Insper
Engenharia Mecatrônica
Relatório Final de Iniciação Tecnológica

João Vitor Oliveira Assunção

Orientador: Prof. André Luiz Maciel Santana
Co-orientador: Prof. Luiz Fernando Cardoso dos Santos Durão

Construção de métricas físicas no futebol: uma abordagem por visão computacional e ciência de dados

São Paulo
2024

Resumo

Desde a sua chegada ao Brasil em 1894, o futebol assumiu um papel protagonista no cenário esportivo nacional. A adesão ao esporte inglês foi tamanha que atualmente o Brasil é denominado como o país do futebol, sendo o maior campeão da principal competição de seleções: a Copa do Mundo. Entretanto, devido aos baixos investimentos, o futebol no Brasil se mostrou cada vez mais ultrapassado e decadente em relação às nações europeias. A Delta Goal surge como uma iniciativa para utilizar a tecnologia na ampliar as análises táticas de uma partida de futebol. Através da elaboração de uma solução computacional que seja capaz de, a partir da transmissão de uma partida de futebol, extrair dados táticos dos jogadores em campo, aplicando métricas previamente estabelecidas entre o time de pesquisa e as equipes parceiras. Para isso, utiliza-se a visão computacional para a extração de dados brutos de uma partida de futebol e o aprendizado de máquina para acessar um modelo cada vez mais fiel à realidade tática observada em uma partida. Dessa forma, os dados dos jogadores extraídos de uma partida de futebol somados com a ciência de dados, responsável por filtrar e analisar os dados brutos obtidos, são uma ótima rede de apoio para o modelo, servindo como dados de entrada, viabilizando um bom treinamento da Inteligência Artificial e maximizando os acertos. Assim, pode-se ter um comparativo entre os dados obtidos pelo programa e dados reais de uma mesma partida, possibilitando a análise da confiabilidade do código gerado. Nesse sentido, pode-se aprofundar nas métricas que exercem maior influência no desempenho dos jogadores e dos clubes, tornando o futebol mais racional.

Palavras-chave: Futebol. Inteligência Artificial. Visão Computacional. Ciência de Dados. Métricas.

Abstract

Since its arrival in Brazil in 1894, football has assumed a leading role in the national sporting scene. The support for English sport was such that Brazil is currently known as the country of football, being the greatest champion of the main national team competition: the World Cup. However, due to low investments, football in Brazil has shown itself to be increasingly outdated and decadent in relation to European nations. Delta Goal appears as an initiative to use technology to expand tactical analysis of a football match. Through the development of a computational solution that is capable of, from the broadcast of a football match, extracting tactical data from players on the field, applying metrics previously established between the research team and partner teams. To achieve this, computer vision is used to extract raw data from a football match and machine learning is used to access a model that is increasingly faithful to the tactical reality observed in a match. In this way, player data extracted from a football match combined with data science, responsible for filtering and analyzing the raw data obtained, are a great support network for the model, serving as input data, enabling good training. of Artificial Intelligence and maximizing successes. Thus, a comparison can be made between the data obtained by the program and real data from the same game, making it possible to analyze the reliability of the generated code. In this sense, we can delve deeper into the metrics that have the greatest influence on the performance of players and clubs, making football more rational.

Keywords: Soccer. Artificial Intelligence. Computer Vision. Data Science. Metrics.

Sumário

1 INTRODUÇÃO 7
1.1 Futebol no Brasil 7
1.2 A tecnologia como alicerce no futebol 7
1.3 Tecnologia e futebol: na prática. 8
1.3.1 *Seleção alemã de 2014* 8
1.3.2 *Brighton & Hove Albion* 8
1.4 Visão Computacional 10
1.5 Objetivos da pesquisa 10
1.5.1 *Métodos de pesquisa* 11
1.5.1.1 Fluxo de Desenvolvimento 11
2 MATERIAIS E MÉTODOS ENVOLVIDOS 12
2.1 Delta Goal: Área e Forma de Atuação 12
2.2 Métodos 12
2.2.1 *Instrumentos de visão computacional* 12
2.2.2 *Usos de Ciência de Dados* 14
2.2.3 *Métodos Ágeis e Kanban* 14
2.3 Materiais 15
2.3.1 *Pesquisa e revisão da literatura* 15
2.3.2 *Análise dos Movimentos de Ruptura* 16
2.3.3 *Análise e Tratamento de Dados* 16
2.3.3.1 Velocidade Média 17
2.3.3.2 Velocidade Máxima 18

2.3.3.3 Distância Percorrida 19

2.3.3.4 Aceleração 20

2.3.3.5 Resultados 21

2.3.3.6 Teste – T..... 24

3 CONCLUSÕES E RESULTADOS..... 25

3.1 Próximos Passos 30

4 REFERÊNCIAS 31

1 INTRODUÇÃO

1.1 Futebol no Brasil

Segundo Freyre (1947) em “O Negro no Futebol Brasileiro”, devido à formação cultural brasileira pautada em elementos primitivos, o futebol, diferentemente dos outros esportes, assumiu uma importância além da esportiva. De maneira natural, o futebol adquiriu uma importância cultural e se tornou uma instituição brasileira, sublimando vários elementos irracionais da formação cultural e social da sociedade brasileira.

Isso evidencia como o futebol, desde sua chegada ao Brasil, tem desempenhado um papel que vai muito além do caráter esportivo, apresentando importância sociocultural e econômica. Além disso, o futebol se tornou o esporte responsável por traduzir as emoções da população brasileira, por servir de canal para manifestações religiosas, políticas e sentimentais, seja dos jogadores ou dos torcedores. Nesse sentido, Carlos Amadeu Botelho Byington diz que “O futebol é a arte do corpo na escola das emoções”.

1.2 A tecnologia como alicerce no futebol

Embora o futebol tenha se popularizado durante a década de 30 no mundo, somente a partir dos anos 2010 o esporte vem sofrendo constantes mudanças visando o aumento da assertividade nas decisões da arbitragem. Durante o Mundial de Clubes de 2012 realizado no Japão, o mundo pôde assistir ao primeiro grande avanço tecnológico experimentado no futebol: a implementação do chip de verificação de gol. O chip implementado é capaz de verificar se toda a circunferência da bola ultrapassou a linha do gol, indicando em um relógio utilizado pelo árbitro se o gol deve ser validado ou não.

Após dois anos da implementação do chip na bola, o mundo do futebol pôde experimentar pela primeira vez a implementação do GPS utilizado por jogadores durante uma partida de futebol. Os dados extraídos pelo sistema de GPS são utilizados posteriormente pelas equipes técnicas para analisar parâmetros esperados durante o jogo.

Mais recentemente, no ano de 2018, o futebol experienciou a implementação do árbitro de vídeo: o VAR. Apesar de todas as polêmicas envolvidas na implementação do VAR, essa nova tecnologia veio para tornar as decisões mais objetivas, minimizando as questões interpretativas e possivelmente manipuláveis de lances polêmicos e determinantes de uma partida.

1.3 Tecnologia e futebol: na prática.

1.3.1 *Seleção alemã de 2014*

Durante a Copa do Mundo de 2014, realizada no Brasil, a seleção alemã se sagrou campeã pela 4ª vez na conceituada competição. Após a conquista, exaltou-se muito a qualidade técnica dos jogadores e como o futebol alemão era capaz de se impor em campo, aplicando o famoso 7 x 1 em cima da seleção brasileira. Entretanto, os placares elásticos e a grande capacidade técnica dessa seleção ofuscaram um trabalho tecnológico muito bem desenvolvido pela comissão técnica alemã. No decorrer da competição, a delegação alemã aplicou um sistema capaz de coletar e analisar os dados de uma partida ou treino. Os dados eram interpretados pela comissão técnica da seleção e os ajustes necessários eram feitos para atingir uma melhor performance. Para efeitos comparativos, a equipe foi capaz de diminuir a velocidade de seus passes de 3,4 segundos para 1,1 segundo (redução de 310%). Além disso, a cada 5,45 finalizações um gol era registrado pela equipe campeã, enquanto a seleção brasileira precisou de 10,1 finalizações para marcar. A seleção argentina, finalista junto da Alemanha, necessitou de 13,13 finalizações para registrar gols.

As estatísticas discrepantes continuam além das apresentadas, tendo como exemplos: 4157 passes registrados (813 passes a mais que a Argentina), melhor ataque com 18 gols em 7 jogos e 4 dos 5 jogadores que percorreram as maiores distâncias vestiam o uniforme alemão.

