INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Liberdade de Expressão ou Racismo/Discurso de Ódio? Um Estudo de Caso para o futebol europeu

Guilherme Afonso de Sousa Vale Mendonça

Mestrado em Ciência de Dados

Orientadores:

Professor Doutor Pedro Sousa Almeida, Investigador Integrado
CRIA
Pólo ISCTE-IUL

Professora Doutora Diana Elisabeta Aldea Mendes, Professora Associada
ISCTE-IUL

Outubro, 2024

Departamento de Métodos Quantitativos para Gestão e Economia

Departamento de Ciências e Tecnologia da Informação

Liberdade de Expressão ou Racismo/Discurso de Ódio? Um Estudo de Caso para o futebol europeu

Guilherme Afonso de Sousa Vale Mendonça

Mestrado em Ciência de Dados

Orientadores:

Professor Doutor Pedro Sousa Almeida, Investigador Integrado
CRIA
Pólo ISCTE-IUL

Professora Doutora Diana Elisabeta Aldea Mendes, Professora Associada
ISCTE-IUL

Outubro, 2024

“You miss 100% of the shots you don’t take”
Wayne Gretzky

Agradecimentos

Queria agradecer, em primeiro lugar, aos meus orientadores, o Professor Pedro Almeida e a Professora Diana Mendes. Foram ambos imprescindíveis para a realização desta dissertação, desde terem a paciência para me responderem aos emails, desde reunirem vezes sem conta, incluindo sábados, para me darem as orientações necessárias, desde terem sabido colaborar entre eles para que fosse possível realizar um projeto que abrange duas áreas e desde terem tido a paciência de me acompanharem durante este longo percurso. Sem vocês não seria possível.

Queria agradecer a toda a minha família, que sempre se mostrou disponível para me ajudar e foi quem me deu a motivação para nunca desistir, mesmo quando essa parecia a única opção possível.

A todos os meus amigos, que, para além de também me motivarem, foram ainda essenciais numa parte igualmente importante: a descontração durante este longo percurso.

Queria agradecer, por último, ao Sporting Clube de Portugal. Sem a existência deste clube, a minha paixão pelo futebol seria diferente e, quem sabe, nunca teria tido inspiração para trabalhar neste tema.

Resumo

Nas redes sociais as pessoas têm tendência a partilhar opiniões e interagir sem um grande nível de moderação por parte das plataformas. Nestes espaços, é comum encontrar comunidades de adeptos de futebol envolvidas em discussões polémicas, que incluem muitas vezes linguagem ofensiva e discriminatória.

Este estudo investiga a natureza da linguagem e dos sentimentos num canal do *Telegram* dedicado à partilha de conteúdo relacionado com adeptos de futebol, dos movimentos *Ultra* e *Hooligan*, centrando-se na deteção de Discurso de Ódio nos comentários dos seus utilizadores.

Foi desenvolvida uma Análise de Sentimentos aos comentários com o modelo *RoBERTa* e aplicou-se o modelo *Toxic BERT* para a deteção e classificação de Discurso de Ódio.

Os resultados indicaram uma clara presença de Discurso de Ódio nos comentários do canal de *Telegram* analisado, com 34% dos comentários sendo classificados como tóxicos, apresentado padrões específicos que sugerem o uso frequente de Discurso de Ódio baseado na discriminação.

Estes resultados sublinham a necessidade de moderar as redes sociais de forma a reduzir a partilha de conteúdo discriminatório e realçam ainda a importância que os grupos *Ultra* e *Hooligan* têm na cultura dos adeptos e da sociedade em geral.

Palavras-chave: Discurso de Ódio; Racismo Futebol; Redes Sociais; Machine Learning; Text Mining;

Abstract

Individuals on social networks often engage and exchange ideas without much oversight from the platforms. Communities of football fans frequently engage in contentious conversations in these settings, frequently using derogatory and discriminatory language.

This study examines the language and sentiment patterns of a Telegram channel that shares football-related content from the Ultra and Hooligan movements, with a particular focus on identifying hate speech in user comments.

A Sentiment Analysis was carried out on the comments using the RoBERTa model and the Toxic BERT model was applied to detect and classify Hate Speech.

The findings show that hate speech was clearly present in the Telegram channel's comments, with 34% of the comments being categorized as toxic. These comments show patterns that suggest the frequent use of hate speech based on discrimination.

These results emphasise the need to moderate social media in order to reduce the sharing of discriminatory content and also highlight the importance that the Ultra and Hooligan groups have in fan culture and society in general.

Keywords: Hate Speech; Racism; Football; Social Networks; Machine Learning; Text Mining

Índice

Índice de Figuras

Índice de Tabelas

Glossário

HS – Hate Speech

BERT – Bidirectional Encoder Representations from Transformers

RoBERTa - Robustly Optimized BERT Approach

NLP – Natural Language Processing

AI – Artificial Intelligence

1. Introdução

O futebol é mais do que um desporto para milhões de pessoas em todo o mundo – é um poderoso fenómeno social e cultural. No entanto, a par da paixão e do entusiasmo que o futebol gera, existe também um lado mais sombrio: o discurso de ódio e o comportamento abusivo entre os adeptos, sobretudo nas redes sociais. As rivalidades entre clubes transformam-se frequentemente em ataques verbais, com os adeptos a utilizarem as plataformas das redes sociais, como o Twitter, o Facebook e o Instagram, para atacar as equipas adversárias e os seus adeptos (e.g., Cleland, 2014; Miranda et al., 2024; Pookpanich & Siriborvornratanakul, 2024). Por vezes, estas interações online acabam por se transformar em violência no mundo real, criando graves problemas sociais.

No domínio das ciências sociais, o discurso de ódio nas redes sociais é entendido como uma forma de comunicação que visa indivíduos ou grupos com base na sua identidade, com a intenção ou o efeito de os rebaixar, desumanizar ou incitar ao mal (Nockleby, 1994; Nobata et al., 2016). As dinâmicas únicas das redes sociais - tais como o anonimato, a viralidade e a criação de câmaras de eco - exacerbam a propagação e o impacto do discurso de ódio (Luo et al., 2021; Törnberg, 2018), tornando-o uma questão crítica tanto para a investigação académica como para o desenvolvimento de políticas. A definição de discurso de ódio nestas plataformas continua a ser um desafio, particularmente à luz das preocupações com a liberdade de expressão e a necessidade de prevenir danos.

No caso do futebol, o discurso de ódio, em particular nas redes sociais, pode manifestar-se de várias formas. Em primeiro lugar, por uma linguagem racista e xenófoba, sendo o futebol frequentemente marcado pelo racismo, com jogadores e adeptos de diferentes origens a serem alvo de abusos com base na sua raça ou etnia (Kassimeris, 2008). Em segundo lugar por homofobia, sendo os insultos homofóbicos igualmente comuns na comunidade futebolística, tanto contra jogadores como contra adeptos rivais (Magrath, 2017). Por último, é praticada uma linguagem violenta ou abusiva em geral, com muitos adeptos de futebol a utilizarem uma linguagem agressiva para insultar ou rebaixar os seus rivais, incitando por vezes à violência.

O anonimato e o alcance global das plataformas de redes sociais facilitam a prática de discursos de ódio sem receio de consequências, permitindo a proliferação de discursos de ódio online (Malta, 2022).

Neste contexto, o *Text Mining* tornou-se uma ferramenta essencial para os cientistas sociais analisarem e abordarem o problema do discurso de ódio nas redes sociais praticado por adeptos de futebol. Ao utilizar técnicas avançadas de extração de texto, os investigadores podem monitorizar as conversas online, detetar o discurso de ódio e compreender os fatores que conduzem a comportamentos hostis (Jahan & Oussalah, 2023; Rini et al., 2020)., podendo através disto monitorizar o conteúdo que é partilhado nas redes para promover ambientes online mais respeitadores. Esta riqueza de dados, combinada com os avanços no processamento de linguagem natural (NLP), faz da extração de texto dos meios de comunicação social uma ferramenta inestimável para que as entidades, públicas ou privadas, compreendam fenómenos sociais e consigam atuar de forma fundamentada.

O problema de investigação desta dissertação situa-se na interseção entre *Data Science* e Sociologia, com o objetivo de investigar a presença de racismo e discurso de ódio no conteúdo partilhado por comunidades de adeptos de futebol em plataformas de redes sociais, particularmente no Telegram. Para isso, serão aplicadas técnicas de Machine Learning e Inteligência Artificial, possibilitando uma análise profunda e uma reflexão sobre um problema profundamente enraizado na nossa sociedade.

O objetivo deste projeto é investigar a presença de racismo e do discurso de ódio no conteúdo partilhado por comunidades de adeptos de futebol nas plataformas de redes sociais, em particular no *Telegram*.

O objetivo inicial do trabalho consistia numa análise dos comentários na rede social *Twitter*, mas, devido às restrições implementadas às técnicas de *Web Scraping* aquando da compra e transformação do *Twitter* em X, foi necessário alterar o objeto de estudo para que o objetivo do trabalho fosse atingido. Desta forma, foi escolhido um canal de *Telegram*, o *GruppaOF*, onde diariamente são postados, pelos seus administradores, conteúdos sobre acontecimentos do mundo dos adeptos do futebol - momentos dos grupos *Ultras* ou *Hooligans* no caminho para o estádio, dentro do estádio, em celebrações, em situações de confrontos físicos, entre vários outros – e onde é possível utilizadores “normais” fazerem comentários dentro de cada publicação.

Através da recolha destes comentários, o estudo pretende analisar os mesmos e detetar a presença de conteúdo que é considerado Racismo/Discurso de Ódio e, com isso, obter conclusões gerais sobre a presença de ódio racial em redes sociais, no mundo dos adeptos de futebol. Dentro deste canal, a esmagadora maioria de utilizadores e intervenientes são adeptos que se identificam ou com o movimento *Ultra* ou com o movimento *Hooligan*, pelo que será fornecido algum contexto sobre as suas origens e o seu impacto na politização do mundo dos

adeptos de futebol, já que, sendo movimentos de massas, as suas manifestações têm um impacto significativo na forma como os adeptos são percebidos.

Será igualmente fornecido um background sobre os temas centrais - Racismo/Discurso de Ódio e *Data Science* - incluindo as técnicas de *Machine Learning*, *Web Scraping* e *Text Mining*, que serão aplicadas neste trabalho.

1.1 Questões de Investigação

- Poderá uma análise de sentimentos a comentários no *Telegram* ser fiável para analisar a forma como as pessoas interagem nas redes sociais?
- Que modelo de *Machine Learning* melhor se aplica a uma base de dados construída a partir de comentários feitos numa rede social, com linguagem de adeptos de futebol?
- Quão presente é o Racismo/Discurso de Ódio no mundo do futebol, nas redes sociais?
- Que conclusões podemos tirar acerca dos movimentos *Ultra* e *Hooligan* na propagação do Racismo/Discurso de Ódio?
- Que papel desempenha o anonimato na prevalência do discurso de ódio entre os adeptos de futebol no *Telegram*?

1.2 Objetivos da Investigação

- Recolher dados de um canal de *Telegram* de interação entre adeptos de futebol para análise;
- Realizar a Análise de Sentimentos de comentários no canal de *Telegram*;
- Detetar a presença ou não de Racismo/Discurso de ódio nos grupos de adeptos de futebol no *Telegram* através de técnicas de *Machine Learning* e *Text Mining*;
- Analisar os limites entre Liberdade de Expressão e Racismo/Discurso de Ódio com base nas interações dos adeptos de futebol no *Telegram*;
- Avaliar o papel dos grupos *Ultras* e *Hooligan* na disseminação do Racismo/Discurso de Ódio.

1.3 Estrutura da dissertação

A dissertação está estruturada em 6 capítulos. O primeiro e presente capítulo, introduz o tema, define o problema, as questões e o objetivo da investigação. O segundo capítulo apresenta

uma revisão de literatura, centrada na análise de artigos recolhidos (principalmente científicos, mas não só) sobre os temas Racismo, Futebol, Discurso de Ódio e *Text Mining*/*Machine Learning* aplicado para o contexto da investigação. O capítulo seguinte contextualiza e reflete sobre a presença de racismo e discurso de ódio no futebol e sua relação com os movimentos *Ultra* e *Hooligan*. O quarto capítulo descreve as metodologias de *Data Science* empregadas na dissertação, detalhando as técnicas específicas utilizadas para investigar o objeto de análise. O capítulo cinco apresenta e discute os resultados da análise empírica, oferecendo uma interpretação crítica dos resultados obtidos. O último capítulo sintetiza as principais conclusões e recomenda direções para pesquisas futuras.

2. Revisão de literatura

2.1 Introdução

A motivação para trabalhar este tema surgiu com a vontade de trazer um pequeno contributo para um problema que se tornou uma preocupação crescente a nível mundial, nomeadamente a prevalência de racismo no futebol.

A dissertação tem como objetivo aplicar técnicas de *Text Mining* a comentários dentro de um canal da rede social *Telegram*, dentro da qual adeptos de futebol interagem entre si. O principal objetivo é perceber se o racismo ainda está muito presente nos conteúdos relacionados com o futebol nas redes sociais e avaliar a diferença entre liberdade de expressão e discurso de ódio/racismo dentro deste tópico.

O racismo no futebol e nas redes sociais está intimamente relacionado com a forma como as plataformas digitais são frequentemente utilizadas para difundir mensagens e conteúdos racistas. As redes sociais oferecem uma plataforma para as pessoas expressarem as suas opiniões e ideias, mas também permitem a propagação de discursos de ódio e discriminação (Matamoros-Fernández & Farkas, 2021). O racismo no futebol é muitas vezes amplificado nestes espaços, onde os comentários racistas podem ser partilhados e vistos por um grande público, e as redes sociais podem também organizar comportamentos racistas, como cânticos racistas coordenados em jogos de futebol. Por outro lado, as redes sociais são também uma ferramenta poderosa para aumentar a sensibilização para o racismo no futebol e mobilizar as pessoas para agirem contra ele. Muitos clubes de futebol, jogadores e organizações antirracismo utilizam as redes sociais para se manifestarem contra o racismo no futebol e para promoverem a igualdade e a diversidade (Lynch (2009).

