UNIVERSIDADE PRESBITERIANA MACKENZIE

RAYAN CRHISTOFER GOMES DA SILVA, GUILHERME SOARES FROTA, BRUNO LUIGI DOS SANTOS TOBIAS, THÁBATA CRISTINA SEIXAS BALBINO.

UMA ANÁLISE EXPLORATÓRIA DE DADOS DO CAMPEONATO BRASILEIRO SÉRIE A (2003-2023)

SÃO PAULO

2024

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

O Campeonato Brasileiro de Futebol, sob a gerência da Confederação Brasileira de Futebol (CBF), é um dos eventos esportivos mais prestigiados e apaixonantes do país, proporciona um rico terreno para a análise de dados. Ao longo das últimas duas décadas, de 2003 a 2023, o torneio testemunhou não apenas emocionantes jogos e competições acirradas, mas também mudanças significativas na dinâmica das equipes e nas estratégias de jogo [1]. Este relatório busca conduzir uma análise exploratória de dados sobre o Campeonato Brasileiro, utilizando um dataset abrangente obtido no Kaggle [2], a fim de desvendar padrões, identificar tendências e fornecer insights valiosos sobre o desempenho das equipes, a evolução do torneio e fenômenos únicos que marcaram este período.

A riqueza dos dados disponíveis nos permite mergulhar nas estatísticas das equipes, resultados das partidas, artilharia, classificações e outros indicadores essenciais, oferecendo uma visão detalhada do cenário futebolístico brasileiro.

A análise proposta não apenas atende ao interesse intrínseco pelos aspectos estatísticos do futebol, mas também serve como uma ferramenta valiosa para gestores, técnicos, e entusiastas que buscam insights estratégicos para entender as dinâmicas competitivas do Campeonato Brasileiro ao longo das últimas duas décadas. Dessa forma, este trabalho visa contribuir para o conhecimento contínuo do futebol brasileiro, enriquecendo a compreensão do esporte e oferecendo uma perspectiva informada sobre seu desenvolvimento ao longo do tempo.

2 DEFINIÇÃO DO CONTEXTO ORGANIZACIONAL E ÁREA DE ATUAÇÃO

O contexto organizacional deste trabalho de ciência de dados abrange o universo do Futebol Brasileiro no período de 2003 a 2023, sob a gestão primordial da Confederação Brasileira de Futebol (CBF). A CBF, enquanto entidade máxima no cenário esportivo nacional, desempenha um papel central na organização, regulamentação e promoção do futebol no Brasil.

A Confederação Brasileira de Futebol (CBF) é a entidade máxima responsável pela organização e administração do futebol no Brasil. Fundada em 1914, a CBF é filiada à Federação Internacional de Futebol (FIFA) e à Confederação Sul-Americana de Futebol (CONMEBOL). Sua sede está localizada na cidade do Rio de Janeiro [1].

A CBF desempenha um papel crucial na regulamentação e gestão de todas as competições de futebol no país, desde os torneios locais até as seleções nacionais. Entre suas responsabilidades estão a definição das regras do jogo, a organização de campeonatos, a coordenação de competições nacionais, a representação do Brasil em eventos internacionais, e a promoção do desenvolvimento e popularização do futebol em território nacional.

Ao longo das últimas duas décadas, a CBF tem sido responsável por coordenar o Campeonato Brasileiro de Futebol, uma das competições mais relevantes e emocionantes do país. Sua atuação vai além da simples administração de torneios, abrangendo a definição de regras, a condução de competições de diferentes categorias, a representação do Brasil em torneios internacionais, e o suporte às federações estaduais.

O Principal produto da CBF, o “Brasileirão Série A” é um produto esportivo gerido como uma empresa que coordena a principal competição de futebol de clubes no Brasil. Operando sob a direção da Confederação Brasileira de Futebol (CBF), essa "empresa esportiva" organiza anualmente um campeonato de elite, atraindo os principais clubes do país em uma competição de pontos corridos. Sua estrutura envolve a definição de regulamentos, logística de partidas, gestão de direitos de transmissão, negociação de patrocínios e promoção de eventos. O Brasileirão Série A é uma marca reconhecida, com uma base sólida de fãs, e contribui significativamente para a economia do futebol brasileiro por meio de receitas diversas. Além disso, como uma "empresa" do esporte, o campeonato tem responsabilidades na promoção do desenvolvimento de jovens talentos, na manutenção da integridade do jogo e na gestão de rivalidades e tradições que enriquecem a experiência dos torcedores. Em resumo, o Brasileirão Série A atua como uma empresa esportiva completa, gerenciando não apenas a competição em si, mas também impactando aspectos econômicos, sociais e culturais do cenário futebolístico brasileiro.

