



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE SOBRAL
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

ANTONIO ERALDO CAETANO MARTINS

**ANÁLISE DA PERCEPÇÃO SOBRE O VIÉS REPRODUZIDO PELOS ALGORITMOS
DE RECONHECIMENTO FACIAL**

SOBRAL

2023

ANTONIO ERALDO CAETANO MARTINS

ANÁLISE DA PERCEPÇÃO SOBRE O VIÉS REPRODUZIDO PELOS ALGORITMOS DE
RECONHECIMENTO FACIAL

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do *Campus* Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Iális Cavalcante de Paula Júnior.

SOBRAL

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

M341a Martins, Antonio Eraldo Caetano.
Análise da percepção sobre o viés reproduzido pelos algoritmos de reconhecimento facial / Antonio Eraldo Caetano Martins. – 2023.
46 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Sobral, Curso de Engenharia da Computação, Sobral, 2023.
Orientação: Prof. Dr. Iális Cavalcante de Paula Júnior.

1. Reconhecimento Facial. 2. Percepção. 3. Análise de Sentimentos. 4. VADER. I. Título.
CDD 621.39

ANTONIO ERALDO CAETANO MARTINS

ANÁLISE DA PERCEPÇÃO SOBRE O VIÉS REPRODUZIDO PELOS ALGORITMOS DE
RECONHECIMENTO FACIAL

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do *Campus* Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Aprovada em: 14 de Dezembro de 2023

BANCA EXAMINADORA

Prof. Dr. Iális Cavalcante de Paula
Júnior (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Antonio Josefran de Oliveira Bastos
Universidade Federal do Ceará (UFC)

Prof. Dr. Wendley Souza da Silva
Universidade Federal do Ceará (UFC)

À minha mãe, Donatila, por tudo o que fez e faz
por mim e pela nossa família.

AGRADECIMENTOS

À minha mãe, Donatila, por sempre fazer o melhor pelos seus filhos e estar aqui para me ajudar, possibilitando cada uma das minhas conquistas. Ao meu pai, Wilson, por sempre estar presente e fornecer o suporte necessário. Sem vocês dois e a fé que têm na educação, eu não conseguiria chegar onde cheguei.

Aos meus irmãos, Rejane, Resende, Elane, Eliana e Erlandia, que prepararam o caminho que me conduziu até aqui e que, de maneira direta ou indireta, sempre me incentivaram a estudar e buscar o melhor.

A todos os meus amigos, em especial Clarisse Mesquita, Jannielly Garcia, Laura Karine Melo e Lucas Sena, que sempre estiveram presentes e me apoiaram nesta caminhada.

Aos amigos e colegas de moradia, Alarilson Mesquita, Brenno Azevedo, Epitácio Fernandes e Wesley Mesquita, com quem pude conversar, compartilhar vivências e risadas, tornando a reta final do curso bem menos árdua.

Aos amigos que a Universidade me proporcionou, em especial João Gabriel Fernandes e Renato Freire, que me ajudaram de várias maneiras, seja nos estudos ou apenas para conversar e nos distrair um pouco da rotina muitas vezes exaustiva.

À Universidade Federal do Ceará, que forneceu apoio financeiro com auxílio moradia e ofereceu uma educação de excelência.

Ao Prof. Dr. Iális Cavalcante de Paula Júnior, pela excelente orientação neste trabalho de conclusão.

Aos demais professores e servidores do *Campus* da UFC em Sobral, que proporcionaram excelentes ensinamentos e sempre estiveram dispostos a ajudar.

“Não há tempo que volte amor, vamos viver tudo
o que há pra viver.”
(Lulu Santos, 1982.)

RESUMO

O avanço do desenvolvimento de algoritmos possibilita a execução de tarefas principalmente realizadas por humanos facilmente reproduzidas por sistemas de Inteligência Artificial (IA). Os sistemas de reconhecimento facial, por exemplo, são adotados por empresas e organizações para a realização de estudos com o propósito de automatizar tarefas. Nesse processo, acontecem problemas referentes aos resultados apresentados por esses sistemas. Esses problemas podem ser causados pelo viés dos algoritmos. Diante disso, este trabalho realiza uma análise da percepção sobre o viés, nos comentários extraídos em duas bases de dados da rede de mídias sociais *Reddit*, onde os usuários comentaram postagens que envolviam reconhecimento facial como assunto principal. A metodologia foi composta pela extração e análise de sentimentos dos comentários, utilizando o analisador de sentimentos VADER da biblioteca *Python Natural Language Toolkit* (NLTK), e pela obtenção dos termos mais frequentes, para posterior análise. Os sentimentos positivos e negativos mostraram-se em quantidades semelhantes e a obtenção de expressões mais citadas mostrou a relação entre violação de direitos e os sentimentos encontrados.

Palavras-chave: reconhecimento facial; percepção; análise de sentimentos; VADER.

ABSTRACT

The advancement in algorithm development enables the execution of tasks primarily performed by humans, easily replicable by Artificial Intelligence (AI) systems. Facial recognition systems, for instance, are adopted by companies and organizations to conduct studies with the purpose of automating tasks. In this process, problems arise concerning the results produced by these systems. These problems can be caused by algorithmic bias. In light of this, this study conducts an analysis of the perception of bias in the comments extracted from two databases on the social media network Reddit, where users commented on posts with facial recognition as the main topic. The methodology consisted of extracting and analyzing sentiments from the comments using the VADER sentiment analyzer from the Python Natural Language Toolkit (NLTK) and obtaining the most frequent terms for subsequent analysis. Positive and negative sentiments were found in similar quantities, and the extraction of the most mentioned expressions revealed the relationship between rights violation and the sentiments identified.

Keywords: facial recognition; perception; sentiment analysis; VADER.

LISTA DE FIGURAS

Figura 1 – Representação de biometria facial	17
Figura 2 – Exemplo de análise de sentimentos	20
Figura 3 – Diagrama da metodologia do trabalho	24
Figura 4 – Fragmento de <i>DataFrame</i> exemplo obtido de um <i>subreddit</i>	26
Figura 5 – Análise de sentimentos do fragmento da Figura 4	27
Figura 6 – Resultado do processo de <i>web scraping</i> em <i>r/technology</i>	29
Figura 7 – Resultado do processo de <i>web scraping</i> em <i>r/worldnews</i>	30
Figura 8 – Gráfico relacionando o valor de <i>compound</i> pelo tempo decorrido em <i>r/technology</i>	31
Figura 9 – Gráfico relacionando o valor de <i>compound</i> pelo tempo decorrido em <i>r/worldnews</i>	33
Figura 10 – Gráfico apresentando a quantidade de comentários por sentimento em <i>r/technology</i>	34
Figura 11 – Gráfico apresentando a quantidade de comentários por sentimento em <i>r/worldnews</i>	34
Figura 12 – Gráfico relacionando a quantidade de comentários por sentimento em <i>r/technology</i> após a subdivisão dos sentimentos	35
Figura 13 – Gráfico relacionando a quantidade de comentários por sentimento em <i>r/worldnews</i> após a subdivisão dos sentimentos	35
Figura 14 – Nuvem de palavras ilustrando os termos com base na frequência em <i>r/technology</i>	37
Figura 15 – Nuvem de palavras ilustrando os termos com base na frequência em <i>r/worldnews</i>	38

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas para <i>r/technology</i>	30
Tabela 2 – Estatísticas descritivas para <i>r/worldnews</i>	32
Tabela 3 – Expressões de duas palavras com mais frequência em <i>r/technology</i>	37
Tabela 4 – Expressões de duas palavras com mais frequência em <i>r/worldnews</i>	38

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
EUA	Estados Unidos da América
IA	Inteligência Artificial
NLTK	<i>Natural Language Toolkit</i>
PLN	Processamento de Linguagem Natural
PRAW	<i>Python Reddit API Wrapper</i>
RNA	Redes Neurais Artificiais
VADER	<i>Valence Aware Dictionary and sEntiment Reasoner</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Justificativa	14
1.2	Objetivos	15
1.2.1	<i>Objetivo Geral</i>	15
1.2.2	<i>Objetivos Específicos</i>	15
1.3	Organização do trabalho	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Inteligência Artificial (IA)	16
2.1.1	<i>Reconhecimento facial</i>	16
2.2	Viés Algorítmico	17
2.3	<i>Web Scraping</i>	18
2.4	<i>Reddit</i>	18
2.4.1	<i>Python Reddit API Wrapper (PRAW)</i>	19
2.5	Processamento de Linguagem Natural (PLN)	19
2.5.1	<i>Análise de Sentimentos</i>	20
2.5.1.1	<i>Valence Aware Dictionary and sEntiment Reasoner (VADER)</i>	21
3	REVISÃO BIBLIOGRÁFICA	22
4	METODOLOGIA	24
4.1	Base de dados e utilização da API	24
4.2	<i>Subreddits utilizados e parâmetros</i>	25
4.3	Dados a serem coletados	25
4.4	Análise do sentimento dos dados coletados	26
4.4.1	<i>Análise descritiva e temporal dos dados após a análise de sentimentos</i>	27
4.5	Categorização dos sentimentos extraídos	27
4.6	Obtenção de termos mais frequentes	28
5	RESULTADOS E DISCUSSÃO	29
5.1	Análise descritiva dos resultados	29
5.2	Resultado da análise de sentimentos	33
5.3	Análise dos termos frequentes	36
6	CONCLUSÃO	40

REFERÊNCIAS 42

1 INTRODUÇÃO

As decisões antes executadas por humanos, desde as triviais até as mais críticas, estão sendo cada vez mais realizadas por algoritmos, impulsionados pelo avanço da Inteligência Artificial (IA) (KHALIL *et al.*, 2020). Nesse contexto, a presença de sistemas que empregam algoritmos de IA para o reconhecimento e identificação facial é frequentemente observada. Esses sistemas são adotados por empresas e organizações para conduzir estudos nos quais os seres humanos são o foco da análise. Esse uso tem como objetivo automatizar processos seletivos, categorizar perfis sociais, identificar suspeitos, auxiliar em contratações, realizar análises de perfil de consumidores e de renda, apoiar a aplicação da lei, entre outros.

