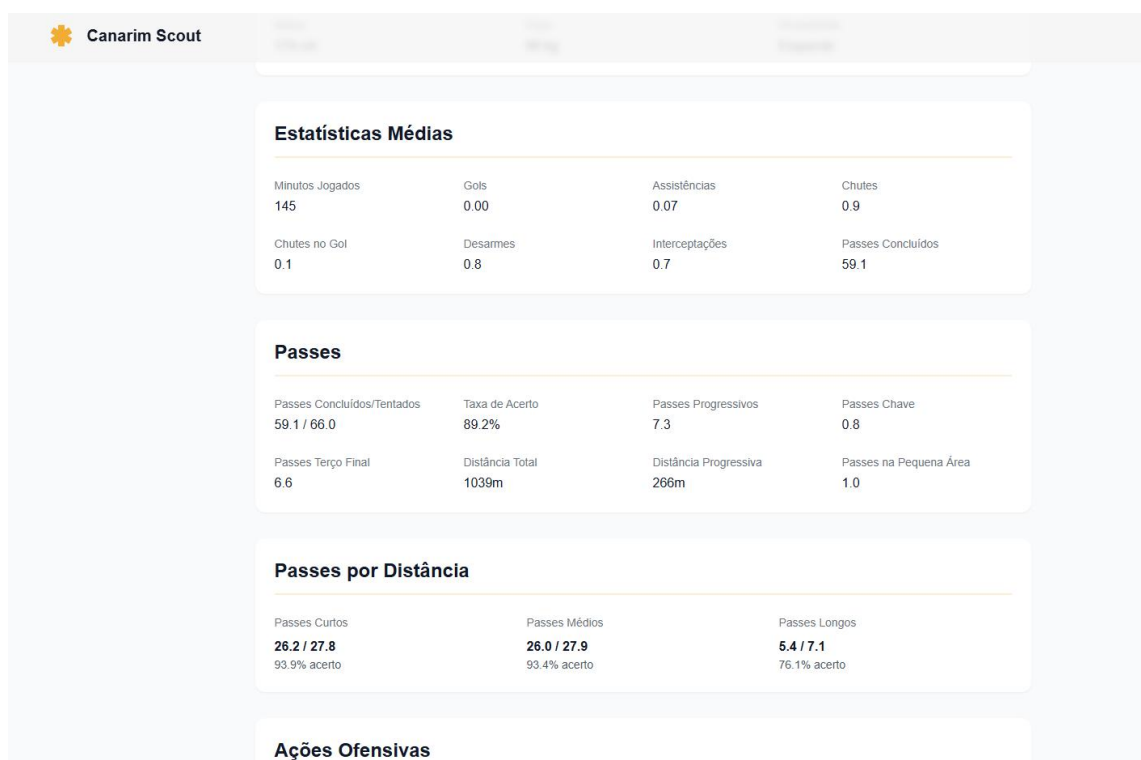
Fonte: Elaborado pelos autores

4.6.3. Média geral

A aba “Média Geral” (Figura 10, 11 e 12) exhibe os valores médios consolidados de desempenho do jogador, permitindo uma visão resumida de como ele se comporta nas principais métricas.

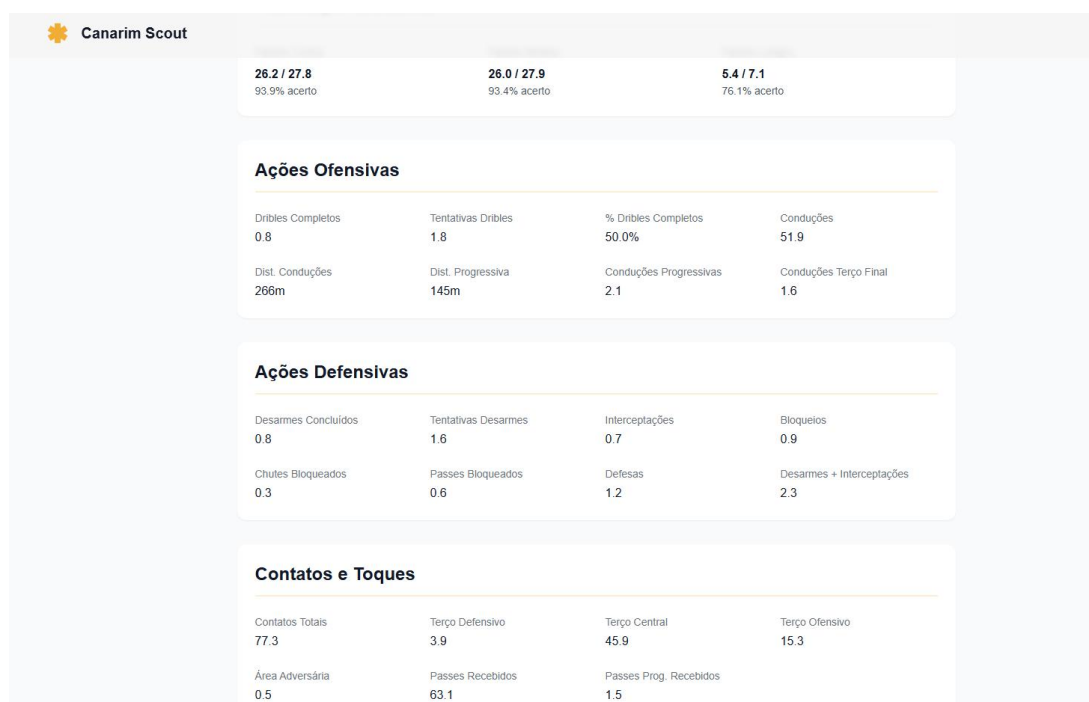
Os dados são apresentados em formato de cards, destacando indicadores como média de gols, assistências, passes, desarmes ou outros valores gerais.

