

RELATÓRIO FINAL

PROJETO INTEGRADOR II

NEXFUT – Plataforma de Inteligência
Esportiva

Trabalho desenvolvido pelos estudantes:
Pedro Aragão; Manuela Borges; Camila Niederauer;
Eduardo Hirle; Matheus Lira; Arthur Rocha.

SUMÁRIO

1. Introdução
2. Contextualização
3. Demanda do Stakeholder
4. Problematização
5. Metodologia
6. Arquitetura e Fluxo do Projeto
7. Desenvolvimento e Evidências Técnicas
8. Resultados Obtidos
9. Conclusões e Recomendações
10. Link do Repositório GitHub

1. Introdução

O futebol ocupa um lugar de destaque no imaginário e na realidade social brasileira. Para milhões de jovens, especialmente aqueles provenientes de regiões periféricas, o esporte representa mais do que lazer: é uma possibilidade concreta de transformação de vida, inclusão social e ascensão econômica. No entanto, apesar do vasto contingente de talentos existentes, uma parcela significativa desses jovens permanece invisível aos olhos de clubes, projetos esportivos e profissionais de scouting. A ausência de mecanismos padronizados de avaliação, a limitação de recursos tecnológicos e a forte dependência de observações subjetivas contribuem para a perda de potenciais atletas que poderiam se desenvolver em ambientes adequados.

Paralelamente, observa-se um avanço significativo no uso de tecnologia nos grandes centros esportivos do mundo. Clubes profissionais têm adotado soluções baseadas em dados, algoritmos de análise de desempenho, visão computacional e estatísticas avançadas para embasar decisões técnicas e estratégicas. Porém, essa realidade ainda está distante das categorias de base, especialmente em ambientes escolares ou comunitários, onde processos de avaliação permanecem rudimentares e pouco sistematizados. Essa lacuna evidencia a necessidade de aproximar ciência de dados e esporte desde as etapas iniciais da formação dos atletas.

Nesse contexto, o projeto **NexFut** surge como uma iniciativa de inteligência esportiva voltada à democratização do acesso a informações relevantes para o desenvolvimento e identificação de jovens jogadores. A proposta consiste em estruturar uma base de dados real, coletada em escolas do Distrito Federal, e utilizá-la como ponto de partida para análises exploratórias capazes de revelar padrões, destacar perfis específicos e oferecer indicadores que subsidiem processos de observação técnica. Assim, o NexFut busca contribuir para um ecossistema esportivo mais transparente, objetivo e inclusivo, no qual decisões relacionadas à captação de talentos sejam fundamentadas em evidências e não apenas em impressões isoladas.

O trabalho desenvolvido está inserido no âmbito do Projeto Integrador II e reflete o esforço do grupo em unir fundamentos de ciência de dados, estatística e análise esportiva. A partir de técnicas de tratamento, limpeza e exploração de dados, foram elaborados gráficos, métricas e interpretações focadas em compreender o perfil dos atletas avaliados. Embora esta etapa não contemple ainda a implementação de um protótipo funcional da plataforma, ela estabelece os alicerces necessários para tal desenvolvimento, permitindo visualizar de forma clara como o NexFut pode evoluir para uma ferramenta real de apoio ao scouting no futuro.

Dessa forma, a introdução do NexFut representa não apenas o início de um relatório técnico, mas a apresentação de um projeto que busca alinhar tecnologia,

esporte e impacto social, oferecendo um caminho promissor para ampliar a visibilidade e o acompanhamento do desempenho de jovens atletas brasileiros.

2. Contextualização

O cenário esportivo contemporâneo tem passado por um processo acelerado de modernização, impulsionado principalmente pelo avanço das tecnologias de análise de desempenho e pela crescente adoção de metodologias baseadas em dados. Em clubes profissionais, decisões relacionadas a contratações, estratégias de jogo, avaliação física e identificação de talentos vêm se tornando cada vez mais orientadas por indicadores quantitativos, diminuindo o espaço antes ocupado exclusivamente por avaliações subjetivas.