Não obstante, um ponto importante a se dar atenção é justamente aos problemas observados nos resultados fornecidos por esses algoritmos no que se refere à generalização e conformidade com a realidade. Esses problemas são muitas vezes ocasionados pelo viés do algoritmo. Esses vieses são conhecidos também como erros sistemáticos, e se repetem de forma previsível em circunstâncias particulares (KAHNEMAN, 2011). Um caso recente de problema gerado pelo viés do algoritmo aconteceu quando uma mulher com traços de herança asiática utilizou o aplicativo Lensa. Esse aplicativo utiliza IA para gerar imagens de pessoas com base em suas *selfies*. No caso citado, o software forneceu imagens sexualizadas e sensuais, enquanto aos colegas homens da mulher forneceu imagens de aspecto heroico e profissional (HEIKKILÄ, 2022).

Nesse aspecto, a IA é construída utilizando-se da inteligência e das habilidades cognitivas humanas como referência para a inteligência da máquina. Assim os modelos de aprendizado de máquina são treinados usando os conjuntos de dados que representam o comportamento humano. Para isso, os dados para treinamento fornecidos aos modelos são considerados como uma verdade inicial, permitindo que o modelo aprenda o comportamento desses dados. No entanto, considerando o viés da natureza humana e do pensamento cognitivo, os dados de treinamento também refletem essas características tendenciosas, o que acaba afetando o aprendizado do modelo ao induzir o viés (ALELYANI, 2021). O viés cognitivo mencionado anteriormente, diz respeito a um erro de raciocínio resultante de estratégias mentais de simplificação que são desenvolvidas na tentativa de processar informações (MACHADO, 2018).

Com base na discussão acima, este trabalho planeja fazer um estudo descritivo sobre a percepção das pessoas sobre o viés reproduzido pelos algoritmos de IA que são frequentemente usados em maior escala para reconhecimento facial. Para isso, será feita a coleta de dados

com *web scraping*, utilizando-se da *Application Programming Interface* (API) da plataforma de mídia social e agregador de conteúdo *Reddit*. Essa mineração de dados objetiva obter relatos, considerações e conclusões sobre casos de viés discriminatório acerca de postagens realizadas na referida rede de mídias sociais. Para tal fim, são utilizadas técnicas de Processamento de Linguagem Natural (PLN), mais especificamente com a ferramenta de análise de sentimentos *Valence Aware Dictionary and sEntiment Reasoner* (VADER), sobre as postagens obtidas no processo de *web scraping*. Além disso, é realizada a obtenção e análise das expressões mais recorrentes nos dados coletados.

1.1 Justificativa

Procedimentos tecnológicos são amplamente usados para obtenção de dados mais confiáveis de populações e usuários. No âmbito empresarial, por exemplo, é de suma importância que existam ferramentas que auxiliem as organizações responsáveis a tomarem decisões. Para isso, metodologias de *Business Intelligence* (BI) quando combinadas a IA proporcionam uma análise bem mais aprofundada e adequada do objeto de estudo, ponto fundamental para o sucesso da empresa, pois qualquer má decisão poderá gerar consequências desastrosas (FONTINHA, 2014).

Além disso, com a evolução das ferramentas tecnológicas, a criação de uma IA vem também passível de erros que podem trazer prejuízos a quem a use. Baseadas primordialmente nos conceitos iniciais fornecidos pelos humanos que as criam, as IAs podem reproduzir comportamentos racistas, discriminação de gênero e preconceitos de classe, eventualmente também de forma combinada.

Essa problemática entra em desacordo com o artigo 2º da Declaração Universal dos Direitos Humanos, onde diz que se caracteriza como uma infração aos direitos humanos oferecer tratamento com distinção de qualquer espécie, seja de raça, cor, sexo, origem nacional ou social, língua ou riqueza (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 1948). Visto isso, analisar a percepção sobre os casos de discriminação dos algoritmos de IA pode evidenciar a insatisfação acerca dos sistemas utilizados e abrir caminhos para a minimização desses problemas.

1.2 Objetivos

1.2.1 *Objetivo Geral*

O objetivo deste trabalho é realizar uma análise sobre a percepção em relação ao viés reproduzido pelos algoritmos de IA utilizados para reconhecimento facial.

1.2.2 *Objetivos Específicos*

- Explorar as bibliografias relacionadas ao viés na implementação de algoritmos de IA;
- Coletar comentários de usuários do *Reddit* acerca de sistemas de reconhecimento facial por meio de *web scraping* utilizando a API da referida rede de mídias sociais;
- Utilizar técnicas de PLN para realizar análise de sentimentos sobre os comentários dos usuários;
- Conduzir uma investigação sobre a evolução temporal e características descritivas das postagens, bem como dos sentimentos expressos nos comentários;
- Identificar a relação entre os termos mais citados na busca e os sentimentos identificados pela análise;

1.3 Organização do trabalho

Este trabalho foi desenvolvido em 6 capítulos. No presente capítulo estão expostos a introdução, o objetivo geral, os objetivos específicos e a justificativa, que apresentam, fundamentam e sintetizam o que se pretende alcançar neste trabalho. O Capítulo 2 aborda os princípios teóricos essenciais que são fundamentais para uma compreensão mais profunda dos elementos do estudo. No Capítulo 3 são apresentados trabalhos relacionados ao contexto ao qual o trabalho aborda. A metodologia encontra-se no Capítulo 4, sendo dividida em cinco seções referentes à base de dados, aos parâmetros do processo de *web scraping*, aos dados coletados, à análise de sentimentos e à obtenção dos termos mais frequentes. Já no Capítulo 5, são apresentados os resultados obtidos, seguidos de discussões de acordo com a metodologia proposta. Por fim, são mostradas no Capítulo 6 as conclusões finais do estudo.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão apresentados os aspectos teóricos fundamentais ao desenvolvimento do estudo, sendo eles os conceitos de IA, reconhecimento facial, viés algorítmico, *web scraping*, análise de sentimentos por PLN, além de abordar o site *Reddit*.

2.1 Inteligência Artificial (IA)

A IA, surgida na década de 1950, é o processo que viabiliza uma máquina, através de algoritmos, possuir capacidade de realizar atividades humanas de maneira autônoma, o que inclui simular a inteligência humana, capacidade de raciocinar, tomar decisões e resolver problemas. O Projeto *Dartmouth Summer* foi o marco fundamental para o início do desenvolvimento das tecnologias que se têm hoje sobre o assunto (MCCARTHY *et al.*, 2006).

Dentre as tecnologias, há o *Machine Learning* (aprendizagem de máquina) que é uma subárea da IA. Nela os sistemas concebidos aprendem padrões contidos em um conjunto de dados com o qual são alimentados previamente, e fazem uso de tais padrões para chegar a resultados não explicitamente programados pelos desenvolvedores, projetando esses padrões em novas situações (ALPAYDIN, 2016).

Outra subárea da IA, também englobada pelo *Machine Learning*, é o *Deep Learning* (aprendizado profundo). Nele usam-se Redes Neurais Artificiais (RNA) de várias camadas para obter sucesso em tarefas como detecção de objetos, reconhecimento de fala, reconhecimento facial e tradução de idiomas, por exemplo. Essa subárea se diferencia das técnicas tradicionais de *Machine Learning*, pois essas técnicas podem aprender automaticamente a representar dados como imagens, vídeos ou textos, sem que sejam explicitadas regras ou conhecimentos definidos por humanos.

Além dessas, há a visão computacional que é um campo da IA que se refere à coleta, análise e reunião de dados visuais através de computadores, com objetivos diversos como a identificação de rostos e biometria, análise representativa de objetos, entidades e conceitos e contextos em imagens (SILVA, 2020).

2.1.1 Reconhecimento facial

O reconhecimento facial aqui citado refere-se ao processo de replicação pelas máquinas computacionais da capacidade natural humana de reconhecer rostos e obter informações

a partir disso. Expressões faciais podem transmitir informações sobre o estado emocional e cognitivo de uma pessoa, por exemplo (ORVALHO, 2019).

Com isso, reconhecimento facial pode ser compreendido como um exemplo de aplicação de IA e visão computacional. Os algoritmos de reconhecimento facial operam mediante o uso de biometria para mapear características faciais ou outros atributos físicos de uma pessoa que esteja presente em uma fotografia ou vídeo (NEGRI *et al.*, 2020), assim como representado na Figura 1. Este processo é realizado com a finalidade de cumprir um objetivo específico: captar a imagem de um rosto humano e estimar a porcentagem de semelhança entre esta imagem e a imagem de um banco de dados de referência, por exemplo (BRANDÃO; OLIVEIRA, 2021).

Figura 1 – Representação de biometria facial



Fonte: Tecnoblog (2021).

2.2 Viés Algorítmico

O viés algorítmico é a tendência apresentada pelo algoritmo em execução. Essa tendência pode fazer com que esse algoritmo tome decisões erradas, assim como possibilitam a ocorrência de discriminações e/ou exclusões indevidas dos objetos de estudo. Existem alguns fatores que contribuem para a criação de algoritmos com viés incorporado, tais como a seleção e o tratamento dos dados de entrada, por exemplo. O fato desses dados estarem incompletos ou inconformes com a realidade também pode gerar vieses indesejados. Outro fator pode ser

a escolha dos desenvolvedores, que pode ser feita sem critérios de diversidade, favorecendo tendências. Além disso, devido à capacidade das IAs de aprender com padrões, se um modelo estiver enviesado, os resultados subsequentes também estarão igualmente equivocados devido ao viés (CARVALHO, 2020).

2.3 *Web Scraping*

O *web scraping* é o processo de extrair dados da web de maneira sistemática. Nesse processo de extração de dados, um agente de software também conhecido como robô web, simula a interação de navegação entre os servidores web e os seres humanos em uma navegação web convencional. Esse robô pode acessar quantos sites forem necessários, analisando o seu conteúdo para encontrar os dados desejados (GLEZ-PEÑA *et al.*, 2013).

Desse modo, esse procedimento pode ser realizado utilizando a linguagem de programação *Python*, na qual a extração de dados poderá ser realizada com maior agilidade e facilidade. Além disso, combinando essa técnica com o uso de APIs, é possível obter a transação de dados utilizando o *web scraping* para extrair os dados desejados e as APIs para ter acesso direto a esses dados.

2.4 *Reddit*

Reddit é um site de mídias sociais e agregador de fóruns onde os usuários criam e compartilham conteúdo. O site possui comunidades chamadas *subreddits* para diferentes interesses onde qualquer usuário pode criá-las. Os *subreddits* são apresentados e normalmente acessados pelo formato *r/nome_do_subreddit*, onde “nome_do_subreddit” deve ser substituído pelo nome real do *subreddit* que se deseja navegar.