Figura 10 - Visualização das Estatísticas Médias, Passes e Passes por Distância do Jogador selecionado



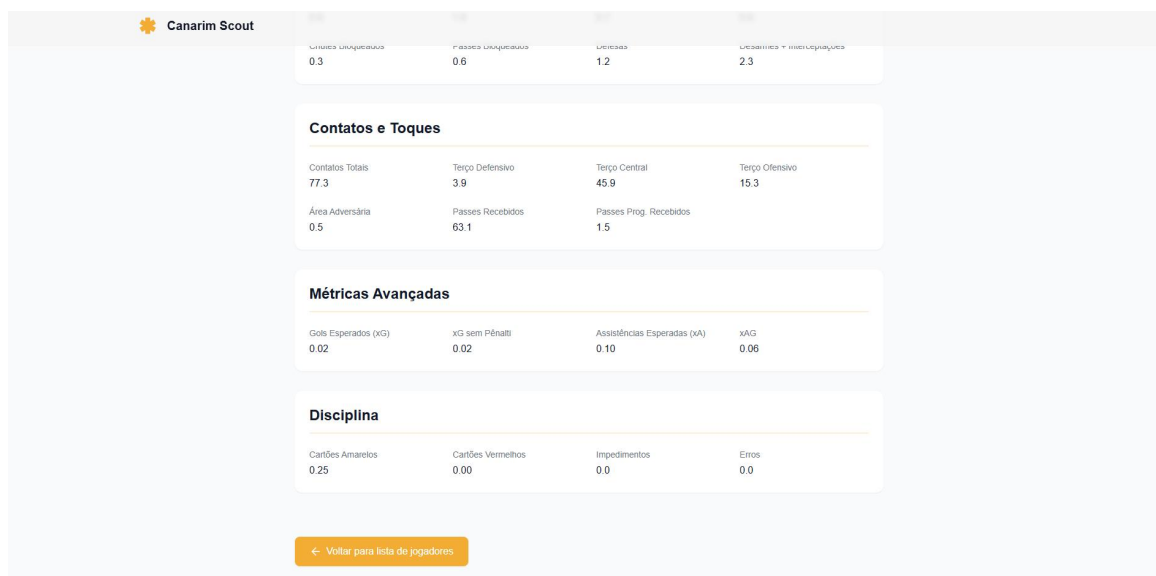
Fonte: Elaborado pelos autores

Figura 11 - Visualização das Ações Ofensivas, Ações Defensivas e Contatos e Toques do Jogador Selecionado.



Fonte: Elaborado pelos autores

Figura 12 - Visualização das Métricas Avançadas e Disciplina do jogador

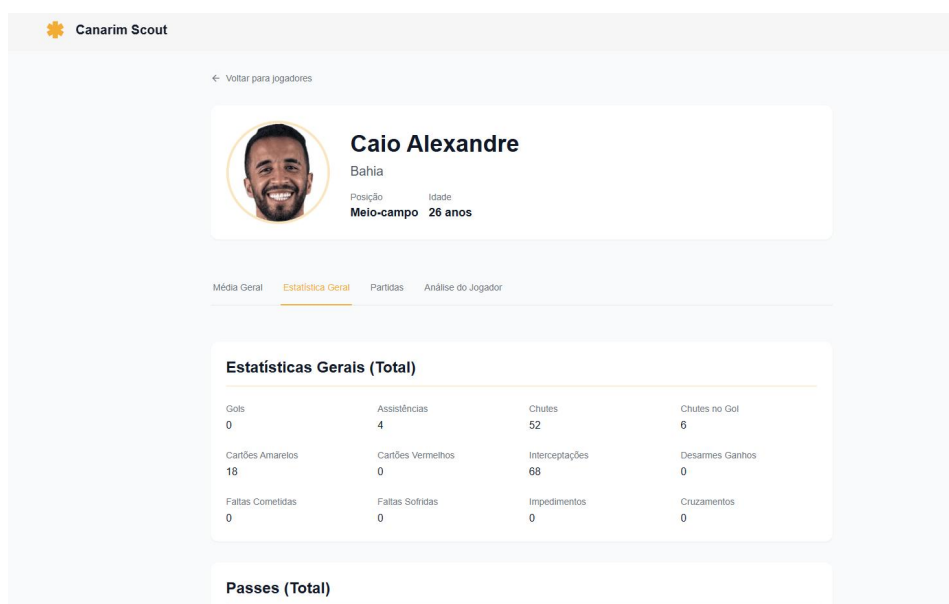


Fonte: Elaborado pelos autores

4.6.3. Estatística geral

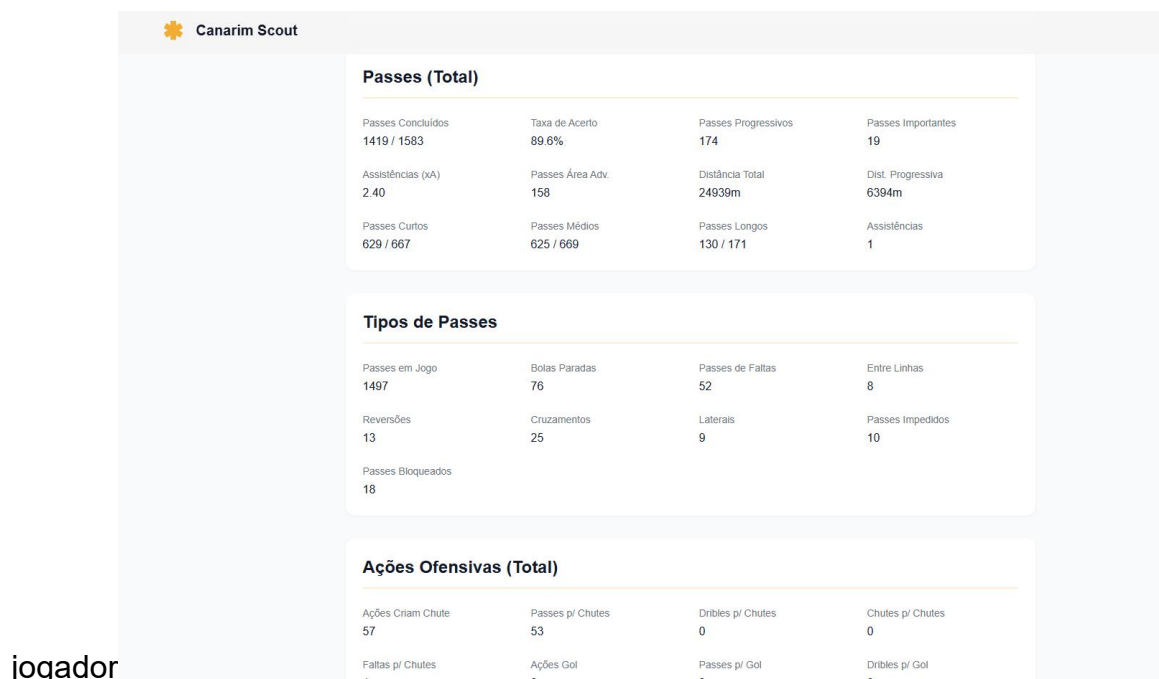
A aba “Estatística Geral” (Figura 13, 14 e 15) mostra as estatísticas detalhadas divididas em categorias: ofensivas, defensivas, posse de bola e passes. Os dados são exibidos em *cards* comparativos, facilitando a identificação de padrões de jogo.

