UNIVERSIDADE FEDERAL DO AMAZONAS

INSTITUTO DE COMPUTAÇÃO

MESTRADO EM INFORMÁTICA

Detecção de Viés Ideológico em Artigos de Notícias Utilizando Aprendizado Métrico Profundo e Representações Contextuais

Jailson Pereira Januário

Manaus - AM

Novembro de 2024

Jailson Pereira Januário

Detecção de Viés Ideológico em Artigos de Notícias Utilizando Aprendizado Métrico Profundo e Representações Contextuais

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito parcial para a obtenção do grau de Mestre em Informática.

Orientador(a)

Prof. Dr André Luiz da Costa Carvalho

Universidade Federal do Amazonas

Instituto de Computação

Manaus - AM

Novembro de 2024

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Januário, Jailson Pereira
J35d Detecção de Viés Ideológico em Artigos de Notícias Utilizando Aprendizado Métrico Profundo e Representações Contextuais / Jailson Pereira Januário . 2024
94 f.: il.; 31 cm.

Orientadora: André Luiz da Costa Carvalho
Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas.

1. Viés Ideológico. 2. Aprendizado Profundo. 3. Aprendizado Métrico. 4. Processamento de Linguagem Natural. I. Carvalho, André Luiz da Costa. II. Universidade Federal do Amazonas III. Título

Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"DETECÇÃO DE VIÉS IDEOLÓGICO EM ARTIGOS DE NOTÍCIAS UTILIZANDO APRENDIZADO MÉTRICO PROFUNDO E REPRESENTAÇÕES CONTEXTUAIS"

JAILSON PEREIRA JANUÁRIO

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA PELA BANCA EXAMINADORA CONSTITUÍDA PELOS PROFESSORES:

Prof. Dr. André Luiz da Costa Carvalho - PRESIDENTE

Prof. Dr. Juan Gabriel Colonna - MEMBRO INTERNO

Profa. Dra. Elloá Barreto Guedes da Costa - MEMBRO EXTERNO

MANAUS, 08 de novembro de 2024.

Documento assinado eletronicamente por **André Luiz da Costa Carvalho**, **Professor do Magistério Superior**, em 04/12/2024, às 11:25, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Juan Gabriel Colonna**, **Professor do Magistério Superior**, em 23/12/2024, às 18:35, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de

outubro de 2015.

Documento assinado eletronicamente por **Elloá Barreto Guedes da Costa**, **Usuário Externo**, em 26/12/2024, às 21:34, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2314225** e o código CRC **FA34F5DA**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

A minha família e amigos.

Agradecimentos

Expressar minha gratidão é uma tarefa que carrega uma grande responsabilidade, pois foram muitas as pessoas que estiveram ao meu lado durante esse período, e espero conseguir mencionar todas de forma justa. Primeiramente, agradeço à minha família, especialmente à minha mãe, que desde o meu primeiro momento neste mundo nunca deixou de acreditar em mim e de sonhar ao meu lado. Em muitas ocasiões, ela sonhou por mim, e com cuidado, carinho, amor e disciplina, me guiou para alcançar todos os objetivos que planejei.

Agradeço também aos meus antigos professores da UEA, que plantaram em mim a semente do desejo de sempre buscar ser um aluno, cientista e profissional com excelência técnica e humana. Em especial, deixo meu agradecimento ao LSI, que foi parte fundamental do início da minha trajetória acadêmica e profissional. A professora Elloá Guedes, o professor Carlos Maurício, o professor Fábio Santos e o professor Flávio Coelho merecem meu sincero obrigado por todo o conhecimento compartilhado e pela ajuda no início deste sonho.

Aos meus amigos, incluindo Juliany Raiol, Jéssica Tavares, Rafaela Sousa, Alice Adativa, Sarah Albuquerque, Paulo Soares e Dandara Sousa, sou imensamente grato. Alguns, mesmo à distância, me inspiraram a seguir em frente e a me tornar um profissional melhor. Um agradecimento especial ao meu querido Paulo Valber, por todo o cuidado e companheirismo nesta reta final.

Por fim, gostaria de expressar minha gratidão ao meu orientador, professor André Carvalho, que me acolheu, apoiou e incentivou durante todo o processo em que estivemos trabalhando juntos. Aos professores e funcionários do ICOMP/UFAM, e a todos que de alguma forma contribuíram para a minha formação, deixo o meu muito obrigado.

Você tem que brilhar tão forte que ninguém poderá te negar

(Blanca Evangelista)

Detecção de Viés Ideológico em Artigos de Notícias Utilizando Aprendizado Métrico Profundo e Representações Contextuais

Resumo

Diante da rápida expansão dos meios de comunicação digitais e da vasta disseminação de informações através de plataformas online, os portais de notícias na Web se estabeleceram como a principal fonte de informação para grande parte da população. A forma como essas informações são apresentadas pode exercer uma influência significativa sobre a opinião pública, moldando o discurso em torno de questões cruciais e, em última instância, impactando os processos de tomada de decisão política. Para aumentar a transparência das informações oferecidas ao público, é essencial desenvolver soluções que permitam a caracterização automática da orientação ideológica nas publicações de artigos de notícias. No entanto, as abordagens recentes têm se mostrado parcialmente eficazes nesse sentido, uma vez que frequentemente dependem de fontes externas e de múltiplos processos de geração de características, o que pode levar a imprecisões. Nesse cenário, este trabalho propõe métodos para a classificação de viés ideológico em artigos de notícias, fundamentados exclusivamente no seu conteúdo. Para isso, exploramos uma abordagem de aprendizado profundo que integra diferentes formas de representação textual, utilizando técnicas de aprendizado métrico profundo. O objetivo é demonstrar a eficácia dessa estratégia em comparação com a literatura existente. A abordagem proposta combina aprendizado contextual e semântico com recursos textuais, validando-os; em um conjunto de dados composto por mais de 30 mil artigos de notícias de temas variados.

Palavras-chave: Viés Ideológico, Aprendizado Profundo, Aprendizado Métrico, Processamento de Linguagem Natural, Contrastive Loss, Triplet Loss.

Detecção de Viés Ideológico em Artigos de Notícias Utilizando Aprendizado Métrico Profundo e Representações Contextuais

Abstract

Given the rapid expansion of digital media and the widespread dissemination of information through online platforms, web news portals have become the primary source of information for a large portion of the population. The way this information is presented and framed can significantly influence public opinion, shaping the discourse around critical issues and ultimately impacting political decision-making processes. To enhance the transparency of information provided to the public, it is essential to develop solutions that enable the automatic characterization of ideological orientation in news articles. However, recent approaches have proven only partially effective in this regard, as they often rely on external sources and multiple feature generation processes, which can lead to inaccuracies. In this context, this work proposes methods for classifying ideological bias in news articles based solely on the content of the articles from these portals. To achieve this, we explore a deep learning approach that integrates different forms of textual representation, employing deep metric learning techniques. The objective is to demonstrate the effectiveness of this strategy compared to the existing literature. The proposed approach combines contextual and semantic learning with textual features, validating it on a dataset composed of over 30,000 news articles covering a wide range of topics.

Keywords: Ideological Bias, Deep Learning, Natural Language Processing, Contrastive Loss, Triplet Loss.

Lista de figuras

Lista de tabelas

Lista de abreviaturas e siglas

BoW Bag of Words

DNN Deep Neural Network

BERT Bidirectional Encoder Representations from Transformers

GPT Generative Pre-trained Transformer

KNN k-Nearest Neighbors

LDA Latent Dirichlet Alocation

LSTM Long Short-Term Memory

PLN Processamento de Linguagem Natural

RoBERTa Robustly Optimized BERT Pretraining Approach

SVM Support Vector Machine

TF-IDF Term Frequency-Inverse Document Frequency

Lista de símbolos

μ Centróide

a, p, n Exemplos de dado âncora, positivo e negativo

C Cluster

d Documento

$d(x, y)$ Distância entre x e y

D Distância

L Função de

k, m Tamanho em unidade interia

Sumário

1 Introdução

Nos últimos anos, a quantidade de informações disponíveis online aumentou exponencialmente, especialmente por meio de portais de notícias. Este crescimento trouxe consigo o desafio de garantir a integridade e a imparcialidade das notícias disseminadas. O viés ideológico, presente em muitos artigos de notícias, pode distorcer a percepção pública dos eventos e influenciar a opinião pública de maneira significativa (GENTZKOW; SHAPIRO, 2006).

Pesquisas indicam que o viés presente nas mídias pode influenciar a maneira como os leitores percebem eventos, decisões políticas e discussões sobre diversos tópicos, chegando até a afetar os resultados eleitorais (CHIANG; KNIGHT, 2011; DELLAVIGNA; KAPLAN, 2007). No entanto, a tarefa de identificar o viés político ou ideológico da mídia é complexa, mesmo para os humanos, devido ao alto grau de subjetividade envolvido (YIGIT-SERT; ALTINGOVDE; ULUSOY, 2016). Nesse contexto, o desenvolvimento de métodos automáticos para abordar esse problema apresenta uma oportunidade de pesquisa significativa.

A identificação automática de viés em notícias é um desafio crucial no domínio do Processamento de Linguagem Natural (PLN) e aprendizado de máquina. Diversas pesquisas têm se dedicado a automatizar a detecção de viés ideológico em portais de notícias, abordando aspectos variados como o conteúdo textual (KRESTEL; WALL; NEJDL, 2012; SPINDE et al., 2022), hiperlinks e citações (EFRON, 2004; LIN; BAGROW; LAZER, 2011), dados provenientes de redes sociais (RIBEIRO et al., 2018) e princípios da Teoria da Informação (AIRES; NAKAMURA; NAKAMURA, 2019).

No entanto, essas abordagens ainda não fornecem um método suficientemente preciso para a detecção de viés ideológico, especialmente em um contexto amplo, pois tendem a focar em caracterizações, métricas específicas e estudos de caso. Além disso, muitas dessas metodologias dependem de fontes externas, o que dificulta o desenvolvimento de uma solução capaz de identificar

automaticamente diferentes tipos de viés em diversos contextos.

Neste estudo, buscamos desenvolver métodos baseados em técnicas de aproximação de *embeddings* para criar espaços de características que minimizem a distância entre artigos de notícias com o mesmo viés ideológico. Diferentemente das abordagens que se limitam à classificação de sentenças, à similaridade de documentos ou a uma única representação de informação, nossa pesquisa explora múltiplos aspectos textuais, incluindo o conteúdo completo dos artigos, a distribuição probabilística de tópicos e a classificação de documentos inteiros em vez de sentenças individuais.

Além disso, examinamos diferentes níveis de viés político/ideológico, incluindo espectros políticos centristas, ultrapassando uma análise meramente polarizada. Nossos métodos são projetados para ser independentes de fontes externas, utilizando apenas o conteúdo textual fornecido pelos portais de notícias para realizar essa classificação. Mostramos que essa abordagem resulta em um método mais eficaz do que as abordagens tradicionais da literatura, possibilitando o desenvolvimento de uma ferramenta automatizada para essa tarefa.

1.1 Problema

O viés ideológico em artigos de notícias é uma questão complexa que afeta a confiança do público nas mídias. A identificação manual desse viés é um processo demorado e sujeito a erros humanos. Portanto, há uma necessidade crescente de métodos automáticos e precisos para detectar e classificar o viés ideológico.

Ao examinar os trabalhos da literatura, podemos destacar métodos desenvolvidos por Baly et al. (2020), Spinder et al. (2022), Efron (2004), Lin et al. (2011), Aires et al. (2019) que criaram técnicas consideradas estado da arte para a detecção de viés ideológico. Esses autores têm como objetivo classificar artigos de notícias com base em sua ideologia. A classificação é

realizada utilizando aprendizado de máquina, considerando o conteúdo textual, links de referências externas ao artigo, ou técnicas de Teoria da Informação.

Observamos que as técnicas descritas pelos autores fornecem apenas caracterizações e métricas para mensurar viés ideológico ou apenas estudos de caso, muitas vezes limitados geograficamente e focados em caracterizar inclinações em relação a um partido específico, principalmente no contexto dos Estados Unidos (republicano/democrata). Outro aspecto notado é a dependência de fontes externas para alcançar uma classificação mais precisa, o que resulta em métodos pouco confiáveis na ausência dessas fontes. Além disso, verificamos que alguns dos métodos descritos utilizam um grande conjunto de atributos comuns no processamento de texto e se baseiam em um cenário exclusivamente polarizado (esquerda/direita, republicano/democrata).

Os fatores mencionados apresentam desafios significativos para o desenvolvimento de uma abordagem robusta e otimizada para classificar o viés ideológico em um contexto mais amplo. Isso inclui a consideração de diferentes níveis de viés (direita, esquerda e centro) em vários contextos e tópicos sociais. Portanto, este trabalho tem como objetivo desenvolver métodos que possam integrar essas características, aprimorando o estado da arte.

1.2 Objetivos

O objetivo geral deste trabalho é propor e avaliar métodos que combinam técnicas supervisionadas de aprendizado de máquina com aprendizado métrico profundo para a classificação de viés ideológico, utilizando aspectos extraídos de artigos de notícias, como conteúdo textual e tópicos relevantes sobre diferentes temas da sociedade.

Para atingir o que está sendo descrito neste trabalho, os seguintes objetivos específicos devem ser cumpridos:

- Propor um método eficaz e eficiente para classificação de viés ideológico baseado na combinação de Apendizagem de Máquina e Aprendizagem de

Métrica Profunda;

- Avaliar qualitativamente a combinação do conteúdo discursivo dos artigos de notícias com as probabilidades de tópicos para a geração de espaços de características;
- Avaliar o desempenho das funções *Contrastive Loss* e *Triplet Loss* na criação de espaços de características;
- Verificar/Avaliar comparativamente se os métodos propostos apresentam resultados superiores às abordagens existentes no estado da arte.