Os usuários podem postar imagens ou mensagens de texto de acordo com as diretrizes de conteúdo do *Reddit* e as regras individuais de cada *subreddit*. Eles também podem comentar nas postagens de outras pessoas, além de votar, funcionalidade na qual os usuários podem votar anonimamente de forma positiva ou negativa.

Ao acessar o *Reddit* os usuários podem optar por acessar comunidades específicas ou navegar na página inicial. A página inicial apresenta postagens de todas as comunidades seguidas pelo usuário. Os usuários também podem navegar no *r/all*, que agrega postagens populares de todos os *subreddits* (MATTERS, 2022).

2.4.1 *Python Reddit API Wrapper (PRAW)*

A *Python Reddit API Wrapper (PRAW)* é uma biblioteca que fornece acesso simplificado à API do *Reddit*. Antes de utilizá-la para realizar o acesso e buscas é necessário que seja criada uma conta do *Reddit* para posteriormente criar uma aplicação de desenvolvimento da API. Essa aplicação fornecerá três credenciais principais (*client id*, *client secret* e *user agent*) que são necessárias para obter acesso às informações desejadas. Para solicitações de apenas leitura, onde serão recuperadas apenas informações públicas, as credenciais acima citadas são suficientes. Para solicitações com mais permissões deve-se acrescentar o *username* e o *password*, que permitirão ações como votar, comentar, postar e editar postagens (GEEKS, 2020). Com isso, deve ser criado um objeto que receberá as credenciais e que será utilizado para cada solicitação à API, deve-se também definir o *subreddit* e os parâmetros da busca desejada.

2.5 **Processamento de Linguagem Natural (PLN)**

O processamento de linguagem natural (PLN) utiliza métodos computacionais para abordar aspectos da comunicação humana, como sons, palavras, sentenças e discursos. Ele analisa estruturas, significados, contextos e usos de palavras e frases. O objetivo principal do PLN é fazer o computador se comunicar em linguagem humana, bem como compreendê-la (GONZALEZ; LIMA, 2003). No que concerne aos métodos computacionais empregados na interpretação da linguagem natural, a seleção do método apropriado dependerá do tipo de dado sob análise. Assim, existem técnicas empregadas no processamento da linguagem falada e outras técnicas para o processamento da linguagem escrita. Na linguagem escrita há ainda a diferenciação das técnicas aplicadas nos diferentes tipos de textos como por exemplo os jornalísticos, científicos e literários (VIEIRA; LOPES, 2010).

A *Natural Language Toolkit (NLTK)* é uma biblioteca *Python* usada para construir programas de PLN. Ela oferece uma infraestrutura completa que inclui classes para representar dados, interfaces padrão para tarefas como tokenização, análise sintática e classificação de texto, além de implementações padrão para essas tarefas. Essas ferramentas podem ser combinadas para resolver problemas de PLN (BARBOSA *et al.*, 2017).

2.5.1 Análise de Sentimentos

A análise de sentimentos objetiva definir técnicas automatizadas que sejam capazes de extrair informações subjetivas de textos em linguagem natural, como opiniões e sentimentos, a fim de obter conclusões sobre esses dados. Com isso, essa análise pode permitir compreender e explicar fenômenos sociais complexos e é uma importante ferramenta de BI que pode ajudar empresas a melhorar seus produtos e serviços (BENEVENUTO *et al.*, 2015), assim como mostrado na Figura 2, onde são apresentados exemplos de classificação de sentimentos de acordo com os textos analisados.

Figura 2 – Exemplo de análise de sentimentos



Fonte: Binds.co (2023).

Essa análise aplica tecnologias de PLN que treinam softwares de computador para entender o texto de maneiras semelhantes aos humanos e passa por duas principais fases: pré-processamento e análise de palavras-chave. No pré-processamento, a análise de sentimentos identifica palavras-chave para destacar a mensagem central do texto. Na análise de palavras-chave, as tecnologias de PLN analisam de forma mais detalhada as palavras-chave extraídas e dão a elas uma pontuação, fornecendo assim uma percepção relativa da emoção expressa no texto para fins analíticos (AMAZON WEB SERVICES, INC., s.d.).

Além disso, alguns conceitos tornam-se importantes para a detecção de sentimentos, como a polaridade, que representa o grau de positividade e negatividade de um texto; força do

sentimento, que representa a intensidade de um sentimento; emoção, que indica um sentimento específico em um texto; e subjetividade e objetividade que podem determinar se uma mensagem tem mais sentenças subjetivas ou objetivas (BENEVENUTO *et al.*, 2015).

2.5.1.1 *Valence Aware Dictionary and sEntiment Reasoner (VADER)*

O VADER (HUTTO; GILBERT, 2014) é uma ferramenta para o *Python* desenvolvida com o auxílio da biblioteca NLTK para uso em processos de análise de sentimentos, principalmente para a análise de dados de mídias sociais. É baseada em regras e em um léxico usado para tratamento de sentimentos expressos nesses tipos de mídias. Ao fazer a análise de sentimentos com o VADER obtém-se uma saída com 4 valores referentes ao texto analisado: “*neg*”, indicando o grau de negatividade; “*neu*” indicando o grau de neutralidade; “*pos*” indicando o grau de positividade; e “*compound*” que representa uma média ponderada das pontuações de sentimento negativo, neutro e positivo. O “*compound*” fornece uma visão geral do sentimento do texto e permite sua interpretação. Ele oferece valores na faixa de -1 a 1 . Quanto mais próximo de -1 for o “*compound*”, mais negativo é o texto. Quanto mais próximo de 1 mais positivo é o texto, e se “*compound*” igual a 0 significa que o sentimento geral do texto é neutro (ELBAGIR; YANG, 2019).

O código de implementação do VADER pode ser consultado neste link.

3 REVISÃO BIBLIOGRÁFICA

Esta seção aborda trabalhos que forneceram embasamento teórico para o presente estudo.

Os autores (KHALIL *et al.*, 2020) apresentam uma revisão da literatura no contexto do viés de sistemas de análise facial. A revisão envolveu 24 estudos, visando identificar os vieses em algoritmos de análise facial e apresentar técnicas implementadas para mitigar esse viés. O trabalho conclui que os algoritmos de aprendizado de máquina e os dados que os alimentam são resultados de dados e cálculos fornecidos por humanos, não estando assim isentos de refletir vieses humanos. O estudo identificou e destacou o viés existente em bancos de dados públicos por meio de uma revisão sistemática e abordou métodos científicos para entender e eliminar o viés.

Os autores (MEHRABI *et al.*, 2021) apresentam duas dimensões principais que afetam sistemas de IA em termos de viés e injustiça, são eles os dados e os algoritmos. Os autores destacam que problemas de injustiça podem causar danos potenciais nas aplicações em sistemas judiciais e de reconhecimento facial. Por fim, concluem que, com a expansão do uso de IA, é importante que os pesquisadores concentrem seus esforços e aprofundem suas pesquisas na área do viés em algoritmos.

Com uma abordagem sobre o Brasil, os autores (COSTA; KREMER, 2022) analisam as formas como as tecnologias de reconhecimento facial afetam direitos fundamentais, especialmente de grupos vulneráveis. O artigo utiliza uma metodologia a partir do uso de técnicas de revisão bibliográfica e análise de legislações e projetos de lei no Brasil. O trabalho evidencia que a implementação de tecnologias de reconhecimento facial envolve o tratamento de dados sensíveis como dados biométricos, raciais, além de outras informações identitárias. Esses dados impactam em dimensões sociais de discriminação e marginalização de pessoas. Diante disso, o artigo abre um lugar de debate sobre como os fatores humanos impactam o uso dessas tecnologias e os objetivos dessa implementação.

Os autores (ELBAGIR; YANG, 2019) realizam um estudo sobre análise de dados da rede social *Twitter*, renomeada em 2023 para *X*, utilizando o VADER para análise de sentimentos. O estudo objetivou realizar uma análise de sentimento nas postagens sobre as eleições presidenciais de 2016 nos Estados Unidos da América (EUA). O VADER se mostrou eficaz e apresentou resultados satisfatórios para análise de sentimentos no tipo de postagem abordada, onde classificou grandes quantidades de dados de maneira fácil e rápida.

Os autores (KHEMANI; ADGAONKAR, 2021) abordam a utilização do VADER para análise de sentimentos em dados extraídos do *Reddit*. A metodologia realiza uma rotulação dos resultados da análise sobre o título de postagens e posteriormente faz uma normalização dos dados apresentando os resultados de forma gráfica. Os resultados apresentaram uma quantidade de aproximadamente 45% dos títulos ficando classificados como neutros, 27% classificados como negativos e 27% classificados como positivos.

4 METODOLOGIA

A metodologia do trabalho consiste na análise de relatos de usuários do *Reddit*, a partir das bases de dados obtidas na recuperação de postagens por *web scraping*. Outra etapa da metodologia consiste na aplicação da ferramenta VADER de PLN para análise de sentimentos nas bases de dados, bem como a realização de uma análise descritiva, de categorização dos sentimentos e de obtenção de termos mais recorrentes sobre essas bases. O diagrama da Figura 3 ilustra os passos que guiaram a metodologia deste trabalho.

Figura 3 – Diagrama da metodologia do trabalho



Fonte: elaborada pelo autor.

Como observado no diagrama da Figura 3, após a obtenção dos dados foi realizada a análise de sentimentos sobre as informações. Paralelamente a isso, foram extraídos os termos mais frequentes das bases de dados, ação que independe da análise de sentimentos. Por fim, após a análise de sentimentos, foram calculadas e apresentadas as características descritivas dos conjuntos de amostras, como também a classificação e cálculo do montante de sentimentos positivos, negativos e neutros, além de suas subclassificações.

4.1 Base de dados e utilização da API

As postagens feitas nos *subreddits* são comumente compostas por opiniões e afirmações acerca de algum fato. Em resposta a essas postagens, outros usuários tecem comentários onde concordam, discordam ou complementam o que acabaram de ler. Considerando isso, foi

utilizada a biblioteca PRAW para recuperar postagens que citavam o termo de busca “*facial recognition*”, reconhecimento facial em português, para posteriormente realizar a análise de sentimentos.