Entretanto, esse movimento ainda não se reflete integralmente na base do futebol brasileiro. Em escolas, centros esportivos comunitários e projetos sociais, grande parte das avaliações se mantém dependente da observação empírica, sem padronização de métricas ou acesso a ferramentas que permitam registrar, comparar e interpretar o desempenho de jovens atletas ao longo do tempo. Essa limitação reforça desigualdades estruturais: enquanto alguns atletas contam com suporte qualificado, outros permanecem invisíveis ao radar de profissionais do meio esportivo.

Foi nesse contexto que o **NexFut** se desenvolveu ao longo do semestre. O grupo teve acesso a uma base de dados real, coletada em escolas do Distrito Federal, composta por variáveis físicas, técnicas e regionais de jovens atletas. A partir desse material, tornou-se possível elaborar análises exploratórias consistentes, identificar padrões relevantes e compreender perfis de jogadores que poderiam despertar interesse de olheiros, clubes e projetos de formação.

Além de seu caráter técnico, o NexFut é dotado de forte dimensão social. Ao estruturar e analisar dados que, em contexto cotidiano, seriam descartados ou subutilizados, o projeto contribui para a construção de um ambiente mais equitativo, ampliando a visibilidade de jovens talentos e oferecendo subsídios objetivos para a continuidade de sua formação esportiva. A contextualização do projeto revela, portanto, uma interseção entre inovação tecnológica e compromisso com a inclusão social, pilares fundamentais para a transformação do ecossistema esportivo brasileiro.

3. Demanda do Stakeholder

O stakeholder fictício definido pelo grupo representa uma **empresa especializada na coleta, organização e disponibilização de dados de jovens atletas**, atuando como intermediária entre escolas/projetos sociais e clubes profissionais. Sua demanda central foi a criação de uma solução inicial que permitisse:

1. **Estruturar adequadamente uma base de dados real**, garantindo limpeza, padronização e coerência entre atributos;
2. **Realizar análises exploratórias capazes de gerar insights relevantes para scouting**, especialmente no que diz respeito a características físicas, técnicas e regionais dos jogadores;
3. **Disponibilizar um ambiente computacional funcional**, por meio de um notebook organizado, que pudesse servir como ponto de partida para uma futura plataforma de inteligência esportiva.

A empresa, enquanto stakeholder, busca uma solução escalável, que possa apoiar decisões de captação e acompanhamento de atletas ainda nas fases iniciais de formação. A demanda contempla não apenas a apresentação de gráficos e métricas, mas também um olhar interpretativo sobre os dados, permitindo extrair conclusões que subsidiem estratégias esportivas e ações de monitoramento contínuo.

4. Problematização

Apesar do vasto potencial esportivo existente no Brasil, especialmente entre crianças e adolescentes, há um conjunto de barreiras estruturais que dificultam o desenvolvimento pleno desses talentos. A ausência de dados organizados é uma das principais limitações do processo de formação. Muitos atletas são avaliados com base em percepções pontuais, condicionadas ao olhar subjetivo do avaliador, ao contexto de um único jogo ou às condições momentâneas de desempenho.

Outro desafio crítico é a desigualdade regional. Projetos e escolas situadas em áreas periféricas frequentemente não contam com recursos para registrar indicadores técnicos de forma sistemática, o que reduz ainda mais a chance de que esses jovens sejam percebidos por clubes ou profissionais capacitados. Assim, oportunidades são perdidas e carreiras promissoras deixam de ser desenvolvidas.

Além disso, sem métricas padronizadas, torna-se difícil comparar atletas, acompanhar evolução, identificar comportamentos recorrentes e reconhecer características que possam ser trabalhadas tecnicamente. O cenário demanda soluções tecnológicas que não apenas registrem dados, mas também gerem interpretação útil, clara e replicável.

O NexFut se propõe justamente a enfrentar esse conjunto de problemas. Ao estruturar uma base real de dados e aplicar análises exploratórias, o projeto evidencia como indicadores simples — como idade, posição, altura, região, disputas vencidas, recuperações e dribles certos — podem revelar padrões valiosos para o processo de scouting. A problematização destaca, portanto, a necessidade de transformar informação dispersa em conhecimento acessível, democratizando o uso de dados no esporte de base.