1.3 Contribuições

As principais contribuições geradas em decorrência do desenvolvimento desta pesquisa foram:

- Desenvolvimento de métodos para a classificação de viés ideológico em artigos de notícias, distinguindo-se das abordagens da literatura ao se basear exclusivamente no conteúdo dos artigos, sem a necessidade de fontes externas;
- Comparação de duas abordagens: uma baseada apenas no conteúdo textual dos artigos e outra que combina o conteúdo com a probabilidade de tópicos relacionados, avaliando qual maximiza a tarefa de classificação de viés ideológico. Aprensentando resultados com Macro F1-score de $45,87\%$ e $84,00\%$ para dois conjuntos de dados diferentes;
- Avaliação do uso de técnicas de aprendizado métrico profundo, especificamente das funções de perda *Contrastive Loss* e *Triplet Loss*, aplicadas a *embeddings* textuais gerados por modelos pré-treinados;
- Demonstração, em experimentos, de que o aprendizado métrico pode aprimorar o contexto e a semântica das representações textuais para a

classificação de artigos segundo sua ideologia, permitindo que os métodos alcancem um *Macro F1-score* superior a 83%, considerando diferentes tópicos.

1.4 Organização do Documento

Esta dissertação está organizada da seguinte forma: no Capítulo 2, apresentamos os fundamentos necessários para a compreensão deste trabalho; no Capítulo 3, realizamos uma revisão dos trabalhos relacionados e uma análise da literatura existente; no Capítulo 4, detalhamos a metodologia, que inclui uma abordagem baseada em conteúdo textual utilizando representações de *embeddings* de modelos pré-treinados, além da combinação dessas representações textuais com a probabilidade de tópicos; no Capítulo 5, descrevemos os resultados obtidos pelos métodos aplicados; e, por fim, no Capítulo 6, apresentamos as considerações finais desta pesquisa.

2 Fundamentação Teórica

Neste capítulo, serão apresentados os fundamentos essenciais para a compreensão desta dissertação. A Seção 2.1 abrange os conceitos relacionados ao problema da detecção de viés ideológico em artigos de notícias. Em seguida, na Seção 2.2, são descritas as técnicas e conceitos sobre representação de texto que serão utilizados neste trabalho. Na Seção 2.3, introduzimos alguns conceitos de aprendizagem de máquina, incluindo o classificador empregado. Posteriormente, na Seção 2.4 e na Seção 2.5, discutimos as técnicas de validação de modelos e as métricas de avaliação. Por fim, na Seção 2.6, apresentamos as considerações finais do capítulo.

2.1 Definições

No decorrer desta proposta, utilizaremos diferentes termos ao nos referirmos a tópicos relacionados ao problema de classificação de alinhamento ideológico em artigos de notícias. Portanto, é necessário definir claramente o que cada um desses termos representa. A seguir, apresentaremos as definições essenciais para a compreensão desta pesquisa.

2.1.1 Artigos em Portais de Notícias

O foco desta pesquisa é a classificação de alinhamento de artigos de notícias de portais de notícias na Web. Portanto, precisamos definir exatamente o que esses termos significam. Elaboramos as seguintes definições:

Definição 1: *Um artigo de notícia, ou simplesmente artigo, é um documento em hipertexto com URL acessível publicamente, divulgado em um portal de notícias.*

Definição 2: *Um portal ou fonte de notícias é um website que distribui*

notícias através da Web. Ele pode ou não possuir uma versão em outras mídias, como jornais tradicionais e televisão.

Os portais de notícias desempenham um papel crucial na sociedade moderna ao fornecer informações atualizadas e variadas para o público em geral (RICH, 1993) . Eles são essenciais para a formação da opinião pública, fornecendo um espaço para reportagens investigativas, análises aprofundadas e comentários sobre eventos atuais. Artigos de notícias, como componentes desses portais, são as unidades básicas de conteúdo que transmitem informações de maneira estruturada e acessível, permitindo que os leitores se mantenham informados sobre os acontecimentos mais recentes.

2.1.2 Viés ou Orientação Ideológica

O viés ou orientação ideológica em artigos de notícias refere-se à tendência de um veículo de comunicação favorecer uma ideologia, partido político ou perspectiva específica ao relatar eventos e questões políticas. Este fenômeno pode influenciar a maneira como as notícias são apresentadas, interpretadas e percebidas pelo público, potencialmente afetando a opinião pública e o comportamento eleitoral.

2.1.2.1 Tipos de Viés Ideológico

Viés ideológico é a inclinação parcial de um veículo de comunicação que resulta na apresentação das notícias de uma maneira que favoreça uma determinada perspectiva política ou ideológica. Shoemaker e Reese (1996) descrevem o viés na mídia como uma perspectiva parcial dos fatos, que pode se manifestar de várias formas, incluindo:

- **Viés de Seleção (*Gatekeeping*)**: A preferência em selecionar certas histórias ou eventos para serem cobertos enquanto ignora outros.
- **Viés de Cobertura**: A dedicação de maior tempo e espaço para certos

acontecimentos ou indivíduos, enfatizando ou minimizando aspectos específicos.

- **Viés de Declaração ou Posição Política**: A apresentação de comentários ou declarações de maneira mais favorável ou desfavorável a uma ideologia ou partido político (SHOEMAKER, 1996).

Diante desses conceitos, em nossa pesquisa estamos interessados em classificar as declarações ou posições políticas nos artigos ou publicações de portais de notícias.

Definição 3 ***Viés ideológico*** *ou* ***orientação ideológica*** *de um artigo de notícia é a preferência por uma ideologia ou partido político expressa através de declarações de texto em portais de notícias ou em meios de comunicação tradicional.*

2.2 Representação de Texto

A representação de texto é um aspecto fundamental do Processamento de Linguagem Natural (PLN), permitindo que os algoritmos de aprendizado de máquina compreendam e manipulem dados textuais. Existem várias abordagens para representar texto, das quais duas das mais proeminentes são as representações baseadas em *embeddings* e em modelos de tópicos.

2.2.1 Representação de Texto com *Embeddings*

A representação de texto com *embeddings* permite que palavras, frases e documentos sejam mapeados para vetores densos em um espaço de alta dimensionalidade. Esses vetores, conhecidos como *embeddings*, capturam as características semânticas e sintáticas do texto, permitindo que algoritmos de aprendizado de máquina compreendam e processem a linguagem natural de maneira mais eficaz (MIKOLOV et al., 2013).

2.2.1.1 *Embeddings* de Palavras

Os embeddings de palavras são vetores densos de baixa dimensionalidade que capturam as semânticas das palavras. Diferente das representações esparsas, como *Bag of Words* (BoW) ou *Term Frequency-Inverse Document Frequency* (TF-IDF), os *embeddings* posicionam palavras semelhantes mais próximas no espaço vetorial. Uma das primeiras e mais conhecidas técnicas de *embeddings* de palavras é o Word2Vec, desenvolvido por Mikolov et al. (2013). Word2Vec utiliza uma rede neural rasa para aprender vetores de palavras baseados em seu contexto em grandes corpora de texto.

A técnica de Skip-gram, uma das abordagens do Word2Vec, maximiza a probabilidade de palavras de contexto dada uma palavra-alvo. Formalmente, o objetivo é maximizar a função:

$$J = \sum_{t=1}^{T} \sum_{c \in C_t} logp(w_c|w_t) \tag{2.1}$$

onde $w_1, ..., w_t$ representa a sequencia de palavras do corpus de treinamento, C_t é o contexto do conjunto de indices das palavras em w_t. Este método permite a captura de relações semâticas e sintáticas entre palavras, como similaridade e analogias.

2.2.1.2 Embeddings de Sentenças e Documentos

Além dos *embeddings* de palavras, técnicas mais avançadas foram desenvolvidas para capturar o significado de sentenças e documentos inteiros. Modelos como BERT (*Bidirectional Encoder Representations from Transformers*), introduzido por Devlin et al. (2019), utilizam a arquitetura Transformer para gerar representações contextuais profundas de texto. BERT é pré-treinado em uma enorme quantidade de dados textuais para aprender representações bidirecionais, capturando o contexto de ambas as direções de uma sentença.

Os *embeddings* gerados por BERT são altamente eficazes em várias tarefas de PLN, como classificação de texto, resposta a perguntas e reconhecimento

de entidades nomeadas. A capacidade de BERT de compreender o contexto bidirecional, modelos ou técnicas de PLN que consideram o contexto de uma sequência de palavras tanto da esquerda para direita quanto da direita para esquerda, resulta em representações mais precisas e ricas, superando métodos tradicionais em *benchmarks* de desempenho.

2.2.2 Representação de Texto com Modelos de Tópicos

Os modelos de tópicos são uma abordagem probabilística para descobrir temas subjacentes em um conjunto de documentos. Um dos métodos mais populares é o *Latent Dirichlet Allocation* (LDA), introduzido por Blei et al. (2003). O LDA assume que cada documento é uma mistura de vários tópicos e que cada tópico é uma distribuição sobre palavras. O objetivo é identificar essas distribuições de forma a revelar os tópicos que estão presentes em um corpus de documentos.

A formulação matemática do LDA pode ser descrita da seguinte maneira: para cada documento d em um corpus, seleciona-se uma distribuição de tópicos t de uma distribuição de *Dirichlet*. Em seguida, para cada palavra no documento, escolhe-se um tópico de uma distribuição multinomial representado por θ_d e, finalmente, escolhe-se uma palavra a partir de uma distribuição multinomial associada ao tópico selecionado. Este processo gera uma mistura de tópicos para cada documento e uma mistura de palavras para cada tópico, permitindo a descoberta de temas latentes nos dados.

Os modelos de tópicos são amplamente utilizados em várias aplicações de PLN. Eles são eficazes na análise de grandes volumes de texto, como artigos científicos, publicações de mídia social e relatórios de notícias, permitindo a extração de temas principais e a agrupamento de documentos por tópicos. Além disso, os modelos de tópicos são utilizados em sistemas de recomendação para sugerir conteúdos baseados nos interesses temáticos dos usuários e na análise de sentimentos para identificar tópicos frequentemente associados a opiniões positivas ou negativas (WANG; BLEI, 2011).

2.3 Aprendizagem de Máquina

A aprendizagem de máquina é uma subárea da inteligência artificial que permite aos computadores aprenderem a partir de dados, sem serem explicitamente programados para cada tarefa específica. Em vez de seguir regras predefinidas, os algoritmos de aprendizagem de máquina analisam grandes conjuntos de dados para identificar padrões e tomar decisões baseadas nesses padrões. Por exemplo, em sistemas de recomendação como os utilizados pela Netflix [1] ou Amazon [2], os algoritmos aprendem as preferências dos usuários com base em seu histórico de visualizações e compras para sugerir novos filmes, séries ou produtos (MURPHY, 2012).

Existem diferentes tipos de aprendizagem de máquina, cada um com suas particularidades e aplicações específicas. A aprendizagem supervisionada, por exemplo, é utilizada para problemas onde os dados de entrada vêm acompanhados de rótulos, como na classificação de e-mails em spam ou não-spam (BISHOP; NASRABADI, 2006). Já a aprendizagem não supervisionada não utiliza rótulos e é frequentemente aplicada em análises exploratórias de dados, como a segmentação de clientes em marketing, onde o objetivo é descobrir grupos de consumidores com comportamentos semelhantes (HASTIE et al., 2009).

Em problemas de classificação, os modelos são chamados de classificadores, cuja função principal é examinar um conjunto de dados para reconhecer padrões que permitam categorizá-los, ou seja, dividi-los em diferentes classes (DIETTERICH, 2000). Entre os classificadores mais usados estão: SVM (*Support Vector Machine*, ou Máquina de Vetores de Suporte), Naive Bayes, kNN (*k-Nearest Neighbors*, ou k-Vizinhos mais Próximos) e RandomForest (Floresta Aleatória) (VINODHINI; CHANDRASEKARAN, 2012). Além de modelos que seguem uma abordagem não supervisionada para clusterização preditiva como *Latent Dirichlet Allocation* (LDA) e o K-means. A seguir, iremos detalhar os conceitos utilizados neste trabalho.

[1] <https://www.netflix.com/>
[2] <https://www.amazon.com.br/>

2.3.1 K-Nearest Neighbors

O *K-Nearest Neighbors* (KNN) é um algoritmo de aprendizado supervisionado amplamente utilizado para problemas de classificação e regressão. Introduzido por Fix e Hodges (1951), o k-NN é conhecido por sua simplicidade e eficácia, especialmente em tarefas de classificação onde a relação entre as características e a classe alvo é complexa ou desconhecida. O princípio básico do KNN é que uma amostra é classificada com base nas classes dos seus k vizinhos mais próximos no espaço de características.

O KNN opera com base base na premissa de que objetos próximos no espaço de características tendem a pertencer à mesma classe. O processo de classificação pelo k-NN pode ser descrito da seguinte forma:

1. Escolha de k: Determinar o número de vizinhos k a serem considerados.

2. Cálculo da distância entre a amostra de teste e todas as amostras de treinamento. A distância euclidiana é a mais comumente usada e definida como:
$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2.2}$$
em que x e y são vetores de características de duas amostras.

3. Identificação dos Vizinhos: Identificar os k vizinhos mais próximos da amostra de teste com base nas distâncias calculadas.