1.3.2 *Brighton & Hove Albion*

Finalizada a edição da Premier League 2022/23, muitos amantes do futebol se surpreenderam com a 6ª colocação da equipe do Brighton, ficando à frente de equipes como a do Chelsea, Leicester City, Tottenham Hotspur, Aston Villa e Nottingham Forest, e se classificando para a Liga Europa pela primeira vez na história.

Tabela 1 – desempenho do Brighton desde 2009

	Terceira divisão	Segunda divisão	Premier League
2008/09	17°		
2009/10	3°		
2010/11	1°		
2011/12		10°	
2012/13		4°	
2013/14		6°	
2014/15		19°	
2015/16		3°	
2016/17		2°	
2017/18			15°
2018/19			17°
2019/20			15°
2020/21			16°
2021/22			9°
2022/23			6°

Fonte: ASSUNÇÃO, 2023

Um dos fatores que afetou positivamente o desempenho do clube foi a aquisição feita por Tony Bloom, no ano de 2009. O ex-jogador de Poker tem hoje uma empresa de consultoria para apostas esportivas. A *Stalizard* se utiliza de modelos estatísticos extremamente complexos para calcular cotações (as "*odds*" do mercado esportivo) mais precisas do que as casas de apostas são capazes de calcular. Esse sistema extremamente eficaz é utilizado a favor do clube inglês no mercado de transferências, utilizando das estatísticas para trazer jogadores pouco visados, mas com ótimos números. Esses investimentos ousados, porém, conscientes, trouxeram grandes retornos para o clube nos últimos anos, sendo que de 2021 até abril de 2024 o clube gastou € 61,8 milhões em alguns jogadores e obteve um retorno de € 233 milhões em vendas.

Tabela 2 – Principais retornos financeiros do clube

Jogador	Valor investido	Valor recebido
Marc Cucurella (Chelsea)	€ 18 mi	€ 65 mi
Ben White (Arsenal)	Formado no clube	€ 58 mi
Yves Bissouma (Tottenham)	€ 16,8 mi	€ 29 mi
Leandro Trossard (Arsenal)	€ 15,5 mi	€ 24 mi
Dan Burn (Newcastle)	€ 3,5 mi	€ 15 mi
Alexis Mac Allister (Liverpool)	€ 8 mi	€ 42 mi

Fonte: ASSUNÇÃO, 2023

1.4 Visão Computacional

A visão computacional é um campo da ciência da computação que se dedica ao desenvolvimento de métodos e algoritmos que possibilitem que máquinas interpretem e compreendam o mundo visual. Esta área visa reproduzir a habilidade humana de extrair informações de imagens e vídeos, podendo ser aplicada em diversas áreas. Atualmente é bastante utilizada em carros autônomos, na medicina, nas indústrias e em jogos (COUTO, 2012).

Para isso, a visão computacional se apoia em princípios matemáticos e estatísticos, incluindo o processamento de imagens, que perpassa por etapas como a filtragem, detecção de bordas e segmentação. Além disso, é possível utilizar-se da inteligência artificial para identificar objetos, rostos e todos os elementos de interesse de uma cena (MOLZ, 2001).

Uma importante etapa dentro da visão computacional é a representação dos dados visuais de forma que possam ser interpretados e manipulados por algoritmos computacionais. Isto é, a imagem de interesse será convertida para um formato digital adequado para a análise a ser feita. Para tal conversão, por vezes se faz necessário a criação de modelos matemáticos adequados para a descrição da imagem ou vídeo (Ibid).

1.5 Objetivos da pesquisa

Por muitas vezes, o futebol é interpretado como um esporte no qual o único aspecto importante é o gol. A verdade é que o futebol vai muito além de simplesmente marcar gols.

Segundo o ex-jogador e bicampeão mundial, Waldir Pereira (1962), “Quem tem de correr é a bola, não o jogador.”

O aclamado ex-jogador e ex-treinador neerlandês, Johan Cruijff (1996) afirmou que, “Em uma partida de futebol, é estatisticamente provado que os jogadores têm a posse de bola por 3 minutos, em média. Então, o mais importante é: o que fazer nos 87 minutos em que você não tem a bola. Isso é o que determina se você é um bom jogador ou não.”

O intuito dessa iniciação tecnológica vai de encontro ao pensamento do brasileiro e do neerlandês, buscando simplificar e racionalizar o futebol, acredita-se que com o uso da visão computacional para o mapeamento e a detecção automáticos de jogadores, pode-se tonar o futebol mais racional e simples, nos levando cada vez mais ao ápice desse esporte: o gol!

1.5.1 *Métodos de pesquisa*

1.5.1.1 Fluxo de Desenvolvimento

O que motiva a realização da pesquisa é um questionamento: Como podemos aplicar visão computacional e análise de dados para auxiliar as comissões técnicas nas tomadas de decisões, com base em um processo de co-design?

Essa pesquisa segue o método *Kanban*, uma abordagem de gestão por meio de elementos visuais que se originou na cultura automobilística japonesa, e tem sido amplamente adotado em contextos não convencionais, como a gestão de projetos e desenvolvimento de softwares. O *Kanban* pode ser entendido como um sistema no qual há um controle do fluxo, visando otimizar a eficiência.

Em essência, o método se utiliza de quadros visuais para apresentação de tarefas que, à medida que vão sendo realizadas ou há um avanço na execução, são movidas para outra coluna de uma etapa mais avançada.

Figura 1 – Método Kanban

2 MATERIAIS E MÉTODOS ENVOLVIDOS

2.1 Delta Goal: Área e Forma de Atuação

A Delta Goal é uma empresa que se utiliza da Inteligência Artificial para inspecionar uma partida completa de futebol. Isto é, todos os momentos do jogo, até mesmo ações sem bola, são analisados de maneira automatizada, extraindo dados preciosos que trazem muitas informações valiosas para as comissões técnicas das equipes parceiras. A Delta Goal é sinônimo de inovação, trazendo para o mercado alternativas e dados que vão além do que pode ser encontrado atualmente. A adaptabilidade da empresa em apresentar dados que se moldam de acordo com o que os times necessitam, trazem um diferencial para a empresa. Tudo isso é possível graças à toda a tecnologia trazida para as análises e ao design pensado nas diferentes necessidades que cada equipe traz.

Segundo a própria Delta Goal no site oficial da empresa, esses são os princípios e objetivos do desenvolvimento do produto:

“Liberte o poder da Inteligência Artificial (IA) e da Visão Computacional (CV) para redefinir o seu jogo de futebol. Cada chute, cada passe, todos os sprints são dados esperando para serem descobertos. Com nossa plataforma de IA de ponta, o desempenho da sua equipe pode ser rastreado, analisado e aprimorado como nunca. Veja como os movimentos dos seus jogadores são registrados e transformados em insights acionáveis.

Não deixe o potencial da sua equipe inexplorado; com nossa tecnologia baseada em IA, seu treinamento pode ser personalizado, suas estratégias podem ser ajustadas e seus jogadores podem atingir seu desempenho máximo. Dê o primeiro passo em direção ao futuro do futebol e deixe-nos guiá-lo em sua jornada rumo ao próximo nível de excelência (DELTA GOAL, 2024).”

2.2 Métodos

2.2.1 *Instrumentos de visão computacional*

Inicialmente, necessita-se extrair os dados que se deseja analisar. Para isso, utilizam-se vídeos das transmissões das partidas e, através do *OpenCV* e do *YoLo* é possível extrair características comuns e identificar objetos. Entretanto, para uma análise computacional fiel à realidade é necessário fazer uma homografia.

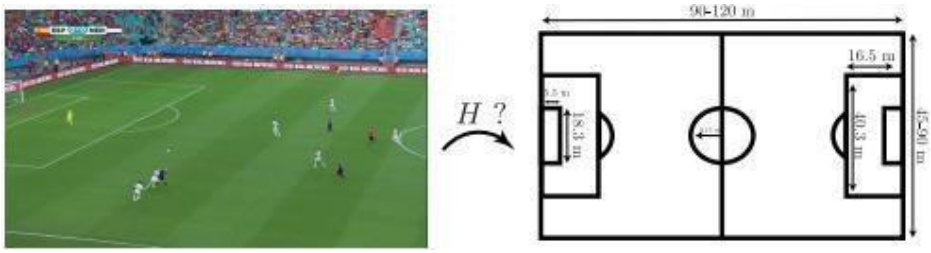

Figura 2 – Processo de homografia

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Figura 3 – Transformação feita pela homografia

A homografia se trata do processo que descreve uma relação entre dois planos ou sistemas de coordenadas distintos, preservando os pontos de fuga, linhas e pontos. Esse processo apresenta algumas importantes propriedades para o futebol, sendo elas: preservação de linhas retas, pontos correspondentes, rotação, translação, escala e projeção perspectiva.