A partir das solicitações feitas juntas à API, podem ser obtidas informações como título, autor, texto da postagem, data de criação da postagem, número de votos positivos, número de votos negativos, nome do *subreddit* no qual está a referida postagem, URL e comentários, bem como informações sobre os comentários como autor, texto, data de criação do comentário, número de votos positivos e número de votos negativos do comentário.

Ademais, a análise de sentimentos realizada não considerou os números de votos positivos e negativos, visto que essas informações geralmente representam apenas a concordância ou discordância com tal comentário, não tendo relação direta com o sentimento exposto.

4.2 *Subreddits* utilizados e parâmetros

O processo de *web scraping* foi realizado nos *subreddits* *r/technology* e *r/world-news*. Esses *subreddits* foram escolhidos por agregar postagens informativas e apresentarem comentários criticamente positivos e negativos, além dos neutros.

Além do termo de busca citado no tópico anterior, é necessário que seja informado o parâmetro de classificação (*sort*) das postagens para o processo de obtenção. O parâmetro de classificação escolhido foi *relevance*, onde os resultados foram classificados com base na relevância, apresentando os conteúdos mais relacionados ao termo de busca. Define-se também o parâmetro para determinar o período de tempo (*time filter*) em que a resposta será buscada. Nesse caso foi utilizado o parâmetro *all*, buscando resultados em todo o período de atividade do referido *subreddit*.

4.3 Dados a serem coletados

Para a coleta dos dados torna-se evidente que determinados atributos dos *subreddits* se destacam como mais relevantes para a extração das informações das postagens. Estes atributos incluem os títulos, o conteúdo textual, as URLs presentes nos textos, os comentários e a data de criação de cada comentário. Os dados obtidos são retornados na estrutura de um *DataFrame* da biblioteca *Pandas* do *Python*.

Cada *subreddit* submetido ao processo de *web scraping* pelo termo de busca, com

os parâmetros *relevance* e *all*, retorna uma base de dados, convertida em *DataFrame*, contendo várias amostras divididas em 5 colunas. Essas colunas são representadas pelos atributos “Título da Postagem”, “Texto da Postagem”, “URL da Postagem”, “Comentário” e “Data do Comentário”.

As postagens de cada *DataFrame* representadas por “Título da Postagem” podem conter vários comentários. Com isso, um exemplo de *DataFrame* obtido exibe a estrutura apresentada na Figura 4. A figura apresenta cinco amostras aleatórias recuperadas do *subreddit r/technology* ao ser realizada a busca pelo termo “*brazil*”. O atributo “Título da Postagem” apresenta os títulos de cada postagem, onde pode-se observar que das cinco amostras, dois títulos se repetem duas vezes cada, significando que para uma mesma postagem com tal título, pelo menos dois comentários foram realizados. O atributo “Texto da Postagem” se encontra vazio pois todas as amostras possuem apenas links no local do texto. O atributo “URL da Postagem” contém os links dos textos. O atributo “Comentário” apresenta o texto de cada comentário, enquanto o atributo “Data do Comentário” apresenta a data de criação desses comentários.

Figura 4 – Fragmento de *DataFrame* exemplo obtido de um *subreddit*

	Título da Postagem	Texto da Postagem	URL da Postagem	Comentário	Data do Comentário
356	IBM Announces 50TB Enterprise Magnetic Tape Dr...		https://www.forbes.com/sites/tomcoughlin/2023/...	LoL. Sure maybe it's only 22 terabyte but You ...	2023-09-06
299	Indian social network Koo gains popularity in ...		https://techcrunch.com/2022/11/22/indian-socia...	It's gaining popularity here in part because "...	2022-11-26
425	AWS: IPv4 addresses cost too much, so you're g...		https://www.theregister.com/2023/07/31/aws_say...	Nah, mobile phones are going to be behind a ca...	2023-08-01
313	IBM Announces 50TB Enterprise Magnetic Tape Dr...		https://www.forbes.com/sites/tomcoughlin/2023/...	Meanwhile some else is looking at a 1 petabyte...	2023-09-06
429	AWS: IPv4 addresses cost too much, so you're g...		https://www.theregister.com/2023/07/31/aws_say...	Yea they're a joke. They offer 8gbps fiber ser...	2023-08-02

Fonte: elaborada pelo autor.

4.4 Análise do sentimento dos dados coletados

O atributo mais interessante para que fosse realizada a análise de sentimentos sobre o mesmo foi o atributo “Comentário”. O analisador de intensidade de sentimento do VADER foi inicializado e aplicado à cada comentário. Com isso, foram geradas pontuações para “*neg*”, “*neu*” e “*pos*” representadas pelo atributo “*compound*” que caracteriza a polaridade do sentimento. Na Figura 5 pode-se observar o resultado dessa análise para o exemplo mostrado na Figura 4. Pode-se observar o atributo “*compound*” apresentando valores maiores, menores e iguais a zero indicando sentimentos positivos, negativos e neutros, respectivamente, para o comentário em

questão. Na amostra em questão as colunas “*neg*”, “*neu*” e “*pos*” foram ocultadas para melhor visualização.

Figura 5 – Análise de sentimentos do fragmento da Figura 4

	Titulo da Postagem	Texto da Postagem	URL da Postagem	Comentário	Data do Comentário	compound
356	IBM Announces 50TB Enterprise Magnetic Tape Dr...		https://www.forbes.com/sites/tomcoughlin/2023/...	LoL. Sure maybe it's only 22 terabyte but You ...	2023-09-06	-0.8991
299	Indian social network Koo gains popularity in ...		https://techcrunch.com/2022/11/22/indian-socia...	It's gaining popularity here in part because "...	2022-11-26	0.8834
425	AWS: IPv4 addresses cost too much, so you're g...		https://www.theregister.com/2023/07/31/aws_say...	Nah, mobile phones are going to be behind a ca...	2023-08-01	-0.1027
313	IBM Announces 50TB Enterprise Magnetic Tape Dr...		https://www.forbes.com/sites/tomcoughlin/2023/...	Meanwhile some else is looking at a 1 petabyte...	2023-09-06	0.0
429	AWS: IPv4 addresses cost too much, so you're g...		https://www.theregister.com/2023/07/31/aws_say...	Yea they're a joke. They offer 8gbps fiber ser...	2023-08-02	0.1366

Fonte: elaborada pelo autor.

4.4.1 Análise descritiva e temporal dos dados após a análise de sentimentos

Uma análise descritiva dos resultados se mostra interessante para que seja possível observar a dimensão de cada base de dados. Esse tipo de observação traz informações dos dados como quantidade de amostras, média, quartis, mínimo, máximo e desvio padrão. Com isso, para cada um dos *DataFrames* foi aplicado o método *Pandas describe()* que fornece as informações citadas para cada atributo de tipo numérico como também para o atributo de data.

Outra abordagem sobre os dados foi a de análise temporal, onde foi observada a disposição da criação dos comentários ao longo do tempo.

4.5 Categorização dos sentimentos extraídos

Para se ter um maior detalhamento na análise de cada comentário, de forma semelhante a (ELBAGIR; YANG, 2019), foi realizada uma manipulação dos resultados dispostos no atributo “*compound*”. Essa manipulação envolve a categorização dos valores coletados em diferentes faixas, além das que caracterizam os sentimentos “Negativo”, “Positivo” e “Neutro”. Tal manipulação objetivou a realização de uma análise que permitiu a atribuição de rótulos mais detalhados aos sentimentos. Com isso, os valores de “*compound*” foram categorizados em “Muito negativo”, “Negativo”, “Neutro”, “Positivo” e “Muito positivo”. Esse processo pode ser

detalhado da seguinte forma:

- a) Muito negativo: $-1,0 \leq compound \leq -0,500$;
- b) Negativo: $-0,500 < compound < 0,0$;
- c) Neutro: $compound = 0,0$;
- d) Positivo: $0,0 < compound < 0,500$;
- e) Muito positivo: $0,500 \leq compound \leq 1,0$.

Ainda, ressalta-se que os dois *DataFrames* contêm amostras com comentários removidos pelos usuários, nesse caso essa operação fica registrada pelos termos “[deleted]” e “[removed]” em substituição aos comentários. Por conseguinte, se faz necessário um tratamento das amostras para desconsiderá-las, pois as amostras em questão são irrelevantes para a análise de sentimentos.

4.6 Obtenção de termos mais frequentes

Após o processo de aplicação de análise de sentimentos utilizando o VADER, foram realizadas análises nos *DataFrames* para extrair as expressões de duas palavras mais frequentes da coluna “Comentário”. Para isso, é necessário que se faça uma tratativa para que sejam excluídos determinados termos irrelevantes como “like”, “https”, “watch”, “org”, “en”, entre outros. Esses termos não agregam valor de entendimento e contexto ao comentário, assim como as *stopwords* da língua inglesa que podem ser removidas com o auxílio da classe *CountVectorizer* da biblioteca *scikit-learn* no *Python*. Essa classe também realiza a contagem das expressões de duas palavras presentes no texto analisado. Em sequência, foram retornadas as 15 expressões de duas palavras que eram mais frequentes na coluna “Comentário” de cada *DataFrame*.

5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os resultados obtidos das análises realizadas, de acordo com a metodologia utilizada, bem como é feita uma discussão dos resultados.

5.1 Análise descritiva dos resultados

Como mencionado no tópico anterior, foram recuperadas duas bases de dados pelo procedimento de *web scraping* buscando pelo termo “*facial recognition*”.

A base de dados recuperada do *subreddit r/technology* retornou o *DataFrame* mostrado na Figura 6. Ele é composto por 46 333 amostras detalhadas pelos atributos “Título da Postagem”, “Texto da Postagem”, “URL da Postagem”, “Comentário” e “Data do Comentário”. Foram recuperados comentários criados no período de 15/08/2012 a 14/10/2023.