4. Classificação: Atribuir a classe mais comum entre os k vizinhos à amostra de teste.

2.3.2 Clusterização Preditiva

A Clusterização preditiva é uma abordagem que combina técnicas de clusterização com aprendizado supervisionado para prever a associação de novas amostras a grupos ou clusters existentes, com base em padrões aprendidos de dados históricos (EZUGWU et al., 2022). Essa técnica é usada principalmente

quando há a necessidade de agrupar dados em categorias latentes (sem rótulos explícitos), mas com a capacidade de prever para novos dados usando os agrupamentos descoberto. Um dos algoritmos mais utilizados é o K-means.

2.3.2.1 K-means

O K-means é um algoritmo de aprendizado não supervisionado amplamente utilizado para problemas de clusterização. Desenvolvido por MacQueen (1967), o K-means tem como objetivo dividir um conjunto de dados em k clusters distintos, onde cada ponto de dados pertence ao cluster com a média mais próxima (centróide). Este algoritmo é comumente utilizado devido à sua simplicidade e eficiência em diversas aplicações, como segmentação de mercado, compressão de imagem e análise de padrões.

O K-means opera em um processo iterativo para minimizar a variabilidade dentro dos clusters. O algoritmo pode ser descrito pelos seguintes passos:

1. Inicialização: Escolher k centros iniciais (centroides) aleatoriamente a partir dos dados.

2. Atribuição: Atribuir cada ponto de dados ao centroide mais próximo, formando k clusters. A distância euclidiana é frequentemente usada para calcular a proximidade:

3. Atualização: Recalcular os centroides como a média dos pontos de dados em cada cluster:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{2.3}$$

   onde μ_i é o centroide do cluster C_i e $|C_i|$ é o número de pontos no cluster C_i.

2.3.3 Aprendizagem de Máquina Profunda

A aprendizagem de máquina profunda, ou *deep learning*, representa um avanço significativo na área de inteligência artificial, permitindo a modelagem de dados com alta complexidade através de redes neurais profundas. Estas redes são compostas por múltiplas camadas de neurônios artificiais que aprendem a representar dados em níveis progressivamente mais abstratos. A capacidade de capturar padrões complexos e não lineares tornou a aprendizagem profunda essencial para diversas aplicações, desde reconhecimento de imagem até PLN (LECUN; BENGIO; HINTON, 2015).

Um dos maiores avanços no campo de PLN foi a introdução dos modelos transformadores, que revolucionaram a maneira como a linguagem é processada por máquinas. O Transformer, proposto por Vaswani et al. (2017), eliminou a necessidade de redes recorrentes, utilizando em seu lugar mecanismos de atenção para processar palavras em paralelo. Este modelo permite que a atenção seja distribuída de forma mais eficiente tais como das palavras em uma sentença, capturando dependências a longo prazo de maneira mais eficaz.

A arquitetura do Transformer foi a base para desenvolvimentos subsequentes, incluindo os modelos BERT (*Bidirectional Encoder Representations from Transformers*) e GPT (*Generative Pre-trained Transformer*), que mostraram desempenho superior em diversas tarefas de PLN (VASWANI et al., 2017).

2.3.3.1 *Bidirectional Encoder Representations from Transformers* (BERT)

BERT é baseado na arquitetura Transformer, introduzida por Vaswani et al. (2017), que utiliza mecanismos de atenção para processar dados sequenciais de forma mais eficaz. Diferentemente dos modelos tradicionais unidirecionais, que analisam o texto em uma única direção (esquerda para direita ou vice-versa), BERT é bidirecional. Isso significa que, ao processar uma palavra em uma sentença, BERT considera tanto o contexto à esquerda quanto à direita da palavra, proporcionando uma compreensão mais profunda do significado e

nuances do texto.

A principal inovação do BERT é seu método de pré-treinamento em duas tarefas principais: *Masked Language Model* (MLM) e *Next Sentence Prediction* (NSP). No MLM, algumas palavras na sentença são mascaradas aleatoriamente, e o modelo deve prever essas palavras mascaradas com base no contexto fornecido pelas outras palavras. Isso força o modelo a aprender relações contextuais mais ricas entre as palavras. Na tarefa de NSP, o modelo recebe pares de sentenças e deve prever se a segunda sentença é a continuação da primeira, o que ajuda a modelar a relação entre sentenças.

A aplicabilidade do BERT se estende a muitos domínios. Em assistentes virtuais e chatbots, por exemplo, BERT melhora a compreensão de consultas complexas dos usuários, proporcionando respostas mais precisas e contextualmente relevantes. Na análise de sentimentos, BERT é capaz de captar nuances emocionais em textos, melhorando a precisão das avaliações de sentimentos em redes sociais e *feedback* de clientes. Além disso, em sistemas de recomendação, BERT ajuda a entender melhor o conteúdo textual e as preferências dos usuários, oferecendo recomendações mais personalizadas (SUN et al., 2019).

2.3.4 Aprendizagem Métrica Profunda

A aprendizagem métrica profunda é uma área emergente dentro do campo de aprendizado de máquina e aprendizado profundo que se concentra na definição e otimização de funções de distância ou similaridade para aprender representações úteis dos dados. Diferentemente dos métodos tradicionais que utilizam métricas pré-definidas como a distância euclidiana, a aprendizagem métrica profunda visa aprender uma métrica diretamente dos dados que é otimizada para a tarefa específica, como classificação, recuperação de informações ou verificação de identidade (KAYA; BILGE, 2019).

Esta modalidade de aprendizado baseia-se na ideia de que representações eficazes dos dados podem ser aprendidas de forma que exemplos seme-

lhantes estejam próximos no espaço de representação, enquanto exemplos diferentes estejam distantes. Esta abordagem é particularmente útil em problemas onde a definição de similaridade é complexa e dependente do contexto, como em reconhecimento de face, emparelhamento de imagens e recuperação de informações.

2.3.5 Distância Euclidiana

A Distância Euclidiana é uma medida matemática que cálcula o comprimento da linha reta entre dois pontos em um espaço multidimensional. Ela é amplamente utilizada em diversas áreas, como análise de dados, aprendizado de máquina e geometria (BISHOP; NASRABADI, 2006). A fórmula geral para dois pontos $P(x_1, y_1)$ e $Q(x_2, y_2)$ no plano cartesiano é:

$$D(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2.4}$$

Essa métrica é particularmente útil qualdo precisamos calcular a aproximidade em aplicações como agrupamentos ou em algoritmos de aprendizado supervisionado, como o KNN.

2.3.5.1 Funções de Perda

Para treinar modelos de aprendizagem métrica profunda, várias funções de perda são utilizadas para guiar o processo de aprendizado. As duas mais comuns são a *Contrastive Loss* e a *Triplet Loss*.

A *Contrastive Loss*, introduzida por Hadsell et al. (2006), utiliza pares de exemplos (âncora e positiva/negativa) para aprender a métrica. A função de perda é definida como:

$$L = (1 - y)\frac{1}{2}D^2 + y\frac{1}{2}\{max(0, m - D)\}^2 \tag{2.5}$$

onde y é o rótulo binário que indica se os pares são semelhantes (0) ou diferentes (1), D é a distância entre as representações dos exemplos, e m é uma margem que separa pares positivos de negativos.

A *Triplet Loss* por sua vez, utiliza de três exemplos: âncora, positiva (similar à âncora) e negativa (diferente da âncora). A *Triplet Loss* é definida como:

$$L = max(0, D(a,p) - D(a,n) + m) \tag{2.6}$$

onde $D(a,p)$ é a distância entre a âncora a e positivo p, $D(a,n)$ é a distância entre âncora a e o negativo n, e m é uma margem que garante que a distância positiva seja menor que a distância negativa por pelo menos m. Nesta definição:

1. *easy triplets* (trigêmio fácil): trigêmio que possuem a perda igual a 0 porque $D(a,p) + m < D(a,n)$.

2. *hard triplets* (trigêmio dificeis): trigêmios que o exemplo negativo n está mais próximo do a do p, ou seja $D(a,n) < D(a,p)$.

3. *semi-hard triplets* (trigêmio semi-dificeis): trigêmio que o n não está mais próximo do a do que do p, mas que ainda assim apresenta perda positiva de forma que $D(a,p) < D(a,n) + m$.

No entanto, às vezes todos os trigêmeos podem ser menos relevantes para o treinamento do modelo. Se o algoritmo for treinado com exemplos trigêmios fácil, o modelo pode ter uma saída abaixo do ideal por não generalizar bem (SONG et al., 2016). Em contraste, o processo de treinamento pode ser ineficiente se o modelo tiver sido treinado em exemplos trigêmios dificeis, nos quais as instâncias âncora e positiva já estão próximas (SCHROFF; KALENICHENKO; PHILBIN, 2015).

Dessa maneira, o emprego da estratégia de selecão de trigêmios ajuda escolher exemplos que contribuem para o aprendizado bem-sucedido. A estratégia

consiste em selecionar trigêmios difíceis que são difíceis para o modelo categorizar corretamente e evitando trigêmio fáceis que o modelo pode classificar com precisão.

2.3.5.2 Arquiteturas

Diversas arquiteturas de redes neurais profundas são empregadas na implementação da aprendizagem métrica, destacando-se as redes siamesas e as redes de *triplet* como as mais comuns. As redes siamesas são compostas por dois ramos idênticos de redes neurais que compartilham os mesmos pesos. Cada ramo é responsável por processar uma das entradas (âncora e positiva/negativa) e gerar uma representação vetorial correspondente. A função de perda contrastiva é então aplicada às saídas dessas representações, com o objetivo de treinar a rede para minimizar a distância entre pares semelhantes e maximizar a distância entre pares distintos.

Por sua vez, as redes de *triplet* representam uma extensão das redes siamesas, sendo projetadas para processar simultaneamente três entradas: âncora, positiva e negativa. A função de perda de *triplet* é utilizada para ajustar as representações de modo que a âncora esteja mais próxima da positiva do que da negativa por uma margem específica. Este método tem se mostrado particularmente eficaz em tarefas como o reconhecimento facial, essencial diferenciar identidades semelhantes com alta precisão.

2.4 Técnicas de Validação de Modelos

A validação de modelos é um componente crucial no desenvolvimento de algoritmos de aprendizado de máquina, garantindo que os modelos generalizem bem para dados não vistos. Sem validação adequada, um modelo pode se ajustar excessivamente aos dados de treinamento, resultando em sobreajuste (*overfitting*), ou pode ser incapaz de capturar padrões subjacentes, resultando

em subajuste (*underfitting*). Nesta seção, descreveremos duas técinas de validação de modelos: a validação cruzada k-fold e validação cruzada estratificada (*Stratified Cross-Validation*).

2.4.1 Validação Cruzada k-fold

A validação cruzada k-fold é uma técnica amplamente utilizada que envolve a divisão do conjunto de dados em k subconjuntos (folds) aproximadamente iguais. O modelo é treinado k vezes, cada vez utilizando k-1 folds para treinamento e o fold restante para validação. A média das métricas de desempenho dos k experimentos fornece uma estimativa mais robusta da eficácia do modelo. Formalmente, o procedimento pode ser descrito da seguinte forma:

1. Dividir o conjunto de dados em k folds.
2. Para cada i-ésimo fold:
    - Utilizar os k-1 folds restantes para treinar o modelo.
    - Avaliar o modelo no i-ésimo fold.
3. Calcular a média das métricas de desempenho ao longo dos k folds.

A validação cruzada k-fold é especialmente útil para conjuntos de dados pequenos, onde maximizar o uso de dados para treinamento e validação é crítico.

2.4.2 Validação Cruzada Estratificada

A validação cruzada estratificada k-fold é uma variante da validação cruzada k-fold que preserva a proporção de classes em cada fold. Este método é particularmente importante para conjuntos de dados desequilibrados, onde a distribuição das classes é desproporcional. A estratificação garante que cada fold seja representativo da distribuição original das classes, proporcionando uma avaliação mais equilibrada do desempenho do modelo. O procedimento

é semelhante ao da validação cruzada k-fold, com a adição de uma etapa de estratificação antes da divisão dos dados.

2.5 Métricas de Avaliação

Existem várias maneiras de avaliar a eficácia dos algoritmos de aprendizado. As medidas da qualidade da classificação são construídas a partir de uma matriz de confusão que registra exemplos corretamente e incorretamente reconhecidos para cada classe (SOKOLOVA; JAPKOWICZ; SZPAKOWICZ, 2006).

A Tabela 2.5 apresenta uma matriz de confusão para classificação binária, onde tp é o número de instâncias da classe positiva que foram corretamente classificadas, fp o número de instâncias que foram clçassificadas como positivas mas perteciam à classe negativa, fn o número de instâncias da classe positiva que foram incorretamente classificadas como negativasfalsos negativos e tn o número de instâncias da classe negativa corretamente classificadas. Nesta seção, descrevemos três medidas comumente utilizadas na litaretura para avaliar a eficácia de modelos preditivos: *precisão*, *revocação*, *acurácia* e *medida F1-score*.

Tabela 1 – Matriz de confusão para classificação binária.

		Predição	
		Positivo	Negativo
Classe real	Positivo	tp	fn
	Negativo	fp	tn

2.5.1 Precisão

Essa métrica corresponde à porcentagem de itens classificados como positivos que, de fato, são positivos (BAEZA-YATES; RIBEIRO-NETO, 2013). É dado pela razão entre o número de verdadeitos positivos e a soma de verdadeiros e falsos positivos, ou seja,

$$prec = \frac{tp}{tp + fp} \tag{2.7}$$

2.5.2 Revocação

Esta métrica consiste na porcentagem de positivos que são classificados como positivos. É formulada como a razão entre os verdaeiros positivos e a soma dos verdadeiros positivos e falsos negativos, ou seja,

$$revoc = \frac{tp}{tp + fn} \tag{2.8}$$

2.5.3 Acurácia

A acurácia representa a proporção de previsões corretas em relação ao total de previsões feitas, fornecendo uma visão geral da capacidade de acertar suas previsões. A fórmula para calcular a acurácia é dada por:

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{2.9}$$

2.5.4 F1-score

F1-score é a média harmônica da precisão e revocação. Relaciona métricas de precisão e recall para obter uma medida de qualidade que equilibre a importância relativa dessa duas métricas (BAEZA-YATES; RIBEIRO-NETO, 2013), e é definida como:

$$F1 = \frac{2 \times (prec \times revoc)}{prec + revoc} \tag{2.10}$$

A macro F1-score (F1-macro) corresponde à média de todos os valores F1 das categorias consideradas.