Dessa forma, fez-se possível partir de uma filmagem que apresenta 3 dimensões e analisá-la em um campo em 2 dimensões. Assim, os resultados desse processo são tratados e organizados em uma base de dados, facilitando a extração de informações táticas por parte das comissões técnicas.

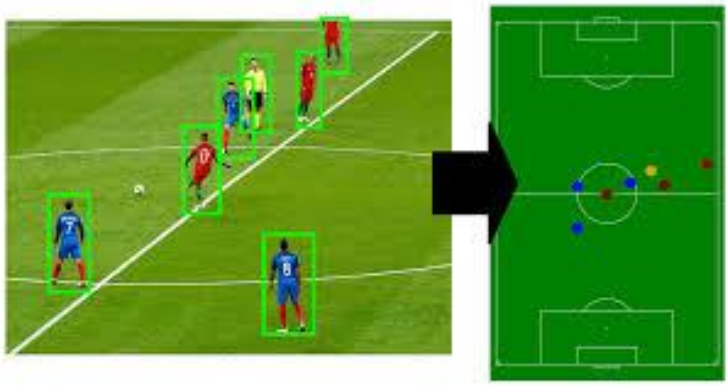

Figura 4 – Aplicação da Homografia

2.2.2 *Usos de Ciência de Dados*

A ciência de dados é uma área interdisciplinar responsável por coletar, tratar, organizar e interpretar dados. Esse processo envolve a aplicação de métodos estatísticos, análises matemáticas, programação e raciocínio lógico.

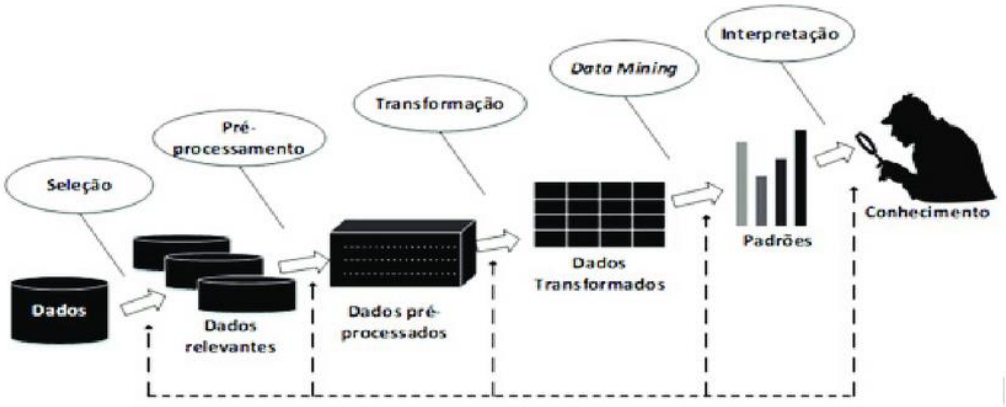

Figura 5 – Fluxo de Análise de Dados

Desse modo, partindo de dados brutos, um cientista de dados é capaz de tratá-los, organizá-los da melhor maneira para trabalhar, identificar padrões, traçar tendências, estabelecer relações matemáticas e correlacionar os dados. Após a etapa de manipulação, organiza-se os dados finais de modo a apresentá-los e torná-los um caminho para que decisões estratégicas sejam tomadas.

Utilizando-se dessa ciência, analisou-se os dados obtidos por meio da aquisição no mercado. Entre os diversos dados apresentados no *DataFrame*, selecionou-se os dados relevantes para a extração de métricas importantes do futebol, processou-se os dados, observou-se as propriedades matemáticas e estatísticas e projetou-se o algoritmo capaz de realizar, no futuro, análises em novos conjuntos de dados.

2.2.3 *Métodos Ágeis e Kanban*

Os métodos ágeis e o *Kanban* compartilham valores fundamentais, mas apresentam vieses diferentes para a abordagem de projetos e processos. Ambos buscam tornar mais flexíveis os métodos para gestão de projetos, adaptando-os de acordo com o avanço de cada setor.

Os métodos ágeis se baseiam na colaboração entre equipes e membros e na capacidade de flexibilização das metas e dos próprios indivíduos. Dessa forma, as interações sobre os processos são adaptadas de acordo com as mudanças de requisitos dos clientes e de acordo com a capacidade de avanço nas tarefas que cada indivíduo foi capaz de apresentar durante o prazo estipulado.

Por outro lado, o *Kanban* é uma abordagem voltada para a organização das metade por meio de elementos visuais, concentrando-se no fluxo contínuo de trabalho e na redução de processos que possam vir a ser desnecessários. Além disso, o *Kanban* exibe para todos os membros de uma equipe a forma de avanço que cada indivíduo apresenta e a capacidade de entrega que tem sido empregada.

O processo de desenvolvimento do algoritmo foi pautado nesses dois métodos, sendo realizado em ciclos iterativos, com apresentações semanais de avanços e a determinação de novas metas de entrega. Foram realizadas reuniões entre os membros da equipe, apresentando os avanços semanais, mostrando as dificuldades inesperadas que surgiram durante a realização da tarefa, relatando os pontos de incertezas e os pontos que apresentaram resultados positivos e, uma vez finalizadas as tarefas atuais, novas metas foram traçadas. Para o registro do *Kanban*, fez-se uso do *Trello*, separando-se os membros por equipes e atrelando a cada indivíduo entregas que se desenrolariam ao longo de uma semana. Além disso, a cada tarefa se atribuiu um valor de complexidade prévia, o qual dizia sobre o tempo de execução esperado e o apoio necessário para a entrega, podendo ser ajustado posteriormente, caso necessário.

2.3 Materiais

2.3.1 *Pesquisa e revisão da literatura*

Durante os 4 meses iniciais da pesquisa, realizou-se, juntamente com o time de desenvolvedores da Delta Goal, a leitura e a revisão de artigos pautados nos processos de homografia, visão computacional e aprendizado de máquina. A finalidade principal de todo esse processo é auxiliar os desenvolvedores sênior na elaboração dos algoritmos utilizados na empresa. Com isso, resumos dos artigos foram feitos a fim de facilitar o processo, filtrando apenas as informações relevantes para o prosseguimento do projeto. Ao longo dessa seção, serão desenvolvidos brevemente os assuntos tratados durante a revisão.

O artigo "*Soccer Field Localization from a Single Image*" (HOMAYOUNFAR; FIDLER; URTASUN, 2016) apresenta a tentativa de representar uma imagem de transmissão de um campo em uma figura de 2 dimensões (conforme o processo de homografia já explicado anteriormente). Dessa forma, divide-se o campo em 4 raios que partem de 4 pontos de fuga calculados automaticamente pelo programa. Após isso, os graus de liberdade da matriz da homografia caem pela metade, facilitando o processo.

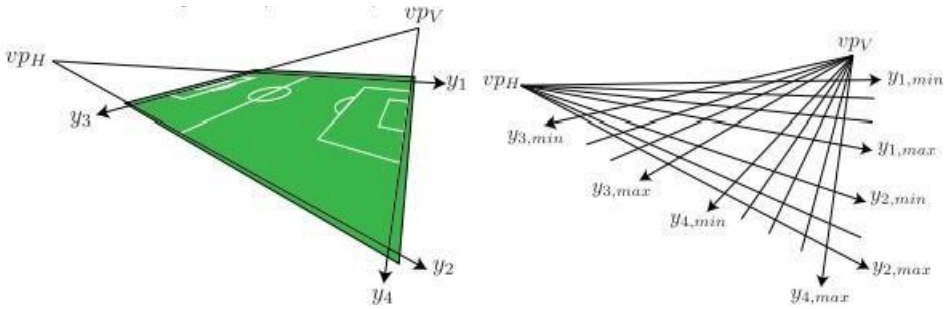

Figura 6 – Exemplo do cálculo dos pontos de fuga

A partir dos resultados obtidos, parte-se para os próximos passos. Primeiramente é necessário estabelecer os pixels de grama, em seguida identifica-se as linhas laterais e de fundo, semicírculo das pequenas áreas, círculo central, linha central, pequenas e grandes áreas e os pontos de fuga da transmissão. Após esse

processo, aplica-se o algoritmo de *Branch and Bound*, no qual são feitas ramificações que diminuem, cada vez mais, as possibilidades. Uma vez finalizado, aplicou-se imagens de transmissões da Copa do Mundo de 2014 para testar o algoritmo e os resultados agradaram.

