

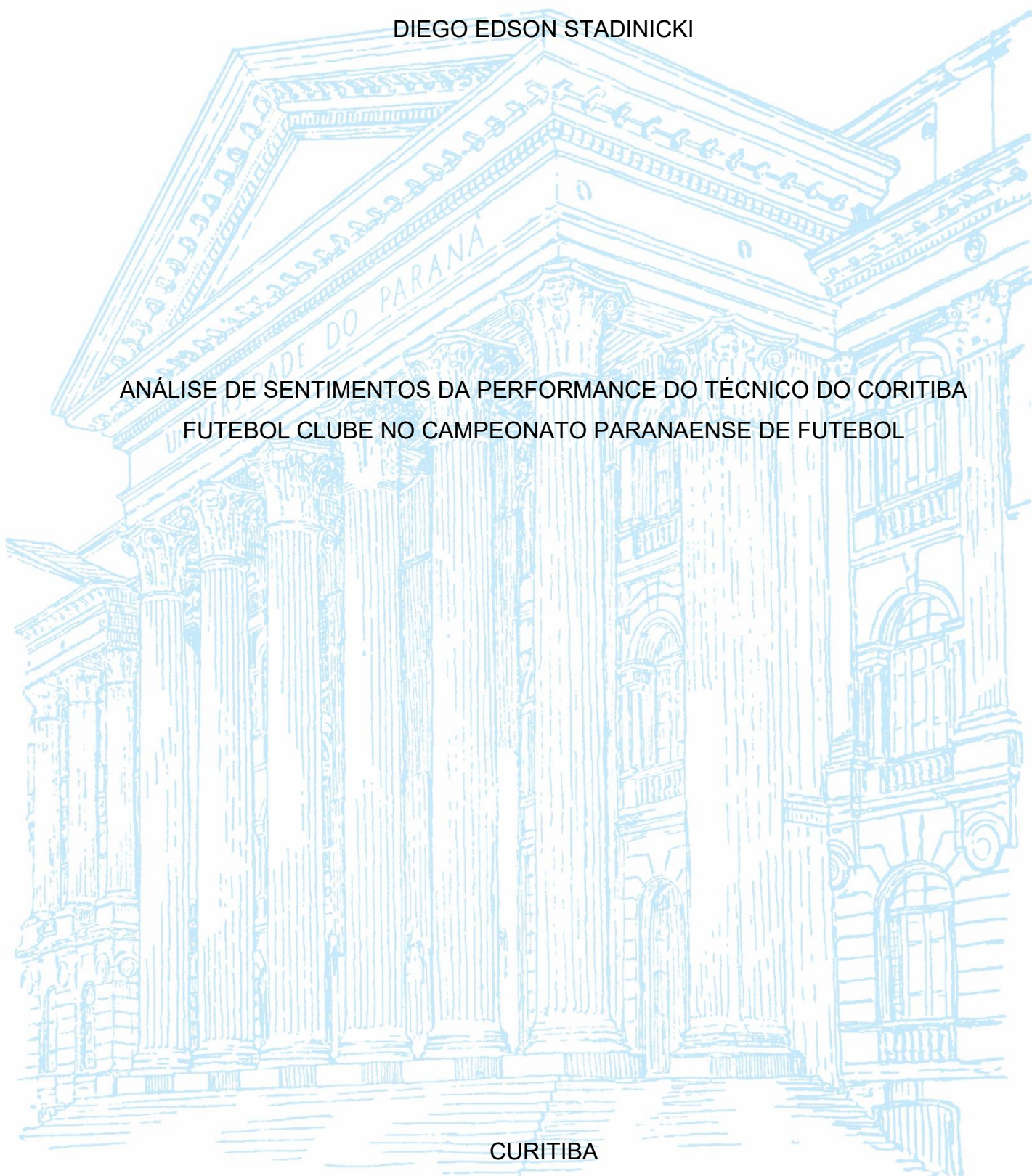
UNIVERSIDADE FEDERAL DO PARANÁ

DIEGO EDSON STADINICKI

ANÁLISE DE SENTIMENTOS DA PERFORMANCE DO TÉCNICO DO CORITIBA  
FUTEBOL CLUBE NO CAMPEONATO PARANAENSE DE FUTEBOL

CURITIBA

2023



DIEGO EDSON STADINICKI

ANÁLISE DE SENTIMENTOS DA PERFORMANCE DO TÉCNICO DO CORITIBA  
FUTEBOL CLUBE NO CAMPEONATO PARANAENSE DE FUTEBOL

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do grau de Bacharel em Gestão da Informação no curso de Graduação em Gestão da Informação, Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná.

Orientador: Prof. José Marcelo Cestari

CURITIBA

2023

## **AGRADECIMENTOS**

Agradeço a Universidade Federal do Paraná pela oportunidade de realizar um ensino superior em um curso que no começo não tinha muitas expectativas, mas que ao longo do percurso foi surtindo interesse em mim.

Sou grato também aos meus amigos que me auxiliaram nesta trajetória, em especial á Isabela, Matheus, Nathália e Julia, pois sem eles presentes no cotidiano deste curso eu não estaria aqui.

Agradeço ao professor Cestari pelo apoio e diversas matérias lecionadas ao longo do curso, com certeza serviram de inspiração e fonte para a realização deste trabalho.

## RESUMO

O *Youtube* é uma das redes sociais mais utilizadas pelos brasileiros e o potencial de divulgação de conteúdo através de vídeos fez com que os clubes de futebol brasileiros dessem uma maior atenção a plataforma, a fim de estabelecerem uma melhor relação com seus torcedores. O trabalho tem como objetivo geral identificar e analisar os sentimentos dos comentários de vídeos da plataforma do *Youtube*, no âmbito da performance do técnico de futebol do Coritiba Futebol Clube, em sua trajetória de começo de temporada no Campeonato Paranaense de futebol de 2023. Como metodologia, foi definido a linguagem *Python* e algumas bibliotecas de apoio como a ferramenta a ser utilizada para os processos de coleta, pré-processamento e análise dos dados extraídos da rede/mídia social. A coleta dos dados foi realizada usando a *Application Programming Interface* do *Youtube*, o pré-processamento usando bibliotecas *do Python* e a análise dos comentários utilizando um léxico adaptado para o português brasileiro denominado “Léxico para Inferência Adaptada”. Os resultados apresentaram que as opiniões dos torcedores com relação ao técnico de futebol e a performance do clube no campeonato foram majoritariamente negativas. Conclui que o objetivo proposto foi contemplado e que com o grande volume de dados presentes nos canais oficiais dos clubes de futebol, é possível mensurar a opinião dos torcedores no *Youtube* para auxiliar os clubes no êxito esportivo. Apresenta como proposta de trabalhos futuros a criação de um modelo baseado nas curtidas dos comentários dos vídeos e a possibilidade da automatização da coleta dos comentários dos vídeos do *Youtube*.

Palavras-chave: Análise de sentimentos. *Youtube*. *Python*. Técnico de futebol. Coritiba Futebol. Clube. Léxico para inferência adaptada.

## ABSTRACT

Youtube is one of the social networks most used by Brazilians and the potential for content dissemination through videos has led Brazilian soccer clubs to give greater attention to the platform in order to establish a better relationship with their fans. The general objective of this work is to identify and analyze the feelings of the video comments on the YouTube platform, in the context of the performance of the Coritiba Futebol Clube soccer coach, in his early season trajectory in the 2023 Paranaense Soccer Championship. As methodology, the Python language and some supporting libraries were defined as the tool to be used for the collection, pre-processing and analysis processes of the data extracted from the network/social media. The data collection was performed using Youtube's Application Programming Interface, the pre-processing using Python libraries, and the analysis of the comments using a lexicon adapted for Brazilian Portuguese called "Lexicon for Adapted Inference". The results showed that the opinions of the fans regarding the soccer coach and the club's performance in the championship were mostly negative. It concludes that the proposed goal was contemplated and that with the large volume of data present in the official channels of the soccer clubs, it is possible to measure the opinion of the fans on Youtube to help the clubs in the sporting success. Presents as a proposal for future work the creation of a model based on the likes of the video comments and the possibility of automating the collection of Youtube video comments.

Keywords: Sentiment analysis. Python. Youtube. Football coach. Coritiba Football Club. Lexicon for Adapted Inference.

## LISTA DE FIGURAS

|  |    |
|--|----|
| FIGURA 1 - TERMOS E RESULTADO DE BUSCA DO CANAL OFICIAL .....                | 12 |
| FIGURA 2 - PRESENÇA DOS CLUBES POR REDE SOCIAL.....                          | 13 |
| FIGURA 3 - FUNCIONAMENTO GENERALIZADO DA ABORDAGEM LÉXICA .....              | 19 |
| FIGURA 4 - FLUXOGRAMA METODOLÓGICO .....                                     | 23 |
| FIGURA 5 - TELA DE OBTENÇÃO DE CHAVE DA API .....                            | 24 |
| FIGURA 6 - COMANDO PARA INSTALAÇÃO DA API .....                              | 24 |
| FIGURA 7 - CANAL OFICIAL DO CORITIBA.....                                    | 25 |
| FIGURA 8 - CÓDIGO DE COLETA DOS COMENTÁRIOS DOS VÍDEOS .....                 | 26 |
| FIGURA 9 - EXECUÇÃO DA FUNÇÃO DE COLETA PARA CADA VÍDEO .....                | 27 |
| FIGURA 10 - EXEMPLO DATAFRAME INICIAL CRIADO .....                           | 28 |
| FIGURA 11 - CÓDIGO DO PRÉ-PROCESSAMENTO DOS DADOS.....                       | 29 |
| FIGURA 12 - DATAFRAME COM A CRIAÇÃO DA COLUNA "CLEAN_COMMENT"<br>.....       | 30 |
| FIGURA 13 – EXEMPLO DE CÓDIGO DA BIBLIOTECA LEIA.....                        | 32 |
| FIGURA 14 - CÓDIGO DA APLICAÇÃO DO LEIA.....                                 | 34 |
| FIGURA 15 - FUNÇÃO DE POLARIDADE DO COMENTÁRIO .....                         | 35 |
| FIGURA 16 - PORCENTAGEM DE CADA POLARIDADE .....                             | 36 |
| FIGURA 17 - NUVEM DE PALAVRAS RELACIONADA AOS COMENTÁRIOS<br>POSITIVOS.....  | 37 |
| FIGURA 18 - NUVEM DE PALAVRAS RELACIONADA AOS COMENTÁRIOS<br>NEGATIVOS ..... | 38 |
| FIGURA 19 - MÉDIA DO COMPOUND POR VÍDEO .....                                | 38 |

## LISTA DE QUADROS

|   |    |
|---|----|
| QUADRO 1 - TRABALHOS CORRELATOS DA BIBLIOTECA LEIA..... | 31 |
|---|----|

## LISTA DE TABELAS

|  |    |
|--|----|
| TABELA 1 - DADOS DOS CANAIS DO <i>YOUTUBE</i> DOS CLUBES BRASILEIROS . | 12 |
| TABELA 2 – QUANTIDADE DE COMENTÁRIOS COLETADOS POR VÍDEO .....         | 27 |

## LISTA DE ABREVIATURAS OU SIGLAS

API – *Application Programming Interface*

CBF – Confederação Brasileira de Futebol

CSV – (*Common Separated Values*)

FC – Futebol Clube

NLP – *Natural Language Processing*

NLTK (*Natural Language Toolkit*)