Figura 6 – Resultado do processo de *web scraping* em *r/technology*

	Título da Postagem	Texto da Postagem	URL da Postagem	Comentário	Data do Comentário
0	Innocent pregnant woman jailed amid faulty fac...		https://arstechnica.com/information-technology...	As an AI / ML practitioner and consultant, the...	2023-08-07
1	Innocent pregnant woman jailed amid faulty fac...		https://arstechnica.com/information-technology...	Class of 09 but in real life. The show may not...	2023-08-07
2	Innocent pregnant woman jailed amid faulty fac...		https://arstechnica.com/information-technology...	Police put way too much faith in technology so...	2023-08-07
3	Innocent pregnant woman jailed amid faulty fac...		https://arstechnica.com/information-technology...	In the US there needs to be a Federal ban on p...	2023-08-07
4	Innocent pregnant woman jailed amid faulty fac...		https://arstechnica.com/information-technology...	Allowing any automated system to make decision...	2023-08-07
...
46328	The DOJ is in the process of creating a massiv...		http://www.allgov.com/usa/ca/news/california-a...	I am pretty sure Google has said their satelli...	2015-03-22
46329	The DOJ is in the process of creating a massiv...		http://www.allgov.com/usa/ca/news/california-a...	All it would take is for one juror to consider...	2015-03-21
46330	The DOJ is in the process of creating a massiv...		http://www.allgov.com/usa/ca/news/california-a...	So why is it crazy to believe that they have l...	2015-03-21
46331	The DOJ is in the process of creating a massiv...		http://www.allgov.com/usa/ca/news/california-a...	its crazy to think the government is going to ...	2015-03-21
46332	The DOJ is in the process of creating a massiv...		http://www.allgov.com/usa/ca/news/california-a...	> So why is it crazy to believe that they have...	2015-03-21

46333 rows × 5 columns

Fonte: elaborada pelo autor.

A base de dados recuperada do *subreddit r/worldnews* retornou o *DataFrame* mostrado na Figura 7. Ele é composto por 17 878 amostras detalhadas também pelos atributos “Título da Postagem”, “Texto da Postagem”, “URL da Postagem”, “Comentário” e “Data do Comentário”. Foram recuperados comentários criados no período de 03/08/2011 a 30/08/2023.

Figura 7 – Resultado do processo de *web scraping* em *r/worldnews*

	Título da Postagem	Texto da Postagem	URL da Postagem	Comentário	Data do Comentário
0	European Parliament calls for a ban on facial ...		https://www.politico.eu/article/european-parli...	It's a bit of a mis-leading headline (unsurpri...	2021-10-06
1	European Parliament calls for a ban on facial ...		https://www.politico.eu/article/european-parli...	In Brazil the government just built a huge dat...	2021-10-06
2	European Parliament calls for a ban on facial ...		https://www.politico.eu/article/european-parli...	That would be really great otherwise you will ...	2021-10-06
3	European Parliament calls for a ban on facial ...		https://www.politico.eu/article/european-parli...	Hell yeah! I don't like voyeurism at all. I do...	2021-10-06
4	European Parliament calls for a ban on facial ...		https://www.politico.eu/article/european-parli...	Everyone in the UK has yet another reason to f...	2021-10-06
...
17873	UK Backs Facial Recognition Project That Helps...		https://www.cityam.com/uk-backs-facial-recogni...	Duck UK	2020-03-16
17874	UK Backs Facial Recognition Project That Helps...		https://www.cityam.com/uk-backs-facial-recogni...	WTF UK?! I get that money seems nice, but mone...	2020-03-16
17875	UK Backs Facial Recognition Project That Helps...		https://www.cityam.com/uk-backs-facial-recogni...	UK-USA-China-Russia \n\nName a more iconic ...	2020-03-16
17876	UK Backs Facial Recognition Project That Helps...		https://www.cityam.com/uk-backs-facial-recogni...	I expect the UK to be the first Western Countr...	2020-03-16
17877	UK Backs Facial Recognition Project That Helps...		https://www.cityam.com/uk-backs-facial-recogni...	They are better at Ducking themselves...	2020-03-16

17878 rows × 5 columns

Fonte: elaborada pelo autor.

Com a tratativa de exclusão de amostras irrelevantes e com a análise de sentimentos realizada pela ferramenta VADER da biblioteca *Python NLTK*, os *DataFrames* de *r/technology* e de *r/worldnews* foram reduzidos a 43 483 e 16 867 amostras, respectivamente. As Tabelas 1 e 2 apresentam uma análise descritiva dos *DataFrames*, fornecendo uma visão geral dos dados.

Tabela 1 – Estatísticas descritivas para *r/technology*

Medidas descritivas	Comentário	Data do Comentário	neg	neu	pos	compound
Quantidade	43483	43483	43483	43483	43483	43483
Média	-	21/11/2019	0,097879	0,796393	0,105615	-0,002911
Mínimo	-	15/08/2012	0,000000	0,000000	0,000000	-0,999500
25%	-	09/07/2019	0,000000	0,703000	0,000000	-0,391900
50%	-	16/01/2020	0,058000	0,812000	0,073000	0,000000
75%	-	29/12/2020	0,156000	0,925000	0,160000	0,391900
Máximo	-	14/10/2023	1,000000	1,000000	1,000000	0,997000
Desvio padrão	-	-	0,126499	0,168377	0,130956	0,503705

Fonte: elaborada pelo autor.

De acordo com a Tabela 1, a data central da distribuição dos comentários no período de 15/08/2012, data inicial (mínimo), à 14/10/2023, data final (máximo), foi 21/11/2019 (média). Ainda para as datas dos comentários, têm-se os quartis apresentando as seguintes informações:

- 25% dos comentários foram criados antes de 09/07/2019, ficando o restante nos 75% posteriores;

- b) 50% dos comentários são de antes de 16/01/2020, sendo os outros 50% posteriores a essa data;
- c) 75% dos comentários foram criados antes de 29/12/2020, ficando o restante posteriores a essa data.

Na descrição de *compound* do *subreddit r/technology*, foi obtido um valor médio igual a $-0,002911$, indicando que o valor médio geral das 43 483 amostras é ligeiramente negativo, muito próximo a zero. Como mínimo dispõe-se o valor $-0,999500$, e como máximo o valor $0,997000$. O intervalo entre esses dois valores equivale à amplitude do valor de *compound*.

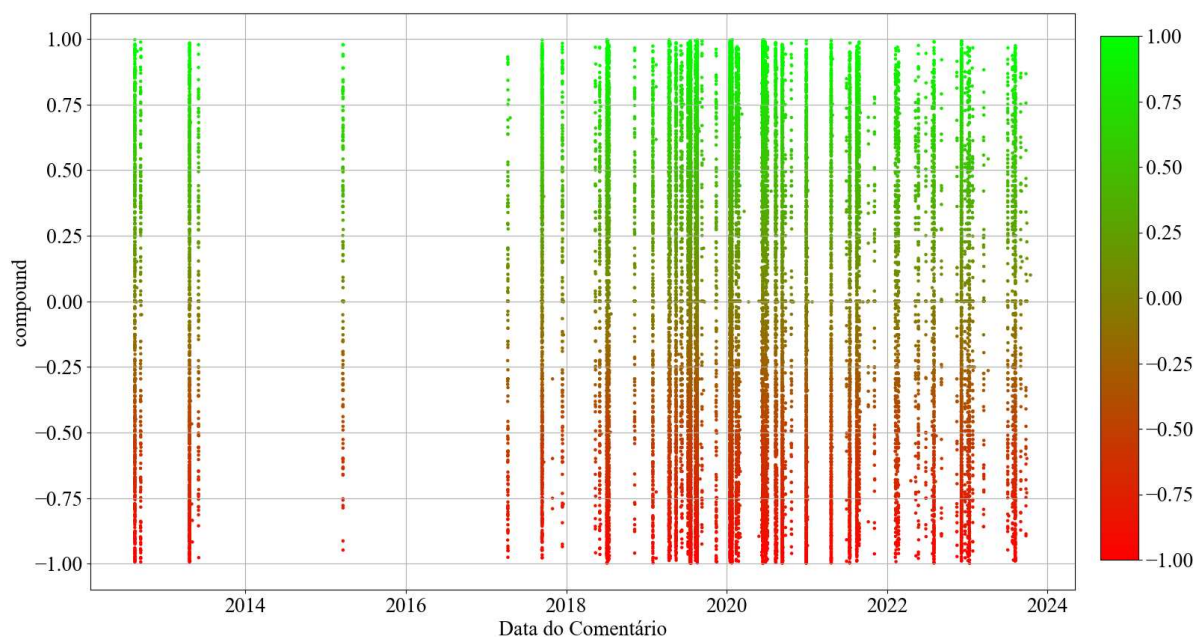
Para os quartis foram obtidas as seguintes informações para os valores de *compound*:

- a) 25% estão abaixo de $-0,391900$;
- b) 50% estão abaixo de $0,000000$ e os outros 50% acima deste valor;
- c) 75% estão abaixo de $0,391900$, ficando o restante acima deste valor.

O desvio padrão de $0,503705$ mostrou que os valores de *compound* estão espalhados em torno da média de forma variável, em maior dispersão.

A distribuição temporal dos comentários e postagens, bem como a disposição dos valores de *compound* do *subreddit r/technology* pode ser observada na Figura 8, onde cada ponto representa um comentário e cada aglomerado vertical de pontos representa a postagem dos respectivos comentários. A cor de cada ponto representa o quanto positivo ou negativo ele é.

Figura 8 – Gráfico relacionando o valor de *compound* pelo tempo decorrido em *r/technology*



Fonte: elaborada pelo autor.

Tabela 2 – Estatísticas descritivas para *r/worldnews*

Medidas descritivas	Comentário	Data do Comentário	neg	neu	pos	compound
Quantidade	16867	16867	16867	16867	16867	16867
Média	-	05/11/2018	0,094663	0,79607	0,107780	0,004604
Mínimo	-	03/08/2011	0,000000	0,000000	0,000000	-0,998000
25%	-	04/01/2017	0,000000	0,700000	0,000000	-0,361200
50%	-	16/02/2019	0,041000	0,814000	0,068000	0,000000
75%	-	17/09/2021	0,153000	0,956000	0,161000	0,388900
Máximo	-	30/08/2023	1,000000	1,000000	1,000000	0,998700
Desvio padrão	-	-	0,131278	0,18168	0,143211	0,493569

Fonte: elaborada pelo autor.

Para a Tabela 2, a data central da distribuição dos comentários no período de 03/08/2011, data inicial (mínimo), à 30/08/2023, data final (máximo), foi 05/11/2018 (média). Ainda para as datas dos comentários, têm-se os quartis apresentando as seguintes informações:

- a) 25% dos comentários foram criados antes de 04/01/2017, ficando o restante nos 75% posteriores;
- b) 50% dos comentários são de antes de 16/02/2019, sendo os outros 50% posteriores a essa data;
- c) 75% dos comentários foram criados antes de 17/09/2021, ficando o restante posteriores a essa data.