2.3.2 *Análise dos Movimentos de Ruptura*

A fim de auxiliar clubes brasileiros e entender a complexidade de tarefas do dia a dia das comissões técnicas, outro importante tarefa foi a análise de partidas completas e a identificação de movimentos de ruptura. Denomina-se um movimento de ruptura quando a equipe de interesse está no último terço do campo, atacando e um jogador inicia uma corrida que é capaz de romper a última linha de defesa da equipe adversária. Após a identificação, registrava-se em uma planilha quais os jogadores envolvidos, qual o momento que a ruptura ocorreu, quanto tempo ela demorou para ocorrer, qual o desfecho da jogada em questão e em qual sub-zona ocorreu a ruptura.

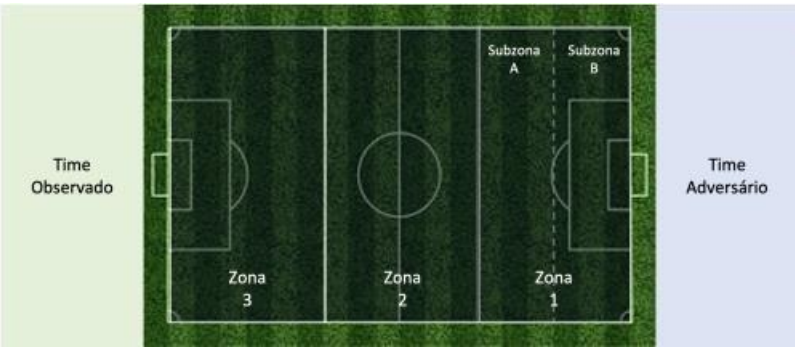

Figura 7 – Divisões e subdivisões do campo

2.3.3 *Análise e Tratamento de Dados*

O futebol é composto por diversas métricas que são importantes para as análises feitas pelas comissões técnicas. Dentre as métricas avaliadas, optou-se por examinar a velocidade média, a velocidade máxima, a distância percorrida e a aceleração de um jogador ao longo de uma partida.

Para isso, dados extraídos por visão computacional, já citado anteriormente neste artigo, foram adquiridos pela Delta Goal para servirem de base para o desenvolvimento de um código capaz de calcular essas métricas e para servirem de comparativo com os dados encontrados.

Esclarece-se que o intuito principal é que esse código posteriormente passe a ser alimentado por dados extraídos pelo processo de visão computacional.

2.3.3.1 Velocidade Média

As informações obtidas estavam organizadas em um *DataFrame* com dados como nome do jogador, tempo de jogo, frame (sendo mais de 1 frame por segundo de jogo), posições nos eixos X (horizontal) e Y (vertical), velocidade, aceleração e características de visibilidade do jogador.

Partindo dos dados organizados no *DataFrame*, iniciou-se o desenvolvimento de um código em Python com o propósito de calcular a velocidade de um jogador em cada instante durante a partida. Para isso, foram necessárias bibliotecas disponíveis em Python para auxiliar na construção do código. Elas são:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from math import *
import sys
```

Figura 8 – Bibliotecas necessárias para implementação do Algoritmo

A partida em análise foi um confronto entre Sport x Atlético Goianiense, em 14/07/2023 que durou 96 minutos. Entretanto, os dados fornecidos tinham como unidade os frames da partida, sendo 25 frames por segundo.

Dessa forma, foram analisados cerca de 76.358 frames para um jogador que tenha atuado durante toda a partida e, devido ao elevado número de dados, fez-se necessário separar o *DataFrame* por jogador.

Sabendo que as posições dos jogadores durante todos os momentos da partida, o método para se calcular o deslocamento durante um frame se baseia no Teorema de Pitágoras. O Teorema de Pitágoras (equação 1) diz que, a soma dos quadrados dos catetos é igual ao quadrado da hipotenusa.

$$h^2 = a^2 + b^2 \tag{1}$$

Trazendo para a realidade de uma partida de futebol, a soma dos quadrados dos deslocamentos em X e em Y é igual ao quadrado do deslocamento de um jogador em campo (equação 2).

$$\Delta S^2 = \Delta S_x{}^2 + \Delta S_y{}^2 \tag{2}$$

Uma vez calculado o deslocamento do jogador, pode-se calcular a velocidade instantânea. Para isso, utiliza-se a equação da velocidade (equação 3) instantânea em todos os deslocamentos obtidos durante o jogo.

$$v = \lim_{\Delta t \to 0} \frac{\Delta S}{\Delta t} \tag{3}$$

Inicialmente, percorreu-se todos os dados através de blocos de repetição e compilou-se as informações obtidas em listas do Python, que posteriormente seriam utilizadas para novos cálculos. Entretanto, após testes, verificou-se um elevado tempo de processamento que, em alguns casos, chegavam a mais de 60 minutos por jogador. A alternativa utilizada para solucionar o problema, foram as listas da biblioteca "*numpy*", que realizam as operações matemáticas de maneira eficaz e rápida, diminuindo o tempo de análise de cada jogador para alguns segundos.

Outra importante etapa necessária nesse processo foi a conversão de medidas. Uma vez que as informações são extraídas do vídeo por visão computacional, as coordenadas de posição em X e em Y são dadas em pixels e, no futebol essa métrica não traz informações significativas para as comissões técnicas. Logo, foi necessário converter as informações de pixel por frame para metros por segundo ou quilômetros por hora. Para isso, foram necessárias informações sobre quantos pixels foram utilizados na análise. Os números são expostos abaixo.