LeIA – Léxico para Inferência Adaptada

GI – Gestão da Informação

VAR – *Video assistant referee*

VADER – *Valence Aware Dictionary for Sentiment Reasoning*

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1 INTRODUÇÃO</b> .....                               | <b>11</b> |
| 1.1 PROBLEMÁTICA .....                                  | 14        |
| 1.2 OBJETIVOS.....                                      | 15        |
| 1.2.1 Objetivo geral .....                              | 15        |
| 1.2.2 Objetivos específicos .....                       | 15        |
| 1.3 JUSTIFICATIVA.....                                  | 15        |
| <b>2 REFERENCIAL TEÓRICO</b> .....                      | <b>17</b> |
| 2.1 ANÁLISE DE SENTIMENTOS .....                        | 17        |
| 2.2 ANÁLISE DE SENTIMENTOS BASEADA EM LÉXICOS .....     | 18        |
| 2.3 TÉCNICO DE FUTEBOL.....                             | 19        |
| 2.4 CAMPEONATO PARANAENSE DE FUTEBOL .....              | 21        |
| <b>3 MATERIAIS E MÉTODOS</b> .....                      | <b>21</b> |
| 3.1 USO DA API DO <i>YOUTUBE</i> .....                  | 23        |
| 3.2 COLETA DOS DADOS .....                              | 25        |
| 3.3 PRÉ-PROCESSAMENTO DOS DADOS.....                    | 28        |
| 3.4 BIBLIOTECA LEIA .....                               | 30        |
| 3.4.1 Trabalhos correlatos da biblioteca Lela.....      | 30        |
| 3.4.2 Classificação com a biblioteca Lela .....         | 32        |
| <b>4 APRESENTAÇÃO DOS RESULTADOS</b> .....              | <b>35</b> |
| <b>5 CONSIDERAÇÕES FINAIS</b> .....                     | <b>39</b> |
| 5.1 TRABALHOS FUTUROS .....                             | 40        |
| <b>REFERÊNCIAS</b> .....                                | <b>41</b> |
| <b>APÊNDICE 1 – LISTA DOS VÍDEOS SELECIONADOS</b> ..... | <b>45</b> |

## 1 INTRODUÇÃO

Com o exponencial crescimento do uso da internet e das redes sociais, um novo contexto interativo entre pessoas e organizações tem sido estabelecido. Usuários publicam suas opiniões em seus perfis a respeito de determinados temas que podem influenciar tanto positivamente, quanto negativamente a imagem das organizações (FILHO CARVALHO, 2014).

Dentre estas redes/mídias sociais, segundo uma pesquisa feita pelo blog Conversion (2023), o *Youtube* foi classificado como a segunda rede social mais usada pelos brasileiros, com 84 milhões de usuários ativos por dia. Fundada em 2005, é uma plataforma de vídeos online que permite seus usuários assistirem, criarem, compartilharem e comentarem sobre vídeos pela internet, tendo mais de um bilhão de usuários cadastrados pelo mundo (SOUZA, 2023); Al-Mosimi *et al* (2017, p. 16173) destacam que esta presença, por sua vez, gera um grande acúmulo de dados opinativos nas redes sociais, que se forem analisados corretamente, podem beneficiar em uma ampla gama de aplicações como rastreamento das opiniões de usuários para identificação de níveis de satisfação ou aspectos a serem melhorados em produtos ou serviços.

Neste contexto, os clubes de futebol notaram o poder da popularização das redes sociais como o *YouTUBE*, devido a oportunidade de compartilhar conteúdo exclusivo e diversificado para seus torcedores, como entrevistas com jogadores, treinamentos, vídeos dos bastidores dos jogos e outros materiais que proporcionam a eles uma perspectiva única do clube e de seus ídolos. A TABELA 1, apresenta os valores de inscritos e número de vídeos publicados ao todo por clubes brasileiros pertencentes ao Campeonato Brasileiro de Futebol do ano de 2023, coletados a partir da busca pelos canais oficiais destes clubes na plataforma do *Youtube*, utilizando os descritores “nome do clube”, por exemplo “Flamengo” mais o termo “oficial”, retornando sempre como primeiro resultado o respectivo canal, ilustrado pela FIGURA 1. Já a FIGURA 2, abrange a presença dos clubes brasileiros nas principais redes sociais.

FIGURA 1 - TERMOS E RESULTADO DE BUSCA DO CANAL OFICIAL



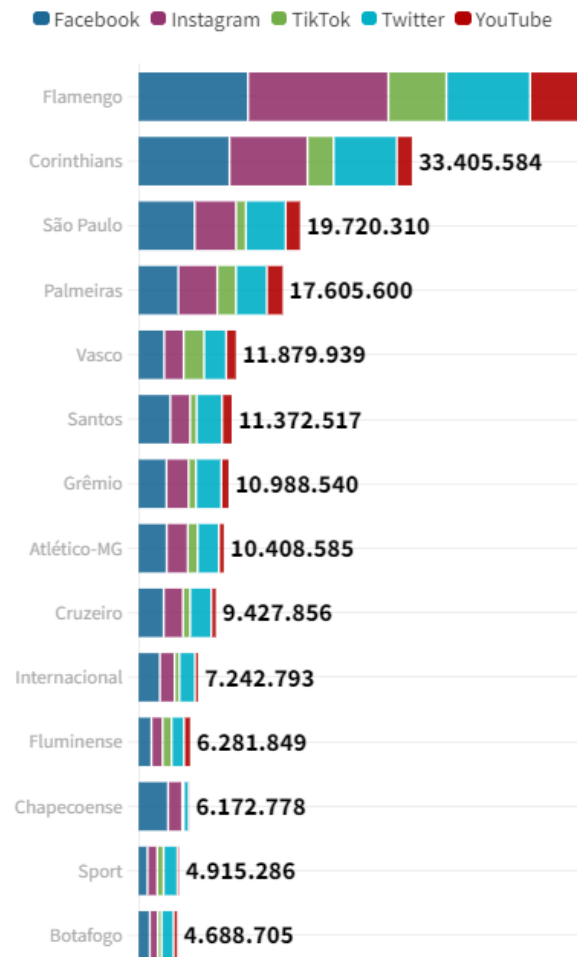
FONTE: O autor (2023).

TABELA 1 - DADOS DOS CANAIS DO YOUTUBE DOS CLUBES BRASILEIROS

| Clube               | Canal do YouTube         | Número de Inscritos (aproximado) | Número de vídeos publicados (aproximado) |
|---------------------|--------------------------|----------------------------------|--|
| América Mineiro     | Coelho TV                | 21,7 mil                         | 3,7 mil                                  |
| Atlético Mineiro    | Galo na Veia             | 602 mil                          | 14 mil                                   |
| Atlético Paranaense | Athletico Paranaense     | 240 mil                          | 4,3 mil                                  |
| Bahia               | TV BAHÉA                 | 219 mil                          | 838                                      |
| Botafogo            | Botafogo TV              | 430 mil                          | 3,4 mil                                  |
| Bragantino          | Massa Bruta TV           | 20,9 mil                         | 948                                      |
| Corinthians         | Corinthians TV           | 1,86 milhões                     | 5,7 mil                                  |
| Coritiba            | Coritibaoficial          | 156 mil                          | 4,5 mil                                  |
| Cruzeiro            | Cruzeiro Esporte Clube   | 551 mil                          | 4,3 mil                                  |
| Cuiabá              | Cuiabá Esporte Clube     | 24,5 mil                         | 1,2 mil                                  |
| Flamengo            | FLA TV                   | 850 mil                          | 8,5 mil                                  |
| Fluminense          | Fluminense               | 734 mil                          | 4,8 mil                                  |
| Fortaleza           | TV Leão Novibet          | 227 mil                          | 2,2 mil                                  |
| Goiás               | Goiás Esporte Clube      | 107 mil                          | 2 mil                                    |
| Grêmio              | Grêmio FBPA              | 884 mil                          | 5,8 mil                                  |
| Internacional       | Sport Club Internacional | 353 mil                          | 2,8 mil                                  |
| Palmeiras           | TV Palmeiras/FAM         | 1,94 milhões                     | 8 mil                                    |
| Santos              | Santos Futebol Clube     | 1,6 milhões                      | 7,1 mil                                  |
| São Paulo           | São Paulo FC             | 1,74 milhões                     | 1,9 mil                                  |
| Vasco da Gama       | VascoTV                  | 1,19 milhões                     | 914                                      |

FONTE: O autor (2023).

FIGURA 2 - PRESENÇA DOS CLUBES POR REDE SOCIAL



FONTE: Lance (2023).

Por mais que redes sociais como *Facebook*, *Instagram* e *Twitter* tenham mais fluxo de usuários, é imprescindível a possibilidade de o *Youtube* gerar informações pertinentes que possam auxiliar na gestão destes clubes e de sua relação com os torcedores devido a quantidade de conteúdo publicado e seus respectivos segmentos, sejam entrevistas coletivas para informar a respeito do gerenciamento do clube, revelações de contratações novas e a possibilidade de analisar o volume de dados gerados por estas publicações, em um contexto opinativo.

Nesta direção, o presente trabalho propõe um estudo no âmbito da análise de sentimentos, focando recuperar comentários de vídeos publicados do canal oficial da plataforma do *Youtube* do Coritiba Futebol Clube, a fim de classificá-los em negativos, positivos ou neutros, com o objetivo de auxiliar o clube e o técnico na tomada de

decisões no começo da temporada de futebol e a entender os sentimentos da torcida no âmbito da performance do time.

## 1.1 PROBLEMÁTICA

De acordo com um levantamento financeiro feito pela EY (2022), os clubes de futebol brasileiro atingiram um faturamento conjunto de R\$ 8,1 bilhões de reais. Levando em consideração este valor relatado, Raja (2006) considera os clubes de futebol um tipo de negócio especial, pois buscam o bom desempenho financeiro e o êxito esportivo. pois a existência de um clube de futebol está condicionada a obtenção de resultados financeiros satisfatórios, todavia, estes somente terão uma trajetória de crescimento sustentado mediante o bom desempenho esportivo.

Logo, neste âmbito da busca pelo êxito esportivo dos clubes de futebol, surge a figura do técnico, que, de acordo com Azevedo, Almeida e Ramalho (2021):

A busca pelo êxito esportivo, torna o técnico uma das figuras centrais no futebol, pelas diversas e importantes funções que exerce dentro de um clube. Cabe ao técnico coordenar a formação da equipe, auxiliando no processo de dispensas, contratações e renovações de jogadores, além da responsabilidade de liderança junto ao elenco e o desenvolvimento das opções táticas dentro dos jogos, selecionando os atletas mais preparados (AZEVEDO C. O; ALMEIDA, A. T. C; RAMALHO, H. M. B, 2021).

Para auxiliar os técnicos de futebol brasileiros nos projetos de desempenho estratégico e esportivo nas temporadas de futebol, surgem os campeonatos estaduais brasileiros, que oferecem uma oportunidade valiosa para os clubes testarem suas equipes e avaliarem o desempenho de seus jogadores e comissão técnica. É uma chance de observar as novas contratações em ação, avaliar o entrosamento do time e ajustar táticas e estratégias. Por fim, os campeonatos estaduais representam uma preparação importante para os times que irão disputar as demais competições nacionais e internacionais, como o Campeonato Brasileiro e a Copa do Brasil.

Conseqüentemente, a possibilidade de analisar grande volume de dados presentes nas redes sociais perante o contexto apresentado, faz se a seguinte pergunta a ser respondida aplicadamente pelo trabalho:

**Como identificar os sentimentos dos torcedores do Coritiba em redes/mídias sociais como o Youtube, com relação à performance do técnico do clube no começo da temporada de futebol de 2023?**

## 1.2 OBJETIVOS

Neste tópico são discorridos os objetivos gerais e específicos de maneira que responda ao problema de pesquisa.

### 1.2.1 Objetivo geral

Identificar e realizar uma análise de sentimentos dos torcedores com relação a performance do técnico do clube com base nos resultados de início de temporada do Coritiba FC na totalidade do Campeonato Paranaense de Futebol, utilizando dos comentários feitos nos vídeos postados pelo clube no *Youtube*.

### 1.2.2 Objetivos específicos

- Buscar e selecionar os vídeos publicados pelo Coritiba Futebol Clube para coleta do campo de comentários;
- Descrever as etapas e as bibliotecas utilizadas para a coleta, pré-processamento e classificação dos sentimentos na linguagem *Python*.
- Relatar trabalhos correlatos que tenham utilizado da mesma biblioteca para a classificação dos sentimentos;
- Aplicar o modelo léxico selecionado para classificar os sentimentos dos torcedores.
- Construir uma base de dados com os comentários criados e suas respectivas métricas de análise de sentimentos;
- Evidenciar e analisar os resultados.