Figura 13 - Visualização das Estatísticas Gerais do jogador



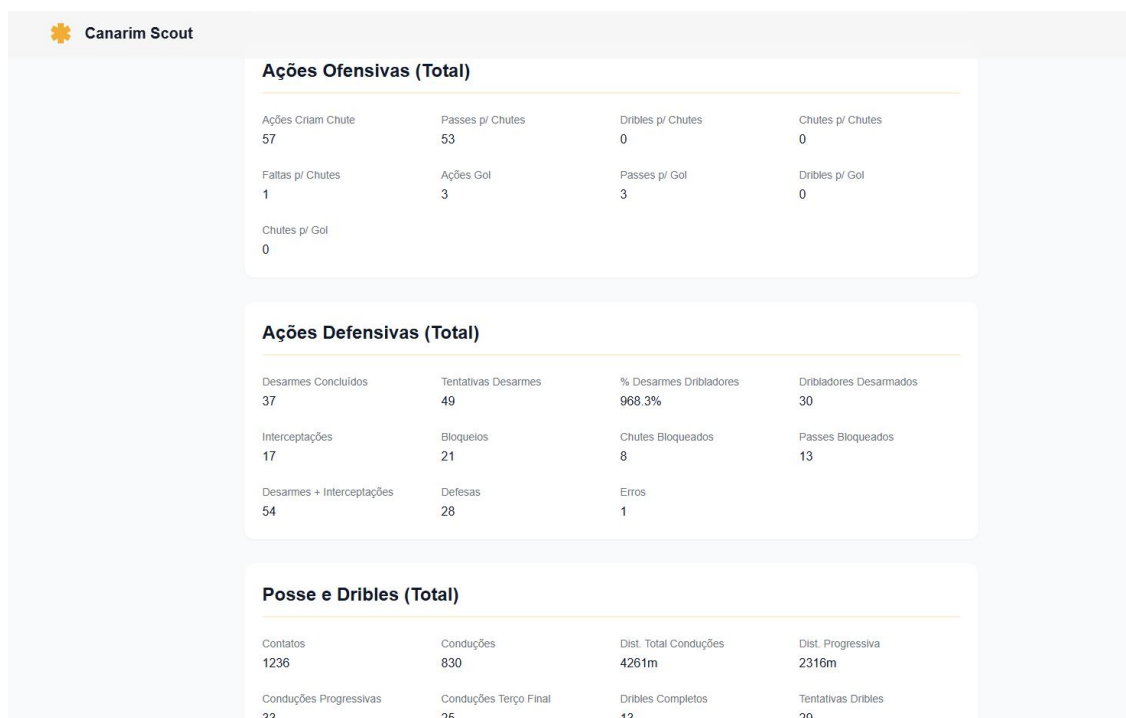
Fonte: Elaborado pelos autores

Figura 14 - Visualização dos Passes, Tipos de Passe e Ações Ofensivas do



Fonte: Elaborado pelos autores

Figura 15 - Visualização das Ações Defensivas e Posses e Dribles do jogador



Fonte: Elaborado pelos autores

4.6.4. Partidas

A aba “Partidas” (Figura 16) apresenta o desempenho individual do jogador em cada jogo. Cada linha representa uma partida, exibindo dados como data, adversário, minutos jogados, gols, passes e outros indicadores relevantes.

Essa tela permite avaliar a consistência do jogador ao longo das partidas, identificando variações de rendimento e possíveis tendências.

Figura 16 - Visualização das partidas do jogador

Canarim Scout

← Voltar para jogadores

Caio Alexandre
Bahia
Posição: Meio-campo | Idade: 26 anos

Média Geral | Estatística Geral | **Partidas** | Análise do Jogador

Últimas Partidas (Página 1 de 6)

Contexto	Data	Torneio	Resultado	Minutos	Gols	Assist.
Casa	2025-08-09	Série A (Matchweek 19)	D 3-3	70'	0	0
Casa	2025-02-25	Libertadores (Second qualifying stage)	W 3-0	90'	0	0
Fora	2025-03-06	Libertadores (Third qualifying stage)	D 0-0	90'	0	0