```
LARGURA_CAMPO_PIXEL = 610
ALTURA_CAMPO_PIXEL = 417
LARGURA_CAMPO_METROS = 118
ALTURA_CAMPO_METROS = 76
FRAMES_POR_SEGUNDO = 25
```

Figura 9 – Parâmetros Físicos do Campo Analisado

Logo a relação metro por pixel foi definida (equação 4 e equação 5).

$$pixel_x = \frac{118}{610}\ m \tag{4}$$

$$pixel_y = \frac{76}{417}\ m \tag{5}$$

Da mesma forma, obteve-se a relação segundo por frame foi definida (equação 6).

$$frame = \frac{1}{25}\ s \tag{6}$$

Sabendo dessas relações, pôde-se obter as velocidades em m/s. A conversão para km/h ocorre de maneira mais simples (equação 7).

$$\frac{km}{h} = \frac{m}{s} * 3{,}6 \tag{7}$$

2.3.3.2 Velocidade Máxima

A velocidade máxima de um jogador durante uma partida de futebol é um indicador crítico que não apenas avalia a qualidade física individual, mas também influencia diretamente o desempenho e os resultados da equipe. Essa métrica quantifica a

capacidade de um jogador em se deslocar rapidamente pelo campo, permitindo-lhe competir de forma eficaz em várias situações de jogo.

Em um contexto competitivo, a velocidade máxima de um jogador é determinante na capacidade de superar a defesa adversária e na habilidade de evitar ser ultrapassado por oponentes. Esse atributo é particularmente crucial em cenários de contra-ataque, onde a capacidade de um jogador em atingir sua velocidade máxima rapidamente pode resultar em oportunidades de gol decisivas. Além disso, durante momentos defensivos, a rapidez em alcançar a velocidade máxima pode ser fundamental para evitar que o adversário se distancie ou crie chances de marcar.

Além de seu impacto direto no aspecto competitivo do jogo, a velocidade máxima de um jogador também desempenha um papel significativo na minimização do risco de lesões. Jogadores que podem se mover rapidamente têm maior capacidade de responder a mudanças repentinas de direção ou de ritmo, reduzindo assim a probabilidade de lesões musculares ou articulares decorrentes de movimentos bruscos.

Além disso, no contexto da análise tática e estratégica, a velocidade máxima de um jogador fornece dados valiosos para treinadores e equipes técnicas. Ao entender as capacidades individuais de velocidade de cada jogador, os treinadores podem adaptar suas estratégias de jogo, posicionamento em campo e substituições de forma mais eficaz, maximizando assim o potencial da equipe.

Em resumo, a velocidade máxima de um jogador de futebol não é apenas uma métrica de desempenho individual, mas sim um componente diferencial para o sucesso coletivo da equipe. Sua importância ultrapassa os aspectos puramente físicos do jogo, influenciando diretamente a dinâmica tática, o resultado da partida e até mesmo a integridade física dos jogadores.

Nesse sentido, jogadores que possuem elevada velocidade em arranques se destacam no futebol mundial, assim como Mbappé, Vinícius Júnior, Kyle Walker, Adama Traoré, Mohamed Salah etc.

Para o cálculo da velocidade máxima o processo se tornou mais simples, uma vez que já se tinham todas as velocidades compiladas em uma lista. Assim, bastou percorrê-la e selecionar o valor mais alto de velocidade registrado.

Inicialmente, o método utilizado foi percorrer a lista, comparar o valor atual com o maior registrado até o momento e, caso seja maior, substituir o maior valor pelo atual. Entretanto, esse processo demandava muito tempo para terminar a análise, devido ao extenso número de dados. Assim, substitui-se o método utilizado por um mais rápido e eficaz, utilizando-se a função “*max*” da biblioteca “*numpy*”, que extrai o maior valor presente na lista de velocidades.

2.3.3.3 Distância Percorrida

A distância percorrida por um jogador durante uma partida é uma métrica que não reflete apenas a capacidade física do atleta, mas também o seu envolvimento em aspectos chaves do jogo, como marcação, pressão, cobertura, criação de jogadas e deslocamentos para a quebra da linha defensiva adversária.

Além disso, a distância percorrida por um jogador por partida desempenha um papel fundamental na avaliação do desempenho individual e coletivo da equipe. Por meio dessa análise, treinadores e comissões técnicas identificam padrões de jogo, áreas que necessitam de melhorias e estratégias mais eficazes contra determinados adversários, permitindo-lhes ajustes táticos e técnicos que maximizem o potencial da equipe em campo.

Considerando a relevância desse parâmetro no futebol moderno, viu-se a necessidade de extrair essa importante métrica dos dados de posicionamento dos jogadores. Sabendo das posições dos jogadores em cada instante e já tendo calculado os deslocamentos durante a partida, pôde-se somá-los de modo a obter a distância total percorrida durante a toda a partida.