2.6 Considerações Finais

Neste capítulo, apresentamos os fundamentos teóricos que embasam o desenvolvimento deste trabalho, abrangendo os conceitos de viés ideológico em artigos de notícias, fundamentos de representação de dados, aprendizagem de máquina e métricas para medir a eficácia dos modelos. No próximo capítulo, serão discutidos os trabalhos relacionados, que também utilizam alguns dos conceitos abordados neste capítulo.

3 Trabalhos Relacionados

Neste capítulo, apresentaremos uma breve revisão dos métodos utilizados para a classificação de viés ou orientação ideológica em portais de notícias. Observamos que, na literatura, diversos aspectos são considerados para realizar essa análise. Assim, listaremos os trabalhos segundo o aspecto focado na classificação de viés, dividindo-os em: baseados em hiperlinks presentes nas páginas web dos portais de notícias (Seção 3.1), no conteúdo textual dos artigos (Seção 3.2) e em dados provenientes de redes sociais (Seção 3.3). Na Seção 3.4, discutimos os trabalhos apresentados. Por fim, na Seção 3.5, apresentamos as considerações do capítulo.

3.1 Métodos Baseados em Hiperlinks

Artigos de notícias são disponibilizados em páginas web e frequentemente contêm hiperlinks, que podem referenciar outras seções do próprio documento ou direcionar para outros documentos. Os estudos descritos a seguir utilizam os hiperlinks presentes em páginas web de artigos de notícias, ou em outros documentos em hipertexto, para a tarefa de detecção de orientação política.

Um dos primeiros estudos sistemáticos sobre viés ideológico em portais de notícias foi realizado por Efron (2004), que introduziu um método baseado em hiperlinks para estimar a orientação política de documentos hipertexto utilizando informações de cocitação. O objetivo era determinar se um documento específico estava associado a comunidades de esquerda ou de direita. Para isso, o método utilizava um modelo probabilístico que avaliava a probabilidade de cocitação entre o documento de interesse e alguns documentos de referência, cuja orientação política era previamente conhecida. Os resultados mostraram que o modelo proposto superou classificadores baseados em léxicos, como Naive Bayes e SVM.

Groseclose e Milyo (2006), por sua vez, exploram o fenômeno do viés na mídia e propõem um modelo teórico para entender como as reputações dos veículos de comunicação influenciam a apresentação tendenciosa das notícias. Os autores argumentam que os veículos de mídia equilibram a necessidade de atrair e manter audiências com a necessidade de preservar sua reputação de imparcialidade. Utilizando um modelo econômico, Gentzkow e Shapiro mostram que os veículos de mídia tendem a fornecer informações que se alinham com as preferências ideológicas de seu público-alvo para maximizar seu lucro e audiência. O método criado pelos autores envolve a análise de padrões de consumo de mídia e a correlação entre as preferências do público e a inclinação ideológica percebida dos veículos de comunicação, fornecendo insights sobre a dinâmica entre viés na mídia e comportamento do consumidor.

Lin et al. (2011) examinam o viés ideológico ao analisar a estrutura e o conteúdo das redes de comunicação. Os autores criaram um método que combina análise de redes sociais com técnicas de processamento de linguagem natural para quantificar o viés presente em diferentes fontes de mídia. Eles construíram grafos que representam as interações entre fontes de notícias e as entidades mencionadas, aplicando algoritmos para detectar comunidades e medir a centralidade das fontes. Com isso, puderam identificar padrões de viés ideológico, demonstrando como certas fontes tendem a se agrupar e promover narrativas semelhantes. Este método oferece uma abordagem quantitativa para entender como o viés se manifesta nas redes de comunicação, proporcionando uma visão mais abrangente e sistemática do viés na mídia.

3.2 Métodos Baseados em Conteúdo Textual

Ao considerar a classificação de orientação ideológica em notícias, o aspecto mais intuitivo é o próprio conteúdo dos artigos, ou seja, o texto presente neles. A seguir, descrevemos estudos que utilizam análise textual para desenvolver métodos de detecção de viés em artigos jornalísticos.

Dallmann et al. (2015) investigaram o viés ideológico em jornais online alemães. Utilizando técnicas de análise de conteúdo e processamento de linguagem natural, os autores examinaram como diferentes jornais apresentam e enquadram notícias de maneira tendenciosa. O estudo avaliou a inclinação política das publicações, levando em consideração fatores como escolha de palavras, ênfase em determinados temas e a frequência de cobertura de assuntos políticos. Os resultados revelaram uma variação significativa na forma como os jornais tratam questões políticas, refletindo suas respectivas tendências ideológicas e influenciando a percepção pública sobre eventos e políticas.

Gangula et al. (2019) exploraram uma abordagem para identificar o viés ideológico em notícias, concentrando-se na análise das manchetes. Utilizando técnicas avançadas de processamento de linguagem natural e aprendizado de máquina, os autores propuseram um modelo que emprega mecanismos de atenção para captar a orientação ideológica implícita nas manchetes. O estudo destaca a importância das manchetes como elementos críticos que podem influenciar a percepção dos leitores antes mesmo de lerem o corpo do texto. Os resultados demonstram que focar nas manchetes pode ser uma ferramenta eficaz para detectar e quantificar o viés político, contribuindo para uma melhor compreensão e combate à desinformação na mídia.

Utilizando o conteúdo textual em um contexto mais amplo, Baly et al. (2020) desenvolveram um método para detecção de viés ideológico em artigos de notícias utilizando técnicas de processamento de linguagem natural e aprendizado de máquina. O estudo emprega um conjunto de dados robusto que foi particionado de duas maneiras: *random*, onde os artigos de texto com as mesmas fontes foram divididos entre as partições de treino e teste, e *media-bias*, onde os artigos de texto de diferentes fontes foram separados nas partições de treino e teste. Os autores demonstraram que essa abordagem foi mais bem-sucedida em classificar viés político em um conjunto de dados contendo artigos de fontes já vistas durante o treinamento, alcançando uma acurácia de $79,83\%$.

3.3 Métodos Baseados em Dados de Redes Sociais

Além dos aspectos presentes nas páginas web contendo artigos de notícias, dados de redes sociais podem ser recursos valiosos para detectar o viés político de um portal de notícias. A seguir, descrevemos trabalhos que utilizam características provenientes de redes sociais em suas abordagens.

Rao e Spasojevic (2016), utilizam *word embeddings* e *Long Short-Term Memory* (LSTM) para classificar o viés político em tweets. A abordagem desenvolvida pelos autores testou diferentes parâmetros das redes neurais e observou seus efeitos na precisão. No geral, os resultados alcançaram 87% de acurácia no melhor modelo para classificar os *tweets* como democratas ou republicanos.

Elejalde et al. (2017), por sua vez, apresentaram uma abordagem para quantificar a inclinação política dos veículos de notícias do Chile com base na automação de um questionário político denominado O Menor Questionário Político do Mundo (ou *The World's Smallest Political Quiz*) [1]. Para isso, os autores utilizaram *tweets* para gerar as previsões das respostas para cada pergunta sobre os assuntos relacionados às questões abordadas no questionário. Dessa forma, os portais são posicionados em um plano cartesiano que informa sobre suas orientações políticas. Além disso, os autores observaram que o viés político se reflete no vocabulário escolhido e nas entidades tratadas pelos portais.

Por fim, Ribeiro et al. (2018) empregaram uma metodologia diferente para inferir o viés político de fontes de notícias em mídias sociais como Facebook e Twitter. A ideia principal é utilizar as interfaces de anúncios, que fornecem informações demográficas sobre a audiência de uma determinada página. Assim, os autores demonstraram que a orientação ideológica (liberal ou conservadora) de uma fonte de notícias pode ser estimada com base em sua audiência, ou seja, a partir da presença de leitores liberais e conservadores entre o público dessa fonte.

[1] <https://www.theadvocates.org/quiz/>

3.4 Sínteses dos Trabalhos Relacionados

Os métodos apresentados foram classificados de acordo com a abordagem escolhida para realizar a detecção de orientação ideológica em portais de notícias e outras fontes, como *tweets*. A Tabela 3.4 resume os trabalhos segundo as estratégias desenvolvidas, o idioma e as classes de viés ideológico abordadas.

Tabela 2 – Resumo dos trabalhos relacionados.

Autor/ano	Estratégia	Idioma	Classes	Desempenho
Efron (2004)	Co-citação/hiperlinks	Inglês	Liberal / conservador	77,50% (acurácia)
Groseclose e Milyo (2006)	Métricas empíricas	Inglês	Não define	-
Lin et al. (2011)	Métricas empíricas	Inglês	Democrata / republicano	-
Dallmann et al. (2015)	Métricas de similaridade	Alemão	Partido alemão	-
Rao e Spajevic (2016)	Word embeddings e LSTM	Inglês	Democrata / republicano	87% (acurácia)
Eljalde et al. (2017)	Diferença rank	Espanhol	Liberal / conservador	-
Ribeiro et al. (2018)	Estatísticas de audiência em mídias sociais	Inglês	Liberal / conservador	-
Gangula et al. (2019)	Headeline Network	Inglês	Não define	-
Baly et al. (2020)	Embeddings com BERT e LSTM	Inglês	Esquerda / direita / centro	79,83% (macro f1-score)

A revisão dos trabalhos indica que muitos oferecem apenas estudos de caso ou caracterizações muito específicas, analisando um conjunto restrito de fontes e não fornecendo uma análise abrangente do desempenho das estratégias. Métodos com uma abordagem mais geral, como os propostos por Efron (2004), Baly et al. (2020) e Rao e Spajevic (2016), também apresentam suas limitações. O método de Efron, por exemplo, não tem um bom desempenho na classificação de páginas web com poucos hiperlinks. Já a abordagem de Baly et al. (2020) não generaliza bem para artigos de notícias provenientes de fontes não incluídas no treinamento. Embora o método de Rao e Spajevic tenha mostrado bom desempenho, ele é limitado a uma abordagem polarizada, restrita a partidos específicos de um determinado país.

Para abordar essas limitações, nesta pesquisa propomos um método para classificação de viés ideológico em artigos de notícias que possui as seguintes características: (i) independência de fontes externas, utilizando apenas o conteúdo textual dos artigos de notícias; (ii) foco na classificação de artigos

inteiros, em vez de apenas sentenças, *tweets* ou comentários; (iii) uma análise menos polarizada, que possa ser generalizada para diferentes casos.

Nesse cenário, consideramos que a abordagem descrita por Baly et al. (2020), em sua versão que utiliza apenas o conteúdo textual, é a mais adequada como linha de base comparativa entre os trabalhos relacionados mencionados anteriormente, por ser a mais próxima ao nosso contexto e por disponibilizar o mesmo conjunto de dados sob as mesmas condições para treinamento e avaliação.

3.5 Considerações Finais

Neste capítulo, apresentamos uma breve revisão dos trabalhos relacionados ao problema da classificação de viés ou orientação ideológica em artigos de portais de notícias. Os trabalhos foram divididos de acordo com os aspectos/estratégias escolhidas: hiperlinks, conteúdo textual e dados de redes sociais. Ao considerar esses estudos, percebemos a oportunidade de desenvolver métodos com desempenho superior ao estado da arte, que dependam apenas do conteúdo dos próprios artigos de notícias e sejam especificamente focados na detecção de viés ideológico. No próximo capítulo, apresentaremos nossa abordagem para a classificação de viés ideológico com base no conteúdo textual.

4 Abordagens Propostas

Neste capítulo, são apresentadas as abordagens propostas para a classificação de viés ideológico em artigos de notícias, baseadas tanto no seu conteúdo textual quanto na combinação de representações textuais com a sua distribuição de probabilidade de tópicos.

O capítulo está organizado da seguinte forma: na Seção 4.1, detalha-se a metodologia desenvolvida para a abordagem fundamentada no conteúdo textual. Em seguida, na Seção 4.2, descreve-se a abordagem que integra representações textuais com a análise de probabilidade de tópicos.

4.1 Método Baseado em Conteúdo Textual

A metodologia adotada neste estudo baseia-se na hipótese de que artigos com o mesmo viés ideológico compartilham formas de discurso semelhantes, ou seja, exibem alinhamento ideológico em sua estrutura e conteúdo. Assim, a metodologia concentra-se na análise do conteúdo textual dos artigos de notícias, com particular atenção às diferenças no discurso entre publicações de fontes com diferentes orientações ideológicas.