Neste contexto, a aplicação da ciência de dados visa extrair insights valiosos a partir de um extenso conjunto de dados relacionados ao Futebol Brasileiro nesse período. A análise exploratória, modelagem estatística e outras técnicas contribuirão não apenas para compreender as dinâmicas esportivas, mas também para oferecer informações estratégicas que possam subsidiar decisões, aprimorar o desempenho das equipes e enriquecer o entendimento geral do fenômeno futebolístico no Brasil.

3 CARACTERIZAÇÃO DO PROBLEMA

O problema a ser explorado nesta análise do Campeonato Brasileiro Série A de 2003 a 2023 reside na compreensão profunda dos fatores que influenciam o desempenho das equipes ao longo dessas duas décadas. Buscamos identificar padrões, tendências e correlações que possam explicar variações nos resultados, classificações e artilharia, indo além dos aspectos superficiais e capturando nuances que moldaram a dinâmica da competição.

O desempenho dos clubes de futebol ao longo de 20 anos de dados analisados [2] pode ser definido por diversos fatores, como:

- Análise Por Ataque e Defesa do Brasileirão: Os clubes que mais gols marcaram e mais gols sofreram. Esses dados podem indicar o número médio de gols por partida de cada clube, número de gols médio sofridos por partida e, comparativamente, podemos analisar qual o impacto dessas estatísticas para o ganho ou perda das partidas.
- Análise de Pontuação do Brasileirão: Análise para exibição dos clubes com melhor e pior pontuação em todas as edições. Esses números podem indicar o número máximo e mínimo de pontuação de clubes nas edições do Brasileirão.
- Análise de Resultados do Brasileirão: É identificado o número de vitórias, empates ou derrotas dos clubes que participaram pelo menos de uma edição do Brasileirão.
- Análise de Campeões do Brasileirão: Clube com maior pontuação em uma edição do Brasileirão é o campeão.

4 APRESENTAÇÃO DOS DADOS (METADADOS)

Os dados a serem analisados são armazenados no site Kaggle, plataforma online dedicada à ciência de dados, aprendizado de máquina e análise estatística. O data-set a ser analisado, intitulado “Campeonato Brasileiro de Futebol” [2], é criado, analisado e mantido pelo autor “Adão Duque”, Analista programador at E4U Software & Internet.

O data-set contém dados de 8404 partidas do Brasileirão do período de 2003 á 2023, era dos pontos corridos. Um script coleta dados de vários sites esportivos e da própria CBF, Globo Esporte, Lance e Bola na Área. Geralmente os dados são atualizados quando o campeonato termina,

O data-set é subdividido em 4 arquivos CSV:

- **Cartões aplicados no Campeonato Brasileiro:**

Legenda - campeonato-brasileiro-cartoes.csv

partida_ID - ID da partida

Rodada - Rodada da partida

Clube - Nome do clube

Cartao - Cor do cartão aplicado

Atleta - Nome do atleta punido pelo cartão

num_camisa - Número da camisa do atleta

Posicao - Posição na partida em que o atleta se encontra

Minuto - Minuto na partida em que o cartão foi aplicado

- **Estatísticas do Campeonato Brasileiro:**

Legenda - campeonato-brasileiro-estatisticas-full.csv

partida_ID - ID da partida

Rodada - Rodada da partida

Clube - Nome do clube

Chutes - Finalizações

Chutes a gol - Finalizações na direção do gol

Posse de bola - Percentual da posse de bola

Passes - Quantidade de passes que o clube deu na partida precisao_passes - Percentual da precisão de passe