Na descrição de *compound* do *subreddit r/worldnews*, foi obtido valor médio igual a 0,004604, indicando que o valor médio geral das 16 867 amostras é ligeiramente positivo, muito próximo à zero. Como mínimo dispõe-se o valor $-0,998000$, e como máximo o valor $0,998700$. O intervalo entre esses dois valores equivale à amplitude do valor de *compound* em *r/worldnews*.

Para os quartis foram obtidas as seguintes informações para os valores de *compound*:

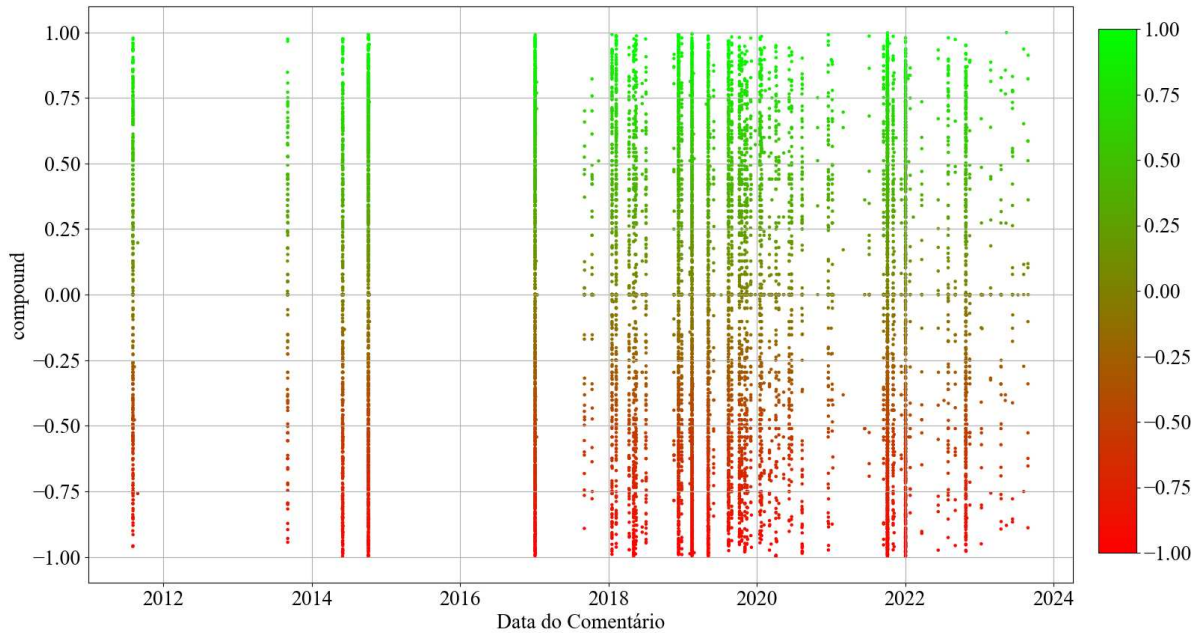
- a) 25% estão abaixo de $-0,361200$;
- b) 50% estão abaixo de $0,000000$ e os outros 50% acima deste valor;
- c) 75% estão abaixo de $0,388900$, ficando o restante acima deste valor.

De forma semelhante ao desvio padrão de *compound* em *r/technology*, o desvio padrão em *r/worldnews* foi de $0,493569$, mostrando que os valores de *compound* se dispuseram de forma dispersa e variável em torno da média.

A distribuição temporal dos comentários e postagens, bem como a disposição dos valores de *compound* do *subreddit r/worldnews* pode ser observada na Figura 9, onde cada

ponto representa um comentário e cada aglomerado vertical de pontos representa a postagem dos respectivos comentários. A cor de cada ponto representa o quanto positivo ou negativo ele é.

Figura 9 – Gráfico relacionando o valor de *compound* pelo tempo decorrido em *r/worldnews*



Fonte: elaborada pelo autor.

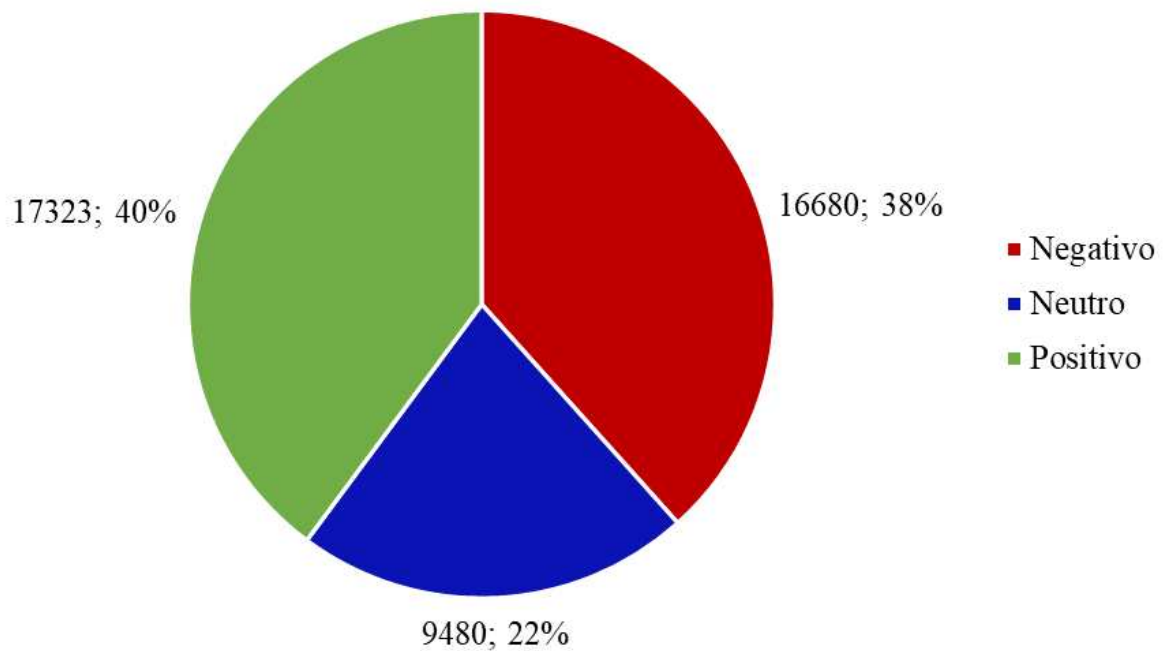
No contexto da criação dos comentários e nos sentimentos destes, pôde-se observar que a realização dos comentários ocorreram, nos dois *subreddits*, majoritariamente no intervalo da segunda metade do período temporal das distribuições. Tal fato pôde ser observado graficamente nas Figuras 8 e 9.

Em relação aos sentimentos, os quartis e os intervalos das duas distribuições demonstraram que os sentimentos dos comentários estão dispostos de forma uniforme para cada postagem realizada no decorrer do tempo. A distribuição observada nos quartis apresenta com mais detalhes o fato observado.

5.2 Resultado da análise de sentimentos

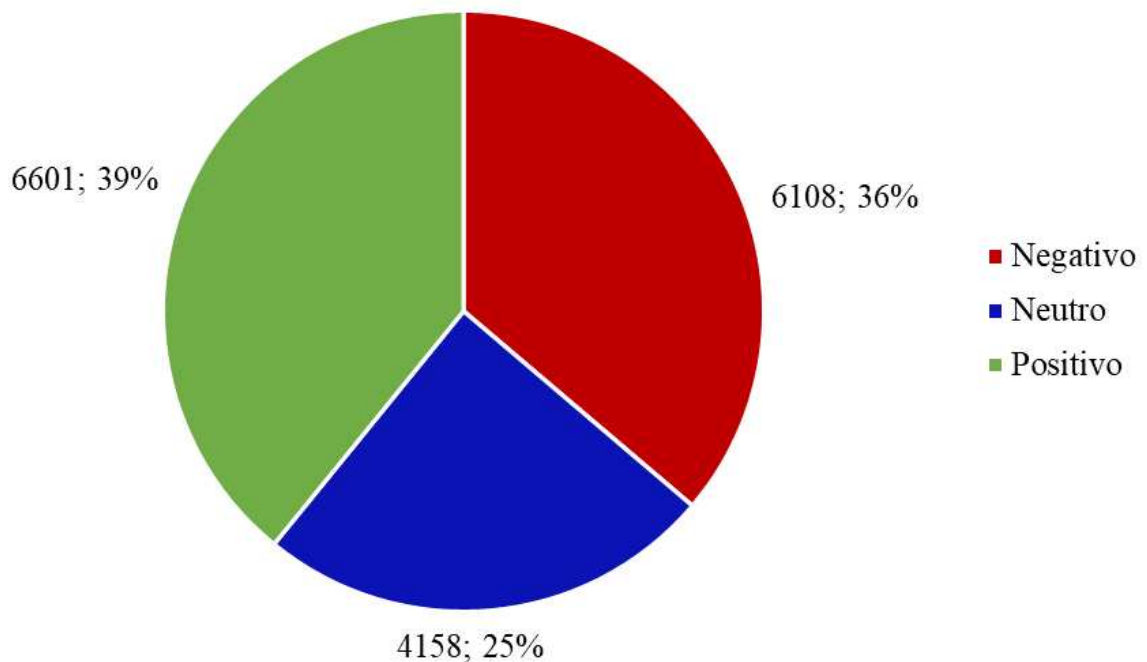
Após a análise de sentimentos das bases de dados estruturadas em *DataFrames*, foram obtidos os seguintes resultados mostrados nos gráficos das Figuras 10 e 11.

Figura 10 – Gráfico apresentando a quantidade de comentários por sentimento em *r/technology*



Fonte: elaborada pelo autor.