## 1.3 JUSTIFICATIVA

Nesta seção são discorridas as justificativas pelo qual o estudo a ser realizado é relevante e importante nos seguintes quatro aspectos: acadêmica, científica (voltado ao curso de gestão da informação), pessoal e social.

A contribuição acadêmica se da prática multidisciplinar a partir de teorias ensinadas a partir de conceitos como Programação de computadores para Gestão da Informação, Tópicos em análise de dados e Mineração de Dados.

A contribuição científica do trabalho, principalmente no âmbito da área e do curso da gestão da informação se deve ao fato de que, por meio de uma breve pesquisa na área da análise de sentimentos realizada na base Acervo UFPR e na busca por assunto do Portal da CAPES. A busca combinada com os descritores “análise de sentimentos” e “youtube” no campo “todos os campos” não retornou resultados no Acervo UFPR, enquanto no Portal da Capes, retornou 13 resultados, mas que olhados a fundo correspondiam a 4 resultados únicos. Logo, a realização do estudo busca preencher esta lacuna como uma forma de documentação de recuperação dos dados do *Youtube* e a utilização de bibliotecas que são constantemente aprimoradas para o processo de análise de sentimentos em redes sociais para conteúdo na língua portuguesa, com o intuito de colaborar com as metodologias de coleta, tratamento e análise dos dados voltados para esta rede social, no caso para o escopo deste trabalho, o futebol.

Do ponto de vista pessoal, o tema escolhido tange o conhecimento adquirido ao realizar as etapas da elaboração do trabalho e a carga de conhecimento profissional que agrega ao extrair, processar e analisar grande volume de dados, além de aprender conforme a execução do projeto ao pesquisar e definir as metodologias e tecnologias/ferramentas que podem ser utilizadas no futuro em um ambiente profissional no contexto de análise de sentimentos. Ademais, houve interesse em definir o tema de pesquisa centrado no futebol, por ser um conteúdo que é bastante ativo no cotidiano do povo brasileiro e consumido pelo próprio pesquisador, principalmente nas redes sociais como *Facebook* e *Youtube*, através de fotos, comentários, textos e vídeos.

Já no cunho social, o trabalho busca demonstrar a relevância que a análise de sentimentos pode continuar tendo no âmbito das redes/mídias sociais, como por exemplo o *Youtube*, como ferramenta estratégica para os clubes de futebol lidarem com opiniões públicas a respeito da performance nos campeonatos ou em um panorama mais geral, visto que estas mídias estão difundidas no cotidiano do brasileiro, sendo a segunda mais usada no Brasil no ano de 2023 (RESULTADOS DIGITAIS, 2023).

## 2 REFERENCIAL TEÓRICO

Neste segmento será abordado o referencial teórico referente ao material e conteúdo encontrado para a elaboração da metodologia e escopo do trabalho, tratando de temas como análise de sentimentos, sua abordagem léxica, a função dos técnicos de futebol para/com o clube, o Campeonato Paranaense de Futebol e sua edição de 2023 para o Coritiba Futebol Clube.

### 2.1 ANÁLISE DE SENTIMENTOS

A análise de sentimentos, ou mineração de opinião, é uma área ativa de estudo no campo do processamento de linguagem natural que analisa as opiniões, sentimentos, avaliações, atitudes e emoções das pessoas por meio do tratamento computacional de subjetividade no texto (HUTTO; GILBERT, 2014).

Tradicionalmente, a análise de sentimentos tem se concentrado na análise de textos usando Processamento de Linguagem Natural e técnicas de Aprendizagem de Máquina baseadas em características. Os avanços em disciplinas como *Big data* e *Deep Learning* têm impactado e beneficiado a evolução do campo. Esta edição especial inclui quatro trabalhos que propõem novas técnicas (IGLESIAS; MORENO, 2019).

Ainda Iglesias e Moreno (2019), os autores abordam que a ampla gama de aplicações da análise de sentimentos tem fomentado sua evolução e que suas técnicas permitiram dar sentido ao grande volume de dados das redes sociais para tomar decisões mais assertivas e baseada em dados, bem como compreender eventos sociais, marketing de produtos ou eventos políticos.

O início e o rápido crescimento da área da análise de sentimentos coincidem com a das redes sociais, por exemplo, resenhas, discussões em fóruns, *blogs*, *microblogs*, Twitter e redes sociais, porque pela primeira vez na história, temos um enorme volume de dados com opinião registrada em formulários digitais (LIU, 2012).

Benevenuto, Ribeiro e Araújo (2015) abordam que as opiniões presentes nas redes sociais, se devidamente recolhidas e analisadas, permitem não só compreender e explicar diversos fenômenos sociais complexos, mas também os prever. Levando em consideração a capacidade de processamento tecnologia dos dias de hoje, é possível o armazenamento e recuperação de um grande volume de dados

eficientemente, pois o foco atual está em desenvolver metodologias para extração de informações e criação de conhecimento a partir de fontes de dados distintas. Ou seja, com o advento de novas ferramentas e metodologias na área, o processo de análise de sentimentos reduz drasticamente a necessidade de se ler documentos para extrair opiniões.

Uma das dificuldades da área está em extrair opiniões e informações úteis expressas através de mensagens curtas de texto, e assim gerar rapidamente dados que devem ser analisados de forma eficiente, para finalmente serem utilizados em tomadas de decisões e em processos de tempo real. Esta dificuldade gera um problema multidisciplinar na computação, envolvendo a mineração de dados, o processamento de linguagem natural e a aprendizagem de máquina (BENEVENUTO; RIBEIRO e ARAÚJO, 2015).

## 2.2 ANÁLISE DE SENTIMENTOS BASEADA EM LÉXICOS

A abordagem léxica utiliza dicionários ou coleções de palavras de opinião com anotações que indicam o sentimento em positivo ou negativo. Essa abordagem pode ser criada de forma manual ou de maneira automática a partir de uma lista de palavras, ademais, algumas pesquisas utilizam adjetivos como indicadores de orientação semântica do texto (JUNQUEIRA; FERNANDES, 2018).

Junqueira e Fernandes (2018) relatam que a classificação de sentimentos baseada em léxicos se divide em duas abordagens, sendo elas a de dicionário e a de *Corpus*. A abordagem baseada em dicionários utiliza um conjunto de palavras com anotações de polaridade e classifica a sentença baseado na polaridade de cada palavra presente no dicionário. São exemplos de dicionários *SentiLex*, *OpLexicon*, *VADER*, *LIWC* e a *LelA*, utilizada no trabalho.

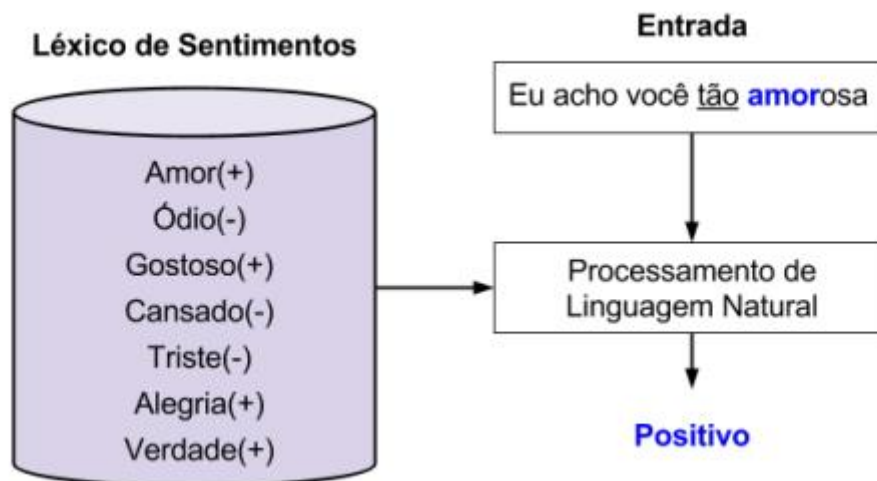
Já a abordagem baseada em *Corpus* depende de padrões sintáticos ou padrões que ocorrem no corpo do texto, sendo *Corpus Corpus* o corpo do texto de linguagem natural utilizado para acumular estatísticas sobre a linguagem natural processada. As informações extraídas incluem anotações de cada palavra que indicam suas partes do discurso como adjetivos, substantivos, advérbios, entre outros.

Taboada *et al* (2011) discorrem que as abordagens léxicas assumem que palavras individuais possuem polaridades prévias, que seriam orientações semânticas que independem de contexto e que pode ser expressada com um valor numérico ou

classe, por exemplo entre -1 e 1 (do qual -1 é o valor sentimental mais negativo e 1 o valor sentimental mais positivo) ou positivo/negativo e/ou feliz/triste.

A FIGURA 3 generaliza o funcionamento de um método de análise de sentimentos com a abordagem léxica. Em suma, o processo de classificação se inicia quando o método recebe um texto de entrada e em seguida é realizado alguns processos de linguagem de processamento natural. Além disso, é feita uma pesquisa no léxico dos termos que compõem a sentença lida e que já possuem uma polaridade previamente definida no método, retratado na parte esquerda da figura. Por último é inferido qual é a polaridade ou sentimento implícito no texto de entrada, sendo por exemplo, positivo, negativo ou neutro.

FIGURA 3 - FUNCIONAMENTO GENERALIZADO DA ABORDAGEM LÉXICA



FONTE: BENEVENUTO, RIBEIRO, ARAÚJO (2015).

Ainda Benevenuto, Ribeiro e Araújo (2015), os autores relatam que a análise de sentimentos baseadas em abordagens léxicas é atualmente uma das estratégias mais eficientes, tanto na questão de predição dos sentimentos, quanto na capacidade computacional em lidar com os códigos e processamento.

### 2.3 TÉCNICO DE FUTEBOL

De acordo com a Confederação Brasileira de Futebol (CBF), técnicos de futebol são profissionais responsáveis por comandar uma equipe em todas as suas

atividades para que ela obtenha o melhor desempenho e resultados nas competições que participa (CBF, 2023). A entidade também relata que a figura do técnico gerencia o grupo de jogadores que são definidos pelo próprio profissional, em conjunto com a diretoria do clube e indicações de contratações que podem ser feitas.

Além disso, os técnicos de futebol são determinantes nos resultados das partidas, visto a constante evolução do futebol, em virtude da necessidade de trabalhos sistemáticos e eficientes em campeonatos longos. A fórmula do sucesso de uma equipe se dá a partir de um esquema tático que aproveita as habilidades dos jogadores que o técnico tem em sua disposição, somadas a uma filosofia de jogo bem estruturada.

Outros fundamentos essenciais para um técnico são voltados para a comunicação entre ele e os jogadores, sendo de suma importância que ela, em meio a competições, seja eficaz para o técnico tendo em vista a conquista da sua liderança perante o grupo de trabalho e para o jogador obter motivação, autoconfiança e a autoestima necessária para obter um rendimento favorável (SANTOS; RODRIGUES, 2008).

A CBF (2023), estabelece algumas das funções exercidas por um técnico de futebol, no seu dia-a-dia:

- O técnico ministra a maioria dos treinos, estabelecendo como deseja que o time atue, tanto em termos técnicos quanto táticos, definindo o esquema de jogo e as alternativas que podem ser preparadas, além de ensaiar algumas jogadas específicas e observar os adversários;
- Escala o time para cada jogo, orienta os atletas e decide as substituições ao longo da partida, junto com sua comissão técnica. Para tanto, recebe informações de diferentes áreas, como preparação física, fisiologia, medicina e logística;
- Contato com os dirigentes do clube para definir as diretrizes do trabalho no dia a dia;
- Responde a questionamentos da imprensa sobre o desempenho do time a partir de coletivas de imprensa organizadas pelo clube, por exemplo.