Fonte: Elaborado pelos autores

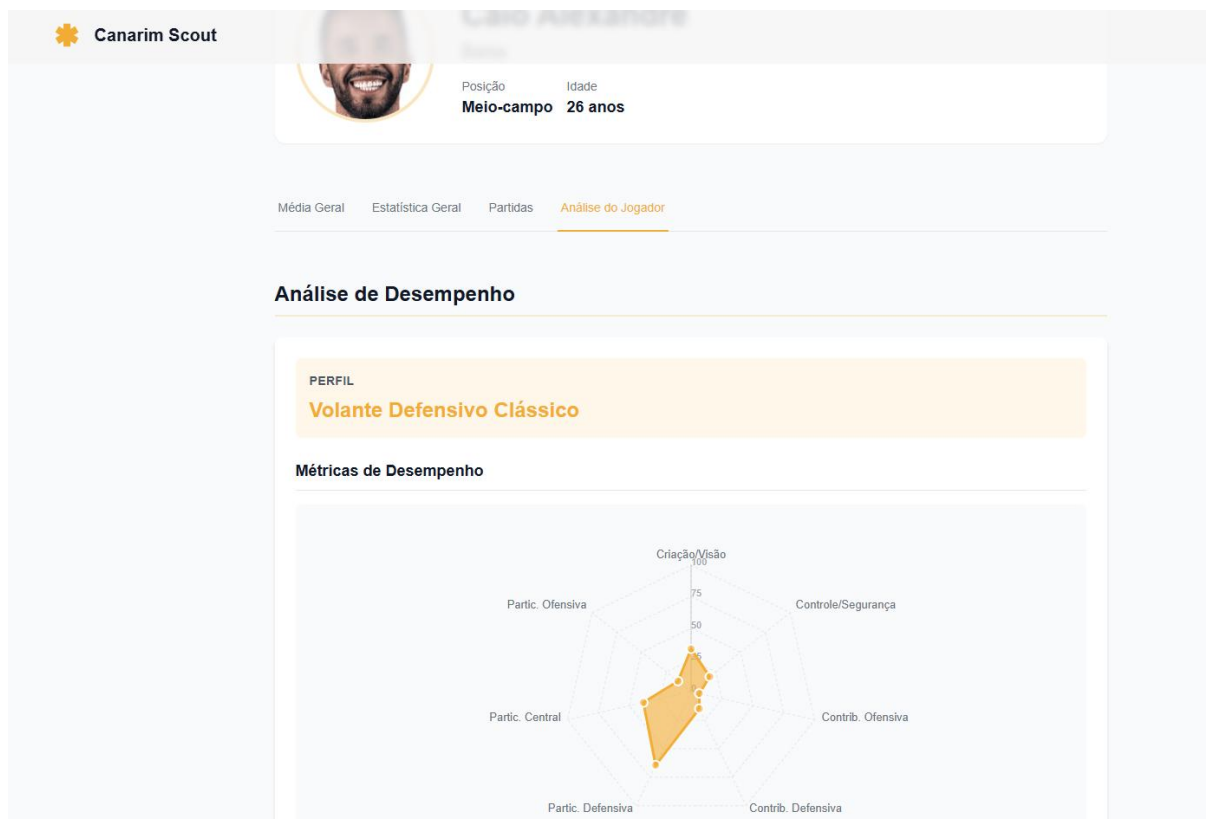
4.6.5. Análise do jogador

A aba “Análise do Jogador” (Figura 17) fornece uma interpretação dos dados do jogador, destacando pontos fortes e pontos fracos com base nas métricas apresentadas nas abas anteriores.

A exibição é visual e direta, podendo incluir um gráfico de radar que ajuda a compreender o perfil de desempenho do atleta.

Essa aba complementa as demais, funcionando como um resumo interpretativo, útil para decisões técnicas ou relatórios de observação.

Figura 17 - Visualização da Análise do Jogador



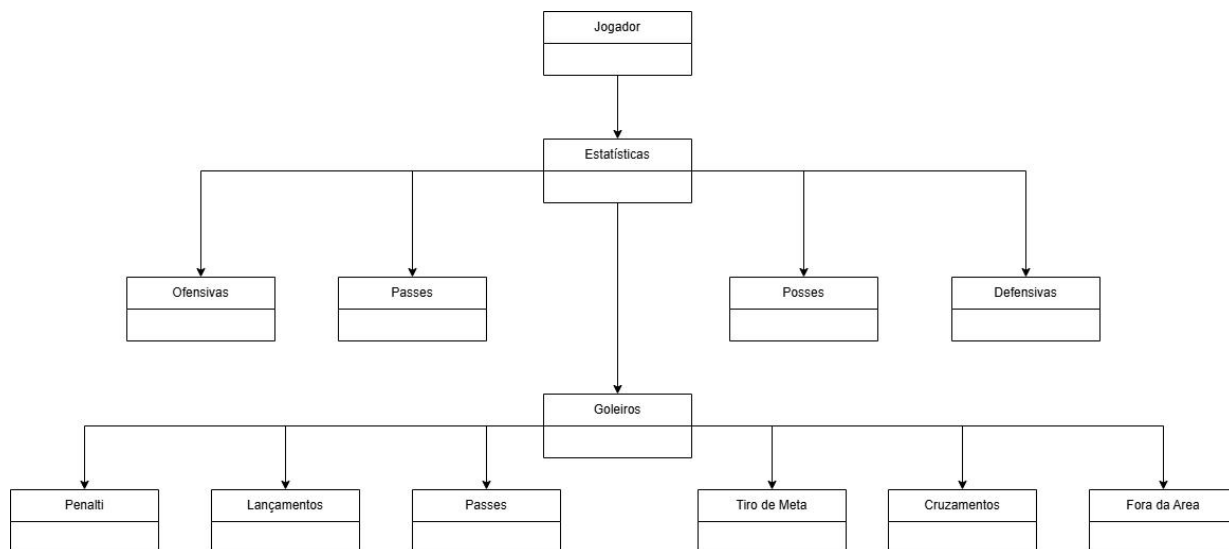
Fonte: Elaborado pelos autores

4.7. Diagrama de classes

É através do diagrama de classes que descrevemos as classes envolvidas, seus atributos e os relacionamentos entre elas, servindo como base para o desenvolvimento posterior das funcionalidades do sistema.

As classes Jogador e Estatísticas representam os objetos centrais da aplicação. Outras classes, como Ofensivas, Passes, Posses e Defensivas complementam a estrutura, permitindo uma separação lógica das estatísticas por tipo de ação em campo. E a classe Goleiro foi feita de forma separada, pois o goleiro possui ações de jogos específicos que não se conectam com os outros tipos de jogadores.

Figura 18 - Diagrama de Classes Simplificado



Fonte: elaborado pelos autores

Classe	Atributos	Métodos
Jogador	Id Nome DtNascimento Altura Peso PeDominante Selecao TimeAtual Posicao Foto	GetPlayers(string position); GetPlayerById(string playerId);
Estatísticas	Id Gols Assistencias MinutosJogados CartaoAmarelo CartaoVermelho	GetMatchStats (string playerId); GetAggregatedStats (string playerId); GetPlayerAnalysisByPosition(string playerId);
Ofensivas	AcaoCriacaoChute FaltaGol AcaoDefensivaGol ChuteGol DribleGol GolBolaParada PasseGol AcaoGol AcaoDefensivaTentChutes FaltaTentChutes ChutesTentChutes DriblesTentChutes PassesTentChutesBolaParada PassesTentChutes AcaoCriacaoChute	
Passes	PassesConcluidos TentativasPasses DistanciaTotal DistPassesProgressivo QtdPassesProgressivos PassesCurtosTentados PerPassesConcluidos PassesCurtosConcluidos PerPassesCurtosConcluidos	