Especificamente, nossa abordagem emprega modelos pré-treinados para gerar *embeddings* textuais, que são subsequentemente refinados por meio de *fine-tuning* utilizando funções de perda como a *Contrastive Loss* e *Triplet Loss*. Essas funções de perda são aplicadas diretamente durante o processo de treinamento para garantir que os *embeddings* gerados capturem de forma eficaz as diferenças ideológicas entre as classes de viés político. Após esse refinamento, os *embeddings* são usados como entrada para um modelo de classificação. A Figura 1 ilustra o processo metodológico adotado no presente estudo, que consiste nos seguintes passos:

1. Aquisição de dados com base em trabalhos relacionados;

2. Geração de *embeddings* de artigos de notícias utilizando modelos pré-treinados e refinados com *Contrastive Loss* e *Triplet Loss*;

3. Treinamento do modelo de classificação utilizando essas representações textuais otimizadas na forma de *embeddings* com modelos de aprendizado de máquina;

4. Avaliação dos resultados.

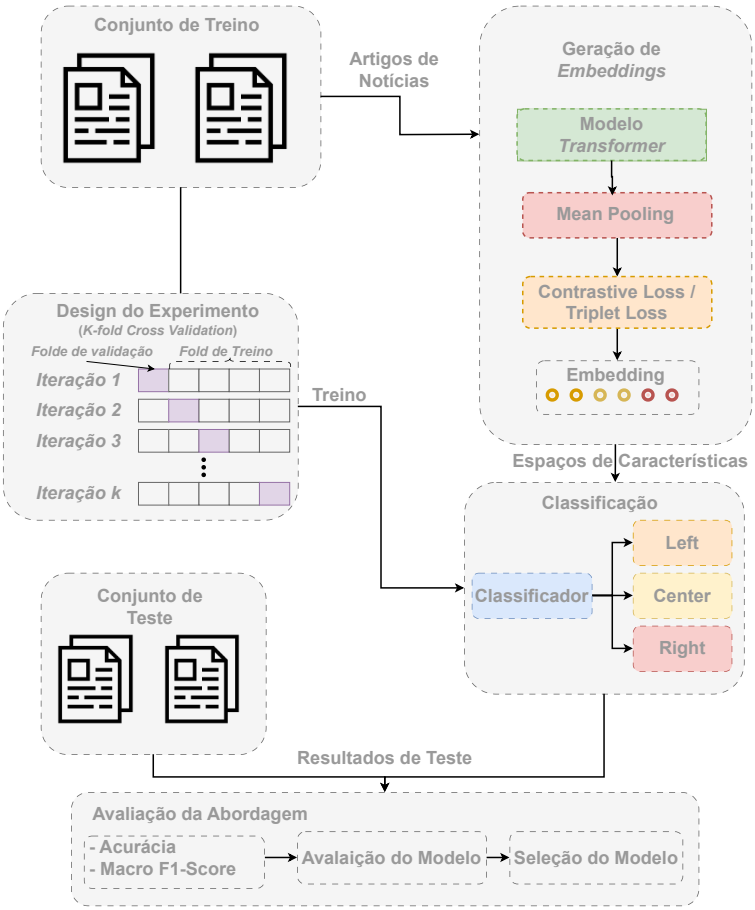


Figura 1 – Visão geral do método de classificação baseado em conteúdo textual.

4.1.1 Conjunto de Dados

A aquisição dos dados de treinamento foi realizada conforme o trabalho de Baly et al. (2020), que desenvolveram um conjunto de dados composto por uma coleção diversificada de artigos noticiosos, selecionados a partir de uma

ampla gama de portais de notícias com diferentes orientações ideológicas. Para garantir a representatividade e a qualidade do conjunto de dados, os autores utilizaram a plataforma AllSides [1], que fornece anotações de ideologia política para artigos individuais.

Estas anotações são obtidas por meio de um processo rigoroso que inclui análises para identificação de viés, revisões editoriais, análises de terceiros, revisões independentes e feedback da comunidade. Este método assegura que os dados sejam altamente confiáveis e adequados tanto para o treinamento quanto para o teste de modelos de aprendizado de máquina, permitindo uma análise precisa das divergências discursivas entre diferentes fontes de notícias (BALY et al., 2020).

Baly et al. (2020), denominaram o cojunto de dados como *Article Bias Prediction* (ABP) e o dividiram em três partições: Treino, Validação e Teste, totalizando 30.246 amostras de artigos de texto extraídos de portais de notícias em língua inglesa. Esses artigos foram rotulados de maneira supervisionada em três classes de viés ideológico: *left*, *right* e *center*. A Figura 2 ilustra um exemplo de como um tópico de notícia foi classificado conforme a ideologia contida no texto.

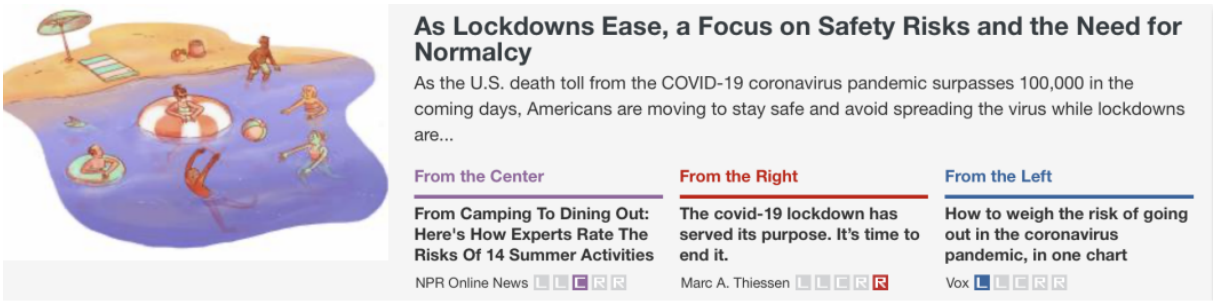


Figura 2 – Exemplo de amostra de artigo de notícia com o tópico sobre a pandemia do coronavírus. (BALY et al., 2020).

Conforme ilustrado na Tabela 3, observam-se características distintivas nas amostras de textos provenientes do ABP, ressaltando a importância desta base de dados. As amostras consistem em publicações de diversos portais de notícias, cobrindo uma ampla gama de tópicos, desde questões relacionadas a

[1] <https://www.allsides.com/media-bias/media-bias-rating-methods>

eleições até discussões sobre racismo. Essa diversidade temática evidencia a extensão dos assuntos contemplados nas amostras de dados.

Tabela 3 – Estatísticas sobre o conjunto de dados.

Viés ideológico	*Quantidade*	*Porcentagem*
Left	12003	34,6%
Center	9743	28,1%
Right	12991	37,3%

Em nossa metodologia, empregamos o conjunto completo de amostras do ABP nas etapas de geração de *embeddings*, extração de tópicos, treinamento e avaliação dos modelos de classificação, conforme as tarefas de aprendizado de máquina.

4.1.2 Geração de *Embeddings*

Para realizar essa tarefa, utilizamos dois modelos pré-treinados baseados em *Bidirectional Encoder Representations from Transformers* (BERT) (DEVLIN et al., 2019): DitilBERT (SANH et al., 2020) e RoBERTa (LIU et al., 2019). Esses modelos oferecem um aprimoramento na modelagem de dependências de longo alcance no texto, capturando relações semânticas entre palavras distantes dentro de uma sentença ou documento. Essa habilidade permite que os modelos compreendam melhor o contexto e a estrutura das sentenças, resultando em representações vetoriais mais precisas e eficazes para tarefas de processamento de linguagem natural (GAO; YAO; CHEN, 2021).

O primeiro modelo avaliado é o DistilBERT (SANH et al., 2020), uma versão compacta e eficiente do BERT, projetada para preservar o desempenho robusto do BERT original, enquanto reduz significativamente os requisitos de memória e computação. O segundo modelo considerado é o *Robustly Optimized BERT Pretraining Approach* (RoBERTa) (LIU et al., 2019), uma variação do BERT que passou por um treinamento estendido com um volume maior de dados, resultando em desempenho superior em diversas tarefas.

Para ambos os modelos, realizamos o processo de *fine-tuning* usando Aprendizado Métrico, aplicando as funções de perda *Triplet Loss* e *Contrastive Loss*. Esse processo nos permitiu otimizar os modelos para gerar embeddings que melhor preservam as relações semânticas nos dados, garantindo que exemplos semanticamente ou contextualmente semelhantes tenham vetores mais próximos no espaço de *embeddings*, enquanto exemplos distintos sejam devidamente separados.

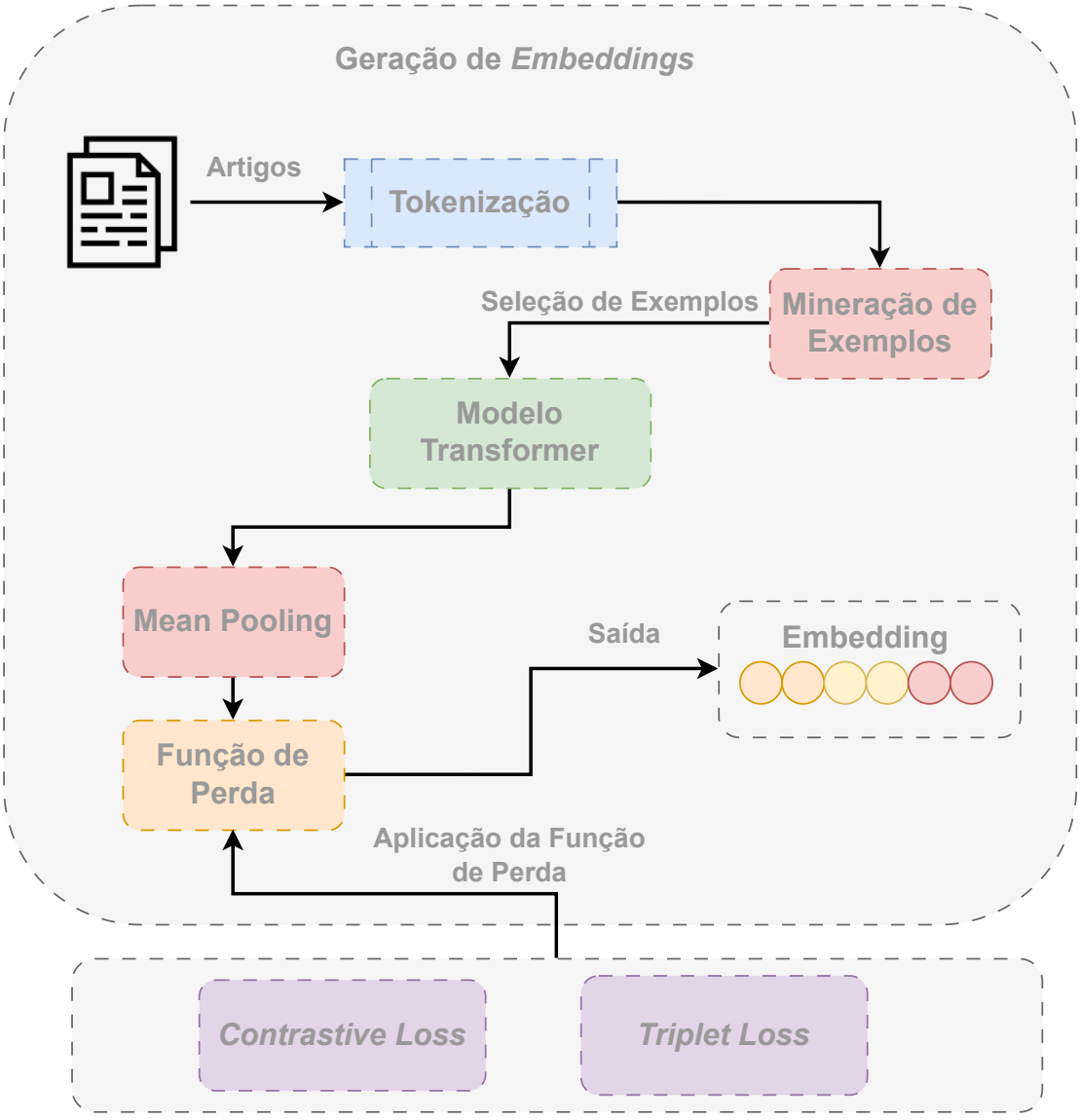


Figura 3 – Processo de treinamento e geração de espaço de características para artigos de notícias.

Durante o processo, os dados foram tokenizados usando o tokenizer

específico de cada modelo, e os vetores tokenizados resultantes foram truncados ou preenchidos para atingir o comprimento de 512 tokens, como ilustrado na Figura 3. É importante destacar que as stopwords foram mantidas, pois os modelos baseados em BERT são capazes de capturar as nuances contextuais em que essas palavras aparecem. Além disso, foi aplicada a *Mean Pooling* ao longo da dimensão dos tokens, resultando em uma única representação vetorial que captura as características mais relevantes de cada dimensão (BALY et al., 2020).

4.1.3 Funções de Perda

Em nossa abordagem, os codificadores (DistilBERT e RoBERTa) foram pré-treinados utilizando as funções de perda *Contrastive Loss* e *Triplet Loss*. A *Contrastive Loss* foi aplicada de maneira convencional, utilizando a distância euclidiana, conforme detalhado no Capítulo 2, em amostras formadas por pares de exemplos positivos e negativos. Por outro lado, o modelo com *Triplet Loss* foi treinado em um conjunto de triplas, cada uma contendo uma âncora, um exemplo positivo e um exemplo negativo. A aplicação da distância euclidiana garantiu que o exemplo positivo estivesse sempre mais próximo da âncora do que o exemplo negativo, como ilustrado na Figura 4.