Figura 11 – Gráfico apresentando a quantidade de comentários por sentimento em *r/worldnews*

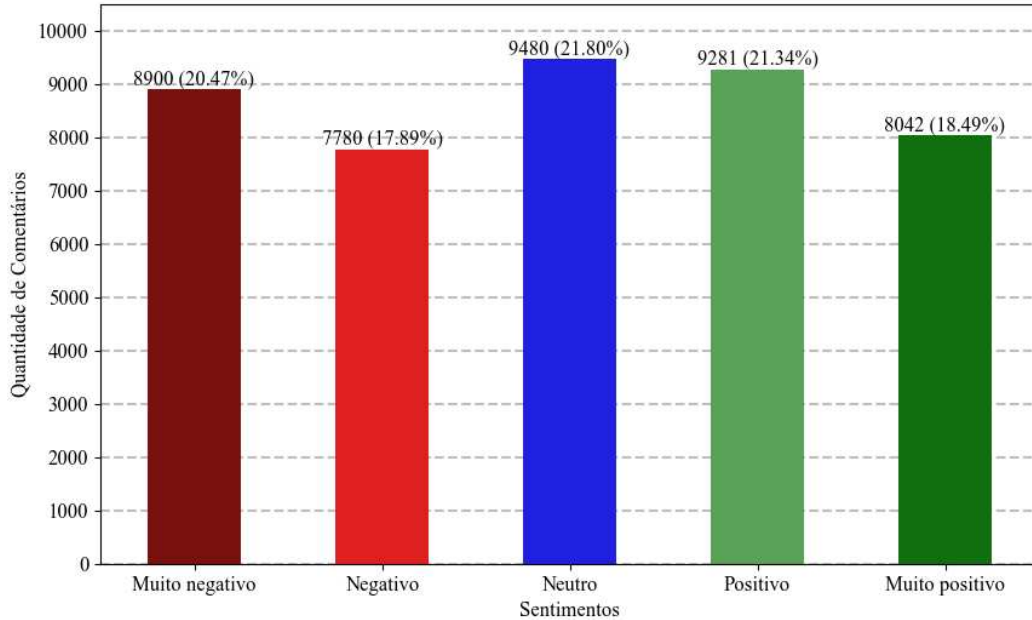


Fonte: elaborada pelo autor.

Ao discretizar os sentimentos “Negativos”, “Neutros” e “Positivos” em “Muito negativo”, “Negativo”, “Neutro”, “Positivo” e “Muito positivo”, as quantidades de comentários

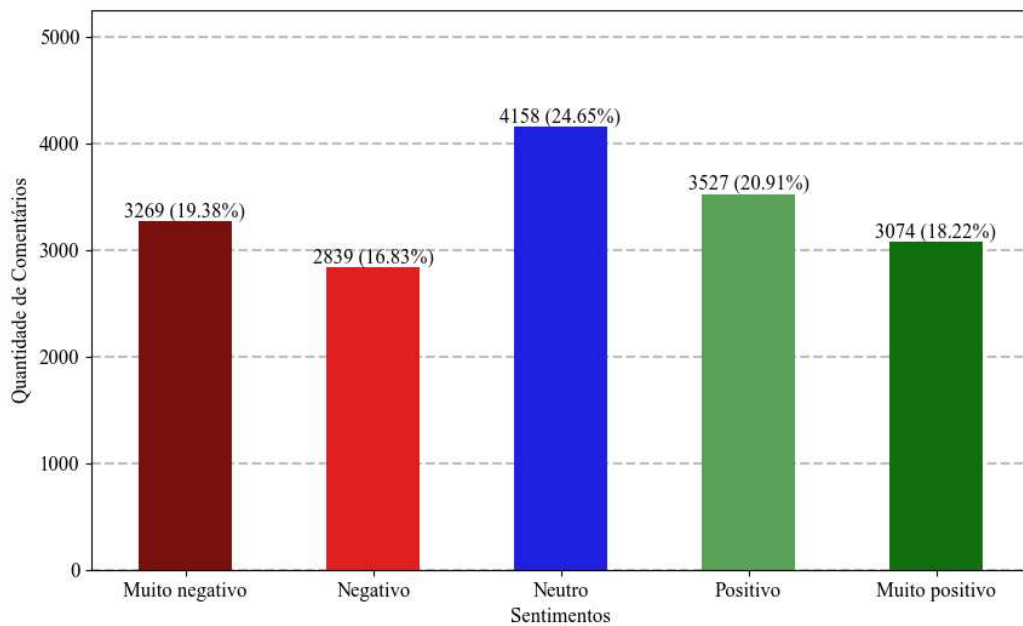
para essas subdivisões foram representadas nos gráficos das Figuras 12 e 13.

Figura 12 – Gráfico relacionando a quantidade de comentários por sentimento em *r/technology* após a subdivisão dos sentimentos



Fonte: elaborada pelo autor.

Figura 13 – Gráfico relacionando a quantidade de comentários por sentimento em *r/worldnews* após a subdivisão dos sentimentos



Fonte: elaborada pelo autor.

Ao realizar a análise dos comentários relacionados à postagens que envolvessem o tema de reconhecimento facial, os dois *subreddits* apresentaram grandes quantidades de comentários positivos e negativos, bem como neutros. Destacaram-se as somas de comentários positivos e negativos, apresentando semelhanças em seus totais.

A Figura 12 mostra a seguinte categorização para o *subreddit r/technology*:

- a) 20,47% dos comentários extraídos foram categorizados como “Muito Negativo”;
- b) 17,89% como “Negativo”;
- c) 21,34% como “Positivo”;
- d) 18,49% como “Muito positivo”.

De forma semelhante, a Figura 13 mostra a seguinte categorização para o *subreddit r/worldnews*:

- a) 19,38% como “Muito Negativo”;
- b) 16,83% como “Negativo”;
- c) 20,91% como “Positivo”;
- d) 18,22% como “Muito positivo”.

Em relação à quantidade de comentários categorizados como neutro, (SOUZA *et al.*, 2017) destaca que o VADER ao se deparar com um comentário que não apresenta valores aceitáveis para “*neg*” ou “*pos*”, ele o classifica como neutro. Com isso, *r/technology* apresentou 21,80% de comentários classificados como “Neutro” enquanto *r/worldnews* apresentou 24,65%.

Observou-se, também, que tanto os gráficos que representam as relações de quantidades de comentários por sentimento, quanto os gráficos que representam as relações de valor de *compound* pelo tempo decorrido, apresentam uma neutralidade geral dos sentimentos, não havendo diferenças discrepantes nas quantidades e nas distribuições dos sentimentos observados.

5.3 Análise dos termos frequentes

A obtenção da frequência de expressões de duas palavras foi realizada com o intuito de extrair os temas mais abordados nos comentários realizados pelos usuários do *Reddit*. Com isso, foi obtido mais um indicador da percepção desses usuários quanto aos possíveis problemas e situações existentes na reprodução do viés em tecnologias de IA. No que compete a área de desenvolvimento e uso de tecnologias de reconhecimento facial, foram encontrados os seguintes dados apresentados nas Tabelas 3 e 4 e nas Figuras 14 e 15.

A Tabela 3 apresenta as 15 expressões mais citadas nos comentários em *r/technology*.

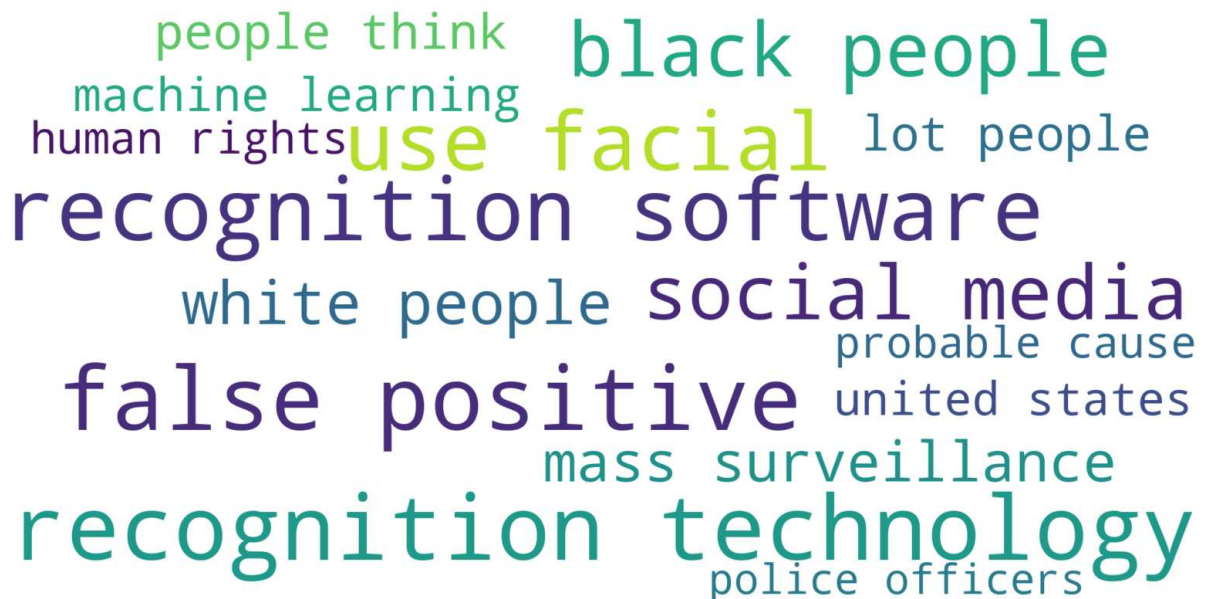
Tabela 3 – Expressões de duas palavras com mais frequência em *r/technology*

Expressões de duas palavras	Frequência
recognition technology	528
false positive	524
recognition software	502
use facial	439
social media	383
black people	376
white people	222
mass surveillance	194
people think	145
lot people	144
machine learning	135
united states	130
human rights	119
probable cause	114
police officers	112

Fonte: elaborada pelo autor

A Figura 14 apresenta uma nuvem de palavras criada a partir da Tabela 3 que facilita a observação dos termos.

Figura 14 – Nuvem de palavras ilustrando os termos com base na frequência em *r/technology*



Fonte: elaborada pelo autor.

A Tabela 4 apresenta as 15 expressões mais citadas nos comentários do *subreddit* *r/worldnews*.

Tabela 4 – Expressões de duas palavras com mais frequência em *r/worldnews*

Expressões de duas palavras	Frequência
recognition technology	203
false positive	172
human rights	154
recognition software	152
use facial	120
minority report	88
social media	78
social credit	75
police state	72
message compose	67
mass surveillance	64
north korea	63
machine learning	57
black people	49
chinese people	45

Fonte: elaborada pelo autor.