5. Metodologia

A metodologia aplicada no desenvolvimento do NexFut foi orientada pelos princípios da ciência de dados e pela necessidade de criar um ambiente analítico consistente, reproduzível e capaz de gerar interpretações úteis do ponto de vista esportivo. O trabalho foi estruturado em etapas sequenciais, que garantiram organização, clareza e rastreabilidade dos processos técnicos.

5.1 Coleta e estruturação dos dados

O ponto de partida foi a disponibilização de uma base de dados real, originada de escolas do Distrito Federal. Essa base incluía atributos demográficos (idade, região), físicos (altura), técnicos (dribles certos, passes decisivos, disputas vencidas, recuperações) e posicionalidade (posição em campo e pé dominante). Por se tratar de dados reais e heterogêneos, a primeira etapa metodológica consistiu na inspeção detalhada da estrutura original e identificação de inconsistências, formatos divergentes ou valores ausentes.

5.2 Tratamento e limpeza da base

Foram aplicadas técnicas de pré-processamento utilizando Python e bibliotecas essenciais como **Pandas**, garantindo:

- Conversão de tipos de dados;
- Padronização de nomes, categorias e unidades;
- Eliminação de duplicidades;
- Correção de inconsistências semânticas;
- Normalização de colunas categóricas para facilitar agrupamentos e comparações.

Essa etapa foi fundamental para assegurar integridade e confiabilidade às análises subsequentes.

5.3 Análise Exploratória de Dados (EDA)

Com a base tratada, iniciou-se a fase de exploração, contemplando cálculos estatísticos e geração de visualizações. Foram empregados **Matplotlib e Seaborn** para criar gráficos de distribuição, dispersão, correlação e comparações por categoria. A EDA permitiu compreender a estrutura do conjunto de dados, identificar padrões, outliers e faixas de desempenho que pudessem ser relevantes para o contexto esportivo.

5.4 Interpretação orientada ao esporte

Diferentemente de uma análise puramente técnica, a interpretação dos gráficos foi conduzida a partir de um olhar esportivo: perfis físicos adequados a determinadas posições, indicadores que sinalizam potencial ofensivo ou defensivo, distribuição de talentos por região e padrões que podem subsidiar decisões de scouting. Essa etapa agregou valor ao trabalho, aproximando dados e prática esportiva.

5.5 Consolidação dos resultados

Ao final das análises, foram sistematizados insights, limitações e oportunidades de evolução do projeto, que servirão como base para futuras etapas do NexFut, incluindo desenvolvimento de dashboards, protótipo de plataforma e eventuais modelos preditivos.

6. Arquitetura e Fluxo do Projeto

Embora esta etapa do projeto não tenha contemplado a implementação de uma plataforma completa, foi estruturado um fluxo metodológico e técnico que serviu como base para a organização do repositório e para o desenvolvimento das análises.

6.1 Estrutura da Arquitetura

A arquitetura pode ser resumida em três camadas principais:

a) Camada de Dados

Responsável pela ingestão e armazenamento do conjunto de dados bruto. Inclui:

- Arquivo CSV com atributos dos atletas;
- Organização da pasta Dados no repositório GitHub;
- Documentação sobre origem e características da base.

b) Camada de Processamento e Análise

Executada via Jupyter Notebook (analise_nexfut.ipynb), engloba:

- Leitura e inspeção da base;
- Pré-processamento e limpeza;

- Geração de gráficos;
- Análises estatísticas e interpretações.

c) Camada de Resultados e Visualização

Inclui:

- Gráficos gerados pela EDA;
- Sumários estatísticos;
- Insights esportivos;
- Preparação das informações para apresentação e relatório.

6.2 Fluxo Operacional

1. **Importação da base de dados**
2. **Tratamento e refinamento**
3. **Exploração de atributos-chave**
4. **Modelagem visual e estatística**
5. **Documentação dos resultados**
6. **Interpretação esportiva e identificação de padrões**

Esse fluxo permite que o projeto seja expandido facilmente para incluir novas camadas no futuro, como APIs de captura automática, dashboards interativos e modelos preditivos.