Esta revisão tem como objetivo analisar a literatura recolhida sobre os temas Racismo, Futebol e/ou Exploração de textos/Aprendizagem automática.

2.2 Análise dos artigos

As redes sociais tornaram-se uma arena importante para o discurso público, mas também servem de terreno fértil para o discurso de ódio, o racismo e a xenofobia. A crescente prevalência de conteúdos de ódio online tem motivado os investigadores a recorrerem à inteligência artificial (AI) e a técnicas de Text Mining para detetar e atenuar estes problemas em tempo real (referência).

O Text Mining, uma ferramenta essencial para extrair padrões e estruturas de grandes volumes de texto, desempenha um papel fundamental na identificação do discurso de ódio e do racismo. Srato, Goularte e Fileto (2020) propõem a utilização de Short Semantic Patterns para reconhecer o discurso de ódio, que consiste na deteção de padrões linguísticos que são frequentemente específicos da linguagem discriminatória. Esta abordagem vai além das palavras-chave, identificando as estruturas linguísticas indicativas de racismo e outras formas de ódio, como insultos ambíguos ou subtis. Al-Hassan e Al-Dossari (2019) alargam esta abordagem, garantindo que as técnicas de Text Mining captam o contexto com precisão em diferentes línguas e culturas.

Além dos Semantic Patterns, a Topic Modelling (identificar Clusters ou grupos de palavras semelhantes num texto) é outra técnica relevante de Text Mining para detetar discursos racistas. Lee e Jang (2021) utilizam Topic Modelling para analisar os discursos em torno do movimento #StopAsianHate no Twitter, tendo descoberto que este método pode identificar temas concretos e narrativas predominantes em grandes conjuntos de dados. A Topic Modelling revela-se particularmente útil para acompanhar a evolução do discurso racista ao longo do tempo, permitindo monitorizar as mudanças no sentimento público e as novas tendências.

Os avanços recentes em Deep Learning conduziram a melhorias significativas na precisão e sofisticação dos sistemas de deteção de discursos de ódio. As redes neuronais, como as Convolutional Neural Networks (CNN), as Long Short-Term Memory (LSTM) e as BERT (Bidirectional Encoder Representations from Transformers), demonstraram capacidades notáveis no processamento de grandes quantidades de dados de texto. Benítez-Andrades et al. (2022) compararam CNN, LSTM e BERT para detetar racismo e xenofobia no Twitter. As suas conclusões indicam que o BERT, devido a ter sido pré treinado em grandes volumes de texto, é excelente na captura de relações contextuais, o que o torna particularmente eficaz na deteção de discursos de ódio ambíguos.

Duwairi et al. (2021) destacam a importância de utilizar estruturas de Deep Learning que sejam sensíveis às complexidades da linguagem na deteção de discursos de ódio em tweets árabes. O seu modelo, que incorpora redes LSTM bidirecionais, demonstra a eficácia do Deep Learning na compreensão das dependências sequenciais dentro do texto, captando o contexto que os métodos tradicionais muitas vezes não conseguem captar. Esta abordagem é particularmente útil para detetar discursos de ódio não explícitos, em que o contexto e as pistas subtis desempenham um papel significativo na transmissão de mensagens discriminatórias.

Um dos principais desafios na deteção de discursos de ódio é garantir que os modelos de AI são eficazes em todas as línguas e contextos culturais. As redes sociais são inerentemente

globais e as expressões de discurso de ódio variam muito entre culturas. Al-Hassan e Al-Dossari (2019) destacam as limitações dos modelos de AI existentes no tratamento de dados multilingues, defendendo o desenvolvimento de algoritmos que englobem a diversidade cultural e linguística. Do mesmo modo, Duwairi et al. (2021) demonstram as complexidades da deteção do discurso de ódio em árabe, onde os métodos de deteção padrão desenvolvidos para o inglês muitas vezes não conseguem captar o contexto linguístico. Os modelos, como o mBERT (Multilingual BERT), oferecem soluções eficazes, ao permitirem a generalização interlinguística, o contexto de cada língua.

Benítez-Andrades et al. (2022) ilustram ainda como os modelos ajustados com dados linguísticos diversos têm um melhor desempenho em todas as línguas, sugerindo que, ao se pré treinar em conjuntos de dados multilingues, a precisão dos modelos na deteção em textos não ingleses aumenta. No entanto, continua a existir uma lacuna significativa na criação de modelos culturalmente sensíveis que evitem a generalização excessiva das normas linguísticas ocidentais e detetem com precisão o discurso de ódio em vários contextos culturais.

A implementação de sistemas de deteção de discursos de ódio baseados em AI levanta desafios éticos e práticos críticos. Os falsos positivos – em que o conteúdo não odioso é erradamente classificado como discurso de ódio – podem levar à censura e potencialmente suprimir o discurso legítimo. Field et al. (2021) discutem os desafios éticos associados ao preconceito racial em NLP, observando que os conjuntos de dados de treino podem reforçar involuntariamente os preconceitos existentes, levando a resultados discriminatórios em modelos de deteção de Discurso de Ódio. Os autores realçam a importância da transparência nos dados de treino e nos parâmetros do modelo usados, defendendo avaliações rigorosas para identificar e mitigar possíveis vieses.

Além disso, a dependência de dados das redes sociais suscita preocupações em matéria de privacidade, uma vez que estes sistemas exigem frequentemente conjuntos de dados extensos dos utilizadores para obterem precisão. Waitzman (2020) apelou a uma supervisão regulamentar, de forma a garantir que os dados dos utilizadores que são recolhidos para a deteção de discursos de ódio são tratados de forma ética e que os direitos de privacidade são respeitados.

A deteção de discursos de ódio através de AI tem aplicações práticas para as plataformas de redes sociais e para os decisores políticos. Os moderadores das redes sociais dependem cada vez mais de sistemas automatizados para pré-filtrar e sinalizar conteúdos potencialmente nocivos antes de estes chegarem a grandes audiências. Balakrishnan, Ng e Arabnia (2022) descrevem uma abordagem semi-supervisionada para identificar a discriminação racial nas

redes sociais, que combina a deteção automática com a supervisão humana para reduzir os erros e garantir uma filtragem contextualmente adequada. Este modelo híbrido sublinha a importância do envolvimento humano nos processos de moderação orientados para a AI, especialmente nos casos em que é necessária uma compreensão do contexto.

Para os decisores políticos, o desenvolvimento de sistemas fiáveis de deteção de discurso de ódio ajuda na adoção de estratégias na de moderação do discurso de ódio online, ao mesmo tempo que se respeita os direitos de liberdade de expressão. Kang e Yang (2022) fornecem uma análise de enquadramento computacional do racismo e da xenofobia nos discursos sobre a construção do muro na fronteira dos EUA com o México, demonstrando como a AI pode apoiar a investigação direcionada para o desenho de políticas, analisando grandes conjuntos de dados para descobrir o sentimento público e narrativas potencialmente prejudiciais. À medida que as tecnologias de AI continuam a evoluir, as estruturas políticas realçam que a transparência e a responsabilidade implementação de algoritmos de deteção de discurso de ódio serão cruciais para manter a confiança do público.

Em suma, a integração da AI, de Machine Learning e Text Mining fez avançar significativamente a deteção do discurso de ódio e do racismo online. A extração de padrões linguísticos, os modelos de Deep Learning, como o BERT, e as técnicas de Topic Modelling oferecem ferramentas robustas para identificar Discurso de Ódio em situações complexas e dependentes do contexto. No entanto, continuam a existir desafios na adaptação destes modelos a diversos contextos culturais e linguísticos. Além disso, as considerações éticas em torno da privacidade, da transparência e do preconceito algorítmico sublinham a necessidade de uma utilização cuidadosa da AI na deteção de Discurso de Ódio.

À medida que os sistemas orientados para a AI se tornam cada vez mais centrais na moderação de conteúdos digitais, a investigação futura deve dar prioridade ao desenvolvimento de modelos culturalmente sensíveis, métricas de avaliação robustas e conjuntos de dados equilibrados. Além disso, a colaboração interdisciplinar entre cientistas sociais, informáticos e decisores políticos será essencial para criar sistemas eficazes de deteção de discursos de ódio que promovam ambientes online seguros e inclusivos, respeitando simultaneamente os direitos individuais.

2.3 Metodologia

A. *Queries* no *Scopus*

A fim de recolher a literatura adequada, foi aplicado um vasto número de consultas à base de dados *Scopus*, que se reportam de seguida. Nesta pesquisa, surgiram inúmeros obstáculos devido ao nível de especificidade do tema que será analisado na dissertação, que também serão aqui abordadas.

No início deste projeto de investigação, como primeiro passo, foram aplicadas as seguintes consultas:

- "data science" AND racism AND football
- "text mining" AND racism AND football
- statistics AND racism AND football

Os resultados desta primeira pesquisa conduziram a um escasso número de três artigos. Entendeu-se que investigação tinha de ser menos específica para encontrar mais artigos relacionados com o tema da dissertação.

Assim, foram efetuadas as seguintes pesquisas, tendo sido encontrado o número de resultados apresentado:

- racism AND machine AND learning, 123 resultados
- racism AND text AND mining, 24 resultados
- racism AND sentiment AND analysis, 155 resultados
- football AND sentiment AND analysis, 74 resultados
- football AND sentiment AND analysis AND racism, 3 resultados
- sports AND racism, 589 resultados
- sports AND racism AND tweets, 5 resultados
- football AND sentiment AND tweets, 26 resultados
- ((racism AND speech) football), 8 resultados
- ((free AND speech)) AND (hate AND speech)), 365 resultados
- racism AND speech AND deep AND learning, 27 resultados
- ((data AND science) AND racism AND football), 2 resultados

Para além de localizar artigos através da pesquisa direta na base de dados Scopus, foram também encontrados alguns artigos/relatórios através de uma pesquisa online no Google. Estes artigos revelaram-se importantes pela forma como abordavam e estudavam o racismo como uma questão social, em vez da abordagem de extração de texto.

B. Critérios de seleção da literatura

Os critérios para que um artigo/relatório fosse considerado relevante para a revisão dependiam do seguinte:

- O tema do artigo

Este critério considerou os artigos/relatórios que estivessem de alguma forma relacionados com o tema da dissertação. Assim, foram considerados artigos/reportagens que tinham como temas "racismo", "racismo no futebol", "discurso de ódio", "discurso de ódio no futebol", "racismo/discurso de ódio nas redes sociais," "*Text Mining* e racismo", ou "modelos de IA/*Machine Learning* e racismo", entre outros.

- A abordagem/metodologia utilizada

Este critério considerou os artigos/relatórios que estão de alguma forma relacionados com a abordagem e as metodologias a serem utilizadas na dissertação. Também podem ser incluídos artigos que não estejam diretamente relacionados com o Racismo no Futebol, mas que utilizem metodologias de Data Science, especialmente abordagens de Text Mining para detetar racismo/discurso de ódio nos discursos, modelos de IA/Machine Learning para detetar racismo/discurso de ódio, ou qualquer coisa relacionada com estes tópicos.

- As conclusões

A qualidade, credibilidade e relevância das conclusões são, por si só, um fator de validação da abordagem utilizada durante o artigo. A robustez das conclusões é crucial para que o artigo seja considerado relevante para o estudo que se pretende realizar e pode, por vezes, ser decisiva na sua seleção.

- Os resultados

Apenas foram considerados artigos cujos resultados apresentavam credibilidade e semelhança nas metodologias aplicadas e caso fossem relevantes para esta dissertação.

2.4 Resultados

Foi construída uma tabela com as informações gerais dos 23 artigos analisados. Cada coluna representa um tema, havendo uma classificação binária quanto ao facto de um tema estar presente/mencionado no artigo (1) ou não estar presente/mencionado no artigo (0), como é possível verificar na Tabela 1 (o número de cada artigo presente na tabela refere-se à ordem dos artigos mencionados no Anexo A):

Tabela 1 - Relevância dos temas por artigo

Artigo	**Racismo/ *Hate Speech***	**Violência**	**Futebol**	***Text Mining***	**Modelos de *Machine Learning***	**Estatística**	**Redes Sociais/ Online**
1	1	1	1	0	0	1	0
2	0	1	1	0	1	1	0
3	1	1	1	0	0	0	0
4	1	0	1	0	1	1	0
5	1	0	1	0	0	1	1
6	1	1	1	0	0	0	1
7	1	0	1	0	0	0	0
8	1	0	1	0	0	1	0
9	1	1	0	1	1	1	1
10	1	1	1	0	0	1	1
11	1	0	0	1	1	1	1
12	1	1	1	0	0	0	0
13	1	0	0	1	1	1	1
14	1	1	0	1	1	1	1
15	1	0	0	1	1	1	1
16	1	1	1	0	1	0	0
17	1	1	1	0	0	1	1
18	1	0	0	1	1	1	1
19	1	1	0	1	1	1	1
20	1	0	0	1	1	1	1
21	1	1	0	1	1	1	1
22	1	0	0	1	1	1	1
23	1	0	0	1	1	1	1
Total %	95.7%	52.2%	52.2%	47.8%	60.9%	78.3%	65.2%

A partir da Tabela 1, é possível perceber que apenas um dos artigos (o número 2) não teve como tema Racismo ou Discurso de Ódio (apesar de ter a Violência como tema). Isso é uma confirmação da questão principal que a pesquisa visava.

A Estatística e Redes Sociais tiveram um papel significativo, seguidas dos modelos de aprendizagem automática (é importante notar que todos os processos relacionados com a extração de texto envolveram a aprendizagem automática).

A percentagem de artigos que utilizaram ou referem "*Text Mining"* (perto de 50%) foi bastante satisfatória, uma vez que a dissertação se debruça sobre a análise da questão social do racismo no futebol (presente nas redes sociais) utilizando metodologias de *Text Mining* para o conseguir.