A Figura 15 apresenta uma nuvem de palavras criada a partir da Tabela 4 que facilita a observação dos termos.

Figura 15 – Nuvem de palavras ilustrando os termos com base na frequência em *r/worldnews*



Fonte: elaborada pelo autor.

A partir das tabelas e das figuras pode-se inferir uma relação entre o aspecto social e o tecnológico. Os dados apresentados evidenciaram a convergência do uso de sistemas tecnológicos de reconhecimento facial com a preocupação e atenção a aspectos sociais como a discriminação racial e a possível violação da privacidade individual, por exemplo.

Uma expressão em destaque é “*false positive*”, falso positivo em português, apresentando uma frequência significativa ao se comparar as duas tabelas e suas outras expressões. A expressão “*recognition software*”, software de reconhecimento em português, também apresentou uma frequência significativa quando comparada. No mesmo passo, observaram-se os termos “*black people*”, “*human rights*” e “*mass surveillance*”, pessoas pretas, direitos humanos e vigilância em massa, em português. Os termos destacados demonstraram a relação entre as postagens dos *subreddits* e os aspectos violatórios de direitos encontrados nos softwares desenvolvidos e abordados nas postagens do *Reddit*.

6 CONCLUSÃO

O uso crescente de tecnologias de IA impulsiona o desenvolvimento de soluções mais acertadas, para que funcionem de forma a auxiliar mais idealmente quem a utiliza, ocasionando o mínimo possível de adversidades. Porém, relatos de problemas apresentados em sistemas de reconhecimento facial revela a existência do viés no algoritmo. Visto isso, o presente trabalho objetivou analisar a percepção sobre o viés encontrado em algoritmos de IA implementados para reconhecimento facial.

As bibliografias estudadas apresentam os malefícios do viés discriminatório, evidenciando que a inevitável interferência humana no desenvolvimento dos algoritmos, propaga ideais e conceitos nos resultados destes, o que pode ser prejudicial a quem utiliza ou é analisado por esses algoritmos.

Com isso, a análise de dados feita sobre o período de tempo aproximado de 12 anos nos *subreddits*, aponta uma neutralidade geral dos comentários analisados em cada postagem, sendo identificados comentários que expressam sentimentos negativos, e também comentários que expressam sentimentos positivos, para uma mesma postagem. Também, os graus de intensidade dos sentimentos dos comentários apresentam dispersão uniforme em seus intervalos de posicionamento. Visto isso, considerando os tamanhos das bases de dados obtidas e a porcentagem de cada categoria de sentimentos, é notório o fato de que uma postagem admitir, também, grandes quantidades de comentários negativos, corrobora com a identificação de um certo incômodo no uso dos sistemas de reconhecimento facial em questão.

Ademais, ao observar a disposição temporal dos comentários realizados, fica explícito que a abordagem e uso de sistemas de reconhecimento facial, em maior escala, se deram no período dos últimos cinco anos, coincidindo assim com a produção de trabalhos que abordam o assunto.

A análise de termos frequentes expõe uma amostra do assunto geral que embasa os sentimentos expostos. Os termos com maior frequência expressam um antagonismo aos comentários positivos e um conformismo aos comentários negativos, relacionando termos sobre condições raciais, direitos humanos, vigilância e utilização de sistemas de identificação. Em relação a este último, apresenta-se como problema frequente o reconhecimento errôneo de faces.

Sendo assim, este trabalho expõe a presença de erros provenientes do viés encontrado nos sistemas de reconhecimento facial ao passo que identifica uma relação entre o que foi comentado e o assunto geral exposto na amostra obtida. A análise realizada apresenta uma

discussão atual que evolui no decurso do tempo, mostrando a preocupação com a violação de direitos causada pelo viés presentes nos algoritmos.

Como trabalhos futuros, sugere-se o desenvolvimento de técnicas que minimizem os efeitos do viés em sistemas de reconhecimento facial que já estão em uso.

REFERÊNCIAS

- ALELYANI, S. Detection and evaluation of machine learning bias. **Applied Sciences**, v. 11, n. 14, 2021. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/11/14/6271>. Acesso em: 23 mai. 2023.
- ALPAYDIN, E. **Machine learning: the new AI**. [S. l.]: MIT press, 2016.
- AMAZON WEB SERVICES, INC. **O que é Análise de sentimento? - Explicação sobre análise de sentimento - AWS**. s.d. Disponível em: <https://aws.amazon.com/pt/what-is/sentiment-analysis/>. Acesso em: 08 jun. 2023.
- BARBOSA, J.; VIEIRA, J. P. A.; SANTOS, R.; JUNIOR, G. V. M.; MUNIZ, M. d. S.; MOURA, R. S. Introdução ao processamento de linguagem natural usando Python. **III Escola Regional de Informatica do Piauí**, v. 1, p. 336–360, 2017.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. **Sociedade Brasileira de Computação**, 2015.
- BINDS.CO. **Análise de Sentimento**. Binds.co, 2023. Disponível em: <https://binds.co/plataforma-de-pesquisa/analise-de-sentimento/>. Acesso em: 30 nov. 2023.
- BRANDÃO, R.; OLIVEIRA, J. L. Reconhecimento facial e viés algorítmico em grandes municípios brasileiros. In: SBC. **Anais do II Workshop sobre as Implicações da Computação na Sociedade**. [S. l.], 2021. p. 122–127.
- CARVALHO, A. P. d. **Viés algorítmico e discriminação: possíveis soluções regulatórias para o Brasil**. TCC (Graduação) – Faculdade de Direito, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020. Disponível em: <http://hdl.handle.net/10183/221914>. Acesso em: 01 dez. 2023.
- COSTA, R. S.; KREMER, B. Inteligência artificial e discriminação: Desafios e perspectivas para a proteção de grupos vulneráveis frente às tecnologias de reconhecimento facial. **Revista Brasileira de Direitos Fundamentais & Justiça**, v. 16, n. 1, 2022.
- ELBAGIR, S.; YANG, J. Twitter sentiment analysis using natural language toolkit and vader sentiment. In: **Proceedings of the international multiconference of engineers and computer scientists**. [S. l.: s. n.], 2019. v. 122, p. 16.
- FONTINHA, R. M. d. C. P. F. **Dados de reconhecimento facial em Business Intelligence**. Dissertação (Mestrado em Engenharia Informática) – Instituto Superior de Engenharia do Porto, Instituto Politécnico do Porto, Porto, 2014. Disponível em: <http://hdl.handle.net/10400.22/6514>. Acesso em: 01 dez. 2023.
- GEEKS, G. for. **Python | PRAW - Python Reddit API Wrapper**. 2020. Disponível em: <https://www.geeksforgeeks.org/python-praw-python-reddit-api-wrapper/>. Acesso em: 29 out. 2023.
- GLEZ-PEÑA, D.; LOURENÇO, A.; LÓPEZ-FERNÁNDEZ, H.; REBOIRO-JATO, M.; FDEZ-RIVEROLA, F. Web scraping technologies in an API world. **Briefings in Bioinformatics**, v. 15, n. 5, p. 788–797, 04 2013. ISSN 1467-5463. Disponível em: <https://doi.org/10.1093/bib/bbt026>. Acesso em: 08 jun. 2023.

GONZALEZ, M.; LIMA, V. L. S. Recuperação de informação e processamento da linguagem natural. In: SN. **XXIII Congresso da Sociedade Brasileira de Computação**. [S. l.], 2003. v. 3, p. 347–395.

HEIKKILÄ, M. The viral AI avatar app Lensa undressed me — without my consent. **MIT Technology Review**, 2022. Disponível em: <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent>. Acesso em: 08 dez. 2023.

HUTTO, C.; GILBERT, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. **Proceedings of the International AAAI Conference on Web and Social Media**, v. 8, n. 1, p. 216–225, Maio 2014. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>. Acesso em: 30 ago. 2023.

KAHNEMAN, D. **Rápido e devagar: duas formas de pensar**. Rio de Janeiro: Objetiva, 2011.

KHALIL, A.; AHMED, S. G.; KHATTAK, A. M.; AL-QIRIM, N. Investigating bias in facial analysis systems: A systematic review. **IEEE Access**, v. 8, p. 130751–130761, 2020.

KHEMANI, B.; ADGAONKAR, A. A review on Reddit news headlines with NLTK tool. In: **Proceedings of the International Conference on Innovative Computing & Communication (ICICC)**. [S. l.: s. n.], 2021.

MACHADO, A. M. O impacto de vieses cognitivos sobre a imparcialidade do conteúdo de inteligência. **Revista Brasileira de Inteligência**, n. 13, p. 1–16, 2018.

MATTERS, I. What is Reddit? - What parents need to know. **internet.matters.org**, 2022. Disponível em: <https://www.internetmatters.org/hub/news-blogs/what-is-reddit-what-parents-need-to-know/>. Acesso em: 31 mai. 2023.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. **AI Magazine**, v. 27, n. 4, p. 12–12, 2006.

MEHRABI, N.; MORSTATTER, F.; SAXENA, N.; LERMAN, K.; GALSTYAN, A. A survey on bias and fairness in machine learning. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 6, p. 1–35, 2021.

NEGRI, S. M. C. de Á.; OLIVEIRA, S. R. de; COSTA, R. S. O uso de tecnologias de reconhecimento facial baseadas em inteligência artificial e o direito à proteção de dados. **Direito Público**, v. 17, n. 93, 2020.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. 1948. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 12 set. 2023.

ORVALHO, V. Reconhecimento facial. **Revista de Ciência Elementar**, Casa das Ciências, v. 7, n. 4, 2019.

SILVA, T. da. Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina. **Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)**, v. 12, n. 31, 2020.

SOUZA, K. F. de; PEREIRA, M. H. R.; DALIP, D. H. Unilex: Método léxico para análise de sentimentos textuais sobre conteúdo de tweets em português brasileiro. **Abakós**, v. 5, n. 2, p. 79–96, 2017.

TECNOBLOG. **Suas fotos foram usadas por sistemas de biometria facial? Este site mostra**. Tecnoblog, 2021. Disponível em: <https://tecnoblog.net/noticias/2021/02/01/suas-fotos-foram-usadas-por-sistemas-de-biometria-facial-este-site-mostra/>. Acesso em: 15 nov. 2023.

VIEIRA, R.; LOPES, L. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. **Em corpora**, p. 183, 2010.