	PassesMediosTentados PassesMediosConcluidos PerPassesMediosTentados PassesLongosTentados PassesLongosConcluidos PerPassesLongosTentados Assistencia XAG XA PassesImportantes PassesPeqAreaAdv PassesAreaAdv CruzamentoPeqArea PassesBloqueados LancamentoEscanteio Cruzamentos PassesBolasParadas PassesAPartirDeFaltas EscanteiosDentroArea PassesEmJogo PassesImpedidos EscanteioForaArea EscanteioReto PassesDeReversao PassesEntreLinhaDefesa LancamentoLateral Passes	
Posses	Contatos ContatosDefensivosPen ContatosTercoDef ContatosTercoCentral ContatosTercoAtaque ContatosAtaquePen ContatoBolaJogo TentativasDribles DriblesCompleto PerDriblesCompleto DriblesInterrompidos PerDriblesInterrompidos Conducoes DistTotal DistProgressivas CarregadaProgressiva CarregadaTercoFinal CarregadaAreaPenalti PerdaDominio	

	Desarmado PassesRecebidos PassesProgressivosRecebidos	
Goleiro	Defesas PartidaSemGol ChutesNoGol GolSofridos PerDefesas PenaltiDefendidos PenaltiSofrido PenaltiCausados PenaltiMarcado Lancamento Penalti Lancamentos PassesGoleiro TiroDeMeta Cruzamento ForaDaArea	GetGoalkeeperStats (string playerId);
Penalti	CobrancaPenalti NaoDefendido PenaltiMarcado ErroAdversario	
PassesGoleiro	Passes TentativasPasses DistanciaMedia	
TiroDeMeta	QtdTirosDeMeta PerTiroDeMetas DistMediaTiroDeMeta DistanciaMedia	
Cruzamento	CruzamentosEnfrentados CruzamentoBloqueado PerCruzamentoBloqueado	
ForaDaArea	AcoesForaDaPequenaArea DistMedia	

4.8. Diagrama de entidade e relacionamento

O diagrama de entidade e relacionamento (DER) serve para representar de forma visual como os dados de um sistema se organizam e se relacionam entre si. Ele é uma ferramenta usada na modelagem de bancos de dados, mostrando as entidades (como clientes, produtos ou pedidos), seus atributos e os relacionamentos existentes entre elas. Dessa forma, o DER facilita o planejamento, a compreensão e a construção de uma base de dados coerente, evitando redundâncias e garantindo que todas as informações necessárias estejam corretamente conectadas.

Gca	Id (INTEGER) date (TEXT) day (TEXT) comp (TEXT) round (TEXT) venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) min (TEXT) sca (TEXT) passlive (TEXT) passdead (TEXT) to (TEXT) sh (TEXT) fld (TEXT) def (TEXT) gca (TEXT) passlive_1 (TEXT) passdead_1 (TEXT) to_1 (TEXT) sh_1 (TEXT) fld_1 (TEXT) def_1 (TEXT) match_report (TEXT) player_id (TEXT)
defense	id (INTEGER)

	<p> date (TEXT) day (TEXT) comp (TEXT) round (TEXT) venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) min (TEXT) tkl (TEXT) tklw (TEXT) def_3rd (TEXT) mid_3rd (TEXT) att_3rd (TEXT) tkl_1 (TEXT) att (TEXT) tkl_ (TEXT) lost (TEXT) blocks (TEXT) sh (TEXT) pass (TEXT) int (TEXT) tkl_int (TEXT) clr (TEXT) err (TEXT) match_report (TEXT) player_id (TEXT) </p>
player_passing_types	<p> id (INTEGER) date (TEXT) day (TEXT) comp (TEXT) round (TEXT) venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) </p>

	min (TEXT) att (TEXT) live (TEXT) dead (TEXT) fk (TEXT) tb (TEXT) sw (TEXT) crs (TEXT) ti (TEXT) ck (TEXT) in (TEXT) out (TEXT) str (TEXT) cmp (TEXT) off (TEXT) blocks (TEXT) match_report (TEXT) player_id (TEXT)
possession	id (INTEGER) date (TEXT) day (TEXT) comp (TEXT) round (TEXT) venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) min (TEXT) touches (TEXT) def_pen (TEXT) def_3rd (TEXT) mid_3rd (TEXT) att_3rd (TEXT) att_pen (TEXT) live (TEXT) att (TEXT) succ (TEXT) succ_ (TEXT) tkld (TEXT) tkld_ (TEXT)

	<p> carries (TEXT) totdist (TEXT) prgdist (TEXT) prgc (TEXT) 1_3 (TEXT) cpa (TEXT) mis (TEXT) dis (TEXT) rec (TEXT) prgr (TEXT) match_report (TEXT) player_id (TEXT) </p>
player_passing	<p> id (INTEGER) date (TEXT) day (TEXT) comp (TEXT) round (TEXT) venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) min (TEXT) cmp (TEXT) att (TEXT) cmp_ (TEXT) totdist (TEXT) prgdist (TEXT) cmp_1 (TEXT) att_1 (TEXT) cmp__1 (TEXT) cmp_2 (TEXT) att_2 (TEXT) cmp__2 (TEXT) cmp_3 (TEXT) att_3 (TEXT) cmp__3 (TEXT) ast (TEXT) xag (TEXT) xa (TEXT) kp (TEXT) </p>

	1_3 (TEXT) ppa (TEXT) crspa (TEXT) prgp (TEXT) match_report (TEXT) player_id (TEXT)
misc	id (INTEGER) date (TEXT) day (TEXT) comp (TEXT) round (TEXT) venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) min (TEXT) crdy (TEXT) crdr (TEXT) 2crdy (TEXT) fls (TEXT) fld (TEXT) off (TEXT) crs (TEXT) int (TEXT) tklw (TEXT) pkwon (TEXT) pkcon (TEXT) og (TEXT) recov (TEXT) won (TEXT) lost (TEXT) won_ (TEXT) match_report (TEXT) player_id (TEXT)
goalkeeper_stats	id (INTEGER) date (TEXT) day (TEXT) comp (TEXT) round (TEXT)

	venue (TEXT) result (TEXT) squad (TEXT) opponent (TEXT) start (TEXT) pos (TEXT) min (TEXT) sota (TEXT) ga (TEXT) saves (TEXT) save_ (TEXT) cs (TEXT) psxg (TEXT) pkatt (TEXT) pka (TEXT) pksv (TEXT) pkm (TEXT) cmp (TEXT) att (TEXT) cmp_ (TEXT) att_gk_ (TEXT) thr (TEXT) launch_ (TEXT) avglen (TEXT) att_1 (TEXT) launch__1 (TEXT) avglen_1 (TEXT) opp (TEXT) stp (TEXT) stp_ (TEXT) _opa (TEXT) avgdist (TEXT) match_report (TEXT) player_id (TEXT)
player	id (INTEGER) player_name (TEXT) full_name (TEXT) position (TEXT) footed (TEXT) height_cm (TEXT) weight_kg (TEXT) birth_date (TEXT)

	birth_place (TEXT) national_team (TEXT) club (TEXT) photo_url (TEXT) player_id (TEXT)
--	---