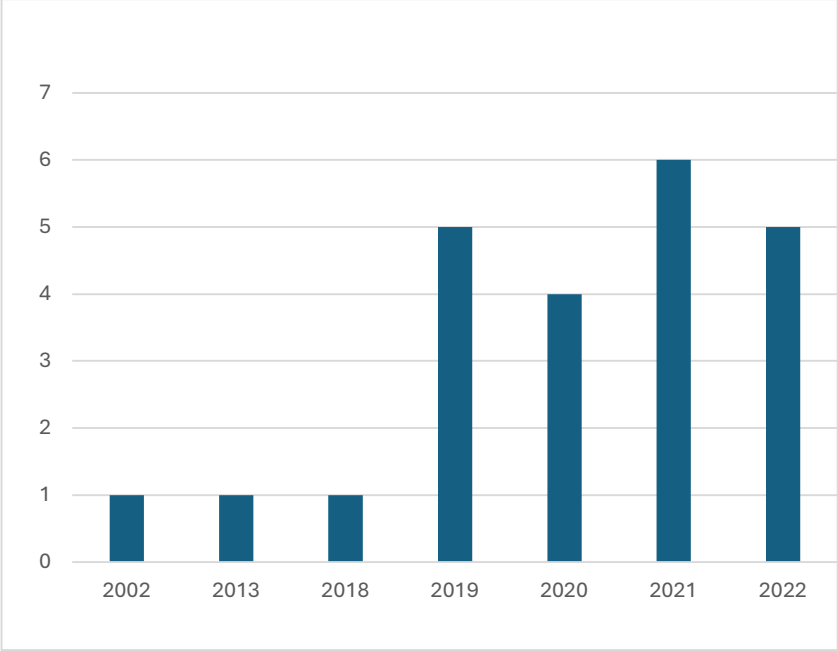

Figura 1 - Ano de publicação dos artigos

A partir do gráfico de barras da Figura 1, é possível observar que a maioria dos artigos recolhidos são recentes, o que é facilmente explicado pela importância crescente que tem sido dada a este problema social e estrutural.

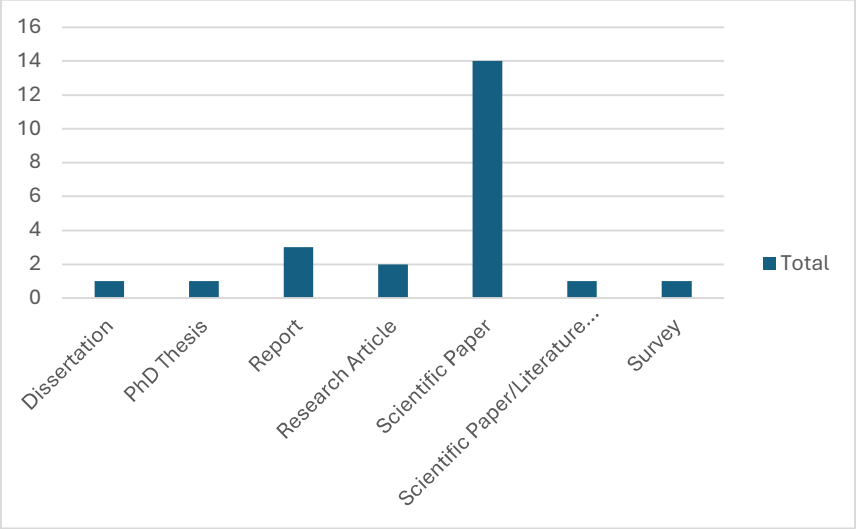

Figura 2 - Tipo de obra

A Figura 2 demonstra o tipo de obra recolhida, o que pode ser facilmente traduzido numa prevalência significativa de artigos científicos.

2.5 Conclusões

Com a recolha e leitura dos artigos recolhidos, foi possível organizar e estruturar o tópico principal da dissertação em dois subtópicos: a questão social crescente que é o racismo no futebol e a abordagem a adotar ao *dataset* extraído através de *Web Scraping* a comentários no canal de *Telegram GruppaOF*, com o objetivo de detetar discurso de ódio/racismo. Isto permitiu dividir o trabalho seguinte em duas fases, que serão determinantes para uma correta elaboração da dissertação.

Observou-se, ainda, que o racismo continua muito presente no futebol e nas redes sociais, sendo crucial que mais medidas sejam tomadas, tanto pelos governos nacionais e instituições europeias, como pelas instituições de futebol nacionais e internacionais, de forma a combater este problema que tem vindo a crescer novamente no futebol europeu.

3. Contexto

3.1 Racismo e discurso de ódio no futebol

O racismo, mais do que um conjunto de crenças pessoais, é um sistema histórico de opressão e dominação que se manifesta em problemas de privilégio e poder e que divide as pessoas em grupos de acordo com critérios raciais e étnicos (Almeida et al., 2023). Embora, possa ser criado ou reproduzido a nível individual, o sistema de práticas racistas é apoiado por algo muito mais profundo do que atos individuais. O racismo pode assumir muitas formas, incluindo preconceitos individuais, discriminação sistémica e racismo institucionalizado. Historicamente, o racismo tem sido utilizado para justificar a escravatura, a colonização e a opressão de grupos raciais marginalizados e continua a ser um problema social significativo nos dias de hoje, com muitas pessoas, de etnias consideradas minorias no local onde habitam (Alves, 2021), a enfrentarem discriminação e desigualdade em áreas como a educação (Araújo, 2016), a justiça criminal (Maeso, 2021) ou, como a dissertação pretende analisar, em desportos como o futebol.

O racismo surge frequentemente associado ao discurso de ódio, considerando-se discurso de ódio aquele que inclui intervenções com teor depreciativo apontadas a um indivíduo ou a um grupo de indivíduos, com motivações que podem ser variadas, normalmente caraterística(s) de um grupo, como a etnia, a cor, a nacionalidade, o sexo, a deficiência, a religião ou a orientação sexual. "O discurso de ódio é motivado por preconceitos e agressões dirigidos a alguém ou a várias pessoas especificamente devido às suas identidades pessoais (…) o autor do discurso de ódio atua com base nas suas opiniões e atitudes discriminatórias e preconceituosas com o objetivo de prejudicar e invalidar as vítimas" (Malta, 2022:4). No caso do racismo, o discurso de ódio pressupõe a ideia de que existe uma "inferioridade racial" (Almeida et al., 2023). Apesar disso, a legislação que define o que é discurso de ódio varia de país para país (Jahan & Oussalah, 2023), não havendo uma definição universal para discurso de ódio racial e apresentando até nuances retóricas na sua caracterização (Almeida et al., 2023), com definições frequentemente consideradas vagas ou contraditórias (Hietanen & Eddebo, 2023).

O futebol é frequentemente marcado por episódios de violência verbal e racismo entre adeptos, sendo o racismo no desporto caracterizado pela discriminação e pelo preconceito dirigidos a jogadores e adeptos com base na sua etnia e/ou religião. Isto pode assumir muitas formas, incluindo cânticos e linguagem racistas, comportamento discriminatório por parte dos

adeptos e jogadores e símbolos ou imagens racistas (Kearns et al., 2023). O racismo no futebol é um problema que está presente há muitos anos e continua a ser uma preocupação em muitos países do mundo. Tal como referido anteriormente, têm sido feitos esforços para combater o racismo no futebol, incluindo educação, campanhas de sensibilização e ações disciplinares, mas o problema persiste.

3.2. Os movimentos *Ultra* e *Hooligan*

Tendo em conta o objeto de análise desta dissertação - o canal de *Telegram*, utilizado na sua esmagadora maioria por adeptos que se identificam com o movimento *Ultra* e/ou *Hooligan* – torna-se necessário discutir os conceitos de *Ultras* e *Hooligans* e até a sua relação com a política.

"*Ultra*", tal como o nome indica, é um tipo de adepto que vive e apoia a equipa de futebol de uma forma mais "além" ou "extrema". Embora exista alguma discussão sobre a origem destes grupos, a verdade é que o termo "*Ultras*" surge em Itália, pelos anos sessenta (Almeida, 2019), integrado no nome de alguns destes grupos. Rapidamente a cultura *Ultra* se difundiu pela Europa fora e o movimento ganhou expressão em vários países, sendo hoje um fenómeno visto à escala internacional. Refira-se que um dos aspetos que tem igualmente marcado a cultura *Ultra* está justamente relacionado com a prevalência de discursos e práticas racistas.

O futebol italiano foi e é incontestavelmente manchado por episódios de teor racista e deve, por isso, ser analisado, de forma a dar contexto e motivação para o tema desta dissertação. Os grupos *Ultras* e *Hooligans*, devido a serem coletivos organizados, permitiram e permitem uma manifestação com maior impacto de certos ideais, sejam eles de teor racista ou antirracista. Entender o movimento *Ultra* implica também explicar em parte a intervenção política no mundo do futebol italiano, visto ter sido pioneiro neste tema e ter tido uma enorme influência para os restantes países europeus, assim como noutros continentes.

Os *Ultras*, definem-se como os adeptos, em grupos organizados, que entoam cânticos, exibem coreografias tanto com cânticos como com bandeiras, faixas e engenhos pirotécnicos, vestem as cores do clube e ficam o decorrer dos 90 minutos (ou mais) de jogo em pé, manifestando o seu apoio. Estas coreografias têm o nome de *tiffos*. Estes grupos demonstram uma paixão ostensiva pelo clube, dedicando horas do seu tempo livre para a preparação de todas estas atividades e deslocando-se mesmo aos jogos em que a sua equipa joga em campo adversário para a apoiar (Bromberger, 1995).

Para além de toda esta paixão pelo próprio clube, os *Ultras* têm lutas comuns: a luta contra a repressão policial e a luta contra o futebol moderno. Futebol moderno é o termo utilizado para o futebol que envolve grandes quantidades de dinheiro, paixão pelo negócio e não pelo jogo por parte dos intervenientes, a falta de "amor à camisola" por parte dos jogadores (movidos por dinheiro), entre outros. É, em resumo, o futebol envolvido pelo mundo capitalista (Bazel & Matthew, 2011)

O movimento *Hooligan*, embora tenha interseções com o movimento *Ultra*, distingue-se em grande parte pela violência. Este movimento nasceu nos anos sessenta, em Inglaterra, entre gangues rivais de skinheads, com raízes nas zonas urbanas mais desfavorecidas (Marivoet, 2009). Embora possa acontecer violência entre grupos *Ultra*, a sua premissa é apoiar o clube, enquanto no movimento *Hooligan* existe a premissa da violência entre grupos de clubes adversários. Existe como que uma disputa entre qual o grupo de um clube que consegue vencer (em confrontos físicos) os grupos de outros clubes. Os *Hooligans* tendem a vestir roupa semelhante, não alusiva ao clube e com tons de preto, de forma a serem discretos e a haver uma maior dificuldade por parte das autoridades em identificá-los (Almeida, 2019).

Os confrontos físicos podem ser espontâneos, em que um grupo ataca outro grupo de surpresa, ou podem ser combinados, em que é definida uma hora e local a acontecer. Existem mesmo países em que a cultura é levada com maior intensidade, existindo ligas formalmente organizadas (com vários confrontos entre os grupos dos vários clubes) com pontuação para que no fim haja um grupo vencedor. Esta forma mais organizada também permitiu que houvesse um maior controlo na violência, havendo regras que tentam minimizar ao máximo vítimas graves ou mortais.

De acordo com alguns autores (Taylor, 1982), o hooliganismo surgiu como uma forma de resistência das classes trabalhadoras contra a "elitização" do futebol. Os *Hooligans* exibem comportamentos considerados "antissociais," como violência, agressividade e, em muitos casos, racismo. Este racismo é visto como uma característica não só associada aos *Hooligans*, mas também a outros grupos organizados de adeptos (Almeida, 2019), sendo estes influenciados por ideais de masculinidade, identidade nacional e ideologias racistas, sexistas, etnocêntricas e xenófobas.

No Reino Unido, os discursos dominantes retrataram os *Hooligans* como símbolo dos problemas da sociedade britânica, o que desviou o foco do debate sobre o racismo numa perspetiva mais ampla, ignorando os processos sociais racializados que se desenrolam na sociedade (Almeida, 2019).

3.3. A política no mundo *Ultra/Hooligan*

O movimento *Ultra* não era apenas uma forma de mostrar fanatismo, mas também uma forma da classe operária italiana se expressar e ganhar identidade. Os primeiros grupos *Ultras* tinham afiliações políticas com a esquerda revolucionária, trazendo alguma política para as bancadas, nas coreografias e demais. Por este motivo, os *Ultras* do futebol italiano são descritos, por vezes, como os "filhos dos trabalhadores" ou "filhos dos imigrantes". No entanto, com a evolução do futebol e a generalização dos grupos *Ultras* durante a década de 1980, estes começaram a ser cada vez mais de direita/extrema-direita (Doidge, 2013). Atualmente, a nível europeu e mundial, embora existam mais grupos que se demonstram apoiantes da extrema-direita do que da esquerda radical, ainda se pode ver muito dos dois mundos. Porém, uma boa parte dos grupos *Ultra* nos dias de hoje tenta mostrar-se apolítica, havendo mesmo campanhas para não misturar a política com o futebol e, mesmo tendo membros de diferentes etnias e religiões, não fazem coreografias ou não se manifestam com frases/faixas de antirracismo ou outras lutas consideradas "políticas" (Doidge et al., 2020). Os grupos que se manifestam contra o racismo ou contra o fascismo (e até a favor da libertação da Palestina, no período mais recente), têm tendência a ser associados à esquerda ou extrema-esquerda. Existem ainda os grupos que se manifestam ativamente contra a imigração, com símbolos fascistas/nazis e com o mote anti "*Antifa*", que são associados à extrema-direita (Almeida, 2019).

Antifa, vem da combinação das palavras "*Anti*" e "*Fascist*". Refere-se (ou os próprios se referem como) a pessoas e grupos que se opõem ao fascismo, ao racismo e a outras ideologias de extrema-direita, recorrendo frequentemente à ação direta para contrariar o que consideram ser movimentos autoritários ou extremistas (Almeida, 2018). O ativismo dos grupos *Ultras* da linha *Antifa*, está geralmente alinhado com a política de esquerda, incluindo perspetivas anticapitalistas, antirracistas e antiautoritárias. Os membros destes grupos defendem ações diretas como protestos, comícios e, por vezes, até confrontos físicos para interromper ou impedir atividades e ideologias de extrema-direita (Bray, 2017).