Outras funções são relatadas por Carter e Bloom (2009), como orientar o treino das técnicas específicas da modalidade esportiva, fornecer instrução e

feedback e acompanhamento da aprendizagem e do desempenho dos atletas. Todas essas funções são planejadas e projetadas para ajudar os atletas a atingirem o seu potencial máximo.

## 2.4 CAMPEONATO PARANAENSE DE FUTEBOL

As temporadas de futebol no Brasil acontecem ao longo do ano e começam com os campeonatos estaduais, sendo o Campeonato Paranaense para o Coritiba. No decorrer do dos estaduais, dá-se início a Copa do Brasil.

O Campeonato Paranaense de Futebol é uma competição que envolve clubes de futebol do estado do Paraná, no sul do Brasil. A primeira edição foi disputada em 1915 e, desde então, tem sido uma das competições mais importantes do calendário esportivo da região. O formato da competição tem variado ao longo dos anos, mas atualmente é disputado por 12 clubes divididos em duas fases. Na primeira fase, os clubes jogam entre si em turno único e os oito primeiros avançam para as quartas de final. Na fase de mata-mata, os confrontos são definidos em jogos de ida e volta, até chegar às finais, também em duas partidas. Dentre os clubes mais tradicionais que já disputaram o Campeonato Paranaense estão o Coritiba, Atlético Paranaense e Paraná Clube. O Coritiba é o maior campeão do estado, com 39 títulos, seguido pelo Atlético Paranaense, com 26 títulos, e pelo Paraná Clube, com sete títulos.

A competição passou por mudanças importantes, com a participação de clubes do interior e a implementação de novas regras, como a utilização do *Video Assistant Referee* (VAR) ou árbitro de vídeo. Na edição de 2023, de acordo com a tabela de resultados fornecidos pelo Globo Esporte (2023), o Coritiba, sob o comando de seu novo técnico para esta temporada Antônio Oliveira, disputou 11 partidas na primeira fase da competição, tendo 6 vitórias, 4 empates e 1 derrota, com um aproveitamento de 66%. Já na fase de mata a mata, caiu nas quartas de final para o FC Cascavel, perdendo o primeiro jogo por 3x1 e empatando o segundo em 0x0.

## 3 MATERIAIS E MÉTODOS

No âmbito metodológico, o trabalho pode ser classificado como um estudo aplicado, interdisciplinar e exploratório, contemplando processos de coleta, pré-processamento, classificação e análise dos dados opinativos dos torcedores.

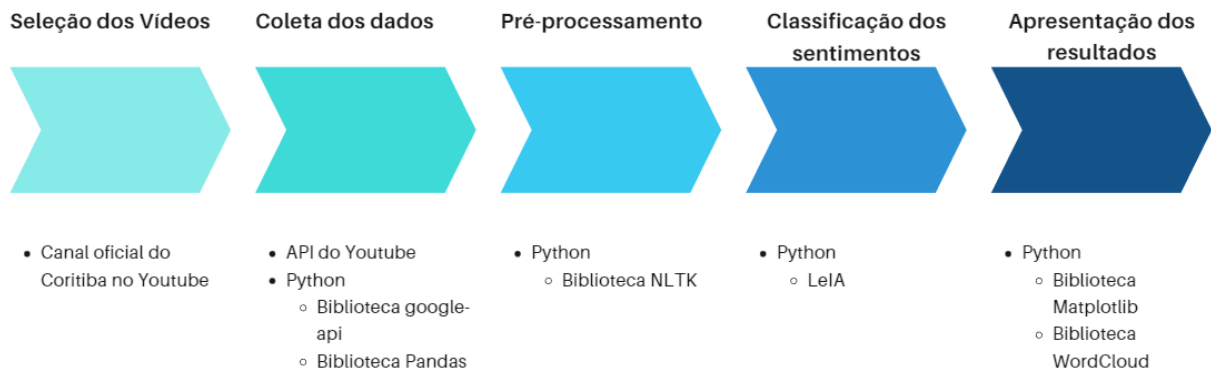
Foi definido como ferramenta principal a linguagem de programação Python, em sua versão 3.10.4 de 23 de março de 2022. Para o ambiente de programação dos códigos, foi utilizado o *Jupyter notebook*, que se trata de uma aplicação *web open-source* que permite criar e compartilhar documentos que possuem códigos ao vivo, equações, visualizações e explicações em *markdown*, sendo amplamente utilizado em ciência de dados, aprendizado de máquina, análise estatística e muitas outras áreas (JUPYTER NOTEBOOK, 2023).

A coleta dos dados foi realizada utilizando a API (*Application Programming Interface*) do *Youtube*, para resgate dos comentários referentes aos vídeos que englobam todas as coletivas de imprensa dos pós-jogos que o Coritiba participou na sua trajetória do Campeonato Paranaense de 2023.

No ambiente do *Jupyter Notebook* foram importadas algumas bibliotecas, sendo elas, a “*google-api-python-client*”, que serve como principal fonte de extração de dados a partir da API do Youtube, a “*Pandas*”, para manipulação e criação de *dataframes* que irão conter os comentários coletados, a “*Matplotlib*” e *Wordcloud*, para a parte de visualização de dados, o conjunto de bibliotecas NLTK (*Natural Language Toolkit*), para lidar com as questões de pré-processamento de linguagem natural dos comentários em português, a fim de deixá-los aptos para análise. Na fase da classificação e análise, foi selecionada e utilizada uma adaptação feita para o léxico em português da biblioteca VADER (*Valence Aware Dictionary and Sentiment Reasoner*) elaborada por Hutto e Gilbert (2014), denominada de LelA (Léxico para Inferência Adaptada), que contém as funções necessárias para a aplicação do modelo léxico de análise de sentimentos não supervisionado, baseado em polaridades positivas, negativas e neutras.

O fluxograma na FIGURA 4 resume a etapa metodológica do trabalho, contemplando as etapas e as bibliotecas utilizadas.

FIGURA 4 - FLUXOGRAMA METODOLÓGICO



FONTE: O autor (2023).

### 3.1 USO DA API DO YOUTUBE

De acordo com o site Red Hat (2022), API significa interface de programação de aplicações, ou seja, um conjunto de definições e protocolos para criar e integrar softwares de aplicações. No contexto de APIs, a palavra Aplicação refere-se a qualquer software com uma função distinta. A interface pode ser pensada como um contrato de serviço entre duas aplicações (AMAZON WEB SERVICES, 2022).

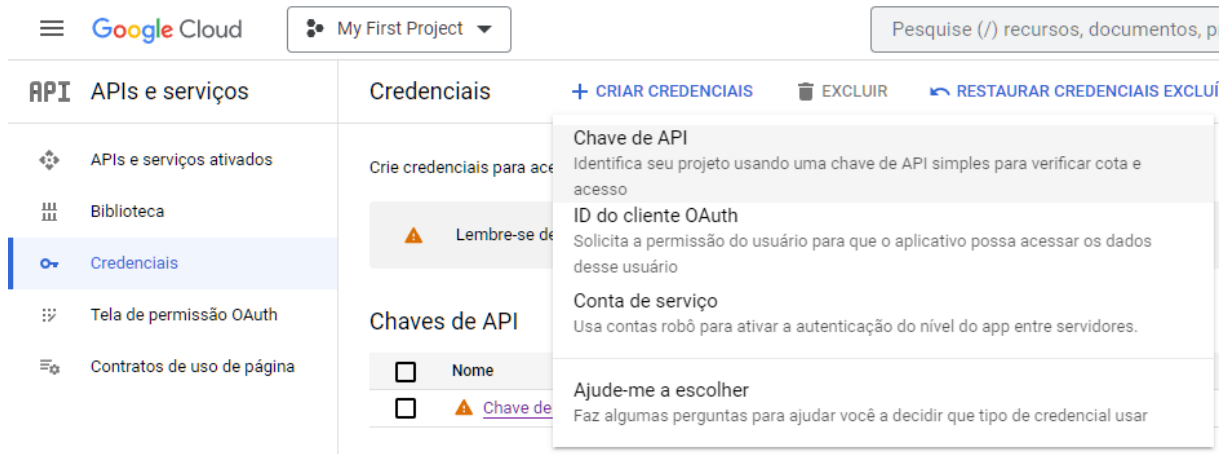
No âmbito da API do *Youtube*, ela permite acesso programático à plataforma de maneira única e avançada, permitindo acessar elementos como vídeos, comentários, lista de reprodução e mais. Atualmente, é disponibilizado a versão 3 da API, que a própria empresa recomenda para desenvolvedores utilizarem em seus projetos e teve sua última atualização em 22 de julho de 2022 (YOUTUBE, 2023).

Para começar usar a API e prosseguir com os objetivos propostos, foi necessário habilitar uma conta *Google* como desenvolvedora, dentro do painel de *Console* do *Google Cloud*<sup>1</sup>. A partir dele, foi possível criar um projeto e adicionar a API do Youtube dentro da biblioteca disponível, chamada de *Youtube Data API v3*, que dispõe dos dados públicos da mídia social.

<sup>1</sup> Site para acesso ao painel de desenvolvedor da Google para acesso a API: <https://console.cloud.google.com/apis/library/browse?project=gold-enterprise-335222&q=youtube>

Dentro do projeto, conforme a FIGURA 5, foi necessário criar uma chave de API para que fosse possível se comunicar com os serviços do Youtube e obter os dados com o código de coleta dos dados.

FIGURA 5 - TELA DE OBTENÇÃO DE CHAVE DA API



FONTE: O autor (2023).

Com as credenciais da API prontas para serem usadas na coleta dos dados, foi necessário instalar as dependências do da API do Google para usufruir dos seus recursos no *Python*, conforme o próprio guia da Google<sup>2</sup> e exemplificado na FIGURA 6.

FIGURA 6 - COMANDO PARA INSTALAÇÃO DA API

### Install the Google client library

- Install the Google client library for Python:

```
pip install --upgrade google-api-python-client google-auth-http2 google-auth-oauthlib
```

FONTE: Google (2023).

<sup>2</sup> Site para download das bibliotecas para uso da API: <https://developers.google.com/docs/api/quickstart/python>

### 3.2 COLETA DOS DADOS

Para a etapa da coleta dos dados referentes aos comentários dos torcedores do Coritiba, foi identificado o canal do Youtube oficial do clube, demonstrado na FIGURA 7, seguindo o levantamento realizado anteriormente dos canais oficiais de cada clube.

FIGURA 7 - CANAL OFICIAL DO CORITIBA



FONTE: Coritibaoficial (2023).

Com 105 mil inscritos, o canal do Coritiba no Youtube é constantemente atualizado com novos vídeos de bastidores, treinos, entrevistas coletivas e novidades relacionadas ao clube. Foram selecionados somente os vídeos disponíveis que pertencem ao cronograma do campeonato paranaense, tendo como referência a página do Globo Esporte<sup>3</sup>, sendo eles coletivas de imprensa postadas pelo clube, com principal enfoque no resultado do jogo e performance da equipe e decisões feitas durante o jogo, relatadas pelo técnico atual. Este tipo de seleção teve como objetivo filtrar ao máximo os comentários que abrangem o resultado da partida e não por exemplo, selecionar vídeos que apresentam o gol ou lance importante no meio de uma partida em andamento, dos quais os comentários podem não refletir o sentimento

<sup>3</sup> Site do Globo Esporte com as rodadas do Campeonato Paranaense de 2023: <https://ge.globo.com/pr/futebol/campeonato-paranaense/noticia/2022/11/11/federacao-divulga-a-tabela-do-campeonato-paranaense-de-2023-confira-as-rodadas.ghtml>

final do torcedor. Para isso, foi elaborado uma lista destes vídeos (APÊNDICE 1) que dispõem das partidas, suas respectivas datas e resultados.

Estes vídeos possuíram os comentários que foram coletados para construir o *dataframe*<sup>4</sup> com os textos a serem analisados no trabalho. Como explicado anteriormente, a linguagem de programação Python dentro do ambiente do *Jupyter notebook* é a ferramenta principal para todas as etapas da análise de sentimentos. O código da FIGURA 8 demonstra as bibliotecas que foram importadas e a função criada para a coleta dos comentários.