5. RESULTADOS E DISCUSSÕES

Grande parte do esforço de desenvolvimento e pesquisa deste trabalho foi pautado na aplicação de técnicas e ferramentas de aprendizado de máquina, com ênfase em métodos de redução de dimensionalidade, padronização de variáveis e agrupamento probabilístico. O objetivo central foi identificar padrões latentes no desempenho dos atletas e, a partir deles, propor uma classificação interpretável dos diferentes estilos de jogo presentes nas principais posições: atacantes, meio-campistas, defensores e goleiros.

Para a análise e classificação dos atletas, foi desenvolvido um *pipeline* de processamento e modelagem de dados voltado à identificação de padrões de desempenho em diferentes posições: goleiros, defensores, meio-campistas e atacantes. Inicialmente, realizou-se a padronização das variáveis por meio da técnica *StandardScaler*, que transforma os dados para que apresentem média igual a zero e desvio padrão igual a um. Essa técnica de normalização é necessária para que todas as métricas tenham a mesma escala, evitando que variáveis com valores numericamente mais altos, como **passes longos**, exerçam influência desproporcional sobre outras de menor amplitude, como **desarmes por 90 minutos**. Dessa forma, todas as variáveis passam a contribuir de maneira equilibrada nas etapas seguintes da análise.

Com os dados devidamente normalizados, aplicou-se a técnica de Análise de Componentes Principais (PCA *Principal Component Analysis*), um método estatístico de redução de dimensionalidade amplamente utilizado em aprendizado de máquina e análise exploratória. O PCA transforma o conjunto original de variáveis em um novo conjunto de variáveis não correlacionadas, chamadas de componentes principais, que são combinações lineares das variáveis originais. Cada componente principal explica uma fração da variância total dos dados, sendo que as primeiras componentes capturam a maior parte dessa variabilidade. Em outras palavras, o PCA permite representar informações complexas de maneira simplificada, reduzindo a quantidade de variáveis necessárias para descrever o comportamento dos jogadores sem perda significativa de informação. Essa redução facilita tanto a visualização dos dados quanto a eficiência dos algoritmos subsequentes.

Além do PCA, foi utilizada a técnica UMAP (*Uniform Manifold Approximation and Projection*), que também tem como objetivo reduzir a dimensionalidade dos dados, porém de forma não linear. Diferentemente do PCA, que trabalha com relações lineares entre variáveis, o UMAP busca preservar a estrutura topológica dos dados, ou seja, mantém a proximidade entre elementos semelhantes no espaço reduzido. Isso é alcançado através da construção de um grafo de vizinhança, que modela as relações locais entre os dados originais, e de sua projeção em um espaço de menor dimensão. Essa técnica é especialmente útil para visualizações e detecção de agrupamentos, pois representa de forma mais fidedigna a relação entre jogadores com características semelhantes, mesmo quando essas relações são complexas ou não lineares.

Antes da aplicação dos métodos de redução de dimensionalidade, foram criados escores agregados a partir das métricas brutas dos atletas. Esses escores representam dimensões específicas de desempenho, como **defesa de chutes**, **distribuição**, **criação de jogadas** ou **finalização**, e foram obtidos por meio da média padronizada das variáveis associadas a cada aspecto técnico. Essa etapa, conhecida como engenharia de atributos (*feature engineering*), tem como finalidade simplificar a análise, reduzir redundâncias e criar indicadores mais interpretáveis, que sintetizam o comportamento de cada jogador em determinadas dimensões do jogo.

A definição dos escores agregados foi guiada por critérios técnicos relacionados às funções e responsabilidades táticas de cada posição em campo. A ideia central foi transformar métricas individuais frequentemente redundantes ou altamente correlacionadas em indicadores compostos que sintetizam aspectos específicos do desempenho, permitindo comparações mais significativas entre atletas. Essa abordagem buscou respeitar as características funcionais de cada papel (goleiro, defensor, meio-campista e atacante), agrupando variáveis que representam dimensões complementares do jogo.

No caso dos atacantes, os escores foram construídos com foco em mensurar sua eficiência e contribuição nas fases finais do campo. O **score de finalização** concentra métricas relacionadas à produção ofensiva direta, como gols, chutes e xG (*expected goals*), refletindo a capacidade de converter oportunidades. O **score de criação de chances** agrega variáveis de passe e construção de jogadas que

antecedem finalizações, como assistências e passes para zonas decisivas, avaliando a habilidade do atacante em gerar perigo coletivo. O **score de pressão** contempla indicadores defensivos, como desarmes e interceptações, traduzindo a contribuição sem bola. Já **disciplina** e **participação no terço ofensivo** mensuram, respectivamente, a conduta tática e o envolvimento em zonas avançadas do campo, enquanto o **score de dribles** avalia a capacidade individual de romper linhas por condução.

Para os meio-campistas, os escores refletem a versatilidade e o papel de transição entre defesa e ataque. O **score de criação e visão de jogo** inclui variáveis de passes decisivos e em profundidade, representando a capacidade criativa. O **score de controle e segurança** foca na precisão e volume de passes, avaliando a regularidade na manutenção da posse. O **score de contribuição ofensiva direta** contempla ações que resultam em finalizações ou gols, enquanto o **score de contribuição defensiva** mede o envolvimento em desarmes, interceptações e bloqueios. As dimensões de **participação por terço** (defensivo, central e ofensivo) descrevem o envolvimento espacial do jogador, permitindo diferenciar perfis de meio-campistas mais recuados, centrais ou ofensivos.