Alguns grupos *Ultras*, devido ao facto de se influenciarem por motivações políticas, assumem-se como grupos *Antifa*. Estes incorporam ideais antifascistas ou antirracistas, posicionando-se contra a discriminação, a xenofobia e a extrema-direita, sendo a favor da aceitação dos imigrantes e de uma política multicultural, assim como o respeito pelos direitos LGBT (Bray, 2017). Os clubes mais famosos com grupos *Ultras* antifascistas conhecidos incluem:

- St. Pauli (Alemanha): Os adeptos do clube são famosos pela sua posição antirracista e antifascista, com mensagens políticas regularmente exibidas durante os jogos.
- Celtic FC (Escócia): Algumas secções da base de adeptos do Celtic, em especial os adeptos da "*Green Brigade*", manifestam opiniões de esquerda, incluindo solidariedade antifascista.
- PSG e Marselha (França): Grupos rivais (o PSG e o Marselha são eternos rivais do futebol francês) mas ambos com ideais semelhantes, antirracistas.

Apesar desta proximidade, os grupos do St Pauli e do Celtic, mais recentemente, passaram a ser rivais, por terem posições distintas quanto ao conflito israelo-palestiniano.

Os grupos *Ultras Antifa* enfrentam frequentemente tensões com grupos *Ultras* de extrema-direita. Em alguns países, estas rivalidades conduzem a confrontos dentro e fora dos estádios. Mesmo dentro de um mesmo clube podem coexistir grupos com ideais distintos – em 2010, um grupo de extrema-direita do PSG teve confrontos físicos com um grupo *Antifa*, tendo mesmo um elemento do primeiro falecido aquando dos mesmos.

Os grupos *Ultras* de extrema-direita são mais predominantes do que os *Antifa*, especialmente em certos países onde o nacionalismo e as ideologias extremistas se cruzam com a cultura do futebol (Kossakowski et al., 2020). Estes grupos defendem frequentemente pontos de vista nacionalistas, xenófobos, racistas e até mesmo assumidamente neonazis, podendo envolver-se em manifestações racistas ou cânticos nos estádios para assinalar as suas posições políticas. Os grupos de extrema-direita estão normalmente organizados entre bases de adeptos com tradições culturais ou políticas de longa data e, muitas vezes, dão ênfase ao patriotismo ou ao orgulho na identidade nacional ou regional (Kossakowski et al., 2020).

Eis três dos exemplos mais conhecidos:

- Lazio (Itália): Os ultras do clube, especialmente um subgrupo conhecido como Irriducibili, são conhecidos pelas suas filiações de extrema-direita e pela exibição de símbolos fascistas. A base de adeptos da Lazio inclui alguns dos casos mais mediáticos de simbolismo de extrema-direita no futebol europeu.
- Chelsea (Inglaterra): No passado, os "Chelsea Headhunters", uma "firma" (nome dado em Inglaterra) de *Hooligans* associada ao clube, tinha ligações a organizações de extrema-direita. Embora o grupo já não esteja tão ativo como no passado, a sua presença histórica influenciou elementos de extrema-direita entre alguns adeptos do Chelsea.

- Legia Varsóvia (Polónia): Os *Ultras* do Legia Varsóvia são bastante conhecidos pelos seus sentimentos nacionalistas, xenófobos e até mesmo racistas, exprimindo-o com coreografias, faixas e cânticos.

Nalgumas ligas, os clubes e as autoridades tomaram medidas contra a manifestação de opiniões extremistas nos estádios, mas a gestão destas questões continua a ser um desafio. Os organismos europeus que regem o futebol, como a UEFA, também aplicaram multas e sanções aos clubes com bases de adeptos que apresentem comportamentos fascistas ou racistas, com o objetivo de combater o extremismo no desporto (UEFA, 2019).

3.4. Racismo e Xenofobia no futebol italiano

O racismo e a xenofobia no futebol italiano surgiram muitos anos antes do nascimento dos grupos *Ultras*, com o rei italiano a expulsar da liga nacional todos os clubes que tivessem jogadores estrangeiros e obrigando-os a jogar numa liga à parte, em 1908 (Kassimeris, 2011). Esta regra foi abandonada no ano seguinte, já que a maior parte dos clubes não abdicou dos seus jogadores, mas anos mais tarde, quando Benitto Mussolini – ditador italiano - ascendeu ao poder, rapidamente impôs regras no futebol como mudança de nomes de clubes com nomes associados ao comunismo, a reaplicação da regra das equipas serem constituídas apenas com jogadores nacionais e ainda, dez anos depois, a proibição de jogadores judeus de praticarem o desporto, alinhando-se com a Alemanha Nazi e a ocupada Áustria (Kassimeris, 2011) - outro dos exemplos do ultranacionalismo italiano foi a tradução de todos os termos ingleses para italiano, tendo mesmo o nome do desporto mudado para "Calcio".

Em Itália, não só existe uma divisão entre os grupos *Ultra*/clubes por motivos políticos (clube de esquerda Associazione Calcio Milan vs clube de direita Football Club Internazionale Milano, Roma (apolítico) vs Lazio (extrema-direita), etc), assim como existe uma divisão por motivos regionais – a *mezzigiorno* – entre o sul e o norte, como é o caso do Napoli vs Atalanta (Kassimeris, 2011). Também os clubes de uma mesma cidade têm lutas territoriais, e muitas vezes até políticas. Os grupos *Ultra* foram e são bastante importantes para esta divisão, pois é através de um coletivo organizado que se consegue melhor expressar ideologias.

Existindo uma divisão política em Itália nos anos 1970 - conflitos políticos e civis agudos, nomeadamente nas relações complexas entre o Estado italiano e os terroristas paramilitares que actuavam na extrema esquerda e direita (Kassimeris, 2011) - os grupos *Ultra* foram também influenciados por aquela, tendo aos poucos se associado a lutas de esquerda/direita -o "Movimento Sociale Italiano" (movimento de extrema-direita) teve mão na organização da

criação dos grupos *Ultras* dos clubes Football Club Internazionale Milano e Società Sportiva Lazio.

De acordo com Guilianoti (apud Kassimeris, 2011:39):

> "The names of ultra groups tend to reflect three identifying elements: the club that they support; the political turmoil and paramilitarization of Italian society during the 1970s and 1980s; and their interests in global youth culture. At Roma there is the Commando Ultra Curva Sud, designating the south curva (end) of the capital's Stadio Olimpico. At Milan, there is the Brigate Rossonere, signifying the red and black colours of the club; similarly at Verona, there is the Brigate Gialloblu, indicating their colours of yellow and blue. Aspects of youth culture are reflected by the Drughi at Juventus, named after the gang from the novel A Clockwork Orange, the Teddy Boys at Udinese; the Skins at Inter; or even the Freak Brothers who follow Ternana."

Apesar de toda esta proximidade dos grupos *Ultra* com a política, a verdade é que nos anos 1980 houve um acalmar da situação, tendo havido uma neutralização política e um crescimento do antagonismo, com o crescimento de situações racistas e xenófobas, incluindo, entre alguns grupos *Ultra* do Norte, um sentimento pronunciado contra os italianos do Sul: "No entanto, os primeiros a serem vítimas de injúrias raciais durante os jogos de futebol não foram visados pela cor da pele ou pelo país de origem. Pelo contrário, foram os italianos brancos (caucasianos) do Sul (talvez um pouco mais escuros). Parece que este racismo era especialmente evidente em jogos que envolviam equipas do Norte e do Sul. " (Kassimeris, 2011) - os italianos do Sul eram chamados de Negro di Merda, por causa da sua proximidade geográfica com África.

Por todos os estádios em Itália havia manifestações de racismo por parte dos adeptos, com coreografias racistas a serem exibidas e alguns atos de violência por motivos de intolerância. A situação, nos dias de hoje, está mais calma, devido ao aumento do controlo policial e a criação de leis que sancionam os clubes pelas atitudes e comportamentos que os adeptos têm nos estádios (UEFA, 2019).

4. Metodologia

4.1. Introdução

Neste capítulo irá ser explicada a metodologia seguida para atingir o objetivo do trabalho: estudar a presença de racismo/discurso de ódio nas redes sociais. Para isso, é apresentada a descrição das seguidas - recolha dos dados, preparação dos dados, aplicação da Análise de Sentimentos, aplicação do modelo de ML e a realização de uma Word Cloud. A Figura 3 sintetiza este processo num fluxograma.

Figura 3 - Fluxograma da Metodologia

4.1.1. Machine Learning

Machine Learning (ML) é um subconjunto da Inteligência Artificial (AI) e da Informática que gira em torno da utilização de dados e aplicação de algoritmos a estes dados, de forma a simular a aprendizagem humana. Estes algoritmos têm a particularidade de aprenderem de forma autónoma, à medida que mais dados são inseridos.

O processo de aprendizagem de um algoritmo de ML pode ser dividido em três partes principais: o processo de decisão, a função de erro e a otimização do modelo. O primeiro, prende-se com o facto de os algoritmos de ML serem normalmente concebidos para fazer previsões ou classificações. Ao serem inseridos dados, quer estejam ou não rotulados, o algoritmo cria uma estimativa com base nos padrões que deteta nos dados. O segundo, serve para medir a precisão das previsões do modelo. Se existirem exemplos, a função de erro compara os resultados previstos para avaliar o desempenho do modelo. O terceiro, foca-se na melhoria do ajuste do modelo aos dados de treino. O algoritmo ajusta os seus parâmetros de forma a minimizar a diferença entre os exemplos reais e as previsões do modelo. O algoritmo repete este ciclo de avaliação e otimização, atualizando-se automaticamente até atingir o nível de precisão desejado (IBM, n.d.; Shinde and Shaw, 2018).

4.1.2. Text Mining

Text Mining é o processo de extrair informações valiosas de dados de texto não estruturados. É uma área de investigação relativamente nova que combina técnicas de recuperação de informação, aprendizagem automática e processamento de linguagem natural. O *Text Mining* é utilizado para uma variedade de tarefas, como a análise de sentimentos, a modelação de tópicos e *clustering* de documentos. Existem muitos desafios na extração de texto, tais como lidar com dados com ruído e não estruturados e a maldição da dimensionalidade. No entanto, os avanços recentes na aprendizagem automática e no processamento de linguagem natural fizeram da extração de texto uma ferramenta cada vez mais poderosa.

O *Text Mining* a comentários nas redes sociais envolve normalmente uma combinação de técnicas de processamento de linguagem natural (NLP), como a análise de sentimentos, a modelação de tópicos e o reconhecimento de entidades nomeadas, bem como algoritmos de Machine Learning. Estas ferramentas são utilizadas para extrair informações relevantes dos comentários e transformá-las num formato que possa ser analisado e visualizado (IBM, n.d.; Gupta & Lehal, 2009).

4.1.3. Natural Language Processing (NLP)

Natural Language Processing é uma subárea da inteligência artificial (AI) que utiliza técnicas de *Machine Learning* para permitir que os computadores compreendam e interajam com a linguagem humana. Esta integra a Linguística Computacional - aquela que modela a linguagem

humana - com a análise estatística, *Machine Learning* e *Deep Learning*, permitindo que os computadores identifiquem, interpretem e produzam texto/discurso.

A investigação com NLP deu início à era da IA generativa, melhorando as capacidades de comunicação dos modelos de linguagem de grande dimensão (Large Language Models - LLM) e permitindo que os modelos que geram imagens interpretem os pedidos dos utilizadores. Para muitas pessoas, a NLP tornou-se parte integrante da vida quotidiana, tendo inúmeras aplicações como motores de pesquisa, chatbots para o serviço ao cliente através de comandos de voz, facilitar a navegação GPS ativada por voz ou ainda apoiar assistentes digitais em smartphones.). Além disso, NLP é cada vez mais importante nas soluções empresariais, ajudando a otimizar e a automatizar operações, a aumentar a produtividade dos funcionários e a simplificar processos essenciais. (IBM, n.d; Nadkarni et al.)

4.2. Recolha de Dados

A recolha de dados está na base do trabalho que será desenvolvido, sendo a objetividade dos dados indispensável para que se obtenham resultados que sejam relevantes. A recolha de dados foi feita através de técnicas de *Web Scrapping*, aplicadas na rede social/plataforma *Telegram Gruppa_OF*. Neste canal são feitas diariamente publicações pelos seus administradores, sobre acontecimentos do mundo dos adeptos do futebol - momentos dos grupos *Ultra* ou *Hooligan* – que ocorrem no caminho para o estádio, dentro do estádio, em celebrações, em situações de confrontos físicos, entre vários outros. O canal permite ainda que utilizadores "normais" façam comentários dentro de cada publicação. O *dataset* recolhido, através das técnicas de *Web Scraping*, tem nele contida informação sobre vários destes *posts*, tendo um campo fundamental para a análise - os comentários feitos a cada publicação.

A amostra em análise nesta dissertação consta nas publicações no canal referido do *Telegram*, referentes ao ano 2023, e compreende cerca de 112480 comentários, dispersos por 2855 publicações.

É importante referir que, inicialmente, a dissertação tinha como objetivo trabalhar apenas com dados do Twitter (atual X). No entanto, tal tornou-se inviável devido à sua aquisição pelo atual CEO, que implementou barreiras à extração de dados do X através de técnicas de *Web Scraping,* originalmente previstas para a realização deste trabalho. Perante esta adversidade, foi necessário alterar o objeto de estudo, tendo-se escolhido a rede social *Telegram*. Embora possa parecer uma mudança significativa, a verdade é que são ambas redes sociais em que não existe grande limitação do conteúdo partilhado, o que é especialmente o caso do *Telegram* -

não existe moderação ou a que existe é muito pouca e, desta forma, os utilizadores têm a liberdade para se expressarem da maneira que entenderem, sem repercussões. A escolha do *Telegram* também se prendeu com o facto de existirem canais, como o selecionado para análise, em que o assunto único é futebol e os seus adeptos, permitindo, assim, um foco maior na extração de comentários que se revelam importantes para o estudo em questão.

4.3. Preparação dos Dados

Nos inúmeros comentários dentro das várias publicações do canal *Telegram*, é possível encontrar diversas línguas/idiomas, como referido anteriormente. Por razões de simplificação e maior objetividade foi escolhido apenas tratar e fazer o estudo dos comentários que se apresentam em língua inglesa, usando para tal a biblioteca de *Python* "LangDetect"[1]. Com este passo, o número de comentários reduziu-se de 112480 para 26788.