Em uma primeira abordagem, fez-se um laço de repetição que percorria e somava termo a termo da lista de deslocamentos, porém, devido ao extenso número de dados, essa análise demandava elevado tempo de operação. Como tentativa de aprimorar essa análise, substituiu-se o método de soma por repetição por uma função disponível na biblioteca *numpy*: a função sum. A função sum é utilizada quando deseja-se somar todos os elementos de uma lista numérica.

2.3.3.4 Aceleração

Uma vez que foram calculadas as velocidades do jogador em todos os frames extraídos da imagem, torna-se possível realizar o cálculo da aceleração do jogador. Para isso, é necessário partir de um conceito físico e matemático que define a aceleração.

A aceleração pode ser entendida como a variação da velocidade no tempo, ou seja, quanto a velocidade está variando de um instante para o instante subsequente. Diferentemente da física, área na qual a aceleração é entendida em uma abordagem vetorial, a análise voltada ao futebol pode ser feita de maneira mais simplista, analisando-se somente a intensidade dessa métrica ao decorrer da partida. Para isso, utiliza-se a seguinte equação matemática (equação 8):

$$a = \frac{dv}{dt} \tag{8}$$

Traduzindo essa análise para a lógica de programação, percorreu-se a lista de velocidades durante todos os instantes da partida, analisou-se a diferença entre os instantes “t” e “t+1” e, após obter essa diferença, dividiu-se pelo intervalo de tempo para se obter a aceleração em cada instante analisado. Incialmente utilizou-se blocos de repetição com o cálculo feito individualmente e armazenado em variáveis. Verificando o elevado tempo de execução, optou-se por modificar a metodologia e utilizar-se da função “`derivate`” presente na biblioteca “`scipy`”. Apesar da mudança de método, observou-se que o tempo de execução continuava acima do ideal. Então, partiu-se para uma terceira forma de análise, utilizando-se da função “`subtract`” presente na biblioteca “`numpy`”, calculou-se a diferença de velocidades entre 2 instantes. Diferentemente das outras tentativas, esse método obteve um tempo de execução dentro do esperado e, essa técnica de análise foi mantida.

2.3.3.5 Resultados

A etapa de validação dos resultados é de extrema importância no método científico, pois através dessa etapa pode-se concluir se há veracidade no método e nos processos utilizados ou se existe a necessidade de se buscar um novo método ou alterar etapas intermediárias. Diante disso, buscou-se comparar os dados obtidos com os dados base.

Comparando-se os valores obtidos entre os códigos de velocidade máxima e de velocidade média com os dados do *DataFrame*, obteve-se os seguintes resultados:

Tabela 3 – Desempenho Comparado do Algoritmo para Velocidade Máxima

NOME	Velocidade Máxima (km/h)		PERFORMANCE
	Código	DataFrame	
JEFFERSON	32,85	32,94	99,73%
IGOR CARIUS	34,78	34,79	99,97%
RHALDNEY	34,62	34,7	99,77%
SHAYLON	34,78	34,86	99,77%
EDINHO	33,4	33,4	100,00%
RAFAEL THYERRE	31,71	31,76	99,84%
RONALDO	32,72	32,82	99,70%
VAGNER LOVE	29,52	29,54	99,93%
LUCAS ANDRE	31,85	31,88	99,91%
RENAN	23,37	23,38	99,96%
RODRIGO SOARES	31,24	31,34	99,68%
HERON	32,52	32,56	99,88%
SABINO	34,74	34,81	99,80%
GUSTAVO COUTINHO	33,31	33,34	99,91%
SALES	33,31	33,33	99,94%
ALEX	32,4	32,41	99,97%
LUCIANO JUBA	33,83	33,87	99,88%
FABIO MATHEUS	34,27	34,29	99,94%
BRUNO TUBARAO	34,15	34,2	99,85%
RONALDO	32,72	32,82	99,70%
EWERTHON	33,35	33,39	99,88%
LUIZ FERNANDO	34,39	34,41	99,94%
ALAN RUIZ	29,69	29,69	100,00%
ITALO	25,74	25,76	99,92%
KELVIN	29,15	29,17	99,93%
PEDRO MARTINS	31,82	31,89	99,78%
FACUNDO LABANDEIRA	32,76	32,83	99,79%
N	33,01	33,12	99,67%
DANIEL	32,22	32,32	99,69%

FELIPINHO	32,31	32,43	99,63%
EMERSON SANTOS	23,78	23,92	99,41%
DAVID FRANCISCO	23,23	23,31	99,66%

FONTE: ASSUNÇÃO, 2024

Apresentando uma acurácia média de 99,83%.

Tabela 4 – Desempenho Comparado do Algoritmo para Velocidade Média

NOME	**Velocidade Média (km/h)**		**PERFORMANCE**
	Código	**DataFrame**	
JEFFERSON	4,8996	4,7628	97,21%
IGOR CARIUS	5,85	5,6484	96,55%
RHALDNEY	6,3432	6,138	96,77%
SHAYLON	6,3468	6,0696	95,63%
EDINHO	5,6916	5,1444	90,39%
RAFAEL THYERRE	5,3208	5,1516	96,82%
RONALDO	2,43	2,1564	88,74%
VAGNER LOVE	5,04	4,7088	93,43%
LUCAS ANDRE	5,8788	5,7204	97,31%
RENAN	2,268	2,0196	89,05%
RODRIGO SOARES	5,346	5,1552	96,43%
HERON	4,8708	4,7772	98,08%
SABINO	4,806	4,6548	96,85%
GUSTAVO COUTINHO	6,2172	6,0444	97,22%
SALES	5,49719	5,2992	96,40%
ALEX	4,86	4,6944	96,59%
LUCIANO JUBA	5,3172	5,13	96,48%
FABIO MATHEUS	5,8536	5,634	96,25%
BRUNO TUBARAO	5,544	5,3712	96,88%
RONALDO	2,43	2,1564	88,74%
EWERTHON	5,1516	4,9536	96,16%
LUIZ FERNANDO	5,9256	5,7708	97,39%
ALAN RUIZ	5,9724	5,7744	96,68%
ITALO	7,0956	5,544	78,13%
KELVIN	6,2316	6,0804	97,57%
PEDRO MARTINS	7,32239	7,0092	95,72%
FACUNDO LABANDEIRA	7,2864	7,1856	98,62%
N	6,9228	6,4044	92,51%
DANIEL	6,9228	6,4548	93,24%
FELIPINHO	6,7428	6,3936	94,82%
EMERSON SANTOS	7,4808	7,2504	96,92%