FIGURA 8 - CÓDIGO DE COLETA DOS COMENTÁRIOS DOS VÍDEOS

```

1 from google.oauth2 import service_account
2 from googleapiclient.discovery import build
3 import pandas as pd
4
5 # define as credenciais de API do YouTube
6 api_key = 'AIzaSyARI5Qhn9JtImDV5-zj8QDznflmnn5dMxk'
7
8 # cria um objeto de serviço para acessar a API do YouTube
9 youtube = build('youtube', 'v3', developerKey=api_key)
10
11 def get_comments_from_videos(video_id, video_name, csv_header):
12
13     # faz a primeira chamada para recuperar os comentários mais recentes
14     response = youtube.commentThreads().list(part='snippet', videoId=video_id, textFormat='plainText').execute()
15
16     # cria uma lista vazia com os dados dos comentários
17     data = []
18
19     for item in response['items']:
20         comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
21
22         data.append({'Video': video_name, 'Comment': comment})
23
24     # continua a busca com paginação
25     while 'nextPageToken' in response:
26         response = youtube.commentThreads().list(part='snippet', videoId=video_id, textFormat='plainText',
27             pageToken=response['nextPageToken']).execute()
28
29         for item in response['items']:
30             comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
31
32             data.append({'Video': video_name, 'Comment': comment})
33
34     # cria um DataFrame com os dados dos comentários, que são um dicionário
35     df = pd.DataFrame(data)
36     df.to_csv('comments.csv', mode = 'a', index=False, header = csv_header)

```

FONTE: O autor (2023).

O código acima coletou os comentários a partir de um “vídeo\_id” que está presente na URL no navegador de internet e armazena eles em um *dataframe*, que posteriormente exporta para um arquivo separado por vírgulas, ou seja, do tipo “CSV”

<sup>4</sup> Um data frame é semelhante a uma matriz mas as suas colunas têm nomes e podem conter dados de tipo diferente.  
([https://www.dcc.fc.up.pt/~ltorgo/SebentaR/HTML/node16.html#:~:text=Um%20data%20frame%20%C3%A9%20semelhante,registo%20\(linha\)%20da%20tabela.](https://www.dcc.fc.up.pt/~ltorgo/SebentaR/HTML/node16.html#:~:text=Um%20data%20frame%20%C3%A9%20semelhante,registo%20(linha)%20da%20tabela.)).

(*Comma Separated Values*) denominado “*comments.csv*”, para que possa ser lido novamente nas partes posteriores da análise.

Detalhe que o código resgatou somente os comentários e não as respostas deles, de maneira que fosse evitado ao máximo coletar registros que remetessem a discussões dos comentários que poderiam fugir do escopo dos vídeos selecionados. A FIGURA 9 contempla a chamada da função criada, com o parâmetro que corresponde a cada ID dos vídeos, apresentado na TABELA 2.

FIGURA 9 - EXECUÇÃO DA FUNÇÃO DE COLETA PARA CADA VÍDEO

```

1 get_comments_from_videos('_en053zxDYM', 'video_01', True)
2 get_comments_from_videos('irV4nbA_Kw0', 'video_02', False)
3 get_comments_from_videos('EvphSYjfBuQ', 'video_03', False)
4 get_comments_from_videos('i_kvrvjq3E', 'video_04', False)
5 get_comments_from_videos('V7LJKp3Jsgw', 'video_05', False)
6 get_comments_from_videos('pk3TgwYE67c', 'video_06', False)
7 get_comments_from_videos('VrKD6AigPBg', 'video_07', False)
8 get_comments_from_videos('lK8EYomU1sU', 'video_08', False)
9 get_comments_from_videos('eDJMG10qtwA', 'video_09', False)
10 get_comments_from_videos('8vKVGdww-DM', 'video_10', False)
11 get_comments_from_videos('Eaxvs9IOJGU', 'video_11', False)
12 get_comments_from_videos('6SVNwY9QRJY', 'video_12', False)
13 get_comments_from_videos('CnYehk4GSvA', 'video_13', False)

```

FONTE: O autor (2023).

A TABELA 2 dispõe da lista de “*vídeo\_id*” usados para coleta e a quantidade de comentários coletados por vídeo, que foram exportados no arquivo *.csv*.

TABELA 2 – QUANTIDADE DE COMENTÁRIOS COLETADOS POR VÍDEO

| Vídeo    | ID do vídeo | Quantidade de comentários coletados | Quantidade total de comentários |
|----------|-------------|-------------------------------------|---------------------------------|
| vídeo_01 | _en053zxDYM | 44                                  | 73                              |
| vídeo_02 | irV4nbA_Kw0 | 20                                  | 36                              |
| vídeo_03 | EvphSYjfBuQ | 29                                  | 45                              |
| vídeo_04 | i_kvrvjq3E  | 22                                  | 26                              |
| vídeo_05 | V7LJKp3Jsgw | 49                                  | 89                              |
| vídeo_06 | pk3TgwYE67c | 37                                  | 58                              |
| vídeo_07 | VrKD6AigPBg | 46                                  | 79                              |
| vídeo_08 | lK8EYomU1sU | 45                                  | 59                              |
| vídeo_09 | eDJMG10qtwA | 30                                  | 36                              |
| vídeo_10 | 8vKVGdww-DM | 59                                  | 69                              |
| vídeo_11 | Eaxvs9IOJGU | 151                                 | 182                             |
| vídeo_12 | 6SVNwY9QRJY | 202                                 | 241                             |

|          |             |     |      |
|----------|-------------|-----|------|
| vídeo_13 | CnYehk4GSvA | 153 | 218  |
| Total:   |             | 887 | 1211 |

FONTE: O autor (2023).

O valor da coluna “Quantidade de comentários coletados” representa a quantidade total que foi resgatada, sem coletar as respostas dos comentários, totalizando 887 registros. A coluna “ID do vídeo” corresponde às ID’s dos vídeos presentes do APÊNDICE 1, pegando os valores depois da expressão “v=” presente no *link* deles.

A FIGURA 10, exemplifica como que ficou o dataframe importando o “CSV” de comentários usando a biblioteca “*Pandas*”, com o método “*pd.read.csv ()*” e selecionado uma amostra dele com o método “*sample ()*”.

FIGURA 10 - EXEMPLO DATAFRAME INICIAL CRIADO

|     | Video    | Comment  |
|-----|----------|--|
| 344 | video_10 | Time ruim do caraio 🤔🤔🤔  |
| 879 | video_13 | Estão mesmo na contra mão com piada de português, perder numa competição agora é estar no caminho certo kkkkkk   |
| 83  | video_03 | maluco tá um "bom dia" ou "boa noite" sorrindo, seja qual for o jornalista/pergunta, ele deve ser mó gente boa kkkkkkkkkk  |
| 121 | video_05 | Não sei que critério utilizaram pra trazer esse técnico, é início de temporada, sim, mas até agora ele não mostrou a que veio, e o time está jogando igual ao do ano passado, perdido em campo, sem jogadas ensaiadas, esquema tático inexistente, e a utilizam das mesmas peças que não deram certo ano passado. Eu acho que o Antônio Oliveira não chega na quinta rodada do brasileiro. |
| 624 | video_12 | SUMAAAAAAAAAAAAAAAA FDP  |
| 419 | video_11 | Coritiba a maior torcida do Pr   |
| 62  | video_02 | As entrevistas do Antônio são ótimas, vão render muito ao longo da temporada, fala o que pensa sem medo. Agora alguns repórteres nem parece que pensam antes de perguntar.   |
| 212 | video_07 | Acho que os dois times vão jogar mais uns cinco joguinhos com torcida feminina   |
| 315 | video_09 | Treinar finalização urgente  |
| 233 | video_07 | A zaga ainda está fraca, precisa melhorar muito.   |

FONTE: O autor (2023).

### 3.3 PRÉ-PROCESSAMENTO DOS DADOS

A FIGURA 11 ilustra o código utilizado para alguma das etapas que vão ser utilizadas para o processamento dos corpos de texto dos comentários coletados e que

irão fazer parte da etapa da análise da frequência de palavras dos vídeos. Como a biblioteca LelA possui integrada a API do VADER, não será necessário o pré-processamento para a classificação dos sentimentos de cada um dos comentários, sendo uma das vantagens das abordagens léxicas.

Esta etapa de pré-processamento dos comentários consiste em técnicas de processamento de linguagem natural como remoção de pontuações, *stopwords* e outros caracteres que podem influenciar na etapa da análise dos comentários a partir da visualização em nuvem de palavras, realizada posteriormente no *corpus* do trabalho.

FIGURA 11 - CÓDIGO DO PRÉ-PROCESSAMENTO DOS DADOS

```
import pandas as pd
import nltk
from nltk.corpus import stopwords
import string

comments_df = pd.read_csv("comments.csv")

#Remoção de textos duplicados nos comentários
comments_df.drop_duplicates(subset="Comment",
                           keep="first", inplace=True)

# Remoção de stopwords
stopwords = set(stopwords.words("portuguese"))

comments_df['Clean_comment'] = comments_df['Comment'].apply(lambda x: " ".
                                                             join(word.lower() for word in x.split()
                                                             if word.lower() not in stopwords))

# Remoção de pontuações exceto '!'
punctuations = string.punctuation.replace("!", "")
comments_df['Clean_comment'] = comments_df['Clean_comment'].apply(lambda x: " ".
                                                                    join(word for word in x if word not in punctuations))

comments_df.head(100)
```

FONTE: O autor (2023).

Primeiramente, a aplicação acima recupera o arquivo “CSV” criado anteriormente contendo todos os comentários recuperados dos vídeos. Em seguida, foi removido entradas duplicadas nos comentários, a partir da função “drop\_duplicates” de maneira que fosse mantido o primeiro registro, mas que os demais fossem apagados da base de dados, com o intuito de não contabilizar o mesmo texto postado por um torcedor mais de uma vez. Este procedimento excluiu 31 registros da base, ficando 855 registros únicos.

Utilizando as bibliotecas e funções importadas do NLTK foi possível remover *stopwords* que são palavras que não necessitam ser indexadas, por possuírem pouco significado, tais como preposições, artigos, conjunções e outros (IBICT, 2015). Além

disso, com o pacote “*string*” é possível remover de maneira automatizada pontuações no texto.

Com isso, foi criada uma coluna no *dataframe* de comentários, denominada “Clean\_comment” para que tivesse um registro separado com o conteúdo pré-processado, exemplificado na FIGURA 12.

FIGURA 12 - DATAFRAME COM A CRIAÇÃO DA COLUNA “CLEAN\_COMMENT”

|     | Video    | Comment  | Clean_comment   |
|-----|----------|--|---|
| 354 | video_10 | Os repórteres têm medo do técnico kkkk pqp só pergunta elogiando o cara  | repórteres têm medo técnico kkk pqp pergunta elogiando cara   |
| 447 | video_11 | O time não tem organização tática, não tem uma jogada ensaiada. O português tomou uma aula de futebol do treinador do Operário e ainda paga de arrogante. Acho que nem começou direito e já deu nel  | time organização tática jogada ensaiada português tomou aula futebol treinador operário ainda paga arrogante acho começou direito deu nel   |
| 464 | video_11 | Cara folgado e arrogante esse mané.  | cara folgado arrogante mané   |
| 702 | video_12 | Tá rindo. Tá rindo da nossa cara. Fora Antônio Oliveira  | tá rindo tá rindo cara antônio oliveira   |
| 440 | video_11 | Esse time tem direito de reclamar do gramado ? Tá pior é time d bairro c brasileiro vai ser Saku de pancada  | time direito reclamar gramado tá pior time d bairro c brasileiro vai saku pancada   |
| 360 | video_10 | Akuma concordo com vc.   | akuma concordo vc   |
| 834 | video_13 | Jogou bem e não marcou gol nunca vi isso   | jogou bem marcou gol nunca vi   |
| 844 | video_13 | Lamentável. Já não confio mais no trabalho desse grupo e nem nesse treinador. Esse treinador é muito cheio de querer, quer determinar o que o repórter vai perguntar. O procurador do TJD estava certo em sua colocação, levou mais um cartão amarelo é muito nervosinho e arrogante igual o Abel e o técnico do flamengo. | lamentável confio trabalho desse grupo nesse treinador treinador cheio querer quer determinar repórter vai perguntar procurador tjd certo colocação levou cartão amarelo nervosinho arrogante igual abel técnico flamengo |
| 562 | video_12 | Ridículo esse Português o time perde de 3 x 1 e ele fica dando rizadinhas na entrevista manda embora esse cara, do outro lado viramos piada pro Atlético.  | ridículo português time perde 3 x 1 fica dando rizadinhas entrevista manda embora cara outro lado viramos piada pro atlético  |
| 202 | video_07 | Olha eu escutando as respostas do técnico português realmente me passaram muita segurança e ele não aceitou o que ocorreu em campo neste jogo o qual foi péssimo exemplo do comportamento dos seus próprios jogadores parabéns   | olha escutando respostas técnico português realmente passaram muita segurança aceitou ocorreu campo neste jogo péssimo exemplo comportamento próprios jogadores parabéns  |

FONTE: O autor (2023).