Para além da remoção de comentários de línguas que não a inglesa, foram também removidos todos os emojis, pois isso envolveria toda uma outra análise que não é a pretendida para esta dissertação.

Antes da aplicação dos modelos, foi necessário proceder à limpeza dos dados, utilizando técnicas de NLP para que se perca a menor informação possível, garantir coerência, reduzir a dimensionalidade dos dados e normalizar os mesmos, garantir uma maior aplicação e, possivelmente uma maior precisão dos modelos (Nadkarni et al., 2011). Para isso, foram utilizadas as seguintes técnicas:

- Remoção de espaços desnecessários, pontuação e emojis

- Tokenização: Processo de decomposição de um texto em palavras individuais ou *tokens*

  Veja-se o exemplo apresentado na Tabela 2 (com o *dataset* utilizado para a dissertação):

[1] https://pypi.org/project/langdetect/

Tabela 2 - Comentários Tokenizados

Comentários	Comentários Tokenizados
Who should be afraid of u	[[Who, should, be, afraid, of, u]
There was a "Fuck Celtic" Banner	[There, was, a, ``, Fuck, Celtic, '', Banner]
Fuck hibs from a rangers fan	[[Fuck, hibs, from, a, rangers, fan]

- *Stop Words*: Remoção de palavras como "the", "of", "is", "and", etc. Estas palavras não representam qualquer utilidade e costumam ser removidas nos processos de NLP.

  Veja-se o exemplo na Tabela 3 (depois de tokenizadas e removidas as Stop Words):

Tabela 3 - Comentários tokenizados com remoção de Stop Words

Comentários	Comentários Tokenizados
Who should be afraid of u	[[afraid, u]
There was a "Fuck Celtic" Banner	[``, Fuck, Celtic, '', Banner]
Fuck hibs from a rangers fan	[[Fuck, hibs, rangers, fan]

- *Lemmatization* e *Stemming*:

  Stemming é preferível para situações em que o contexto não se revela importante para a análise, sendo que a *Lemmatization* é mais precisa, preserva o significado e a função gramatical, mas é mais lenta, isto é, precisa de mais tempo de computação, pois avalia a frase inteira e não apenas palavra a palavra (Khyani, 2021). Posto isto, foi decidido aplicar a *Lemmatization*, pois representa a técnica que maior utilidade irá ter para a aplicação dos modelos posteriormente, garantindo uma preservação do contexto lexical.

- Correção de palavras mal digitadas:

  Embora seja algo que se mostre importante para garantir a fiabilidade e coerência do texto analisado, de forma a se conseguir uma melhor precisão na aplicação dos modelos,

a verdade é que com o tipo de conteúdo que o *dataset* apresenta (nomes de clubes, abreviaturas com esse mesmo propósito, palavras com contexto específico - conteúdo partilhado por adeptos de futebol - , entre vários outros), a aplicação desta técnica não se apresentou como algo fundamental, tendo mesmo sido contraproducente por alterar diversas palavras, o que não era o objetivo. Sendo assim, decidiu-se não utilizar esta técnica para a preparação do texto.

4.4. Análise de Sentimentos com *RoBERTa*

4.4.1 Análise de Sentimentos e o modelo *RoBERTa*

Análise de Sentimentos (SA) é uma técnica de NLP utilizada para determinar o tom emocional subjacente a uma série de palavras. Esta desempenha um papel crucial na compreensão de opiniões, sentimentos e emoções expressos em dados textuais, o que a torna particularmente valiosa para analisar conteúdos de utilizadores de redes sociais (Yue et al., 2019).

Assim, após a preparação dos dados, foi feita uma Análise de Sentimentos aos comentários. Os comentários *tokenizados* foram unidos novamente (mantendo apenas os tokens) e foi aplicado o modelo *RoBERTa*[2] *(Robustly Optimized BERT Approach*), mais especificamente o *CardiffNLP*[3], modelo este que foi obtido através do *HuggingFace* e se baseia na arquitetura *RoBERTa* para criar modelos especificamente ajustados a determinadas aplicações, como a análise de sentimentos nas redes sociais. Este modelo específico foi otimizado num grande conjunto de dados de tweets com *labels* em função do sentimento. O processo de *Fine-Tuning* ajusta o modelo *RoBERTa* pré-treinado para compreender melhor as nuances da linguagem e das expressões utilizadas no Twitter. Ao treinar com dados das redes sociais, o modelo aprende a reconhecer a linguagem informal, as abreviaturas, os emojis (não incluídos nesta dissertação) e outras caraterísticas únicas dos tweets que podem influenciar o sentimento.

O *RoBERTa* é um modelo que deriva do BERT (*Bidirectional Encoder Representations from Transformers*) e foi desenvolvido pelo *Facebook AI*. Este é um modelo mais otimizado que o modelo BERT original, treinando em conjuntos de dados maiores, removendo o *Next Sentence Prediction* (NSP) e utilizando tempos de treino mais longos. O *RoBERTa* demonstra,

[2] https://huggingface.co/docs/transformers/model_doc/roberta
[3] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

assim, melhor desempenho em várias tarefas de NLP, quando comparado com o *BERT* (Liu, 2019)

Como este modelo reconhece linguagem informal em contexto de redes sociais, a análise de sentimentos foi feita diretamente aos comentários, ao invés de ser feito aos comentários pré-processados (comentários após os passos referidos em 4.3).

O *dataset* é composto por um total de 2855 linhas, onde cada linha representa um post e dentro desse post há um X número de comentários. O número total de comentários analisados foi de 26788, o que dá uma média de aproximadamente 9,4 comentários por post. Para correr o código, foi utilizado o *Google Colab* – plataforma *Cloud-based* fornecida pela *Google* que permite aos utilizadores escrever e executar código Python num navegador *Web* – e a performance foi bastante positiva, tendo demorado 14 minutos e 43 segundos a fazer a Análise de Sentimentos a todos os comentários. A Tabela 4 resume estes dados.

Tabela 4 - Tempos de processamento da Análise de Sentimentos

Nº de Linhas (Posts)	Nº Total de comentários	Nº médio de comentários por Post	Tempo de Processamento	Tempo de Processamento médio por comentário
2855	26788	9,4	14m43s	0.033s

4.4.2 Discussão de resultados da Análise de Sentimentos

Assim, o modelo CardiffNLP/RoBERTa realizou análise de sentimentos a 26788 comentários, tendo recebido como inputs os comentários no seu estado *Raw* e devolvendo como output uma classificação, com os valores: Negativo [0], Neutro [1] ou Positivo [2]. Veja-se o exemplo da Tabela 5 e Figura 4.

Tabela 5 - Exemplo da classificação da Análise de Sentimentos

Comentário	Valor numérico	Classificação
That's surprising! A lot of people in Lecco province is supporting Atalanta	2	Positivo
In february was Lecce	1	Neutro
Don't have the needs to make an agenda with a faggot like you	0	Negativo
Yeah I hope Russia invade you scummy robbing cunts and rape your mothers	0	Negativo
reclaim your country from niggers and arab	0	Negativo

Figura 4 - Resultados da Análise Sentimentos

Os resultados demonstram uma esmagadora minoria para os sentimentos Positivos (2037), que representam apenas 7,6% dos comentários analisados. Este valor demonstra claramente que os adeptos de futebol, nas trocas de mensagens que fazem entre si nas redes sociais, não revelam uma “simpatia” uns pelos outros. Levando em consideração que a maioria dos posts no canal de Telegram analisado serem relativos a confrontos físicos entre adeptos, é "compreensível” que o tipo de intervenções feita pelos variados adeptos de futebol seja de carácter ofensivo. Existem vários posts que são apenas de momentos no estádio, no caminho para o estádio ou em meras celebrações - que são momentos de festa – onde o sentimento partilhado pelos adeptos nos comentários pode ser mais positivo, mas os mesmos não são suficientes para que este seja substancial.

O sentimento mais sentido foi o Neutro, com 49%. No entanto, vários destes comentários podem ter sido assim classificados pelo facto da linguagem ser ambígua. Vejamos dois exemplos na Tabela 6.

Tabela 6 - Exemplos de comentários neutros que contêm conteúdo negativo

Comentário	Valor Numérico	Classificação
Biris never show	1	Neutro
50 x 5 ?? Well done.. anyone can kick someone who is down..	1	Neutro

No primeiro comentário podemos ver uma situação do mundo *Hooligan* em que, quando existem confrontos físicos agendados e um dos grupos não comparece, existe a tendência para serem considerados um grupo "cobarde". Esta expressão de "never show" representa por isso um insulto, mas não é considerado um comentário negativo pela ambiguidade que o próprio comentário apresenta. No segundo comentário podemos ver uma clara ironia, mas o modelo não consegue interpretar como sendo algo negativo.

Os comentários negativos tiveram uma representatividade considerável, com 43,4% (com 11625 ocorrências), o que permite concluir que existe um ambiente constante de provocação, insulto e ódio entre os adeptos de futebol que intervêm neste canal de Telegram. Os adeptos mais fervorosos de futebol tendem a ter uma paixão tão grande pelo clube e um ódio pelos seus rivais, que entram em discussões uns com os outros e o sentimento de raiva é evidentemente presente. Se o modelo conseguisse detectar todos os comentários negativos que foram classificados como neutros, pela sua ambiguidade, o valor total seria certamente bastante mais elevado.

Outro valor interessante a observar, é o rácio Negativo para Positivo, valor que é dado pela divisão entre o primeiro e o segundo e que deu um resultado de 5,71. Este valor indica que, a cada 6 comentários negativos, ocorre um comentário positivo. Se se observar o rácio com Neutro + Negativo para Positivo, o valor é de 12,15. Significa isto que, em média, apenas após

12 comentários de teor negativo ou neutro, é que podemos finalmente encontrar um comentário com sentimento positivo.

4.5 Modelação

Para atingirmos o objetivo principal, que consiste na deteção de *Hate Speech* (pesquisado em inglês) nas redes sociais num canal de Telegram, utilizado por adeptos de futebol, é necessário aplicarmos um modelo de *Machine Learning* aos dados recolhidos. Neste caso, após alguma pesquisa, foi escolhido um modelo de *Machine Learning* que é também derivado do algoritmo BERT, através da plataforma *HuggingFace*.

De entre vários modelos observados, escolheu-se o modelo Toxic BERT, do utilizador *Unitary* na plataforma *Hugging Face*, que se pode encontrar na mesma através da pesquisa *unitary/toxic-bert*. Este apresenta 6 classes de classificação - desde racismo, identidade, sexismo, entre outros - que o tornam menos ambíguo quando comparados com outros modelos. Alguns dos modelos observados tinham apenas três classes, como *Hate Speech*/*Not Hate Speech*/*Toxic*, e outros apenas duas classes como *Hate Speech*/*Not Hate Speech*, o que levaria a que se tivesse pouca informação para estudar o problema em causa e tirar conclusões, tornando o estudo pouco relevante.

Este modelo foi especificamente concebido para classificar dados textuais em categorias de Toxicidade, utilizando uma arquitetura pré-treinada baseada em BERT (*Bidirectional Encoder Representations from Transformers*).

A plataforma onde se processou o modelo foi novamente o *Google Colab*, tendo o modelo tido os tempos de processamento que se reportam na Tabela 7.

Tabela 7 - Tempo de processamento do modelo Toxic BERT

Nº Total de comentários	Tempo de Processamento	Tempo de Processamento médio por comentário
26788	57m07s	0.128s

Inicialmente, tentou-se correr o modelo aos comentários tokenizados mas, como o modelo não estava a conseguir correr da forma que se pretendia e a dar resultados nada esclarecedores, decidiu-se não se aplicar o modelo aos comentários Tokenizados, isto porque, tal como na

Análise de Sentimentos, o modelo foi desenvolvido com base em redes sociais e, portanto, já foi treinado para reconhecer linguagem natural. O contexto é muito importante para a deteção de qual o tipo de Hate Speech que está a ser utilizado, daí ser importante manter a estrutura dos dados recolhidos, assim como é importante que as expressões utilizadas não sofram qualquer alteração, como se observou nos passos desenvolvidos na Seccção 4.3.

O modelo Toxic BERT apresenta 6 classes possíveis para a classificação de um comentário com sendo tóxico, como se descreve na Tabela 8.

Tabela 8 - Tipos de classes do modelo Toxic BERT

Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Class 6
Toxic	*Insult*	*Obscene*	*Identity Hate*	*Threat*	*Severe Toxicity*

Ao contrário de alguns outros modelos, este não prevê explicitamente um rótulo "não tóxico". Em vez disso, concentra-se em identificar a qual das classes tóxicas pertence um determinado comentário. Procedeu-se a uma alteração no código para identificar os comentários não tóxicos, conforme descrito na próxima secção

4.5.1 Implementação e ajustes no código

Como foi referido anteriormente, a linguagem utilizada foi *Python* e utilizou-se a biblioteca *Transformers*[4] para a aplicação do modelo. Inicialmente, experimentou-se com uma pequena amostra de 100 linhas com X comentários do conjunto de dados, para que demorasse menos tempo a correr e fosse possível ir fazendo as alterações consideradas adequadas. Foram seguidos os seguintes passos:

1. **Configuração da *Pipeline***: Inicializámos o modelo unitário/*toxic BERT* através da função pipeline() – O texto de entrada é convertido para o formato exigido pelo modelo – com a biblioteca *transformers*.

[4] https://huggingface.co/docs/transformers/index

2. **Implementação de um *Threshold***: Como o modelo não inclui uma etiqueta *Non Toxic*, aplicou-se um *Threshold* personalizado. Se nenhuma das seis categorias tóxicas de um comentário ultrapassasse o limite definido (fixado em 0,1), atribuiu-se manualmente a etiqueta *Non Toxic*. Isto permitiu filtrar os comentários com níveis de toxicidade muito baixos, garantindo que apenas os comentários significativamente tóxicos eram categorizados. O *Threshold* é baixo pois pôde-se observar vários comentários com uma classificação baixa para qualquer uma das classes, mas que continham discurso de ódio ou algum nível de toxicidade. O exemplo da Tabela 9 permite ilustrar esta questão.