7. Desenvolvimento e Evidências Técnicas

O desenvolvimento do NexFut ocorreu de maneira iterativa ao longo do semestre, com foco em consolidar uma base de dados confiável e extrair informações relevantes para a avaliação esportiva de jovens atletas.

7.1 Organização do repositório

O GitHub foi utilizado como ferramenta central para versionamento e colaboração. A estrutura inclui:

- Pasta **Dados**: arquivos CSV brutos e tratados;
- Pasta **Código**: notebook com todas as análises;
- Pasta **Documentação**: materiais referentes ao PI2;

- README contendo instruções gerais.

7.2 Notebook Analítico

O notebook `analise_nexfut.ipynb` funciona como núcleo técnico do projeto. Ele contém:

- Importações bibliotecárias;
- Carga da base;
- Estatísticas iniciais;
- Geração dos principais gráficos;
- Interpretações relevantes.

7.3 Evoluções realizadas durante o semestre

- **Ampliação da base:** recebimento de dados reais com mais atributos que o previsto inicialmente.
- **Refinamento:** ajustes necessários para padronização, categorização e consistência dos dados.
- **Manutenção do escopo:** decidiu-se manter a entrega centrada na análise, postergando o protótipo da plataforma para o próximo semestre.

7.4 Evidências técnicas

Incluem:

- Gráficos de distribuição (idade, altura, posição, região).
- Gráficos de radar para perfis ofensivos, defensivos e criativos.
- Correlação entre métricas defensivas (recuperações × disputas vencidas).
- Rankings individuais (dribles certos).
- Estatísticas descritivas de cada grupo de atletas.

Essas evidências reforçam a confiabilidade da análise e a conexão com o contexto de scouting esportivo.

8. Resultados Obtidos

Os resultados alcançados permitem uma visão multifacetada do perfil dos atletas analisados e oferecem diretrizes claras que podem beneficiar processos de captação e acompanhamento técnico.

8.1 Perfil demográfico e físico

As análises revelaram:

- Concentração etária entre 12 e 14 anos — faixa crítica para desenvolvimento cognitivo e motor.
- Altura média próxima de 155 cm, com variação esperada para a idade.
- Domínio de jogadores destros, com minoria canhota — característica taticamente valiosa.

8.2 Distribuição posicional e regional

- Predomínio de meias e atacantes, indicando maior interesse dessas posições no contexto escolar.
- Regiões como Gama, Ceilândia e Samambaia apresentaram maior densidade de participação, sugerindo potenciais polos de captação.

8.3 Desempenho técnico

- Jogadores ofensivos se destacam por volume de dribles e passes decisivos.
- Defensores apresentam correlação clara entre recuperações e disputas vencidas — indicador de leitura de jogo.
- Os gráficos de radar facilitaram a comparação de perfis para análise de perfis completos ou especializados.

8.4 Insights relevantes para scouting

- Idade influencia diretamente o pico de criatividade e tomada de decisão.
- Atletas que combinam bom drible e boa visão de jogo tendem a apresentar maior potencial de projeção.
- Métricas defensivas apontam perfis promissores para posições de volante e zagueiro.

9. Conclusões e Recomendações

O NexFut demonstrou que a análise orientada por dados pode transformar significativamente a maneira como jovens atletas são avaliados. Mesmo com recursos simples e uma base inicial, foi possível estruturar indicadores consistentes, criar visualizações informativas e extrair insights com valor real para o contexto esportivo.

A conclusão geral do projeto é que **há enorme potencial na aplicação de ciência de dados ao esporte de base**, tanto do ponto de vista de desempenho quanto de inclusão social.

10. Repositório do Projeto

Todo o código-fonte, documentação complementar, arquivos de dados utilizados e histórico de versionamento da NexFut encontram-se disponibilizados de forma pública no repositório oficial do grupo no GitHub. O repositório reúne o notebook de análise exploratória, a base de dados organizada, os gráficos gerados durante o desenvolvimento e demais artefatos utilizados ao longo do semestre.

O acesso pode ser realizado por meio do seguinte link:

https://github.com/projeto-integrador-cdml/documenta-o-projetos-integradores-template-projeto_2_eduardo