### 3.4 BIBLIOTECA LEIA

Nesta seção, foi realizado uma busca de artigos científicos que tenham utilizado a biblioteca LeIA como método de análise de sentimentos e descreve esta etapa no contexto do trabalho proposto.

#### 3.4.1 Trabalhos correlatos da biblioteca LeIA

Com o intuito de trazer um repertório do uso científico da biblioteca escolhida para a análise dos sentimentos, foi realizada uma busca pelo portal da Capes e do *Google Scholar* utilizando os descritores “Léxico para inferência adaptada” mais “*github*” (plataforma onde está disponibilizada a biblioteca para uso), filtrando por publicações na língua portuguesa.

O Portal da Capes não retornou resultados, enquanto que no *Google Scholar* foram encontrados sete trabalhos que utilizaram o léxico proposto por Almeida (2018), apresentados no QUADRO 1.

QUADRO 1 - TRABALHOS CORRELATOS DA BIBLIOTECA LEIA

| <b>Título</b>  | <b>Autor(es)</b>  | <b>Data de publicação</b> |
|--|---|---------------------------|
| Análise de sentimento como apoio à seleção de livros: um estudo aplicado à plataforma Skoob        | MARINHO, Ronnie Shida;<br>PEREIRA, Clayton Martins;<br>SANTAREM SEGUNDO, José Eduardo | 2022                      |
| Processamento de linguagem natural para detecção de padrões no mercado de ações brasileiro         | PEREIRA, Francielle Vasconcellos  | 2022                      |
| Análise da subjetividade em língua portuguesa no <i>Twitter</i> relacionada a pandemia da COVID-19 | MALAKIN, Lucas Alexandre;<br>REZENDE, Solange Oliveira                                | 2022                      |
| Suporte à análise interativa de discussões online combinando técnicas de mineração de dados        | ALMEIDA, Rafael José de Alencar   | 2020                      |
| Detecção de Bots no Twitter através de técnicas de Processamento de Linguagem Natural              | MIRANDA, Gabriel Alves  | 2022                      |
| Tiradentes no TripAdvisor - O que se fala sobre essa simpática cidade histórica?                   | ALVES, Antônio P. S. <i>et al</i>   | 2022                      |
| Utilizando Pistas Linguística para Detectar Conteúdo Enganoso em Português                         | RODRIGUES, Rodrigo F.;<br>FREITAS, Larissa A. de                                      | 2021                      |

FONTE: O autor (2023).

Os trabalhos resgatados são atuais e contemplam as mais variadas aplicações, não tendo seus escopos restringidos às redes sociais. Abordando alguns dos trabalhos resgatados, Marinho, Pereira e Santarem Segundo (2022) realizam uma análise de sentimentos da plataforma Skoob como uma ideia de propor um novo parâmetro de avaliação que auxilie na tomada de decisões para escolha de leitura dos usuários presentes nesta rede social voltada aos livros.

Já o trabalho “Processamento de linguagem natural para detecção de padrões no mercado de ações brasileiro” (Pereira, 2022) aborda a biblioteca LeIA como um dos possíveis métodos de lidar com o processamento de linguagem natural em dados extraídos de redes sociais, mas não restringe sua análise a biblioteca em si, citando outros métodos de análise de sentimentos para este meio. No entanto, acaba por

utilizar a biblioteca LeIA em comparação a outros métodos de análises mencionados no trabalho, devido ao fato de que ela é baseada na VADER, sendo mais adaptada para o contexto das redes sociais.

O trabalho de Malakin e Rezende (2022) traz uma breve análise de sentimentos voltada ao período da pandemia de COVID-19 e as repercussões na rede social do *Twitter*.

Foi verificado que nenhum dos trabalhos resgatados propõem a aplicação do modelo no contexto do *Youtube* e futebol.

### 3.4.2 Classificação com a biblioteca LeIA

A etapa da análise dos sentimentos, foi realizada com os métodos da biblioteca ou pacote LeIA que é um *fork*, ou seja, uma implementação nova baseada na original, do léxico e ferramenta para análise de sentimentos VADER, mas adaptado para textos em português, com suporte para *emojis* e foco na análise de sentimentos de textos expressos em mídias sociais (ALMEIDA, R. J. A., 2018). A FIGURA 13, a seguir, exemplifica uma análise de sentimento com frases pré-estabelecidas para uso na linguagem *Python*.

FIGURA 13 – EXEMPLO DE CÓDIGO DA BIBLIOTECA LEIA

```
from leia import SentimentIntensityAnalyzer

s = SentimentIntensityAnalyzer()

# Análise de texto simples
s.polarity_scores('Eu estou feliz')
#{'neg': 0.0, 'neu': 0.328, 'pos': 0.672, 'compound': 0.6249}

# Análise de texto com emoji :)
s.polarity_scores('Eu estou feliz :)')
#{'neg': 0.0, 'neu': 0.22, 'pos': 0.78, 'compound': 0.7964}

# Análise de texto com negação
s.polarity_scores('Eu não estou feliz')
#{'neg': 0.265, 'neu': 0.241, 'pos': 0.494, 'compound': 0.4404}
```

FONTE: Almeida, R.J.A (2018).

Por ser baseada no método implementado por Hutto e Gilbert (2014), sua execução retorna um dicionário *Python* com 4 pontuações de sentimentos:

- quão negativa (“*neg*”) é a sentença entre 0 e 1;
- quão neutra (“*neu*”) é a sentença entre 0 e 1;
- quão positiva (“*pos*”) é a sentença entre 0 e 1;
- uma pontuação composta (“*compound*”) entre -1 e 1.

De forma geral se usa a pontuação composta para representar o sentimento nas análises. A soma entre “*neg*”, “*neu*” e “*pos*” deve ser igual a 1 ou próxima a ele (em caso de float). O *compound* (composto) corresponde à soma da pontuação de valência de cada palavra no léxico e determina o grau do sentimento, e não um determinado valor geral como os itens anteriores fazem, representando uma média normalizada e ponderada do sentimento relacionado ao texto. Seu valor está entre -1 (sentimento negativo mais extremo) e +1 (sentimento positivo mais extremo). Neste contexto, usar o “*compound*” pode ser suficiente para determinar o sentimento subjacente de um texto, pois, em suma, tem-se o seguinte:

- Um sentimento positivo corresponde a um composto com valor maior ou igual a 0,05;
- Um sentimento negativo corresponde a um composto menor ou igual a -0,05;
- Um sentimento neutro corresponde a um composto com valor está entre -0,05 e 0,05.

Para a biblioteca estar disponível para importação e análise no ambiente do Jupyter Notebook, foi preciso baixar a pasta com os arquivos necessários fornecida por Almeida (2018) e disponibilizada na plataforma *Github*<sup>5</sup>. No âmbito da aplicação da biblioteca para a execução da análise, o código da FIGURA 14, ilustra a etapa de criação de uma nova coluna no *dataframe* dos comentários, de maneira que a partir de cada um dos textos processados retornem o valor correspondente ao “*compound*”.

---

<sup>5</sup> Link para download da biblioteca: <https://github.com/rafjaa/LeIA>

FIGURA 14 - CÓDIGO DA APLICAÇÃO DO LEIA

```

1 from leia import SentimentIntensityAnalyzer
2
3 s = SentimentIntensityAnalyzer()
4
5 comments_df['Compound'] = comments_df['Comment'].apply(lambda x: s.polarity_scores(x)['compound'])
6 pd.set_option('max_colwidth', None)
7 comments_df.head(50)

```

|   | Video    | Comment   | Clean_comment  | Compound |
|---|----------|---|--|----------|
| 0 | video_01 | ❤️❤️✅   | ❤️❤️✅  | 0.0000   |
| 1 | video_01 | Moreno e Pinho mudaram o time .   | moreno pinho mudaram time  | 0.0000   |
| 2 | video_01 | Jesus trindade titular caramba.   | jesus trindade titular caramba   | 0.0000   |
| 3 | video_01 | Verdade esse e o caminho  | verdade caminho  | 0.3182   |
| 4 | video_01 | Marcelino titular ❤️❤️  | marcelino titular ❤️❤️   | 0.0000   |
| 5 | video_01 | Já deu no meio do reporte gostei hahahahah técnico ponta firme tem que dar no meio msm o cara atrapalhando na hora da coletiva. | deu meio reporte gostei hahahahah técnico ponta firme dar meio msm cara atrapalhando hora coletiva | 0.5859   |
| 6 | video_01 | Já já vai ter padeiro de Portugal vindo pro Brasil dirigir time.  | vai ter padeiro portugal vindo pro brasil dirigir time   | 0.0000   |
| 7 | video_01 | Que enrabada no repórter, maluco sem noção demais   | enrabada repórter maluco noção demais  | 0.2023   |
| 8 | video_01 | Chato e arrogante.  | chato arrogante  | -0.6486  |
| 9 | video_01 | O time do aruco é muito bom! Bem aplicado em campo por isso que não foi fácil!  | time aruco bom! bem aplicado campo fácil!  | 0.7339   |

FONTE: O autor (2023).

O código criou uma coluna denominada “*Compound*”, que utilizou do texto da coluna “*Comment*” para aplicar o modelo do método “*SentimentIntensityAnalyser ()*” e atribui nela somente o valor da “*compound*”. Percebeu-se que ao executar o método, os textos que continham emojis não foram atribuídos um valor, por mais que dentro do código e dos arquivos da biblioteca possuíssem os dados para lidar com este tipo de comentário.

Em seguida, o valor do “*compound*” foi utilizado como parâmetro para a aplicação demonstrada na FIGURA 15, para ser criado uma coluna do *dataframe* contendo a polaridade do sentimento predominante naquele comentário, sendo negativo, positivo ou neutro.