Tabela 9 - Exemplos de comentários tóxicos com probabilidades baixas

Comentário	Probabilidades de Toxicidade
White power	[0.09131096303462982, 0.010609108954668045, 0.0017106940504163504, 0.0011430863523855805, 0.0010651270858943462, 0.00044287994387559593]
Football Hooligans = nationalists	[0.12475244700908661, 0.00720358919352293, 0.003369562327861786, 0.0009983344934880733, 0.0006231367588043213, 0.0002939120167866349]
roleplay freaks	[0.26274698972702026, 0.01526623871177435, 0.002983727725222707, 0.0021511977538466454, 0.000536814914084971, 0.0003110144170932472]

3. **Ordenação das classes previstas**: Além de identificar a classe mais provável (a primeira pontuação mais alta), extraiu-se e classificou-se todas as classes previstas para cada comentário, juntamente com as suas probabilidades. Isto permitiu uma visão completa de como o modelo distribuiu a sua confiança entre diferentes categorias, o que é útil para uma análise mais profunda. Uma das razões desta alteração foi o facto de 100% das vezes os comentários apresentarem a classificação mais alta como sendo *Toxic*, não fornecendo uma visão pormenorizada de que tipo de toxicidade cada comentário apresenta.

4. **Aplicação do modelo no conjunto de dados completo**: Depois de testar a abordagem numa amostra mais pequena, o conjunto de dados completo foi processado utilizando a mesma lógica. Cada comentário foi classificado e os resultados foram guardados num ficheiro Excel, incluindo não só a classe mais tóxica para cada comentário, mas também a distribuição de probabilidade completa para todas as classes.

Em resumo, este processo permitiu-nos classificar automaticamente grandes conjuntos de dados de comentários de acordo com o tipo e o grau de toxicidade, mantendo a flexibilidade através da aplicação de limiares. Estas classificações e pontuações de probabilidade constituem a base para a análise posterior neste capítulo.

5. Discussão de Resultados

5.1 Discussão de Resultados Binários

Como referido anteriormente, devido à ausência de uma Classe *Non-Toxic*, o modelo classificou 100% dos comentários como pertencentes à Classe *Toxic,* por ser a categoria com maior probabilidade de ocorrência. Sendo assim, foi necessária a implementação de um *Threshold* para que os dados pudessem ser melhor analisados relativamente ao seu conteúdo.

Assim, tendo em conta a metodologia adotada, obtiveram-se os resultados que se apresentam na Figura 5.

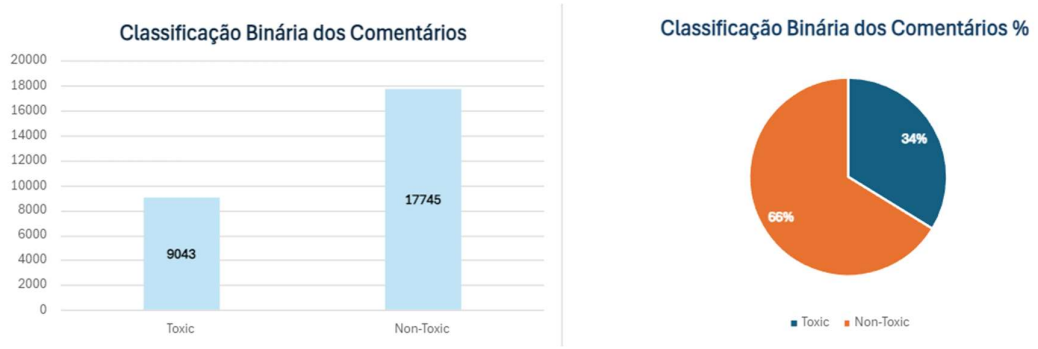

Figura 5 - Classificação binária dos comentários

Os resultados indicam que 34% dos comentários (9 043) foram classificados como tóxicos e 66% (17 745) como não tóxicos. Recorda-se que estes comentários foram recolhidos num grupo de Telegram utilizado por adeptos de futebol, em que estes participam em conversas, discussões e interações ocasionalmente controversas. O facto de um terço de todos os comentários serem tóxicos é significativo e aponta para uma cultura de hostilidade e agressão nestas comunidades dentro desta rede social. Também se infere a partir deste resultado que existe alguma relação entre a percentagem de comentários tóxicos e a percentagem de comentários negativos (43%), obtida aquando da Análise de Sentimentos na Secção 4.4.

Os comentários tóxicos são caracterizados por uma linguagem que inclui insultos, ameaças e várias formas de discurso de ódio. Embora seja comum os adeptos de desporto expressarem emoções fortes e rivalidades, esta elevada percentagem de interações tóxicas sugere que o

discurso dos adeptos nestes grupos passa frequentemente da discordância para uma comunicação prejudicial e destrutiva. Isto pode refletir a forma como as plataformas online, com menos barreiras à linguagem abusiva, se podem tornar câmaras de eco para discursos de ódio, ataques pessoais e linguagem divisiva.

No contexto dos grupos de adeptos de futebol, a toxicidade pode ir para além dos meros insultos, sendo provável que estes dados representem uma tendência preocupante no que toca a racismo, xenofobia e outras formas de discriminação que estão, muitas vezes, incorporadas em discurso de adeptos. Ao longo da história, o racismo tem sido um problema no futebol, como se pode verificar no capítulo 2 desta dissertação e, neste ambiente de redes sociais, o discurso de ódio baseado na identidade e os estereótipos nocivos podem ser mais comuns, escondidos sob o pretexto de rivalidade entre adeptos.

Estas conclusões levantam questões importantes sobre a normalização da liberdade de expressão sem filtros em certas redes sociais. Com 33% dos comentários classificados como tóxicos, existe a preocupação de que este comportamento possa influenciar ações no mundo real ou perpetuar um ambiente hostil, tanto online como offline, sendo que o volume significativo deste tipo de linguagem pode indicar uma eventual necessidade de moderação mais forte nas plataformas das redes sociais, com o objetivo de atenuar a propagação do discurso de ódio e manter a integridade destas comunidades dentro das redes.

Além disso, esta análise realça a questão mais vasta do papel das redes sociais na amplificação de comportamentos nocivos. O anonimato e o alcance de plataformas como o *Telegram* permitem que os utilizadores se envolvam em comportamentos tóxicos com menos consequências, contribuindo para um ciclo em que a linguagem hostil se torna enraizada na cultura dos adeptos. Abordar esta questão é fundamental não só para criar espaços online mais inclusivos e respeitosos, mas também para combater o racismo, a discriminação e a intolerância no próprio desporto.

5.2 Discussão de Resultados das Classes de Toxicidade

Assim, para que se possa observar de uma forma mais pormenorizada o tipo de toxicidade presente nos diversos comentários, procedeu-se à contagem de cada classe com 2ª maior probabilidade de ocorrer em cada comentário, isto porque, como referido anteriormente, a

primeira classe foi 100% classificada nos comentários como *Toxic*. Os resultados são apresentados na Figura 6.

Figura 6 - Toxicidade dos comentários por classes

Uma das primeiras observações que se pode fazer é a ausência de qualquer Classe *Severe Toxic* como sendo mais provável. Tendo em conta que existe bastante conteúdo muito agressivo neste canal de *Telegram*, a explicação para este resultado pode prender-se com o facto de, mesmo sendo conteúdo muito agressivo, as outras classes sendo mais predominantes, não satisfazendo uma maior probabilidade para *Severe Toxic*. Também poderia ter sido criada uma *Threshold* para classificar como *Severe Toxic* caso a probabilidade passasse aquela, mas, tendo em conta o objetivo de estudo desta dissertação, não se considerou que tal fosse necessário, continuando o trabalho analisando os quatro tipos de discurso *Toxic* que se obteve nos resultados.

A maior proporção de comentários tóxicos foi classificada como *Obscene* (50,64%), o que representa um nível significativo de linguagem ofensiva, e reflete o intenso envolvimento emocional dos adeptos. Estes comentários podem incluir palavrões ou linguagem imprópria utilizada nas discussões, muitas vezes dirigidas a outros indivíduos ou grupos, o que os torna uma caraterística comum nestas discussões. Atente-se à Tabela 10 onde se apresentam estes resultados.

Tabela 10 - Exemplos de comentários classificados como Obscene

Comentário	Classe
Behave yerself mate nothing like going down the ground very early on arsehole	Obscene
scum of the earth	Obscene
This list is 1000% bullshit	Obscene
I think hansie loves to suck there dicks	Obscene
somebody up licking gruppaof ass to get this results	Obscene

A classe *Insult* (29,26%), com 2646 ocorrências, foi a segunda categoria mais frequente, mostrando como os ataques pessoais e os comentários depreciativos são uma parte essencial do discurso. Esta análise vai de acordo ao entendimento geral de que ambientes competitivos como o futebol podem fomentar uma atmosfera altamente antagónica em que se trocam constantemente insultos entre adeptos rivais, tanto por apoiarem um clube diferente, como pelo ódio que sentem devido a, por exemplo, terem existido confrontos físicos anteriores (ou tentativas de) entre os grupos *Hooligan* ou *Ultra* – segundo comentário da Tabela 11. Na mesma tabela, podemos observar mais exemplos do mesmo.

Tabela 11 - Exemplos de comentários classificados como Insult

Comentário	Classe
Obsessed with us yous cunts	Insult
Yous were hiding behind police ya mad clown	Insult
Pedophile march	Insult
You’re the biggest gobshite on this chat	Insult
roleplay freaks	Insult

Como quarta colocada, temos a classe *Threat*, com 595 ocorrências (cerca de 7%). Estas correspondem, tal como a palavra indica, a ameaças feitas entre si pelos utilizadores do canal, mostrando mais uma vez o quão hostil é o ambiente vivido dentro das redes sociais entre adeptos de futebol. Alguns dos comentários que vemos nos exemplos da Tabela 12 demonstram

o ambiente politizado referido no Capítulo 3 (Contexto) desta dissertação, no mundo de futebol. Os membros de grupos *Ultra* e *Hooligan*, que têm afiliações políticas em termos de ideias, têm a tendência para mostrar as suas posições e ódios subjacentes às mesmas nas interações que fazem com outros utilizadores neste canal de *Telegram*. A política está sempre muito presente, o que leva à intensificação das divisões e a um aumento da polarização não só no mundo do futebol, mas também na sociedade como um geral - são repercussões inevitáveis.

Tabela 12- Exemplos de comentários classificados como Threat

Comentário	Classe
Attack a House with one Batterie	Threat
Palestina will be burned to the ground	Threat
Germany is going to pieces	Threat
In Russia we would kill this cops as a pigs	Threat
Death to all refugees!!!!!	Threat
Glory to Greece death to USrael !!	Threat

Em último, analisa-se a categoria que mais se identifica com o tema desta dissertação: *Identity Hate*. Esta categoria apresenta 1.224 ocorrências (cerca de 13% do total de comentários Tóxicos), sendo especialmente preocupante, pois é aquela que mais gera *Hate Speech* relacionado com discriminação e que mais representa e fomenta as situações de preconceito. Com base nos exemplos apresentados na Tabela 13 é possível concluir que esta categoria reflete casos de racismo, xenofobia e homofobia, entre outros, sendo esta a classe que mais relevância tem para o tema desta dissertação: o racismo no futebol nas redes sociais. A existência destes comentários revela um lado mais sombrio da cultura dos adeptos de futebol, em que se observa que o *Hate Speech* é utilizado para rebaixar os outros através da discriminação das características de certos indivíduos ou grupos, indo muito além das rivalidades típicas que até são por muita gente consideradas "saudáveis" no futebol.

Tabela 13 - Exemplos de comentários classificados como Identity Hate

Comentários	Classe
Best at being jews and communists. And now all West Russia is muslim lol. Thanks to their KGB jew cabbad leader Jewtin.	Identity Hate
Gays v jews Hope both lose	Identity Hate
Europeans? psg fans are 70% africans	Identity Hate
help to France for cleaning niggers and arabs	Identity Hate
Turkish not europeans. Stay at home.	Identity Hate
Haha you must be the white guy on the picture...no wonder you defend niggas when you are in relationship with one	Identity Hate
Only good communist is a dead one	Identity Hate
Nazi people on internet say to other nazis online to hide in person that they are nazi to people	Identity Hate
Celtic side is gaypride.	Identity Hate

Os comentários revelam um nível acentuado de hostilidade centrada na raça, religião, nacionalidade e convicções políticas, mostrando uma vez mais os preconceitos profundamente enraizados. Os pontos de vista anti-semitas e anticomunistas são combinados com homofobia e insultos raciais, especialmente dirigidos a indivíduos negros, muçulmanos e LGBTQ+. A juntar com isto, observa-se também a presença de xenofobia, que considera certos grupos étnicos como "não europeus" ou indesejáveis. A linguagem é inflamatória e desumanizante, apelando frequentemente à violência ou à "limpeza" de grupos específicos, revelando um alinhamento com ideologias de extrema-direita e de exclusão. No seio de uma comunidade de adeptos de futebol, esta retórica fomenta a divisão e a hostilidade, normalizando o HS e, potencialmente, encorajando a agressão no mundo real. A presença destes comentários num fórum público sublinha a necessidade de uma moderação eficaz para travar este tipo de discurso tóxico e atenuar os danos que pode causar tanto em espaços online como offline.

É importante notar que o modelo não consegue sempre detetar a presença de racismo ou xenofobia em todos os comentários que os contêm, isto porque existem outras classes que podem estar mais presentes no discurso, como a *Insult*, *Threat* ou *Obscene*. Os comentários classificados como alguma destas classes podem também conter elementos de retórica

discriminatória ou de ódio, mas não prevalecerem sobre a probabilidade de serem um *Insult*, *Threat* ou *Obscene* - por exemplo, um comentário do tipo "insulto" pode ter uma carga racial ou um comentário do tipo "obsceno" pode conter insultos homofóbicos. Vejam-se os exemplos na Tabela 14.