DAVID FRANCISCO	10,1088	9,5508	94,48%

FONTE: ASSUNÇÃO, 2024

Apresentando uma acurácia média de 94,81%.

Apesar de um código para o cálculo da distância percorrida ter sido elaborado, o *DataFrame* inicial não apresentava esse dado, não tendo como compará-lo.

Tabela 5 – Distância Percorrida Calculada pelo Algoritmo

NOME	Distância Percorrida (km)
JEFFERSON	8,66
IGOR CARIUS	10,26
RHALDNEY	11,09
SHAYLON	10,73
EDINHO	8,39
RAFAEL THYERRE	9,29
RONALDO	4,3
VAGNER LOVE	7,48
LUCAS ANDRE	5,21
RENAN	4,02
RODRIGO SOARES	9,06
HERON	8,61
SABINO	8,39
GUSTAVO COUTINHO	7,95
SALES	9,57
ALEX	8,52
LUCIANO JUBA	8,59
FABIO MATHEUS	10,18
BRUNO TUBARAO	8,85
RONALDO	4,3
EWERTHON	9,02
LUIZ FERNANDO	9,51
ALAN RUIZ	5,16
ITALO	2,13
KELVIN	2,99
PEDRO MARTINS	2,03
FACUNDO LABANDEIRA	2,02
N	1,07
DANIEL	1,01
FELIPINHO	0,93
EMERSON SANTOS	0,47
DAVID FRANCISCO	0,64

FONTE: ASSUNÇÃO, 2024

Apesar do método de análise da aceleração ter sido executado no tempo tolerável, os resultados obtidos não foram validados pelo *DataFrame* utilizado como base. Acredita-se que os dados obtidos possam incluir erros de oscilação nos pontos obtidos ou não foram completamente expostos ao vídeo quando se trata da métrica aceleração, pois a análise feita com fundamentos físicos e matemáticos não condizem

com os dados fornecidos. Nesse sentido, decidiu-se manter a análise fundamentada nestas ciências.

2.3.3.6 Teste – T

O teste – t foi desenvolvido em 1908 por William Sealy Gosset, como uma ferramenta utilizada para comparar duas diferentes amostras. A aplicação desse método é amplamente difundida em diversas áreas da ciência, incluindo os teste necessários para a validação de um algoritmo.

Uma vez desenvolvido o algoritmo que extrai importantes métricas do futebol de um *Dataframe,* um processo crucial para a aplicação é a validação, verificando a eficácia e a confiabilidade que o algoritmo apresenta.

O teste – t apresenta uma abordagem que permite determinar se a diferença entre duas diferentes amostras é significativa do ponto de vista estatístico. Esse método é capaz de oferecer uma avaliação objetiva e quantitativa da performance do algoritmo comparado aos dados do mercado, fornecendo uma análise que pode ser interpretada e resultar no descarte do algoritmo ou na validação.

Frente à capacidade do teste – t de determinar a eficácia e a confiabilidade que se tem do algoritmo desenvolvido, utilizou-se desse método para a etapa de validação.

Os resultados são exibidos abaixo:

Tabela 6 – Desempenho Comparado do Algoritmo para Velocidade Média

	Variável 1	*Variável 2*
Média	5,749199375	5,462775
Variância	2,338904562	2,163427798
Observações	32	32
Variância agrupada	2,25116618	
Hipótese de diferença de média	0	
gl	62	
Stat t	0,763600471	
P(T<=t) uni-caudal	0,224000728	
t crítico uni-caudal	1,669804163	
P(T<=t) bi-caudal	0,448001456	
t crítico bi-caudal	1,998971517	

FONTE: ASSUNÇÃO, 2024

Tabela 7 – Desempenho Comparado do Algoritmo para Velocidade Máxima

	Variável 1	*Variável 2*
Média	31,673125	31,7275
Variância	10,8020673	10,7909935
Observações	32	32
Variância agrupada	10,7965304	
Hipótese de diferença de média	0	
gl	62	
	-	
Stat t	0,06619378	
P(T<=t) uni-caudal	0,47371827	
t crítico uni-caudal	1,66980416	
P(T<=t) bi-caudal	0,94743654	
t crítico bi-caudal	1,99897152	

FONTE: ASSUNÇÃO, 2024

Estabelecendo-se α = 0,05 (tolerância de 5% de erro), obteve-se os testes – t apresentados, acima e, pôde-se concluir que o algoritmo desenvolvido apresenta eficácia e confiabilidade suficiente para implementação em análises futuras de outros conjuntos de dados, pois apresenta o "*P-value*" menor que α.

3 CONCLUSÕES E RESULTADOS

Nesta seção apresenta-se os principais aprendizados durante a condução desse estudo. Ao longo das seções anteriores, buscou-se relatar as principais atividades executadas durante a pesquisa e, agora busca-se evidenciar os principais resultados obtidos e suas interpretações.

Os cálculos de velocidade média, velocidade máxima e distância percorrida foram validados pela comparação dos dados fornecidos pelo *DataFrame*. Alguns gráficos que evidenciam melhor a comparação dos dados é exposta abaixo:

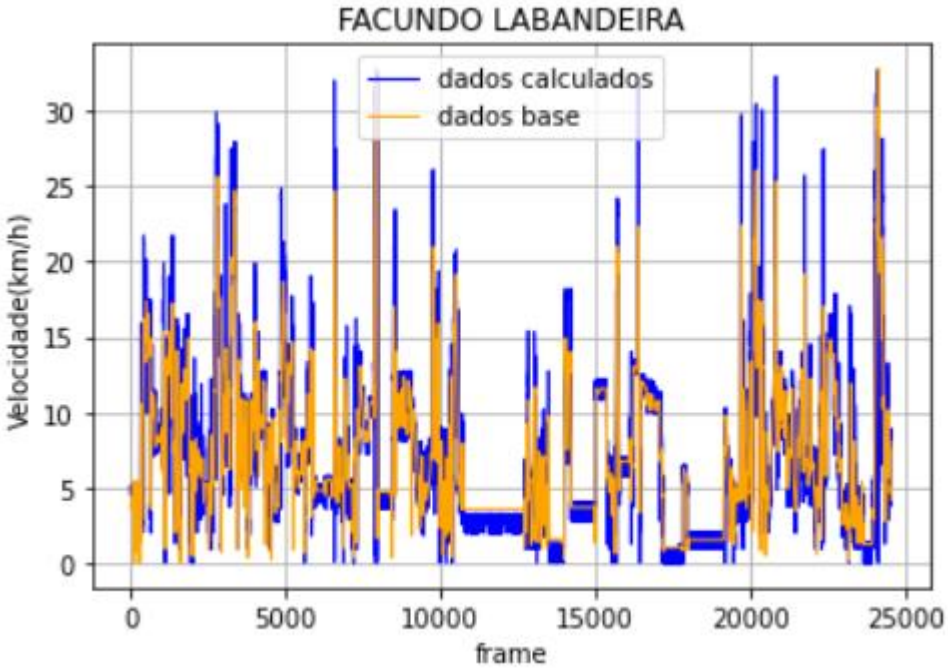

Figura 10 – Melhor Desempenho Comparativo entre Algoritmo Implementado e o Comercial

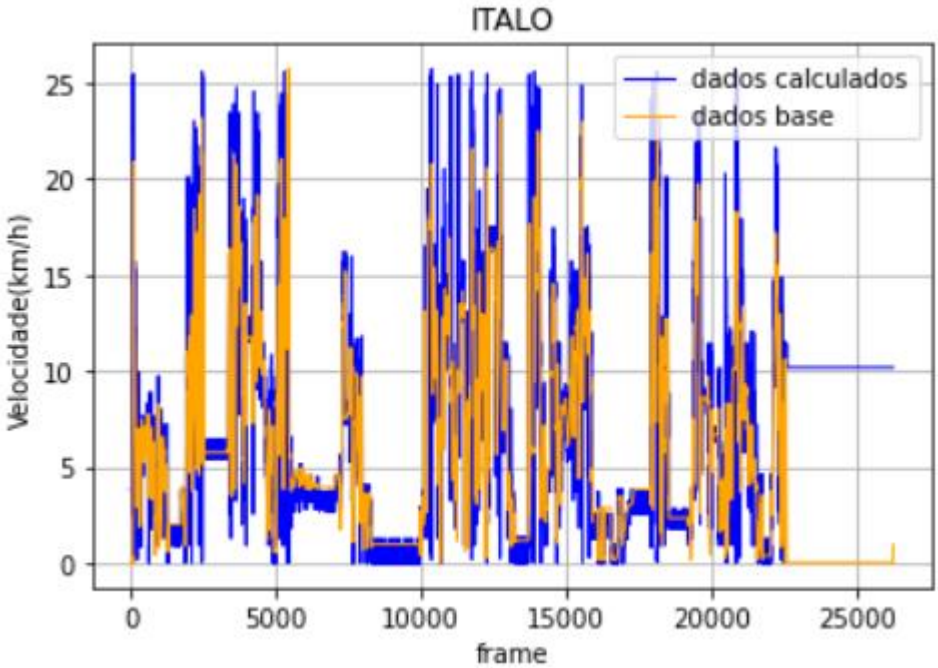

Figura 11 - Pior Desempenho Comparativo entre Algoritmo Implementado e o Comercial

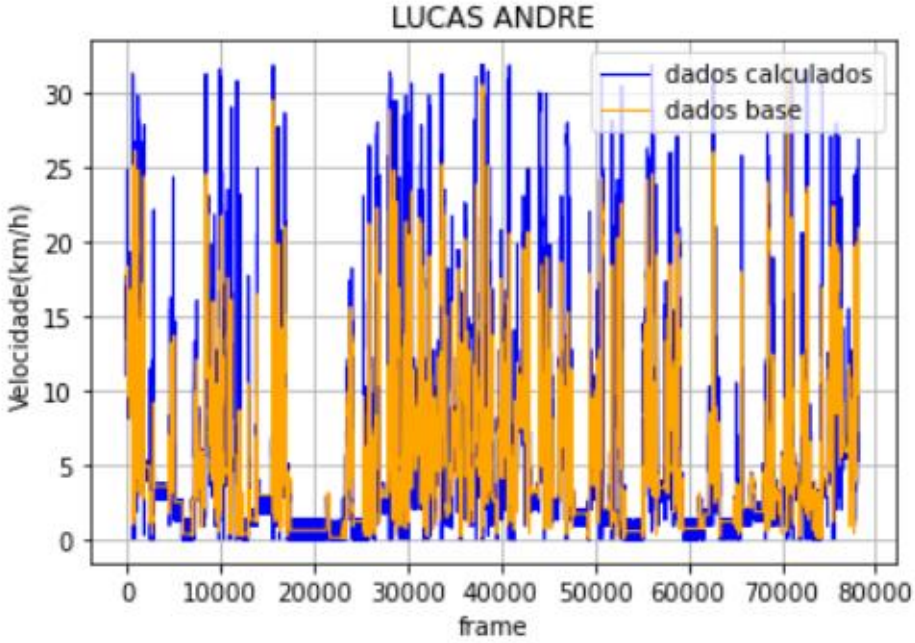

Figura 12 – Média do Desempenho Comparativo entre Algoritmo Implementado e o Comercial

Velocidade Média

Código: 2.024 (m/s) | DataFrame: 1.996 (m/s)
Código: 7.2864 (km/h) | DataFrame: 7.1856 (km/h)

Velocidade máxima

Código: 32.76 (km/h) | Dataframe: 32.83 (km/h)

Figura 13 – Melhor Resultado Numérico Comparativo entre a Velocidade Média do Algoritmo Implementado e do Comercial