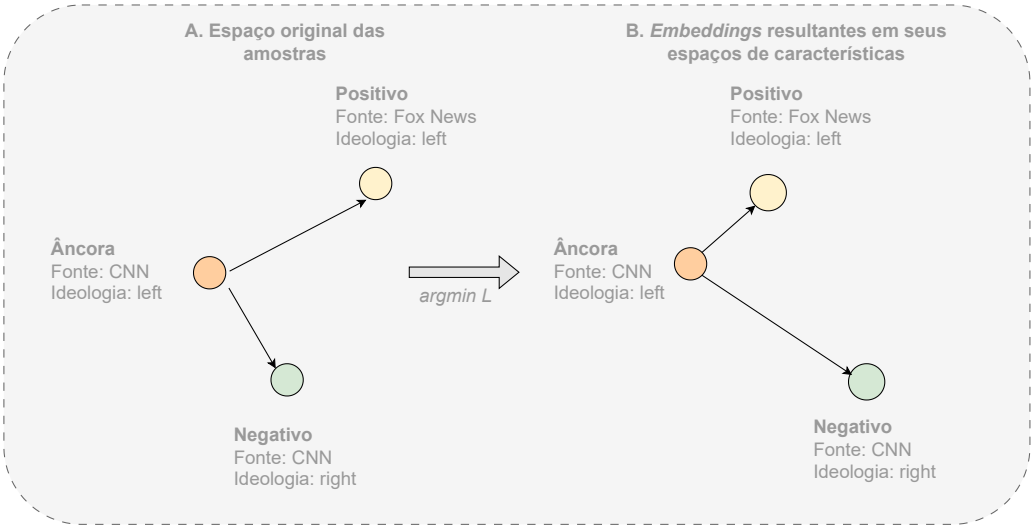


Figura 4 – Exemplo de uso da *Triplet Loss*.

A Figura 4 apresenta um exemplo da aplicação da *Triplet Loss*. Nesse exemplo, o caso positivo possui a mesma ideologia do âncora, porém é publicado por um meio de comunicação diferente. Já o caso negativo, embora publicado pelo mesmo meio de comunicação, apresenta uma ideologia distinta do âncora. Dessa maneira, o codificador tende a agrupar exemplos com ideologias semelhantes de forma próxima, independentemente de sua origem.

Além disso, utilizamos a mineração de exemplo para capturar amostras negativas semi-difíceis, descritas no Capítulo 2, que são suficientemente próximas do âncora para serem consideradas negativos, conforme ilustrado na Figura 5. A seleção de triplas negativas semi-difíceis oferece informações mais valiosas ao modelo e reduz o risco de *overfitting* em relação a amostras com negativos difíceis. Esse processo melhora a capacidade do modelo em discriminar entre textos similares e dissimilares (KERTéSZ, 2021). Consequentemente, os *embeddings* gerados passam a refletir com maior precisão as relações semânticas.

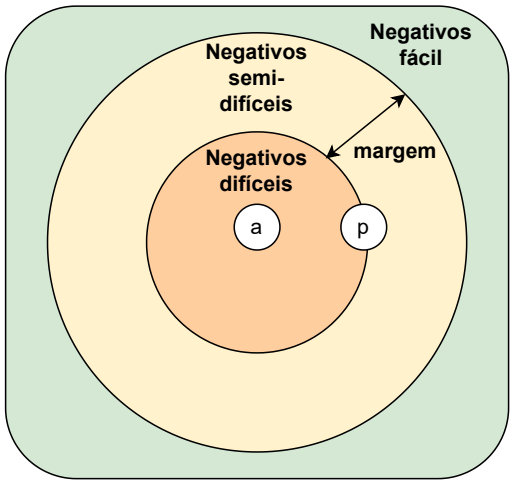


Figura 5 – Regiões dos espaços dos exemplos negativos difíceis, semi-difíceis e fáceis. O exemplo âncora é representado pela letra a e o exemplo positivo pela letra p.

Após o treinamento dos modelos pré-treinados e a geração dos *embeddings*, esses *embeddings* serão submetidos à etapa de classificação, conforme o método proposto.

4.1.4 Modelo de Classificação

Após a geração dos *embeddings* nos respectivos espaços de características, foram utilizadas três abordagens de classificação: *K-Nearest Neighbors* (KNN), *K-Means* e *Deep Neural Network* (DNN). O objetivo principal destas abordagens é classificar artigos de notícias com base nos espaços latentes gerados, permitindo a avaliação da proximidade entre instâncias semelhantes. Partiu-se da premissa de que artigos posicionados próximos no espaço latente apresentam uma orientação ideológica similar, facilitando a classificação com base nessa proximidade.

O modelo KNN foi configurado com uma gama de valores para k, pois essa abordagem permite determinar o número ótimo de vizinhos a serem considerados na tarefa de classificação, garantindo maior precisão na modelagem. A métrica de distância adotada foi a distância euclidiana, escolhida porque oferece simplicidade no cálculo e é amplamente eficaz na análise de dados em espaços de alta dimensão, onde diferenças entre ponstos são significativas (DUARTE; SILVA, 2019). Já o algoritmo *K-means* foi aplicado com um número fixo de três clusters, porque o objetivo era segmentar os dados em três grupos distintos, refletindo a quantidade de classes existentes no conjunto de dados.

O classificador baseado na DNN foi projetado para processar a saída dos modelos de transformadores, sendo composto por dois blocos de camadas densas totalmente conectadas, com 512 e 256 neurônios, respectivamente. Essa arquitetura foi escolhida para reduzir gradualmente a dimensionalidade do vetor de entrada, preservando informações relevantes e facilitando a generalização do modelo. A função de ativação *ReLU* foi empregada nas camadas internas para mitigar o problema do desaparecimento do gradiente (ZHANG; RAO, 2020). A camada de saída é ajustada para a tarefa de classificação de viés ideológico e utiliza a função de ativação *softmax*, como ilustrado na Figura 6. O treinamento da DNN foi conduzido utilizando o algoritmo de otimização *Adam*, e a função de perda foi calculada através da entropia cruzada categórica.

A validação cruzada com 5 folds foi empregada para avaliar a perfor-

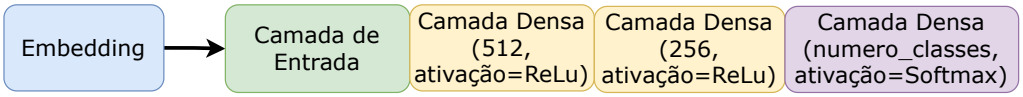


Figura 6 – Modelo de DNN elaborado para a tarefa de classificação de viés ideológico.

mance dos três modelos, assegurando que cada instância do conjunto de dados fosse utilizada tanto para treinamento quanto para validação, o que possibilita uma estimativa confiável da acurácia e da robustez dos modelos no conjunto de tete. Além disso, foi adotada a validação cruzada estratificada (*Stratified Cross-Validation* para garantir que cada fold de treinamento e validação mantivesse aproximadamente a mesma proporção de classes presente no conjunto de dados original. Essa abordagem foi fundamental para evitar viés na avaliação e assegurar que a distribuição das classes fosse representativa em todos os folds, promovendo uma avaliação mais precisa da capacidade de generalização dos modelos.

4.1.5 Avaliando Desempenho

A performance dos modelos de classificação serão medidas utilizando duas métricas principais: acurácia e *Macro F1-score*. A acurácia será empregada para determinar a proporção de previsões corretas feitas pelo modelo em relação ao total de casos analisados, enquanto o *Macro F1-score* fornecerá uma média harmônica entre a precisão e a sensibilidade, oferecendo uma visão mais equilibrada da performance do modelo, especialmente em situações de classes desbalanceadas. Caso a hipótese seja verdadeira e os discursos de cada artigo de notícia estejam relacionados ao viés ideológico, o classificador deve ser capar de atribuir o viés de cada artigo corretamente.

4.2 Método baseado em Conteúdo Textual e Probabilidade de Tópicos

A segunda abordagem que investigamos para a classificação do viés político baseia-se tanto no conteúdo textual quanto nas informações latentes dos tópicos presentes nos artigos de notícias. Kim et al. (2022) destacam que a modelagem de tópicos pode ser uma ferramenta eficaz para identificar e quantificar temas específicos que indicam uma orientação ideológica. Com base nisso, combinamos as representações textuais geradas por modelos pré-treinados com as probabilidades dos tópicos para caracterizar o viés ideológico dos artigos. A Figura 7 ilustra detalhadamente as etapas dessa abordagem, que segue um processo semelhante ao descrito na Seção 4.1, utilizando o mesmo conjunto de dados e metodologia para geração dos embeddings de texto com modelos pré-treinados. A seguir, descrevemos os detalhes de cada etapa.

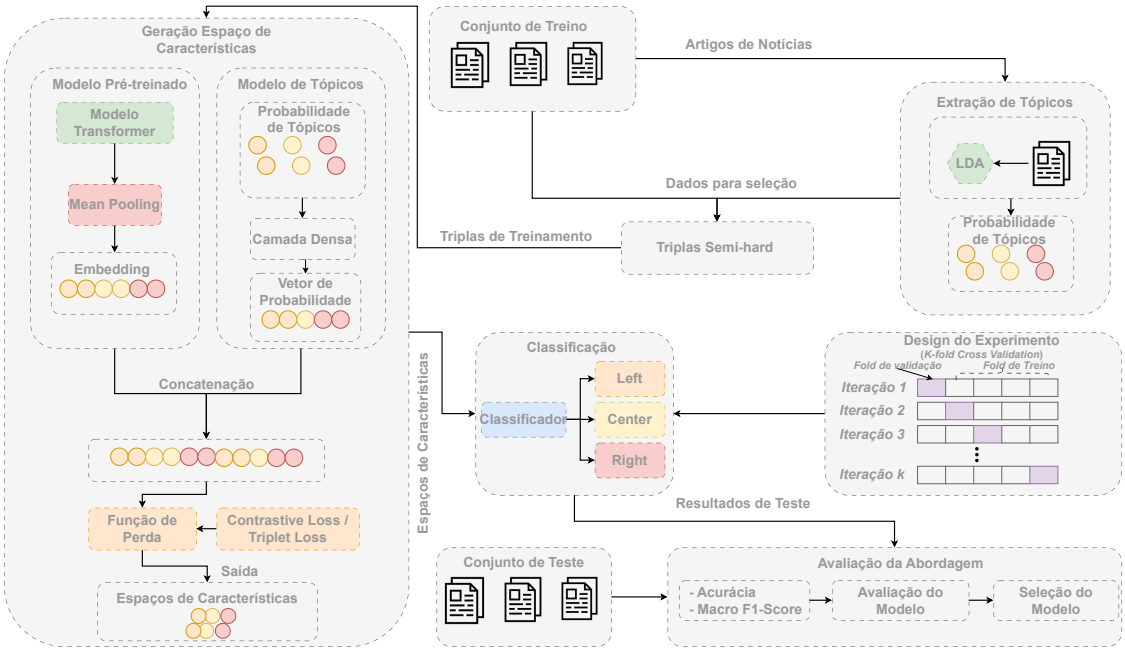


Figura 7 – Visão geral do método de classificação baseado em conteúdo textual e probabilidade de tópicos.

4.2.1 Extração de Tópicos

Para a extração de tópicos de texto, utilizou-se o *Latent Dirichlet Allocation* (LDA), uma técnica de modelagem de tópicos amplamente adotada em análise de texto. O LDA foi aplicado para identificar conjuntos de palavras que tendem a co-ocorrer em documentos, revelando os tópicos subjacentes ao corpus textual.

Foram conduzidas três rodadas de extração de tópicos, especificando-se 10, 15 e 20 tópicos, respectivamente, para explorar a granularidade temática em diferentes níveis. A escolha de múltiplos números de tópicos permitiu a comparação e avaliação da coesão e distinção dos tópicos identificados, proporcionando uma compreensão mais detalhada e robusta das estruturas temáticas presentes no conjunto de dados analisado.

4.2.2 Geração de Espaços de Características

Nesta etapa, a arquitetura combina as entradas das informações contextuais e temáticas, conforme ilustrado na Figura 7. As entradas consistem em dois componentes:

1. **Representações contextuais**: Um vetor de 768 dimensões correspondente às representações textuais com a aplicação de *Mean Pooling*.

2. **Vetores de probabilidade de tópicos**: Um vetor de dimensão T, representando as probabilidades dos tópicos geradas pelo modelo LDA, conforme descrito na Seção 4.2.1.

Essas duas entradas foram concatenadas, resultando em um vetor de dimensão $768 + t$, onde t representa o número de tópicos. Este vetor é então passado por uma camada densa da mesma dimensão. Durante o treinamento da rede neural, usamos as funções *Contrastive Loss* e *Triplet Loss*, com o objetivo de minimizar a distância entre representações de dados similares e maximizar a distância entre exemplos distintos. Além disso, adotamos a mesma estratégia de mineração de exemplos semi-difíceis descrita anteriormente na

Seção 4.1.3, onde a preparação dos triplos de treinamento se concentra em exemplos negativos desafiadores que ainda são classificados incorretamente.

Após o treinamento do modelo neural que combina representações textuais e probabilidades de tópicos, as representações geradas em um espaço latente comum, na forma de embeddings, foram submetidas à etapa de classificação.

4.2.3 Modelo de Classificação e Avaliação de Desempenho

De forma análoga à abordagem baseada em conteúdo textual, os modelos de classificação KNN, K-Means e DNN adotados nesta abordagem seguem os mesmos princípios descritos na Seção 4.1.4. Esses modelos foram avaliados com as métricas de acurácia e *Macro F1-Score*, conforme a definição apresentada no Capítulo 2. Após o treinamento, o modelo foi avaliado no conjunto de teste para medir sua eficácia em classificar corretamente o viés ideológico dos artigos de notícia.

4.3 Considerações Finais

Neste capítulo, foram apresentadas duas abordagens distintas para a classificação de viés ideológico em artigos de notícias, baseadas na análise de conteúdo textual e na combinação de representações textuais com a probabilidade de tópicos. A primeira abordagem fundamenta-se na hipótese de que artigos com o mesmo viés ideológico compartilham discursos semelhantes, explorada por meio da geração de *embeddings* contextuais utilizando modelos pré-treinados, como o DistilBERT e o RoBERTa. A segunda abordagem complementa essa análise ao incorporar a modelagem de tópicos através do LDA, proporcionando uma caracterização mais detalhada dos conteúdos ideológicos.