Tabela 14 - Exemplo de comentários com Identity Hate mas em que as outras classes se sobressaem

Comentário	Classe
Fenian scum only out with the police escort	Insult
Palestina will be burned to the ground	Threat
free palestina PUTIN WILL HELP US WITH THAT and al you greedy JEWS WIL BE KILLED	Threat
It's just Moroccans and Algerians who are the problem.	Insult
Muhammed was pedophile	Insult
Fuck Muslims. Us West need to stand up!!	Obscene
Your mother is a gypsy.	Insult
fucking albanians dirty scum only with knife and sticks they can fight fucking pussies Kosovo je Srbija	Obscene
Go fuck yourself pussy typer gay	Obscene

5.3 *Word Clouds*

5.3.1 *Word Cloud* dos comentários tóxicos

De forma a se obter uma visão das palavras que mais ocorrem dentro dos comentários classificados como tóxicos (qualquer um que esteja dentro das quatro categorias da Figura 6 – sem os comentários *Non-Toxic*), foi desenvolvida uma *Word Cloud*, tendo apresentado as palavras que se observam na Figura 7 – a *Word Cloud* foi feita aos comentários após a remoção de Stop Words, de forma a retirar palavras que aparecem muito, mas que não apresentam relevância para o estudo.

Figura 7 - Word Cloud dos comentários tóxicos

Através da Figura 7 é possível concluir que:

As rivalidades são intensas e existe um sentimento de identidade/pertença a um grupo. A proeminência de palavras como “ultra”, “hooligan” e vários nomes geográficos ou específicos de clubes (como “Twente”, “Roma”, “English”, “Italy”) é uma caraterística clara das rivalidades profundamente enraizadas que existem no seio dos adeptos de futebol.

A utilização frequente de termos depreciativos, palavrões e referências à violência física (“fight”, “weapon”, “atack”) sugere que as interações entre os utilizadores deste canal não ocorrem meramente online, podendo vir a haver ou tendo já havido confrontos físicos entre

eles. Os canais de Telegram, tal como as restantes redes sociais, podem funcionar como amplificadores da rivalidade e ódio entre pessoas/grupos. Dada a utilização de linguagem agressiva e discriminatória, existe o risco de a hostilidade online contribuir para tensões na vida real, especialmente no movimento Hooligan.

A presença de termos como “faggot”, “nazi”, “muslim” e “jew” confirma o preconceito subjacente (em algumas pessoas) que não se limita apenas às rivalidades, mas que se estende a preconceitos sociais mais amplos. Isto reflete a necessidade de sensibilização e, potencialmente, até de moderação e regulação para evitar que estes canais promovam um ambiente hostil que vise identidades baseadas na raça, religião ou orientação sexual.

A presença da palavra 'antifa' na word cloud não apenas reflete o contexto polarizado das discussões, mas também evidencia como o termo é frequentemente usado de forma pejorativa para denegrir indivíduos ou grupos associados a movimentos antifascistas, destacando a carga ideológica presente nos discursos dos adeptos.

5.3.2 *Word Cloud* dos comentários classificados como *Identity Hate*

Foi realizada a mesma análise, mas apenas para os comentários que estão classificados como sendo de *Identity Hate*, como se pode observar na Figura 8.

Figura 8 - Word Cloud dos comentários Identity Hate

Através da Figura 8, é possível concluir que:

As palavras como “gay”, “jew”, “muslim”, “nazi”, “albanian”, entre outras, destacam-se de forma clara, realçando a linguagem baseada na discriminação. Estes termos confirmam a tendência neste canal para interações onde são destacados grupos específicos com base na religião, nacionalidade, etnia ou orientação sexual.

Termos como “german”, “polish”, “italian”, “albanian”, “turk” e “romanian” estão associados a estereótipos negativos relacionados com a nacionalidade. Isto reflete as rivalidades e preconceitos étnicos que são comuns em alguns grupos *Ultra* ou *Hooligan*, especialmente quando as tensões internacionais entre diferentes grupos são elevadas.

Palavras como “nigger”, “terrorist”, “slave” e “fascist” reflectem racismo explícito e outras formas de discriminação. Estes termos indicam que o discurso neste subconjunto tem uma quantidade significativa de linguagem depreciativa e incendiária centrada na raça, religião ou filiações políticas.

As palavras relacionadas com confrontos físicos, como “fight”, “weapon”, “knife” e “atack”, continuam a ser predominantes, mesmo após o filtro de *Identity Hate*.

A presença de palavras como “fascist”, “comunist”, “terrorist” e “antifa” são uma confirmação da aproximação política que muitos grupos ou indivíduos têm, o que demonstra uma mistura entre os preconceitos ideológicos e as rivalidades relacionadas com o futebol.

5.4 Reflexões

5.4.1 Presença de discurso de ódio nos grupos *Ultras* e *Hooligan*: problemas

Os grupos de adeptos de futebol *Ultras* e *Hooligan*, especialmente nas comunidades online, podem tornar-se ambientes onde a agressão, a hostilidade e o tribalismo florescem. Nestas comunidades, a expressão de lealdade para com a própria equipa conduz frequentemente a conflitos com grupos de adeptos adversários, com discussões que rapidamente se transformam em insultos ou ameaças e, posteriormente, culminam em confrontos físicos. Tendo em conta que estas interações nos canais de *Telegram* não são controladas, elas incluem muitas vezes, como vimos, sentimentos racistas, xenófobos ou homofóbicos, levando a um aumento da

violência verbal ou física entre grupos que defendem ideais conservadores e preconceituosos (grupos de extrema-direita) e grupos que têm como sua premissa o antirracismo e a defesa dos direitos humanos (grupos *Antifa*). Este estudo mostra que uma parte notável do discurso dos adeptos não é simplesmente por motivos de paixão pelo seu clube, mas abertamente hostil e com conteúdo bastante controverso, refletindo questões sociais mais amplas de intolerância e discurso de ódio.

No contexto da cultura futebolística, cada Grupo *Ultra* ou *Hooligan* representa uma identidade coletiva que tem uma influência muitas vezes profundamente enraizada no tecido social que rodeia os clubes e as comunidades de adeptos. Estes grupos, que são normalmente muito organizados e operam com um sentido de solidariedade, têm um grande impacto social, tanto na arena desportiva como fora dela, especialmente quando transmitem sentimentos de racismo, xenofobia ou HS.

Os *Ultras* e os *Hooligans* desempenham um papel fundamental e preponderante na atmosfera e no ambiente que é vivido no exterior e interior dos estádios, servindo frequentemente de porta-voz da comunidade de adeptos, sendo que a identidade coletiva referida em cima amplifica a sua influência, fazendo com que as suas mensagens - positivas ou negativas - tenham impacto num panorama mais vasto, isto é, o seu impacto não é unicamente no mundo do futebol. Quando estes grupos se mostram como apoiantes de um HS discriminatório, as suas palavras e ações podem legitimar e difundir atitudes discriminatórias., normalizando e alimentando o preconceito, não só no seio da comunidade futebolística, mas também na sociedade em geral.

Os resultados deste conjunto de dados revelam uma presença acentuada de ódio identitário (13,54% dos comentários tóxicos) e insultos (29,26%), que, como foi possível observar nos exemplos apresentados, incluem expressões de hostilidade contra identidades ou grupos específicos, algo que nas mãos de grupos coletivos como os *Ultras* ou *Hooligans*, ganha um maior peso, pois reforça em grande escala mentalidades de exclusão e encoraja outros a adotar atitudes semelhantes. Sendo o futebol um desporto internacional, mundialmente conhecido, o mais preocupante neste tipo de influência passa no impacto que tem nos jovens (os adultos do futuro), pois são estes quem mais se deslumbra com o tipo de cultura *Ultra* ou *Hooligan* e, a partir do momento que admira um destes grupos, também sente que deve admirar os ideais que estes defendem, ganhando desde jovens também o sentimento de ódio por outras pessoas

baseado em características discriminatórias e sentindo-se realizados ao imitarem e alinharem nestas atitudes demonstradas.

5.4.2 Liberdade de Expressão vs. Limites da Liberdade de Expressão

O equilíbrio entre a liberdade de expressão e a moderação de conteúdos nocivos é uma questão fundamental na regulamentação de redes sociais como o *Telegram*. Embora as plataformas visem incentivar a comunicação aberta, há um reconhecimento crescente de que certos tipos de discurso – especialmente os que incitam à violência ou promovem a discriminação, frequentemente desumanizando outros – podem ter consequências graves no mundo real.

A liberdade de expressão, um direito humano fundamental, permite que os indivíduos expressem opiniões e ideias sem receio de censura, mas isto é algo que não deve ser tido como uma verdade absoluta. As redes sociais têm uma enorme responsabilidade no que conta a estabelecer a fronteira entre o que constitui um discurso digno de estar publicado e o que é um discurso que deve ser moderado ou mesmo removido. No contexto do presente estudo, os comentários que foram classificados como *Identity Hate*, *Threat* ou *Insult* (os comentários obscenos são mais intervenções gerais, mais "normais" num ambiente de estádio) ultrapassam a fronteira do território nocivo, degradando a qualidade do discurso que é partilhado e tendo repercussões que causam os danos no mundo real: as atitudes discriminatórias tornam-se cada vez mais toleráveis, estando mais á vista de pessoas e, portanto, ganhando cada vez mais seguidores destes ideais, encorajando comportamentos violentos discriminatórios contra indivíduos ou grupos de indivíduos.

Embora a linguagem obscena possa, por vezes, ser defendida como parte da cultura dos adeptos de futebol, há um forte argumento de que qualquer tipo de *Hate Speech*, especialmente quando dirigido a alguém, não tem lugar no discurso público. Os moderadores destes canais devem ter a responsabilidade de avaliar o conteúdo que é partilhado e encontrar um equilíbrio entre a normalização que permite a expressão apaixonada (mesmo que isso implique que se permita conteúdo considerado agressivo) e a prevenção da normalização do *Hate Speech* discriminatório.

O alto número de comentários tóxicos neste conjunto de dados destaca a necessidade de melhores mecanismos de moderação nos grupos do *Telegram*, especialmente aqueles com um grande número de seguidores. Embora a censura completa do discurso negativo não seja prática nem desejável, ferramentas de moderação automatizadas, como o modelo *Toxic BERT* usado neste estudo, podem ajudar as plataformas a identificar e abordar o conteúdo mais prejudicial com mais eficiência.

Em suma, o objetivo não deve ser silenciar toda a dissidência ou emoção, mas sim garantir que as redes sociais continuem a ser um espaço seguro onde os utilizadores possam participar num diálogo sem medo de serem visados pela sua etnia, nacionalidade, religião, orientação sexual ou outro tipo de identidade. Plataformas como o *Telegram* devem evoluir continuamente as suas políticas de moderação de conteúdo para refletir esse equilíbrio, garantindo que a liberdade de expressão seja mantida e, ao mesmo tempo, protegendo os indivíduos contra danos.

6. Conclusão

A ampliação das redes sociais como veículo de expressão de ideias e opiniões, aumenta a quantidade de informação que diariamente se produz e que está disponível para as pessoas.

Os adeptos de futebol, em particular os *Ultras* e *Hooligans*, usam as redes sociais e estes canais de Telegram como uma forma de interagirem com pessoas com os mesmos interesses e, assim, manifestarem as suas opiniões, raivas, descontentamentos, rivalidades ou até partilha de emoções positivas. Nos *posts* nas redes sociais, em particular no canal *GruppaOF* no *Telegram*, não é invulgar encontrar discurso de ódio, com elementos de forte violência, nomeadamente contra indivíduos ou grupos rivais, adquirindo frequentemente carácter discriminatório e racista.

Ao analisar o sentimento dos comentários do *Telegram* entre os adeptos de futebol, a utilização do modelo *RoBERTa* revelou-se bastante eficaz. Este modelo categorizou de forma fiável os comentários em classes negativas, neutras e positivas, o que permitiu obter informações sobre o tom geral e as interações no seio do canal de *Telegram GruppaOF*. Dado que as interações nestes espaços de adeptos são frequentemente pesadas, com uma tendência para uma linguagem de confronto, esta Análise de Sentimentos ofereceu uma forma simplificada de avaliar as tendências gerais de sentimentos.

A elevada presença de sentimentos negativos revelou a intensidade das emoções e das opiniões fortes partilhadas pelos utilizadores, muitas vezes com tendência para a hostilidade, rivalidades, provocações e discriminações. No entanto, os comentários positivos e neutros também serviram como contraste, refletindo que, mesmo nestes espaços, existe por vezes um sentimento geral de pertença a uma identidade comum, apreciado por todos, e um equilíbrio ocasional nas interações. O desempenho do *RoBERTa* sublinha a importância da análise de sentimentos como uma ferramenta viável para acompanhar a dinâmica emocional nos espaços das redes sociais, onde a paixão e a raiva influenciam as interações.

Os grupos *Ultra* e *Hooligan*, com a sua forte presença tanto nos estádios como em comunidades online como o *Telegram*, moldam significativamente a cultura dos adeptos e influenciam as atitudes da sociedade. Os cânticos e os comportamentos discriminatórios - quer tenham origem no racismo, na homofobia ou noutros preconceitos - tornam-se normalizados

através das suas ações, afetando os adeptos, especialmente os mais jovens, e perpetuando atitudes prejudiciais na cultura dos mesmos.

Nos grupos de *Telegram* e nas restantes redes sociais, onde as discussões podem intensificar-se rapidamente, o discurso de ódio não controlado aumenta os sentimentos de divisão e reforça os preconceitos. Este ambiente online, combinado com as manifestações no estádio, acarreta uma profunda responsabilidade: tanto os *Ultras* como os *Hooligans* têm o poder de promover tanto ideologias nocivas como de respeito e inclusão. Abraçar esta última opção ajuda a alinhar a cultura dos adeptos com os valores do desporto, neste caso o futebol, criando um espaço onde os adeptos de todas as origens podem sentir-se bem-vindos e seguros, tanto nas bancadas como online.

O anonimato no Telegram permite que os utilizadores expressem opiniões e sentimentos sem revelar a sua identidade, diminuindo muitas vezes o risco de repercussões sociais ou legais. Este anonimato pode encorajar os utilizadores a partilhar declarações ou ideias que, de outra forma, poderiam manter privadas, incluindo as racistas, xenófobas ou homofóbicas. Neste ambiente, a linguagem discriminatória pode circular livremente, normalizando e potencialmente amplificando o discurso de ódio entre indivíduos que partilham a mesma opinião. Para os membros de grupos *Ultra* ou *Hooligan*, esta liberdade permite que a retórica negativa se espalhe mais facilmente, contribuindo para uma atmosfera tóxica tanto online como nos estádios.