```
Velocidade Média

Código: 1.971 (m/s)   |  DataFrame: 1.54 (m/s)
Código: 7.0956 (km/h) |  DataFrame: 5.5440000000000005 (km/h)

Velocidade máxima

Código: 25.74 (km/h)  |  Dataframe: 25.76 (km/h)
```

Figura 14 – Pior Resultado Numérico Comparativo entre a Velocidade Média do Algoritmo Implementado e do Comercial

```
Velocidade Média

Código: 1.633 (m/s)   |  DataFrame: 1.589 (m/s)
Código: 5.8788 (km/h) |  DataFrame: 5.7204 (km/h)

Velocidade máxima

Código: 31.85 (km/h)  |  Dataframe: 31.88 (km/h)
```

Figura 15 – Resultado Numérico Médio do Comparativo entre a Velocidade Média do Algoritmo Implementado e do Comercial

```
Distância percorrida: 11.09 km
```

Figura 16 – Distância Percorrida Calculada pelo Algoritmo Implementado

Entretanto, a validação não foi possível para os cálculos de aceleração, que divergiram muito dos valores de referência. Devido a isso, os cálculos não foram considerados corretos, mesmo que, diante de uma abordagem física e matemática, estejam fiéis a uma análise científica.

Para efeito comparativo, os gráficos de aceleração dos jogadores que apresentaram o melhor, o pior e o desempenho médio – medidos com base na concordância com os dados base – nos cálculos das velocidades, são exibidos abaixo, evidenciando a divergência entre os conjuntos de dados.

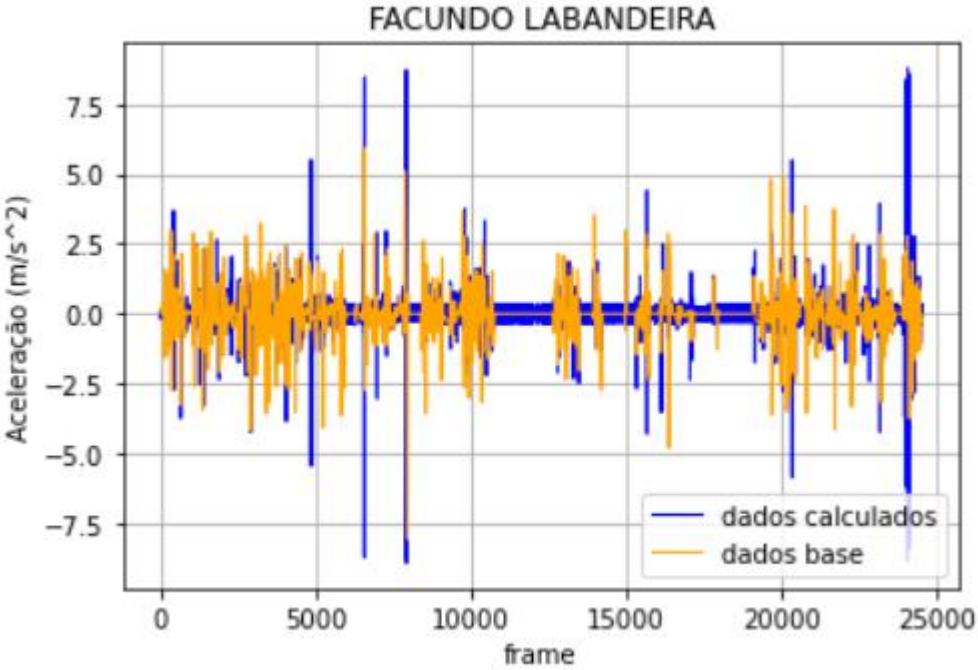

Figura 17 – Comparativo da Aceleração para o Melhor Desempenho da Análise da Velocidade Média

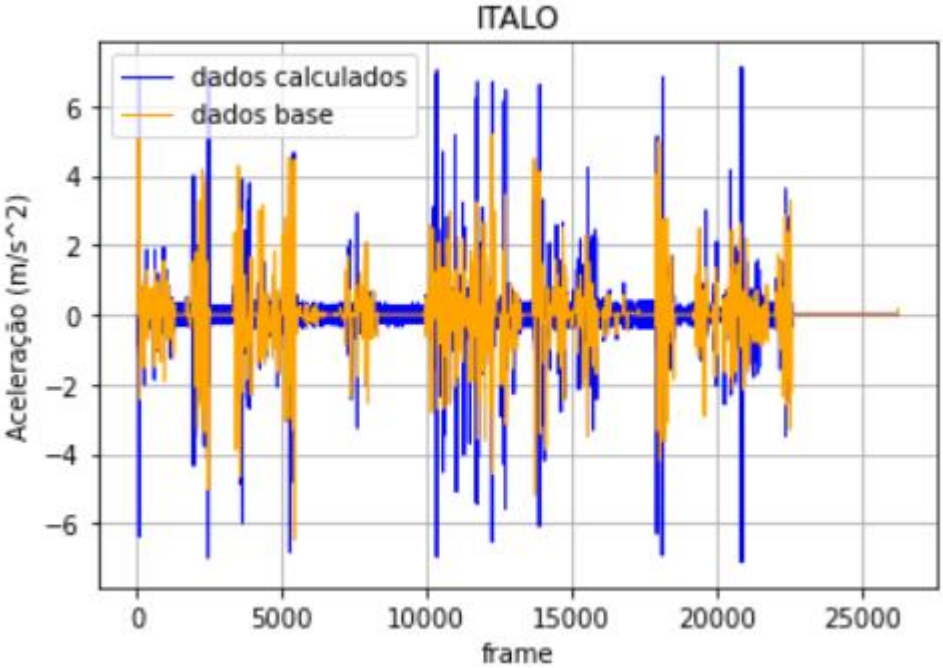

Figura 18 – Comparativo da Aceleração para o Pior Desempenho da Análise da Velocidade Média

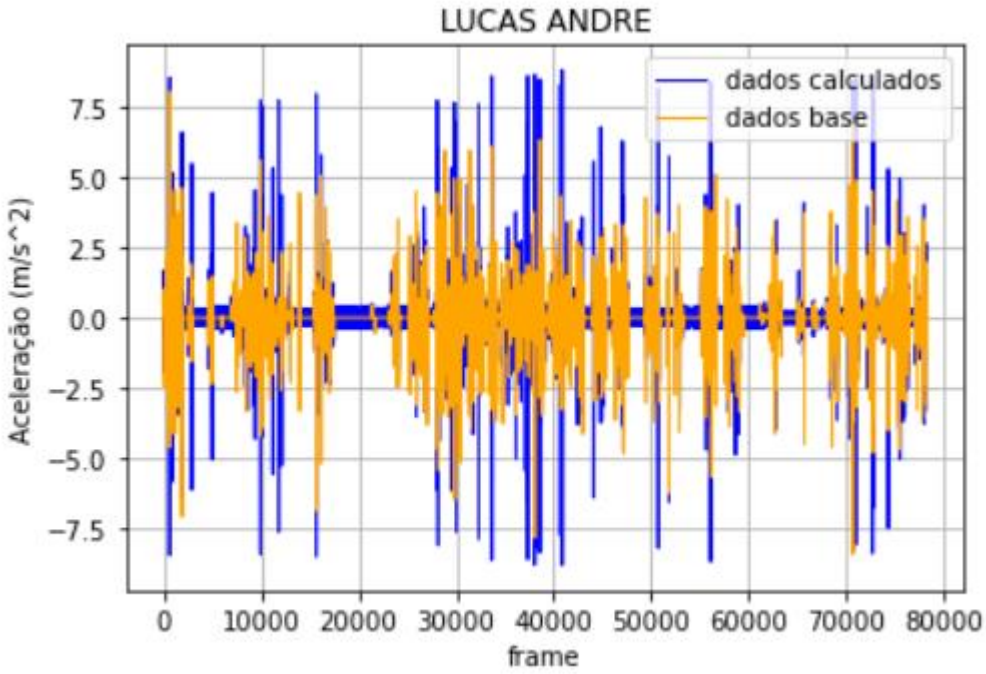

Figura 19 – Comparativo da Aceleração para o Desempenho Médio da Análise da Velocidade Média

O desvio-padrão entre os conjuntos de dados foram calculados para os três jogadores e são exibidos abaixo:

- Desvio-padrão (Facundo Labandeira): 0,71
- Desvio-padrão (Italo): 0,67
- Desvio-padrão (Lucas André): 0,69

Dessa forma, concluiu-se que os códigos utilizados para os cálculos de velocidade média, velocidade máxima e distância percorrida estão aptos para serem implementados. Por outro lado, o código responsável pelo cálculo da aceleração não foi aprovado para a implementação.

3.1 Próximos Passos

Através do entendimento das principais necessidades no dia a dia das comissões técnicas, definiu-se as principais métricas e estas foram estudadas afundo. Entende-se que a definição, o estudo e a implementação das principais métricas relevantes para uma partida de futebol foram feitas com sucesso. Além disso, o estudo dos impactos que as redes sociais geram na saúde psicológica de um jogador são promissores e podem evidenciar a importância do cuidado psicológico para tratamento e prevenção de problemas desenvolvidos por ataques do público.

Dessa forma, planeja-se dar continuidade nesse projeto através do estudo estatístico de quanto cada métrica extraída de uma partida de futebol impacta na nota

atribuída a um jogador, considerando a posição de atuação, o país que atua e a competição em análise.

Além disso, busca-se produzir criteriosamente uma classificação automática de comentários veiculados nas redes sociais referentes a um jogador, podendo avaliá-los numericamente e analisar uma possível correlação com os desempenhos subsequentes.

Nesse sentido, será possível continuar impactando positivamente os clubes brasileiros, trazendo mais tecnologia e assertividade nas ações técnicas dentro das equipes.

4 REFERÊNCIAS

ACERVO O GLOBO. Didi, o genial mestre da folha-seca, aos 72 anos. O Globo, 27 abr. 2021. Disponível em: https://acervo.oglobo.globo.com/em-destaque/didi-genial-mestre-da-folha-seca-aos-72-anos-24989982. Acesso em: 15 maio 2024.

CARTAS, A.; BALLESTER, C.; HARO, G. A graph-based method for soccer action spotting using unsupervised player classification. 11 2022.

COUTO, L. N. Sistema para localização robótica de veículos autônomos baseado em visão computacional por pontos de referência. 2012. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

DELTA GOAL. Introducing AI to reach the next level of performance. Disponível em: https://www.deltagoal.ai. Acesso em: 15 maio 2024.

DELTA GOAL. LinkedIn. Disponível em: https://br.linkedin.com/company/delta-goal. Acesso em: 31 agosto 2023.

FREYRE, Gilberto. **Como o futebol moldou a identidade cultural do brasileiro**. O Negro no Futebol Brasileiro. Disponível em: < Como o futebol moldou a identidade cultural do brasileiro (observatorioracialfutebol.com.br)>. Acesso em: 27 ago. 2023.

GLOBO ESPORTE. Filosofia Cruyff: frases do ídolo holandês sobre o mundo do futebol. Globo Esporte, 24 mar. 2016. Disponível em: https://ge.globo.com/futebol/futebol-internacional/noticia/2016/03/filosofia-cruyff-frases-do-idolo-holandes-sobre-o-mundo-do-futebol.html. Acesso em: 15 maio 2024.

GUIMARÃES, Leonardo. Como a Análise de Dados ajudou a Alemanha na vitória da Copa do Mundo. 2014. Disponível em: < Como a Análise de Dados ajudou a Alemanha na vitória da Copa do Mundo - Know Solutions>. Acesso em: 29 ago.

2023.

GUSMÃO, Gustavo. Solução de big data é um dos segredos da Alemanha na Copa. 2014. Disponível em: < Solução de big data é um dos segredos da Alemanha na Copa | Exame>. Acesso em: 29 ago. 2023.

HOMAYOUNFAR, N.; FIDLER, S.; URTASUN, R. Soccer field localization from a single image. Computer Vision and Pattern Recognition, 04 2016.

LANDI, Bruno. Melhor da Copa, Messi chutou menos que C.Ronaldo; veja n°s. 2015. Disponível em: < Melhor da Copa, Messi chutou menos que C. Ronaldo; veja nºs (terra.com.br)>. Acesso em: 30 ago. 2023.

MOLZ, R. F. Uma metodologia para o desenvolvimento de aplicações de visão computacional utilizando um projeto conjunto de hardware e software. 2001. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre

MOTA< Cahê. Fifa aprova uso de bolas com chip: Mundial de Clubes terá a tecnologia. 2023. Disponível em: < Fifa aprova o uso de bolas com chip: Mundial de Clubes terá a tecnologia | globoesporte.com>. Acesso em: 29 ago. 2023.

PESSOA, Lucas. Futebol brasileiro apresenta alta de público e bilheteria no primeiro semestre de 2023. 2023. Disponível em: <https://www.lance.com.br/lancebiz/torcidas/futebol-brasileiro-apresenta-altadepublico-e-bilheteria-no-primeiro-semestre-de-2023.html>. Acesso em: 27 ago. 2023.