Faltas - Quantidade de faltas cometidas na partida

cartao_amarelo - Quantidade de cartões amarelos para o clube na partida cartao_vermelho - Quantidade de cartões vermelhos para o clube na partida

Impedimentos - Quantidade de impedimentos para o clube na partida Escanteios - Quantidade de escanteios para o clube na partida

- **Gols no Campeonato Brasileiro:**

Legenda - campeonato-brasileiro-gols.csv

partida_ID - ID da partida

Rodada - Rodada da partida

Clube - Nome do clube

Atleta - Nome do atleta que fez o gol

Minuto - Minuto na partida em que o gol foi marcado

- **Diversos dados do Campeonato Brasileiro:**

Legenda - campeonato-brasileiro-full.csv

ID - ID da partida

Rodada : Rodada que aconteceu a partida

Data : Data que ocorreu a partida

Horário : Horário que ocorreu a partida

Dia : Dia da semana que ocorreu a partida

Mandante : Clube mandante

Visitante : Clube Visitante

formacao_mandante: Formação do mandante

formacao_visitante: Formação do visitante

tecnico_mandante: Técnico do mandante

tecnico_visitante: Técnico do visitante

Vencedor : Clube vencedor da partida. Quando tiver "-", é um empate Arena : Arena que ocorreu a partida Mandante

Placar : Gols que o clube mandante fez na partida Visitante

Placar : Gols que o clube visitante fez na partida

Estado Mandante : Estado do clube mandatorio

Estado Visitante : Estado do clube visitante

Estado Vencedor : Estado do clube vencedor. Quando tiver "-", é um empate

5 REPOSITÓRIO GITHUB

Todos os arquivos e dados utilizados neste trabalho serão armazenados no GitHub [3].

6 CRONOGRAMA

6.1 Cronograma Geral

Cronograma Projeto Aplicado I	
Etapa 1	Data de Entrega: 06/03/2024
Apresentação do projeto, objetivos, metas e milestones	
Atividades	Foi realizada?
Montagem do grupo	Ok
Escolha da temática	Ok
Orientação dos grupos (encontro sícrono)	Ok
Comunicação	Ok
Organização do Material	Ok
Cronograma Inicial do projeto	Ok
Etapa 2	Data de Entrega: 03/04/2024
Definição do produto analítico	
Atividades	
Etapa 3	Data de Entrega: 27/04/2024
Apresentação de produtos	
Atividades	
Etapa 4	Data de Entrega: 31/05/2024
Encerramento do projeto	
Atividades	

Tabela 1: Cronograma Geral para cumprimento do Projeto Integrador I.

6.2 Cronograma de Atividades

Cronograma de Atividades			
Atividade	**Responsável**	**Data de Entrega**	**Observação**
Definir tema	Todos	2/29/2024	
Criar grupo Whatsapp e Discord	Thábata	2/28/2024	
Criar corpo documento do relatório	Crhistofer	2/28/2024	
Pesquisar metadados	Luigi	3/2/2024	
Criar cronograma	Guilherme	2/28/2024	
Repositório Github	Crhistofer/Guilherme	3/4/2024	
Desenvolver documento relatório	Todos	3/4/2024	
Revisão de relatório para entrega	Todos	3/6/2024	
Reunião de próximos passos	Todos	3/11/2024	

Tabela 2: Cronograma de Atividades a serem realizadas para cumprimento do Projeto Integrador I.

8 REFERENCIAS BIBLIOGRÁFICAS

[1] Wikipedia. Campeonato Brasileiro de Futebol. Disponível em: https://pt.wikipedia.org/wiki/Campeonato_Brasileiro_de_Futebol

[2] Duque, João. Campeonato Brasileiro de Futebol. Disponível em: https://www.kaggle.com/datasets/adaoduque/campeonato-brasileiro-de-futebol

[3] SILVA, R. C. G. Projeto Aplicado 1 – Uma análise exploratória de dados do campeonato brasileiro série A. Disponível em: https://github.com/RayanCrhistofer/PROJETOAPLICADO1--UMA-AN-LISE-EXPLORAT-RIA-DE-DADOS-DO-CAMPEONATO-BRASILEIRO-S-RIE-A-2003-2023-.git