Em relação à questão de investigação sobre qual modelo de *Machine Learning* é mais adequado para analisar os comentários de adeptos de futebol numa rede social, o modelo *Toxic BERT* mostrou-se altamente eficaz. Este modelo classificou com êxito vários tipos de toxicidade, dentro de um contexto de conversa (por comentários) feitos por adeptos (principalmente adeptos pertencentes a grupos *Ultra* ou *Hooligan*) de futebol. Dada a linguagem e a intensidade únicas encontradas nos comentários dos adeptos, o *Toxic BERT* foi capaz de analisar e categorizar variadas expressões com linguagem hostil ou discriminatória com uma precisão considerável.

Com a utilização do *Toxic BERT*, a análise destacou efetivamente os tipos (distinguindo-os) e a prevalência de linguagem tóxica nas redes sociais, oferecendo informações valiosas sobre a natureza das discussões nestes espaços e sendo fundamental para compreender a dinâmica das interações entre os adeptos de futebol. O desempenho do *Toxic BERT* neste

contexto sublinha a sua relevância e adaptabilidade para análises semelhantes, tornando-o uma excelente escolha para avaliar a toxicidade em conteúdos gerados por adeptos dentro de redes sociais.

Em suma, a análise feita nesta dissertação, através de várias técnicas, mas principalmente com a aplicação do modelo Toxic BERT, demonstra que existe efetivamente um elevado número de interações baseadas no Discurso de Ódio discriminatório, nomeadamente racismo, xenofobia e homofobia entre adeptos de futebol nas redes sociais, em particular no Telegram. O estudo comprova a necessidade equilibrar cuidadosamente a liberdade de expressão com a necessidade de moderação, especialmente nestes espaços onde o anonimato permite que todos os tipos de intervenções sejam feitos por parte dos utilizadores sem medo de repercussões. Embora a liberdade de expressão seja algo essencial e vista como dado adquirido na nossa sociedade, é fundamental estabelecer limites para evitar que o discurso de ódio se torne em algo ainda mais comum e se espalhe sem controlo pela sociedade. A introdução de políticas de moderação e regulação responsáveis, tanto no Telegram como em todas as plataformas de redes sociais, pode ajudar a promover comunidades online mais saudáveis e respeitadoras, replicando-se para a sociedade na vida real.

Como sugestões para investigações futuras, poderá ser interessante uma análise comparativa do Discurso de Ódio entre as algumas plataformas de redes sociais, como o Twitter, Facebook ou YouTube – estudar padrões, semelhanças ou diferenças no Discurso de Ódio observado em cada uma. Poderá ser também interessante o estudo do impacto das técnicas de moderação e regulação no discurso de ódio, investigando de que forma as diferentes abordagens de moderação (por exemplo, deteção automática, denúncia de utilizadores ou políticas específicas da plataforma) têm impacto na prevalência e no tipo de discurso de ódio nas plataformas utilizadas pelas comunidades de adeptos de futebol.

Referências Bibliográficas

Almeida, Pedro, Pereira, Janainna, & Candido, Diego. (2023). Online hate speech on social media in Portugal: extremism or structural racism? Social Identities, 29(5), 419-435. DOI: 10.1080/13504630.2024.2324277

Alves, Rita. (2021). Quando Ninguém Podia Ficar: Racismo, Habitação e Território. Lisboa: Tigre de Papel.

Araújo, Marta. (2016). 'A very 'prudent integration': white flight, school segregation and the depoliticization of (anti) racism'. Race Ethnicity and Education, 19(2), 300-323.

Bazel, Matthew. (2011). Theatre of Silence: the lost soul of football. Cambridge, Pegasus.

Bray, M. (2017). Antifa: The anti-fascist handbook. Melville House.

Bromberger, Christian. (1995). Le Match de Football: Ethnologie d'une Passion Partisanne à Marseille, Naples et Turin. Paris: Édition de la Maison des Sciences de l'Homme.

Cleland, J. (2014). Racism, football fans, and online message boards: How social media has added a new dimension to racist discourse in English football. Journal of Sport and Social Issues, 38(5), 415-431.

Doidge, M., Kossakowski, R., & Mintert, S. (2020). Ultras: The passion and performance of contemporary football fandom. In Ultras. Manchester University Press.

Goal. (n.d.). What is a football ultra? Serie A hardcore fan culture explained. Goal.com. Retrieved October 28, 2024, from https://www.goal.com/en/news/what-is-a-football-ultra-serie-a-hardcore-fan-culture-explained/aohlkilvcywp1v3c8e1f1a37w

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, 1(1), 60-76.

Hietanen, M., & Eddebo, J. (2023). Towards a Definition of Hate Speech—With a Focus on Online Contexts. Journal of Communication Inquiry, 47(4), 440-458. https://doi.org/10.1177/01968599221124309

IBM. (n.d.). What is machine learning (ML)? IBM. Retrieved October 27, 2024, from https://www.ibm.com/topics/machine-learning

IBM. (n.d.). What is natural language processing (NLP)? IBM. Retrieved October 27, 2024, from https://www.ibm.com/topics/natural-language-processing

IBM. (n.d.). What is text mining? IBM. Retrieved October 27, 2024, from https://www.ibm.com/topics/text-mining

Jahan, M. S., & Oussalah, M. (2023). A systematic review of hate speech automatic detection using natural language processing. Neurocomputing, 546, 126232.

Kearns, C., Sinclair, G., Black, J., Doidge, M., Fletcher, T., Kilvington, D., ... & Rosati, P. (2023). A scoping review of research on online hate and sport. Communication & Sport, 11(2), 402-430.

Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). An interpretation of lemmatization and stemming in natural language processing. Journal of University of Shanghai for Science and Technology, 22(10), 350-357.

Kossakowski, R., Nosal, P., & Woźniak, W. (2020). Politics, ideology and football fandom: The transformation of modern Poland. Routledge.

Liu, Y. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Luo, H., Cai, M., & Cui, Y. (2021). Spread of misinformation in social networks: Analysis based on Weibo tweets. Security and Communication Networks, 2021(1), 7999760.

Lynch, Danny. (2009). 'Kick It Out'. In Christos Kassimeris (ed.), Anti-Racism in European Football: Fair Play for All. Plymouth: Lexington Books, 67-104.

Magrath, R. (2017). The intersection of race, religion and homophobia in British football. International Review for the Sociology of Sport, 52(4), 411-429.
Maeso, Silvia. (2021). O Estado do Racismo em Portugal: racismo antinegro e anticiganismo no direito e nas políticas públicas. Lisboa: Tinta da China.
Malta, Margarida. (2022). Media Narratives of Hate Speech and Crimes in Portugal. Tese de Mestrado em Estudos Internacionais.
Marivoet, S. (2009). Subculturas de adeptos de futebol e hostilidades violentas–O caso português no contexto europeu. Configurações. Revista Ciências Sociais, (5/6), 279-299.
Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. Television & new media, 22(2), 205-224.
Miranda, S., Gouveia, C., Di Fátima, B., & Antunes, A. C. (2024). Hate speech on social media: behaviour of Portuguese football fans on Facebook. Soccer & Society, 25(1), 76-91.
Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.
Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016, April). Abusive language detection in online user content. In Proceedings of the 25th international conference on world wide web (pp. 145-153).
Nockleby, J. T. (1994). Hate speech in context: The case of verbal threats. Buff. L. Rev., 42, 653.
Pedro Almeida, Janainna Pereira & Diego Candido (2023) Online hate speech on social media in Portugal: extremism or structural racism?, Social Identities, 29:5, 419-435, DOI: 10.1080/13504630.2024.2324277
Pookpanich, P., & Siriborvornratanakul, T. (2024). Offensive language and hate speech detection using deep learning in football news live streaming chat on YouTube in Thailand. Social Network Analysis and Mining, 14(1), 18.
Rini, R., Utami, E., & Hartanto, A. D. (2020, October). Systematic literature review of hate speech detection with text mining. In 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1-6). IEEE.
Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6). IEEE.
Spaaij, R. (2007). Football hooliganism as a transnational phenomenon: Past and present analysis: A critique – More specificity and less generality. The International Journal of the History of Sport, 24(4), 411–431. https://doi.org/10.1080/09523360601157156
Sportskeeda. (n.d.). Modern football: It's all about the money. Sportskeeda.com. Retrieved October 28, 2024, from https://www.sportskeeda.com/football/modern-football-its-all-about-the-money
Taylor, Ian. (1982). 'On the Sports Violence Question: Soccer Hooliganism Revisited'. In Jennifer Hargreaves (ed.), Sport, Culture and Ideology. London: Routledge, 152-196.
Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. PLoS one, 13(9), e0203958.
UEFA. (2019). UEFA safety and security regulations. Union of European Football Associations.
Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. Knowledge and Information Systems, 60, 617-663.

Anexo A – Referências Revisão de Literatura

Armenia, S., Papathanasiou, J., Pompei, A., & Tsaples, G. (2020). A Systemic Perspective on Racism in Football: The Experience of the BRISWA Project. PuntOorg International Journal, 4(2), 56–70. https://doi.org/10.19245/25.05.pij.4.2.1

Dinesh Jackson, S. R., Fenil, E., Gunasekaran, M., Vivekananda, G. N., Thanjaivadivel, T., Jeeva, S., & Ahilan, A. (2019). Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM. Computer Networks, 151, 191-200. doi:10.1016/j.comnet.2019.01.028

Newman, J. A., Warburton, V. E., & Russell, K. (2021). Conceptualizing bullying in adult professional football: A phenomenological exploration. Psychology of Sport and Exercise, 54 doi:10.1016/j.psychsport.2021.101883

Business Bliss Consultants FZE. (November 2018). Racism in Football. Retrieved from https://ukdiss.com/examples/racial-bias-premier-league-9863.php?vref=1

U.I.S.P Unione Italiana Sport per Tutti; Comitato Regionale Emili-Romagna, Balestri, C. B., Grasselli, G. G., Dembowski, G. D., & Diener, S. D. (2002, April). Racism, Football and the Internet. European Union Agengy for Fundamental Rights. https://fra.europa.eu/en/publication/2002/racism-football-and-internet-thematic-report#publication-tab-1

Cleland, J. (2014). Racism, football fans, and online message boards: How social media has added a new dimension to racist discourse in english football. Journal of Sport and Social Issues, 38(5), 415-431. doi:10.1177/0193723513499922

Newman, J. A., Warburton, V. E., & Russell, K. (2022). Whistleblowing of bullying in professional football: To report or not to report? Psychology of Sport and Exercise, 61 doi:10.1016/j.psychsport.2022.102177

Garå, J. (2018, November 28). International fan survey report on racism in football. Forza Football. https://blog.forzafootball.com/kick-it-out/

Manzini, T., Lim, Y. C., Tsvetkov, Y., & Black, A. W. (2019). Black is to criminal as caucasian is to police: Detecting and removing multiclass bias in word embeddings. Paper presented at the NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, , 1 615-621.

Fighting discrimination in sport | Think Tank | European Parliament. (n.d.). https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)696163

Srato, D., Goularte, F. B., & Fileto, R. (2020). Short semantic patterns: A linguistic pattern mining approach for content analysis applied to hate speech. International Journal on Artificial Intelligence Tools, 29(2) doi:10.1142/S0218213020400023

Almeida, P. A. (2018). FUTEBOL RAÇA E NAÇÃO EM PORTUGAL [PhD Thesis]. Universidade de Coimbra.

Alotaibi, A., & Abul Hasanat, M. H. (2020). Racism detection in twitter using deep learning and text mining techniques for the arabic language. Paper presented at the Proceedings - 2020 1st International Conference of Smart Systems and Emerging Technologies, SMART-TECH 2020, 161-164. doi:10.1109/SMART-TECH49988.2020.00047

Alhejaili, R. M., Yafooz, W. M. S., & Alsaeedi, A. A. (2022). Hate speech and abusive laungage detection in twitter and challenges: Review. Paper presented at the Proceedings of International Conference on Computational Intelligence and Sustainable Engineering Solution, CISES 2022, 86-94. doi:10.1109/CISES54857.2022.9844317

Duwairi, R., Hayajneh, A., & Quwaider, M. (2021). A deep learning framework for automatic detection of hate speech embedded in arabic tweets. Arabian Journal for Science and Engineering, 46(4), 4001-4014. doi:10.1007/s13369-021-05383-3

Digital, C. (n.d.). Church Court Chambers | RACISM AND FOOTBALL – WHAT ARE THE POSSIBLE SOLUTIONS? – AN ARTICLE BY YASIN PATEL. https://churchcourtchambers.co.uk/article/racism-and-football-what-are-the-possible-solutions-an-article-by-yasin-patel/

Waitzman, E. M. (2020, January 14). Racism in Football: Tackling Abusive Behaviour. House of Lords Library. https://lordslibrary.parliament.uk/research-briefings/lln-2020-0012/

Balakrishnan, V., Ng, K. S., & Arabnia, H. R. (2022). Unravelling social media racial discriminations through a semi-supervised approach. Telematics and Informatics, 67 doi:10.1016/j.tele.2021.101752

Lee, C. S., & Jang, A. (2021). Questing for justice on twitter: Topic modeling of #StopAsianHate discourses in the wake of atlanta shooting. Crime and Delinquency, doi:10.1177/00111287211057855

Field, A., Blodgett, S. L., Waseem, Z., & Tsvetkov, Y. (2021). A survey of race, racism, and anti-racism in NLP. Paper presented at the ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 1905-1925.

Kang, Y., & Yang, K. C. C. (2022). Communicating racism and xenophobia in the era of donald trump: A computational framing analysis of the US-mexico cross-border wall discourses. Howard Journal of Communications, 33(2), 140-159. doi:10.1080/10646175.2021.1996491

Benítez-Andrades, J. A., González-Jiménez, Á., López-Brea, Á., Aveleira-Mata, J., Alija-Pérez, J. -., & García-Ordás, M. T. (2022). Detecting racism and xenophobia using deep learning models on twitter data: CNN, LSTM and BERT. PeerJ Computer Science, 8 doi:10.7717/PEERJ-CS.906

Al-Hassan, A., & Al-Dossari, H. (2019). DETECTION OF HATE SPEECH IN SOCIAL NETWORKS: A SURVEY ON MULTILINGUAL CORPUS. Computer Science &Amp; Information Technology(CS &Amp; IT). https://doi.org/10.5121/csit.2019.90208