Nos defensores, a construção dos escores priorizou o equilíbrio entre solidez defensiva e contribuição construtiva. O **score de solidez defensiva** agrega métricas de interrupção e bloqueio de jogadas, medindo a eficiência em neutralizar ataques adversários. O **score de duelos de borda** captura o desempenho em confrontos diretos, especialmente relevantes para as laterais. Já a **distribuição curta/média** e **progressão de jogo** avaliam a qualidade da saída de bola, a primeira pela precisão em passes curtos e médios, e a segunda pela capacidade de progredir o jogo por passes longos ou conduções. O **score de participação ofensiva** mede o envolvimento criativo de defensores que apoiam o ataque, enquanto **envolvimento na posse** reflete o volume de ações com a bola. Por fim, o **score de disciplina e risco** quantifica faltas e erros, indicando estabilidade e tomada de decisão sob pressão.

Para os goleiros, os escores foram desenhados de acordo com as principais dimensões técnicas da posição. O **score de defesa de chutes** (*shot stopping*) agrupa métricas como número de defesas e gols esperados pós-chute (*PSxG*), mensurando a eficiência sob pressão. O **score de controle aéreo e saída da área**

representa a capacidade do goleiro de atuar como **líbero**, interceptando cruzamentos e antecipando jogadas fora da pequena área. Por fim, o **score de confiabilidade e pênaltis** avalia consistência geral e desempenho em situações críticas, incluindo jogos sem sofrer gols e defesas de penalidade.

A construção desses agrupamentos foi orientada tanto por fundamentos teóricos da análise de desempenho esportivo quanto por estudos aplicados em *scouting* e métricas avançadas do futebol. Dessa forma, cada *score* sintetiza uma dimensão tática mensurável, facilitando a interpretação e a comparação entre atletas que exercem funções semelhantes, além de servir como base para as etapas seguintes de redução de dimensionalidade e clusterização.

Na etapa de agrupamento, foi empregado o modelo GMM (*Gaussian Mixture Model*), um algoritmo de *clustering* probabilístico que identifica grupos de elementos com base em suas similaridades estatísticas. O GMM assume que os dados podem ser modelados como uma combinação de múltiplas distribuições normais (gaussianas), cada uma representando um *cluster*. A partir disso, o modelo estima a probabilidade de cada jogador pertencer a determinado grupo. Diferentemente de métodos mais tradicionais, como o *K-Means*, que realizam uma atribuição rígida (um jogador pertence apenas a um grupo), o GMM adota uma abordagem probabilística, permitindo que um jogador apresente diferentes graus de associação a múltiplos *clusters*. Essa característica é particularmente vantajosa em contextos esportivos, onde as fronteiras entre estilos de jogo são frequentemente difusas.

Para determinar o número ideal de grupos, foram empregados três critérios de avaliação: AIC (*Akaike Information Criterion*), BIC (*Bayesian Information Criterion*) e o *Silhouette Score*. Tanto o AIC quanto o BIC são métricas baseadas em teorias de informação que medem o equilíbrio entre o ajuste do modelo e sua complexidade. O AIC busca minimizar a perda de informação ao ajustar o modelo aos dados, penalizando modelos excessivamente complexos. Já o BIC, de natureza mais conservadora, aplica uma penalização mais severa para modelos com muitos parâmetros, privilegiando soluções mais simples e generalizáveis. Em ambos os casos, valores menores indicam modelos mais adequados. O *Silhouette Score*, por sua vez, é uma métrica geométrica que avalia o quão bem os elementos estão agrupados. Ele mede a coesão interna de cada *cluster* e o grau de separação entre *clusters* distintos, assumindo valores entre -1 e 1. Valores

próximos de 1 indicam que os jogadores estão bem agrupados dentro de seus respectivos *clusters* e bem separados dos demais; valores próximos de 0 indicam sobreposição; e valores negativos sugerem classificações incorretas.

A análise final dos resultados obtidos com o modelo de clusterização revelou uma segmentação clara entre os diferentes perfis de atletas, permitindo compreender melhor os papéis desempenhados dentro de campo e as características predominantes em cada grupo. A combinação entre padronização, redução de dimensionalidade (PCA e UMAP) e agrupamento probabilístico possibilitou uma separação coerente e interpretável dos jogadores, evidenciando tanto padrões de desempenho quanto comportamentos táticos específicos por posição.

No grupo dos atacantes, foram identificados nove perfis distintos. Os jogadores classificados como: Finalizadores Oportunistas demonstram excelente posicionamento e capacidade de converter chances em gols, atuando de forma mais direta e com poucos toques. Já os Pressionadores Intensos se destacam pela alta intensidade sem bola, forçando erros defensivos e contribuindo para a recuperação de posse no campo adversário. O perfil Criador Mestre & Driblador representa os atacantes mais técnicos e criativos, com destaque para assistências, dribles e criação de jogadas. Os Atacantes All-Around exibem versatilidade, combinando finalização, movimentação e participação na construção, funcionando como peças completas no setor ofensivo.

Entre os perfis de menor impacto, há os Reservas: Baixo Rendimento e Reservas: Impacto Neutro, jogadores que têm menor contribuição em métricas ofensivas e defensivas, atuando de forma discreta quando acionados. O Reserva: Artilheiro de Áreas representa um subgrupo com boa eficiência em pequenas amostras de minutos, mas limitado fora da área. Já o Finalizador Hiper-Eficiente aparece como um outlier, reunindo alta precisão nas finalizações e desempenho acima da média em poucas oportunidades. Por fim, o perfil Anomalia Disciplinar identifica atacantes com boas métricas ofensivas, mas penalizados por comportamento irregular, como número elevado de faltas, cartões ou perdas de bola em excesso.

Entre os meio-campistas, seis perfis principais emergiram: O Meia Armador de Elite (Camisa 10) representa os criadores centrais de alto impacto, responsáveis

por passes decisivos, controle de ritmo e contribuições diretas para gols. O Volante Defensivo Clássico concentra suas ações na proteção da defesa, com alto volume de desarmes e interceptações. O Box-to-Box / Meia de Ligação é um perfil dinâmico, que atua em ambas as fases do jogo e se destaca pela transição entre defesa e ataque. Já os perfis Neutro / Leve Inclinação Ofensiva e Neutro / Sem Inclinação Ofensiva abrangem jogadores equilibrados, que participam de forma estável, mas sem grande destaque em nenhuma fase. Por fim, o *Outlier* Centralizado reúne casos atípicos, de meio-campistas com métricas incomuns, seja por papel tático específico ou desempenho fora da curva.