FIGURA 15 - FUNÇÃO DE POLARIDADE DO COMENTÁRIO

```

1 #função de verificação de polaridade de acordo com as métricas do LEIA
2 def see_polarity(df):
3     if df['Compound'] >= 0.05:
4         polarity = "positive"
5
6     elif df['Compound'] <= -0.05:
7         polarity = "negative"
8
9     else:
10        polarity = "neutral"
11
12    return polarity

1 comments_df['Polarity'] = comments_df.apply(lambda df: see_polarity(df), axis=1)
2 comments_df.head(10)

```

|   | Video    | Comment   | Clean_comment  | Compound | Polarity |
|---|----------|---|--|----------|----------|
| 0 | video_01 | ♥♥✓   | ♥♥✓  | 0.0000   | neutral  |
| 1 | video_01 | Moreno e Pinho mudaram o time .   | moreno pinho mudaram time  | 0.0000   | neutral  |
| 2 | video_01 | Jesus trindade titular caramba.   | jesus trindade titular caramba   | 0.0000   | neutral  |
| 3 | video_01 | Verdade esse e o caminho  | verdade caminho  | 0.3182   | positive |
| 4 | video_01 | Marcelino titular ♥♥  | marcelino titular ♥♥   | 0.0000   | neutral  |
| 5 | video_01 | Já deu no meio do reporte gostei hahahahah técnico ponta firme tem que dar no meio msm o cara atrapalhando na hora da coletiva. | deu meio reporte gostei hahahahah técnico ponta firme dar meio msm cara atrapalhando hora coletiva | 0.5859   | positive |
| 6 | video_01 | Já já vai ter padeiro de Portugal vindo pro Brasil dirigir time.  | vai ter padeiro portugal vindo pro brasil dirigir time   | 0.0000   | neutral  |
| 7 | video_01 | Que enrabada no repórter, maluco sem noção demais   | enrabada repórter maluco noção demais  | 0.2023   | positive |
| 8 | video_01 | Chato e arrogante.  | chato arrogante  | -0.6486  | negative |
| 9 | video_01 | O time do aruco é muito bom! Bem aplicado em campo por isso que não foi fácil!  | time aruco bom! bem aplicado campo fácil!  | 0.7339   | positive |

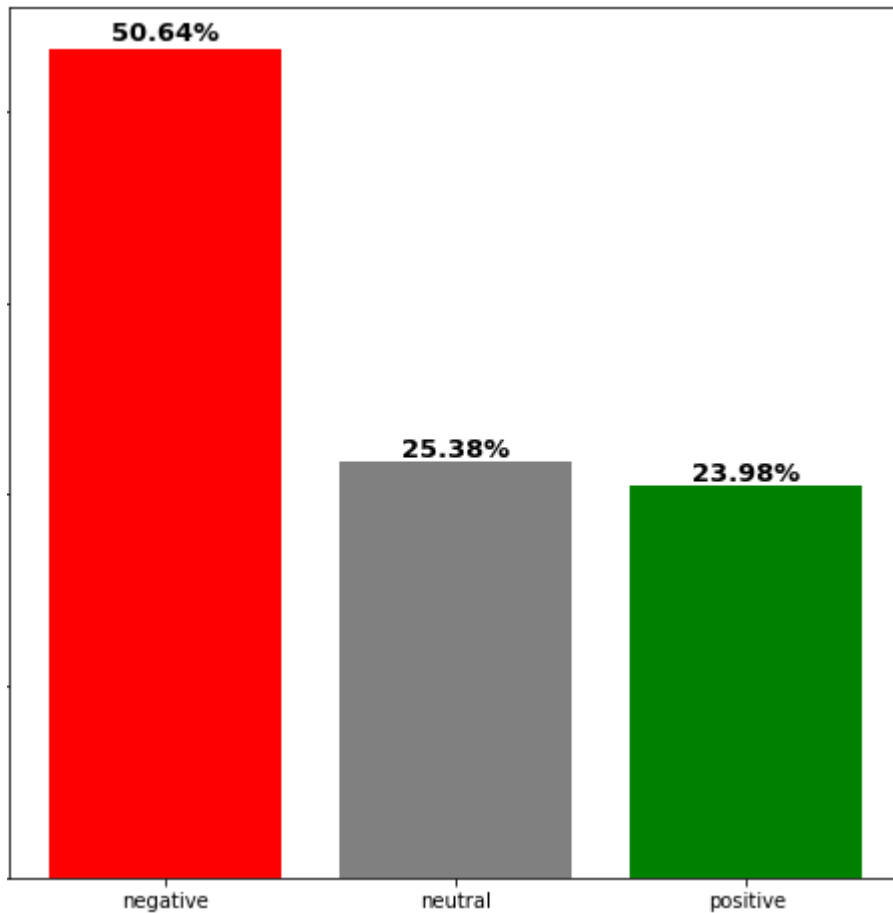
FONTE: O autor (2023).

A função criada segue o funcionamento da análise de sentimentos proposta por Almeida (2018), levando em consideração os intervalos do valor do “*compound*” para que possa ser classificado qualitativamente os comentários, sendo apresentado na coluna “*Polarity*”.

#### 4 APRESENTAÇÃO DOS RESULTADOS

Nesta secção são evidenciados os resultados atingidos com a execução do modelo léxico da biblioteca Lela nos comentários extraídos dos vídeos, a partir de gráficos com a biblioteca *Matplotlib* e *WordCloud*. A FIGURA 16 apresenta a porcentagem de comentários separados por cada uma das três polaridades do conjunto de dados utilizado.

FIGURA 16 - PORCENTAGEM DE CADA POLARIDADE



FONTE: O autor (2023).

A partir da visualização gerada acima, foi identificado que os comentários classificados com o sentimento negativo correspondem a mais da metade dos registros dos comentários coletados, demonstrando em um grande descontentamento por parte dos torcedores em relação a performance do clube e à condução das entrevistas coletivas por parte do técnico do time de maneira geral. Não houve muita divergência entre a proporção dos comentários neutros e positivos classificados pela biblioteca.

Além desta análise geral caracterizada por apresentar a partir de gráficos os sentimentos dos torcedores do Coritiba, foi realizada o mapeamento da frequência das palavras que mais apareceram e conseqüentemente contribuíram para a classificação de sentimentos negativos e positivos ao longo de todas as coletivas de imprensa, com o intuito de identificar os pontos positivos e negativos da performance do time e as razões para tais. Esta análise foi feita a partir de nuvem de palavras com as bibliotecas *Wordcloud*, tendo como fonte de dados as colunas pré-processadas,





clube nas quatro últimas partidas foi determinante para uma opinião negativa dos torcedores com relação a performance demonstrada nos jogos. Este fato faz necessário uma intervenção por parte do clube para identificar os fatores que levaram a este nível de performance que foi desaprovado pela torcida, para auxiliar na tomada de decisões do clube nos campeonatos posteriores da temporada.

## 5 CONSIDERAÇÕES FINAIS

Os técnicos de futebol representam uma figura de suma importância nos clubes de futebol brasileiros nos últimos anos, especialmente os que buscam obter êxito esportivo além do financeiro. Sua capacidade de definir uma partida a partir de esquemas táticos, escalações e contratações podem ser o diferencial nas equipes que almejam este objetivo. Levando em consideração o grande volume e diversificação de conteúdo postado pelos clubes de futebol brasileiros nas redes sociais, os clubes podem mensurar os sentimentos dos torcedores de maneira que verifiquem as opiniões deles ao longo do trabalho dos técnicos, para determinar se é válido continuar com o profissional atual ou deve trocar, seguindo como parâmetros o êxito esportivo e o sentimento da torcida.

O trabalho realizado pode servir como fonte de dados e inspiração no contexto da aplicabilidade da análise de sentimentos no *Youtube*, pois é uma rede social que está em constante crescimento e conta com uma grande participação do povo brasileiro, sendo a segunda rede social mais acessada.

O objetivo geral deste trabalho era identificar e analisar os sentimentos dos torcedores do Coritiba com relação a performance do técnico no comando do clube no Campeonato Paranaense de Futebol, a partir da coleta de dados em redes sociais como o *Youtube*. Para alcançar o objetivo proposto, foi utilizado API da rede social e bibliotecas que lidam com processamento de linguagem natural e análise de sentimentos, em especial o modelo léxico Léxico para inferência adaptada (LeIA). Por não necessitar de um pré-processamento dos dados para classifica-los utilizando a abordagem léxica proposta por Almeida (2018), o modelo auxilia na eficiência de lidar com os volumes de dados imediatamente coletados e o trabalho buscou contribuir com a aplicação deste modelo no contexto das redes sociais, desta vez no *Youtube*, que não tinha sido usado até então, a partir do levantamento feito.

Foi evidenciado que os sentimentos da torcida com relação ao técnico e aos resultados alcançados por este na campanha do clube no Campeonato Paranaense de Futebol foram majoritariamente negativos.

A partir da coleta dos dados e das aplicações dos métodos das bibliotecas disponíveis na linguagem *Python*, ficou evidente como a análise de sentimentos pode auxiliar clubes de futebol brasileiros como o Coritiba em analisar comentários de seus torcedores na rede social.

## 5.1 TRABALHOS FUTUROS

Existem pontos de melhoria que podem ser aprofundados na realização de trabalhos futuros, como a possibilidade de implementar um modelo de coleta dos dados automatizado baseado nos títulos dos vídeos publicados pelos clubes brasileiros, focando em vídeos de entrevistas coletivas ou contratações de maneira automatizada, deixando este processo mais eficiente se for considerar um grande volume de dados, dependendo da presença dos torcedores em cada clube.

Ademais, como o *Youtube* possui um sistema de curtidas para os comentários postados, seria interessante transformar esta métrica em algo relevante na etapa da análise dos dados utilizando dos dicionários léxicos, com o intuito de estabelecer um “peso” para comentários que tenham mais curtidas que outros, por exemplo.

## REFERÊNCIAS

- ALMEIDA, Rafael José de Alencar. **LeIA - Léxico para Inferência Adaptada**. 2018. Disponível em: <https://github.com/rafjaa/LeIA>. Acesso em: 20 jan. 2022.
- ALMEIDA, Rafael José de Alencar. **Suporte à análise interativa de discussões online combinando técnicas de mineração de dados**. 2020. 109 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de São João del-Rei, São João del-Rei, 2020. Disponível em: <http://ppgcc.ufsj.edu.br/wp-content/uploads/2021/06/Dissertacao-Rafael-Jose-de-Alencar-Almeida.pdf>. Acesso em: 13 fev. 2023.
- AL-MOSLMI *et al.* Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review. **IEEE Access**, n. 99, mar. 2017. [https://www.researchgate.net/publication/315799563\\_Approaches\\_to\\_Cross-Domain\\_Sentiment\\_Analysis\\_A\\_Systematic\\_Literature\\_Review](https://www.researchgate.net/publication/315799563_Approaches_to_Cross-Domain_Sentiment_Analysis_A_Systematic_Literature_Review). Disponível em: Acesso em: 1 jun. 2023.
- ALVES, Antônio P. S. *et al.* Tiradentes no TripAdvisor - O que se fala sobre essa simpática cidade histórica? In: Brazilian workshop on social network analysis and mining (BRASNAM), 11., 2022, Niterói. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 145-156. Disponível em: <https://sol.sbc.org.br/index.php/brasnam/article/view/20524>. Acesso em: 1 jun. 2023.
- AMAZON WEB SERVICES. **O que é uma API**. 2022. Disponível em: <https://aws.amazon.com/pt/what-is/api/#:~:text=API%20significa%20Application%20Programming%20Interface,de%20servi%C3%A7o%20entre%20duas%20aplica%C3%A7%C3%B5es>. Acesso em: 20 dez. 2022.
- AZEVEDO, C. O.; ALMEIDA, A. T. C. de; RAMALHO, H. M. de B. Rotatividade de treinadores e o desempenho das equipes de futebol no Brasil. **Economia Aplicada**, v. 25, n. 1, p. 5-32, 2021. Disponível em: <https://www.revistas.usp.br/ecoa/article/view/171487>. Acesso em: 20 jun. 2023.
- BENEVENUTO, Fabrício; RIBEIRO, Filipe; ARAÚJO, Matheus. Métodos para Análise de Sentimentos em Mídias Sociais. In: WebMedia2015 (minicurso). Disponível em: <https://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf>. Acesso em 1 jun. 2023.
- CARTER, Adam D.; BLOOM, Gordon A. Coaching knowledge and success: going beyond athletic experiences. **Journal of Sport Behavior**, New York, v. 32, n. 4, p. 419-437, 2009.
- CBF. **Técnico de futebol**: o que é preciso para virar um profissional no brasil? 2023. Disponível em: <https://www.cbf.com.br/cbfacademy/en-us/noticias/292-tecnico-de-futebol-o-que-e-preciso-para- virar-um-profissional-no-brasil>. Acesso em: 1 jun. 2023.

EL-KHAIR, Ibrahim Abu. Effects of stop words elimination for Arabic information retrieval: a comparative study. **International Journal of Computing & Information Sciences**, v. 4, n. 3, p. 119-133, 2006.