Os resultados dessas abordagens são discutidos no Capítulo 5, onde foram realizados testes utilizando o conjunto de dados particionado de duas

maneiras distintas, baseando-se nas fontes de origem dos artigos de notícias.

5 Resultados Obtidos

Este capítulo apresenta os resultados obtidos a partir das metodologias propostas para a classificação de viés ideológico em artigos de notícias. Inicialmente, na Seção 5.1, são descritos os dados experimentais utilizados para o treinamento, validação e teste dos modelos de aprendizado. Em seguida, na Seção 5.2, detalhamos os parâmetros e configurações adotados em cada abordagem. Na Seção 5.3 e na Seção 5.4, são apresentados os resultados das abordagens baseadas em conteúdo textual e na combinação de representações textuais com a probabilidade de tópicos, respectivamente. Finalmente, na Seção 5.5, realizamos uma sumarização e comparação dos experimentos conduzidos.

5.1 Dados Experimentais

O conjunto de dados utilizado neste estudo, conforme mencionado na Seção 4.1.1, é o *Article Bias Prediction* (ABP), desenvolvido por Baly et al. (2020). Uma característica distintiva desse conjunto é sua representatividade em relação a diferentes perspectivas políticas sobre vários tópicos e eventos. A maioria dos rótulos dos artigos no ABP está fortemente alinhada com os rótulos das fontes de onde os artigos foram extraídos. Isso sugere que os classificadores de aprendizado de máquina poderiam acabar modelando a fonte, em vez de captar a ideologia política expressa nos artigos.

Para mitigar esse problema, os autores adotaram um pré-processamento dos artigos, removendo marcadores explícitos, como o nome dos autores e o nome do meio de comunicação, que frequentemente aparecem no conteúdo dos textos. Esse cuidado visa garantir que a ideologia política seja modelada com base nas informações linguísticas presentes nos artigos. Além disso, o conjunto de dados foi dividido de duas maneiras distintas para permitir uma análise mais robusta: (i) *media-bias split* e (ii) *random split*.

O *media-bias split* foi estruturado para incluir 1.300 artigos provenientes de 12 meios de comunicação no conjunto de teste. Dessa forma, os conjuntos de treino e validação são compostos por artigos de 61 outros meios de comunicação, garantindo que todos os artigos de um mesmo meio de comunicação estejam presentes apenas em um dos conjuntos: treinamento, validação ou teste. Esse arranjo assegura que os modelos sejam treinados e testados em artigos cujas fontes não foram previamente expostas durante o treinamento. A Tabela 4 apresenta a distribuição da quantidade e porcentagem de exemplos por classe nesse particionamento.

Tabela 4 – Estatísticas sobre a partição *media-bias split*.

Treino			**Validação**			**Teste**		
Viés	*Total*	*Porcentagem*	*Viés*	*Total*	*Porcentagem*	*Viés*	*Total*	*Porcentagem*
left	8861	33.32%	left	1640	69.60%	left	402	30.92%
center	7488	28.16%	center	618	26.23%	center	299	23.0%
right	10241	38.51%	right	98	4.15%	right	599	46.07%

O *random split* mantém o mesmo conjunto de exemplos de artigos da partição de teste do *media-bias split*. No entanto, as partições de treino e validação contêm exemplos de fontes que também aparecem na partição de teste. Isso assegura que os modelos sejam ajustados e avaliados apenas em artigos cujas fontes foram observadas durante o treinamento. A Tabela 5 apresenta as estatísticas dos exemplos nas partições de treino, validação e teste.

Tabela 5 – Estatísticas sobre a partição *random split*.

Treino			**Validação**			**Teste**		
Viés	*Total*	*Porcentagem*	*Viés*	*Total*	*Porcentagem*	*Viés*	*Total*	*Porcentagem*
left	9750	34.84%	left	2438	34.84%	left	402	30.92%
center	7988	28.55%	center	1998	28.55%	center	299	23.0%
right	10240	36.60%	right	2560	36.59%	right	599	46.07%

5.2 Estratégia de Treinamento, Parâmetros e Hiperparâmetros

Seguindo as partições definidas pelo ABP, os exemplos de artigos de notícias nas partições de treino e validação foram utilizados tanto para a geração de embeddings quanto para o treinamento dos modelos de classificação. Os modelos DistilBERT e RoBERTa foram retreinados utilizando os seguintes hiperparâmetros: o otimizador *Adam* foi empregado para a estimação do gradiente descendente no algoritmo de *backpropagation*, com uma taxa de aprendizado fixada em $0,0001$, um tamanho de *batch* de 16 e um número máximo de épocas igual a 100. A escolha desses hiperparâmetros foi conduzida de maneira empírica. Para mitigar o *overfitting*, o treinamento incorporou a técnica de *early stopping* com um valor de *patience* igual a 30, monitorando-se a variação da perda no conjunto de validação.

A estratégia de validação cruzada estratificada foi adotada para o treinamento dos modelos de classificação, conforme descrito na Seção 4.1.1. Adicionalmente, foi utilizada uma busca em grid para realizar a otimização sistemática dos hiperparâmetros do classificador KNN, onde o número de vizinhos variou entre $5, 10, 15, 20, 25$ e 30. Para o K-Means, o algoritmo foi treinado considerando um número de clusters equivalente ao número de classes presente no conjunto de dados do ABP. A DNN foi treinada com os mesmos hiperparâmetros utilizados na geração dos *embeddings*, diferenciando-se na função de perda, onde foi empregada a *cross entropy*, uma função de perda que mede a divergência entre duas distribuições de probabilidade.

Para a implementação do modelo proposto, foi adotado um conjunto de bibliotecas na plataforma Python, a citar: Numpy [1], Pandas [2], Scikit-Learn [3] e PyTorch [4]. Além disso, a execução do treinamento e teste do modelo proposto fez uso de um desktop com processador Intel Xeon W-2235, 128 GB de memória

[1] <http://www.numpy.org/>
[2] <https://pandas.pydata.org/>
[3] <https://scikit-learn.org/stable/>
[4] <https://pytorch.org/>

principal, 2 TB de memória segundária e uma placa de vídeo NVIDIA Quadro RTX 8000 com 48 GB de memória.

5.3 Classificação do Viés Ideológico Baseada em Conteúdo Textual

Nesta abordagem, os classificadores foram avaliados exclusivamente com base nas representações textuais dos conjuntos *media-bias split* e *random split*. Para cada artigo, o modelo treinado atribui uma classificação de viés ideológico, categorizando-o como *center*, *left* ou *right*. Para avaliar o desempenho dos modelos nessa tarefa, utilizamos duas métricas principais: Acurácia e *Macro F1-score*. Além disso, empregamos a matriz de confusão para verificar a frequência com que os modelos classificam corretamente o viés ideológico de um artigo de notícia. Os resultados para todos os métodos testados, com os melhores valores destacados em negrito, são apresentados na Tabela 6.

Tabela 6 – Performance das métricas no conjunto de teste na tarefa de classificação de viés ideológico.

Modelo	**Split**	**Representação**	**Função de Perda**	**Classificador**	***Macro F1-score***	**Acurácia**
Baly et al. (2020)		GloVe - embeddings	-	LSTM	31,51	32,30%
Baly et al. (2020)		BERT - embeddings	-	**BERT**	**35,53**	**36,75**%
1		DistilBERT - embeddings	*Contrastive Loss*	KNN	39,41	42,33%
2		DistilBERT - embeddings	*Contrastive Loss*	Kmeans	37,86	40,21%
3		DistilBERT - embeddings	*Contrastive Loss*	DNN	33,92	33,83%
4		RoBERTa - embeddings	*Contrastive Loss*	**KNN**	**45,44**	**48,89%**
5	*Media-bias split*	RoBERTa - embeddings	*Contrastive Loss*	Kmeans	29,42	33,36%
6		RoBERTa - embeddings	*Contrastive Loss*	DNN	39,62	39,14%
7		DistilBERT - embeddings	*Triplet Loss*	KNN	41,92	45,41%
8		DistilBERT - embeddings	*Triplet Loss*	Kmeans	23,29	28,76%
9		DistilBERT - embeddings	*Triplet Loss*	DNN	36,61	36,73%
10		RoBERTa - embeddings	*Triplet Loss*	KNN	42,91	45,61%
11		RoBERTa - embeddings	*Triplet Loss*	Kmeans	36,37	40,75%
12		RoBERTa - embeddings	*Triplet Loss*	DNN	39,65	39,52%
Baly et al. (2020)		GloVe - embeddings	-	LSTM	65,50	66,17%
Baly et al. (2020)		BERT - embeddings	-	**BERT**	**80,19**	**79,83**%
13		DistilBERT - embeddings	*Contrastive Loss*	KNN	78,39	78,01%
14		DistilBERT - embeddings	*Contrastive Loss*	Kmeans	81,86	81,85%
15		DistilBERT - embeddings	*Contrastive Loss*	DNN	77,92	76,87%
16		RoBERTa - embeddings	*Contrastive Loss*	KNN	82,26	82,01%
17	*Random split*	RoBERTa - embeddings	*Contrastive Loss*	Kmeans	29,41	42,33%
18		RoBERTa - embeddings	*Contrastive Loss*	**DNN**	**83,90**	**83,89**%
19		DistilBERT - embeddings	*Triplet Loss*	KNN	78,91	77,84%
20		DistilBERT - embeddings	*Triplet Loss*	Kmeans	73,87	71,61%
21		DistilBERT - embeddings	*Triplet Loss*	DNN	77,88	77,24%
22		RoBERTa - embeddings	*Triplet Loss*	KNN	83,84	83,30%
23		RoBERTa - embeddings	*Triplet Loss*	Kmeans	79,92	80,92%
24		RoBERTa - embeddings	*Triplet Loss*	DNN	81,89	80,66%

Como demonstrado, o modelo KNN utilizando *embeddings* gerados pelo

RoBERTa com *Contrastive Loss* obteve os melhores resultados gerais nas duas métricas avaliadas para o conjunto *media-bias split*. Esse modelo superou os demais modelos testados, bem como o melhor *baseline* desenvolvido por Baly et al. (2020), com uma melhoria de $9,91\%$ no *Macro F1-score*. Em relação ao conjunto *random split*, a DNN com *embeddings* gerados pelo RoBERTa e *Contrastive Loss* apresentou desempenho superior em comparação com outras variações testadas e também com o melhor *baseline*, resultando em uma melhoria de $3,73\%$.

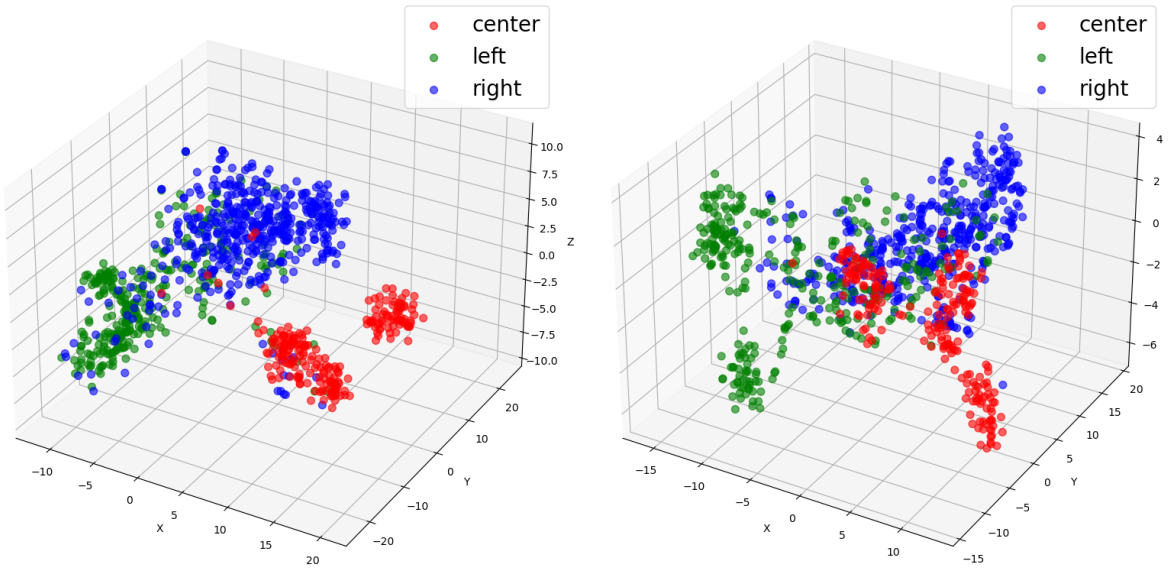
Ao considerar o desempenho dos modelos, observa-se uma ligeira superioridade nos experimentos que utilizaram o classificador KNN no conjunto *media-bias split*. As evidências sugerem que a simplicidade deste algoritmo não paramétrico facilita sua interpretação e ajuste, permitindo um aproveitamento eficaz dos *embeddings* de alta qualidade fornecidos pelos modelos pré-treinados.

A similaridade nos resultados entre o KNN e a DNN no conjunto *random split* pode ser atribuída a essa alta qualidade dos *embeddings* gerados pelos modelos pré-treinados. Essas representações vetoriais capturam de forma eficiente as nuances semânticas e contextuais dos textos, o que pode diminuir a necessidade de técnicas de classificação mais complexas. Tanto o KNN quanto a DNN conseguem utilizar essas representações ricas para identificar padrões nas classes ideológicas. No caso do KNN, ele baseia suas decisões de classificação diretamente na proximidade no espaço vetorial, enquanto a DNN, apesar de sua capacidade de modelar relações mais complexas, pode não proporcionar um ganho substancial em desempenho quando os *embeddings* já são altamente discriminativos. Assim, ambos os métodos apresentam desempenho semelhante, beneficiando-se das representações robustas oferecidas pelos modelos pré-treinados.