Na defesa, a clusterização identificou sete perfis. O Defensor Sólido e de Baixo Risco caracteriza atletas consistentes, com boa leitura de jogo e baixo índice de erros. O Defensor Agressivo (Alto Risco) se destaca pela intensidade nos duelos e interceptações, mas pode comprometer o sistema com entradas precipitadas. Os Laterais/Alas Ofensivos de Elite são defensores com participação significativa no ataque, contribuindo com cruzamentos e passes-chave. Já os Construtores Criativos/Ofensivos e Construtores Sólidos de Posse representam zagueiros e laterais com papel na iniciação de jogadas, variando entre estilos mais criativos e seguros.

Os Reservas Passivos/Lentos e Reservas Agressivos/Inclinado formam subgrupos de atletas com menor minutagem, sendo o primeiro mais conservador e o segundo mais propenso a ações físicas intensas e faltas.

Por fim, entre os goleiros, cinco perfis principais foram encontrados. O Goleiro Líbero (*Sweeper-Keeper*) é caracterizado por sua participação ativa na construção e nas coberturas fora da área, em linha com o estilo moderno da posição. O Goleiro Consistente / *All-Rounder* apresenta equilíbrio entre fundamentos defensivos e jogo com os pés, mantendo desempenho estável. O perfil O Muro / Estrela na Defesa Pura engloba arqueiros com alto índice de defesas difíceis e impacto direto na prevenção de gols. Já o Goleiro Mediano / Focado na Linha representa jogadores mais conservadores, com menor envolvimento na saída de bola. Por último, o Goleiro em Risco / Ineficiente agrupa atletas com desempenho abaixo da média, seja por baixa taxa de defesas ou erros recorrentes.

Estes resultados consolidam uma visão detalhada e objetiva do desempenho dos atletas por função, permitindo comparar estilos, identificar talentos promissores

e compreender como diferentes perfis se complementam dentro de uma equipe. A abordagem baseada em dados mostrou-se eficiente não apenas para classificar jogadores, mas também para traduzir o comportamento em campo em métricas interpretáveis.

6. CONSIDERAÇÕES FINAIS

O presente Trabalho de Conclusão de Curso (TCC), intitulado Sistemas de Análise de Scout no Futebol Brasileiro, propôs o desenvolvimento de uma ferramenta digital orientada a dados, visando aprimorar o processo de identificação de talentos no Brasil. A proposta surgiu como uma alternativa viável para mitigar as limitações enfrentadas por clubes de pequeno e médio porte no Brasil, que tradicionalmente dependem de métodos de *scouting* baseados em observação subjetiva.

O objetivo central foi coletar, organizar e analisar métricas de desempenho de jogadores com o intuito de subsidiar decisões estratégicas e promover maior objetividade na seleção de atletas. O trabalho adotou uma abordagem exploratória qualitativa e utilizou dados estatísticos acessíveis da plataforma *FBref*, acessados por técnicas de *scraping*.

O desenvolvimento da aplicação foi integralmente concluída, focada em oferecer módulos de coleta, processamento e exibição de dados garantindo consultas, filtros e uma interpretação automatizada das métricas. As funcionalidades do sistema incluem a visualização de fichas de jogadores com navegação por abas como "Média Geral", "Estatística Geral", "Partidas" e "Análise do Jogador".

O maior esforço de pesquisa e desenvolvimento concentrou-se na aplicação de técnicas de aprendizado de máquina para a análise de desempenho. Foi implementado um pipeline de processamento que incluiu a padronização das variáveis através do *StandardScaler* e a redução de dimensionalidade utilizando PCA (*Principal Component Analysis*) e UMAP (*Uniform Manifold Approximation and Projection*).

Um resultado crucial foi a engenharia de atributos (*feature engineering*), que permitiu a criação de escores agregados específicos para cada posição (atacantes, meio-campistas, defensores e goleiros). Esses escores, como "score de finalização" para atacantes e "score de solidez defensiva" para defensores, sintetizaram o comportamento dos jogadores em dimensões táticas mensuráveis, facilitando a interpretação.

A aplicação do modelo GMM (*Gaussian Mixture Model*), um algoritmo de clustering probabilístico, revelou uma segmentação clara e coerente entre os diferentes perfis de atletas. Essa clusterização permitiu a identificação de perfis distintos em cada setor, como os Finalizadores Oportunistas e Criadores Mestres & Dribladores entre os atacantes, o Meia Armador de Elite (Camisa 10) e o Volante Defensivo Clássico entre os meio-campistas, e o Goleiro Líbero e O Muro / Estrela na Defesa Pura entre os goleiros.

Em suma, a abordagem baseada em dados demonstrou ser eficiente para classificar e traduzir o comportamento em campo em métricas interpretáveis. Isso possibilita a comparação de estilos, a identificação de talentos promissores e o acompanhamento contínuo da evolução dos atletas, contribuindo para a profissionalização do *scouting*.

O desenvolvimento deste sistema representa um passo significativo para a democratização do acesso à análise de dados no futebol brasileiro. Este TCC reforça que, quando os dados são coletados, tratados e analisados de forma estruturada, eles se transformam em informação confiável, tornando o processo de identificação de talentos mais eficiente e justo para clubes com recursos limitados.

7. REFERÊNCIAS

BERGKAMP, Tom L. G. et al. How soccer scouts identify talented players. *European Journal of Sport Science*, v. 22, n. 7, p. 994–1004, 2022.

CARLING, C., BLOOMFIELD, J., Nelsen, L. et al. The Role of Motion Analysis in Elite Soccer. *Sports Med* 38, 839–862, 2008.

CHANDRA, B.; JENNET SHINNY, D.; KESHAV ADHITYA, M. Prediction of football player performance using machine learning algorithm. *Research Square*, 2024.

FBREF. *FBref.com – Football Statistics and History*. Sports Reference LLC. Disponível em: <https://fbref.com/>. Acesso em: 06 maio. 2025.

MACEDO; LEITE. Scout como um instrumento avaliativo do treinamento esportivo nas categorias de base do futebol. *Revista Brasileira de Futebol*, 2009

REIS, Pedro Henrique Ribeiro. A realidade das categorias de base do Clube Atlético Mineiro: os desafios de se tornar profissional. 2019. Monografia (Bacharelado em Jornalismo) – Centro Universitário do Sul de Minas, Varginha, 2019.

FÁVERO, Luiz Paulo. Manual de análise de dados. São Paulo: Elsevier, 2017.