GLOBO ESPORTE. Campeonato Paranaense. 2023. Disponível em: <https://ge.globo.com/pr/futebol/campeonato-paranaense/>. Acesso em: 1 abr. 2023.

GOOGLE WORKSPACE. **Guia de início rápido do Python**. 2023. Disponível em: <https://developers.google.com/docs/api/quickstart/python>. Acesso em: 1 jun. 2023.

HUTTO, C.; GILBERT, E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. **Proceedings of the International AAI Conference on Web and Social Media**, v. 8, n. 1, p. 216-225, 2014. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>. Acesso em: 20 jan. 2023.

IBICT. **Stop words e sinônimos**. 2015. Disponível em: [http://wiki.ibict.br/index.php/Stop\\_words\\_e\\_sinonimos#:~:text=As%20Stop%20Words%20s%C3%A3o%20palavras,torna%20a%20busca%20mais%20eficaz](http://wiki.ibict.br/index.php/Stop_words_e_sinonimos#:~:text=As%20Stop%20Words%20s%C3%A3o%20palavras,torna%20a%20busca%20mais%20eficaz). Acesso em: 1 jan. 2023.

IGLESIAS, C. A.; MORENO, A. Sentiment Analysis for Social Media. **Applied Sciences**, v. 9, n. 23, nov. 2019. Disponível em: <https://www.mdpi.com/2076-3417/9/23/5037#B1-applsci-09-05037>. Acesso em: 1 jan. 2023.

IVO, Diego. **Ranking Social Media no Brasil: as redes sociais mais usadas no país**. 2023. Disponível em: <https://www.conversion.com.br/blog/redes-sociais-mais-usadas-no-brasil/>. Acesso em: 1 jun. 2023.

JUNQUEIRA, Kássio T. C.; FERNANDES, Animta Maria da Rocha. Análise de Sentimento em Redes Sociais no Idioma Português com Base em Mensagens do Twitter. In: Computer on the Beach. 2018. Vale do Itajaí. **Anais...** Itajaí, Centro de Ciências Tecnológicas da Terra e do Mar. 2018. p. 681-690.

JUPYTER NOTEBOOK. **Jupyter.org**. 2023. Disponível em: <https://jupyter.org/>. Acesso em: 20 jan. 2023.

LANCE. **Ranking digital do futebol brasileiro: veja clubes com mais seguidores nas redes sociais**. 2023. Disponível em: <https://www.lance.com.br/lancebiz/ranking-digital-do-futebol-brasileiro-veja-clubes-com-mais-seguidores-em-2023.html>. Acesso em: 1 jun. 2023.

LIU, Bing. **Sentiment analysis and opinion mining**. Chicago: Morgan & Claypool Publishers, 2012. Disponível em: <https://www.cs.uic.edu/~liub/FBS/liub-SA-and-OM-book.pdf>. Acesso em: 1 jan. 2023.

MALAKIN, Lucas Alexandre; REZENDE, Solange Oliveira. Análise da subjetividade em língua portuguesa no Twitter relacionada a pandemia da COVID-19. **Mecai 2022**. Disponível em:

[https://cemeai.icmc.usp.br/WMECAI/wp-content/uploads/2022/09/WMECAI\\_2-Lucas-Alexandre-Malakin.pdf](https://cemeai.icmc.usp.br/WMECAI/wp-content/uploads/2022/09/WMECAI_2-Lucas-Alexandre-Malakin.pdf). Acesso em: 1 jun. 2023.

MIRANDA, Gabriel Alves. Detecção de Bots no Twitter Através de Técnicas de Processamento de Linguagem Natural. 2022. 48 f. Trabalho de Conclusão de Curso (Graduação de Estatística) - Instituto de Matemática e Estatística, Universidade Federal Fluminense, Niterói, 2022. Disponível em: <https://app.uff.br/riuff/handle/1/28633>. Acesso em: 1 jun. 2023.

PAGOTTO, D. *et al.* Sentimentos Negativos de Empreendedores e a Covid-19: Uma Análise de Tweets. **Revista Inteligência Competitiva**, v. 12, n. 1, p. 1-24, 2022. Disponível em: <https://iberoamericanic.org/rev/article/view/414>. Acesso em: 1 jan. 2023.

PEREIRA, Francielle Vasconcellos. Processamento de linguagem natural para detecção de padrões no mercado de ações brasileiro. (Bacharel em Ciência da computação) – Setor de Ciências Biológicas, Universidade Federal de Santa Maria, Curitiba (PR), 2022. Disponível em: [https://repositorio.ufsm.br/bitstream/handle/1/24612/TG500\\_Francielle%20Vasconcelos.pdf?sequence=1](https://repositorio.ufsm.br/bitstream/handle/1/24612/TG500_Francielle%20Vasconcelos.pdf?sequence=1). Acesso em: 1 jun. 2023.

RAJA, Guzmán Isidoro. Measuring Efficiency and Sustainable Growth in Spanish Football Teams. **European Sport Management Quarterly**, v. 6, n. 3, set. 2006. Disponível em: [https://www.researchgate.net/publication/49292066\\_Measuring\\_Efficiency\\_and\\_Sustainable\\_Growth\\_in\\_Spanish\\_Football\\_Teams](https://www.researchgate.net/publication/49292066_Measuring_Efficiency_and_Sustainable_Growth_in_Spanish_Football_Teams). Acesso em: 1 jun. 2023.

RED HAT, **O que é API**. 2023. Disponível em: <https://www.redhat.com/pt-br/topics/api/what-are-application-programming-interfaces>. Acesso em: 1 jan. 2023.

RODRIGUES, Rodrigo F.; FREITAS, Larissa A. de. Utilizando Pistas Linguística para Detectar Conteúdo Enganoso em Português. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL), 13., 2021, Evento Online. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 447-451. Disponível em: <https://sol.sbc.org.br/index.php/stil/article/view/17827>. Acesso em: 1 jun. 2023.

SANTOS, Alexandre Crispin; RODRIGUES, José de Jesus Fernandes. Análise da Instrução do Treinador de Futebol. Comparação entre a Prelecção de Preparação e a Competição. **Fitness & Performance Journal**, Rio de Janeiro, v. 7, n. 2, p. 112-122, 2008.

SOUZA, Miguel. **YouTube**. 2023. Disponível em: <https://brasilecola.uol.com.br/informatica/youtube.htm>. Acesso em: 20 abr. 2023.

TABOADA, Maite. *et al.* Lexicon-based methods for sentiment analysis. **Computacional Linguist**, v. 37, n. 2, p. 267-307, 2011. Disponível em: <https://direct.mit.edu/coli/article/37/2/267/2105/Lexicon-Based-Methods-for-Sentiment-Analysis>. Acesso em: 1 abr. 2023.

VOLPATO, Bruno. **As redes sociais mais usadas no Brasil e no mundo em 2023, com insights, ferramentas e materiais**. 2023. Disponível em:

<https://resultadosdigitais.com.br/marketing/redes-sociais-mais-usadas-no-brasil/>

Acesso em: 1 jun. 2023.

YANG, D.G; YU, G. A method of feature selection and sentiment similarity for chinese micro-blogs. **Journal of Information Science**, v. 39, n. 4, p. 429-441, 2013.

YOUNG, T. *et al.* Recent trends in deep learning based natural language processing. **Journal of Ambient Intelligence and Humanized Computing**, v. 9, n. 4, p. 170-191, 2018.

## APÊNDICE 1 – LISTA DOS VÍDEOS SELECIONADOS

| Nome do vídeo   | Partida (Resultado)            | Data do jogo | Link do vídeo   |
|---|--------------------------------|--------------|---|
| CORTIBA 1 X 0 ARUKO   Coletiva António Oliveira                             | Coritiba x Aruko (1x0)         | 15/01/2023   | <a href="https://www.youtube.com/watch?v=en053zxDYM">https://www.youtube.com/watch?v=en053zxDYM</a>   |
| COLETIVA ANTÓNIO OLIVEIRA   Foz do Iguaçu 0 x 3 Coritiba                    | Foz do Iguaçu x Coritiba (0x3) | 18/01/2023   | <a href="https://www.youtube.com/watch?v=irV4nbA_Kw0">https://www.youtube.com/watch?v=irV4nbA_Kw0</a> |
| COLETIVA ANTÓNIO OLIVEIRA   Coritiba 1 x 0 Rio Branco                       | Coritiba x Rio Branco (1x0)    | 22/01/2023   | <a href="https://www.youtube.com/watch?v=EvphSYjfBuQ">https://www.youtube.com/watch?v=EvphSYjfBuQ</a> |
| CÂMERA COXA   Coletiva António Oliveira                                     | Coritiba x Maringá (1x0)       | 25/01/2023   | <a href="https://www.youtube.com/watch?v=i_kvmvjqj3E">https://www.youtube.com/watch?v=i_kvmvjqj3E</a> |
| COLETIVA ANTÓNIO OLIVEIRA   Coritiba 2 x 2 Azuriz                           | Coritiba x Azuriz (2x2)        | 29/01/2023   | <a href="https://www.youtube.com/watch?v=V7LJKp3Jsgw">https://www.youtube.com/watch?v=V7LJKp3Jsgw</a> |
| COLETIVA ANTÓNIO OLIVEIRA   Coritiba 3 x 1 Cianorte                         | Coritiba x Cianorte (3x1)      | 02/02/2023   | <a href="https://www.youtube.com/watch?v=pk3TwgYE67c">https://www.youtube.com/watch?v=pk3TwgYE67c</a> |
| COLETIVA ANTÓNIO OLIVEIRA   Athletico 1 x 1 Coritiba                        | Athletico x Coritiba (1x1)     | 05/02/2023   | <a href="https://www.youtube.com/watch?v=VrKD6AigPBq">https://www.youtube.com/watch?v=VrKD6AigPBq</a> |
| COLETIVA ANTÓNIO OLIVEIRA   Coritiba 0 x 0 São Joseense                     | Coritiba x São Joseense (0x0)  | 12/02/2023   | <a href="https://www.youtube.com/watch?v=IK8EYomU1sU">https://www.youtube.com/watch?v=IK8EYomU1sU</a> |
| COLETIVA ANTÓNIO OLIVEIRA   Coritiba 1 x 0 Londrina   Campeonato Paranaense | Coritiba x Londrina (1x0)      | 15/02/2023   | <a href="https://www.youtube.com/watch?v=eDJMGI0qtWA">https://www.youtube.com/watch?v=eDJMGI0qtWA</a> |
| COLETIVA ANTÓNIO OLIVEIRA   FC Cascavel 0 x 0 Coritiba                      | Cascavel x Coritiba (0x0)      | 19/02/2023   | <a href="https://www.youtube.com/watch?v=8vKVGdww-DM">https://www.youtube.com/watch?v=8vKVGdww-DM</a> |

|   |                               |   |            |   |
|---|-------------------------------|---|------------|---|
| COLETIVA<br>ANTÓNIO<br>OLIVEIRA  <br>Operário 3 x<br>0 Coritiba  <br>Campeonato<br>Paranaense | Operário<br>Coritiba<br>(3x0) | x | 26/02/2023 | <a href="https://www.youtube.com/watch?v=Eaxvs9IOJGU">https://www.youtube.com/watch?v=Eaxvs9IOJGU</a> |
| FC<br>CASCAVEL<br>3 X 1<br>CORITIBA  <br>Coletiva<br>Antônio<br>Oliveira                      | Coritiba<br>Cascavel<br>(1x3) | x | 05/03/2023 | <a href="https://www.youtube.com/watch?v=6SVNwY9QRJY">https://www.youtube.com/watch?v=6SVNwY9QRJY</a> |
| COLETIVA<br>ANTÓNIO<br>OLIVEIRA  <br>Coritiba 0 x 0<br>FC Cascavel                            | Coritiba<br>Cascavel<br>(0x0) | x | 11/03/2023 | <a href="https://www.youtube.com/watch?v=CnYehk4GSvA">https://www.youtube.com/watch?v=CnYehk4GSvA</a> |