Ao analisar os componentes dos experimentos, nota-se que o desempenho um pouco inferior do DistilBERT em relação ao RoBERTa na geração de *embeddings* para a classificação de viés ideológico pode ser atribuído à

diferença na capacidade de representação entre os dois modelos. O DistilBERT, por ser uma versão mais compacta e eficiente do BERT, possui menos parâmetros, o que reduz sua capacidade de capturar nuances complexas no texto. Essa limitação pode impactar sua eficácia em tarefas que requerem uma compreensão aprofundada do contexto, como a detecção de viés ideológico. Apesar dos valores de *Macro F1-score* serem próximos, o RoBERTa, com sua arquitetura otimizada e treinamento em um volume maior de dados, consegue gerar *embeddings* mais ricos e detalhados, o que resulta em um desempenho superior.

Em relação à *Contrastive Loss* e à *Triplet Loss*, observa-se um desempenho semelhante, o que pode ser atribuído à alta qualidade dos *embeddings* gerados pelos modelos pré-treinados, que já promovem uma boa separação nos espaços vetoriais, bem como à limitação do conjunto de dados em termos do número de exemplos negativos ou da diversidade entre eles. Essas circunstâncias impediram que as vantagens teóricas de uma função de perda sobre a outra se traduzissem em uma diferença significativa na prática para a tarefa de classificação de viés ideológico.



(a) Embeddings with Contrastive Loss

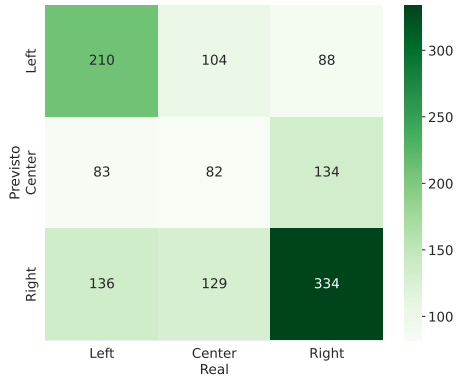
(b) Embeddings with Triplet Loss

Figura 8 – Visualização dos embeddings gerados com o RoBERTa no conjunto do *random split*.

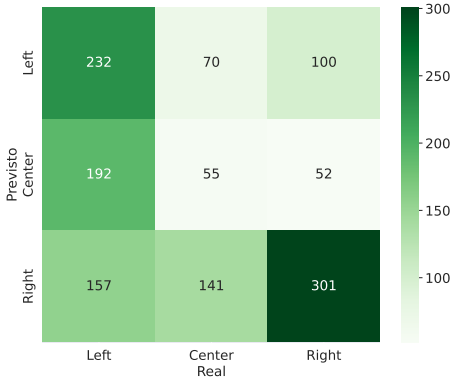
A Figura 8a, gerada pelo t-SNE [5], revela que os *embeddings* obtidos utilizando *Contrastive Loss* exibem uma separação mais clara entre as classes. As classes *left* (verde), *center* (vermelho) e *right* (azul) mostram menos sobreposição, indicando uma discriminação mais acentuada entre as classes no espaço vetorial. Isso sugere que o uso de *Contrastive Loss* preserva a proximidade dos pontos dentro de cada classe de forma mais coesa. Em contraste, a *Triplet Loss* resulta em uma maior dispersão entre os artigos dentro da mesma classe, o que pode ser prejudicial para a classificação de viés ideológico

Ao examinar o desempenho em cada conjunto de dados, a Figura 9 apresenta quatro matrizes de confusão que comparam os quatro melhores modelos para a classificação de viés ideológico no conjunto *media-bias split*. A combinação KNN com RoBERTa e *Contrastive Loss* (Figura 9a), exibe uma separação eficaz entre as classes, contribuindo para um desempenho elevado no *Macro F1-score*. Em contraste, a combinação KNN com DistilBERT e *Triplet Loss* (Figura 9b) melhora a classificação da classe *left* em relação ao uso do RoBERTa com *Contrastive Loss*. A combinação KNN com RoBERTa e *Triplet Loss* se destaca por atingir o maior número de previsões corretas na classe *right*, mantendo um desempenho competitivo em comparação com o *baseline*. Por fim, a combinação que utiliza uma *Deep Neural Network* (DNN) com RoBERTa e *Triplet Loss* demonstra um bom desempenho geral, embora enfrente maior dificuldade na classificação da classe *center*.

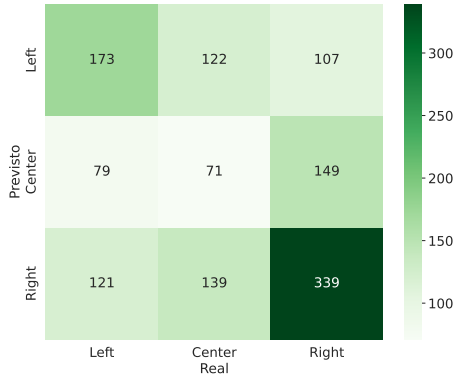
[5] t-SNE (t-distributed Stochastic Neighbor Embedding) é uma técnica de redução de dimensionalidade amplamente utilizada para visualizar dados de alta dimensão.



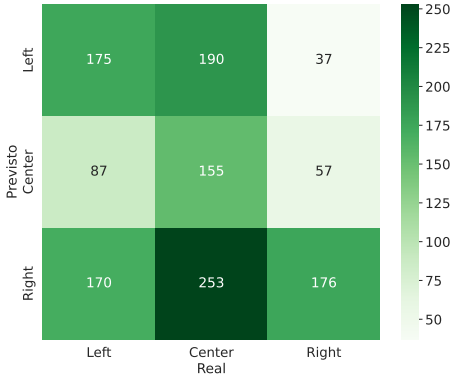
(a) KNN + RoBERTa + *Contrastive Loss*.



(b) KNN + DistilBERT + *Triplet Loss*.



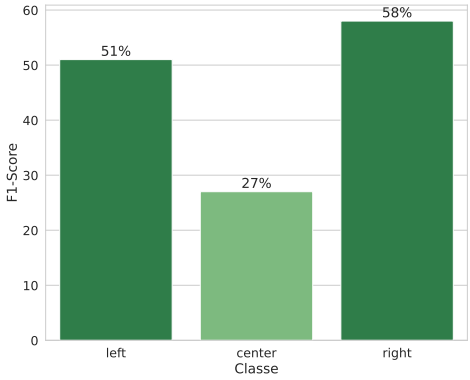
(c) KNN + RoBERTa + *Triplet Loss*.



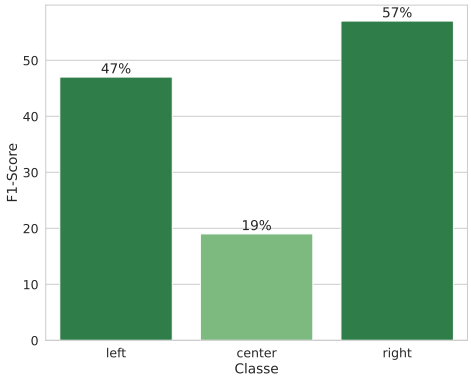
(d) DNN + RoBERTa + *Triplet Loss*.

Figura 9 – Matriz de confusão dos quatro melhores experimentos no conjunto do *media-bias split*.

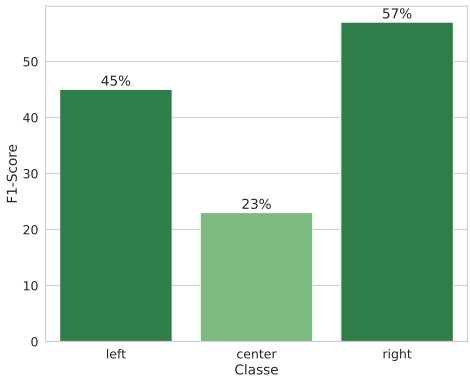
Os resultados sugerem que a utilização do RoBERTa com a *Contrastive Loss*, particularmente em combinação com o modelo KNN, proporciona uma melhor separação das diferentes ideologias políticas, especialmente para as classes *left* e *right*. Além disso, observou-se que as diferentes combinações de modelos e técnicas de *embedding* possuem pontos fortes e fracos específicos em relação à classificação das diversas classes ideológicas. As combinações que envolvem o RoBERTa tendem a apresentar um desempenho geral superior, particularmente para as classes *left* e *right*. Entretanto, a classe *center* continua a ser um desafio para todos os modelos, com F1-scores consistentemente mais baixos, o que indica que artigos centristas são mais difíceis de classificar corretamente, conforme ilustrado na Figura 10.



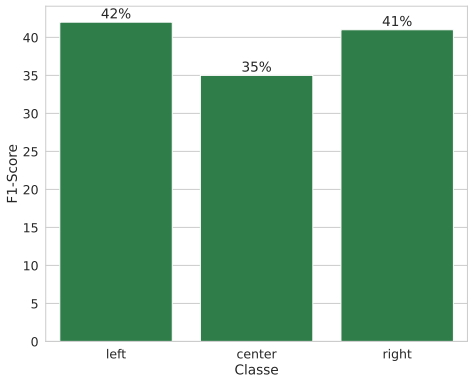
(a) KNN + RoBERTa + *Contrastive Loss*.



(b) KNN + DistilBERT + *Triplet Loss*.



(c) KNN + RoBERTa + *Triplet Loss*.

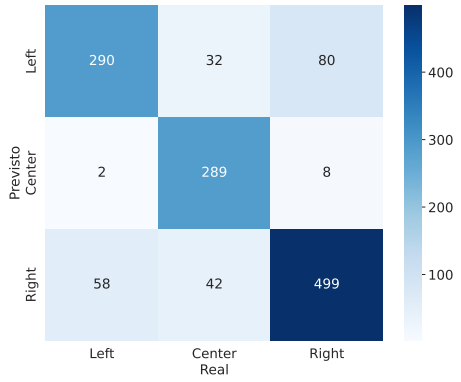


(d) DNN + RoBERTa + *Triplet Loss*.

Figura 10 – Desempenho das medidas de Macro F1-score por classe dos melhores modelos no conjunto *media-bias split*.

Em relação ao desempenho no conjunto do *random split*, observa-se que a combinação KNN com RoBERTa e Triplet Loss (Figura 11a) apresenta uma alta precisão na classificação da classe *right*, com 499 predições corretas, enquanto mantém um bom desempenho nas classes *center* e *left*. A combinação DNN com RoBERTa e Contrastive Loss (Figura 11b) também apresenta um desempenho robusto, especialmente nas classes *center* e *left*, embora com uma leve redução na classe *right*. A combinação KNN com RoBERTa e Contrastive Loss (Figura 11c) mantém um equilíbrio entre as três classes, com resultados ligeiramente inferiores na classe *left*. Por outro lado, a combinação DNN com RoBERTa e Triplet Loss (Figura 11d) exibe uma maior dificuldade na classificação da classe *right*, com 436 predições corretas, mas compensa com uma alta precisão na classe *left*. Esses resultados indicam que, embora todas as combinações

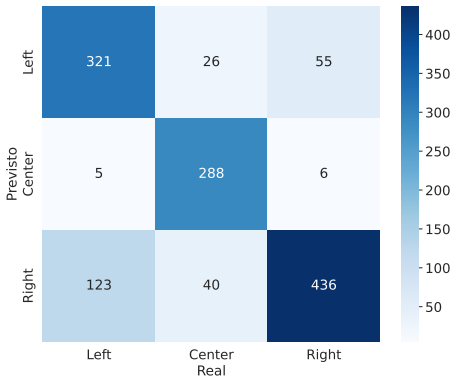
obtenham um bom desempenho geral, o KNN com Triplet Loss tende a ser mais eficaz para a classe *right*, enquanto a DNN com Triplet Loss demonstra maior eficácia para a classe *left*.



(a) KNN + RoBERTa + *Triplet Loss*.

(b) DNN + RoBERTa + *Contrastive Loss*.

(c) KNN + RoBERTa + *Contrastive Loss*.

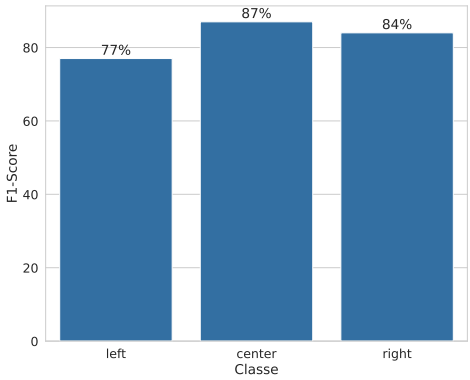


(d) DNN + RoBERTa + *Triplet Loss*.

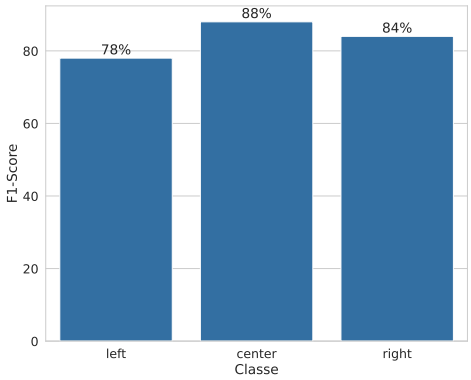
Figura 11 – Matriz de confusão dos quatro melhores experimentos no conjunto do *random split*.

Ao analisar os resultados de *F1-score* por classe apresentados na Figura 12, observa-se que as combinações de KNN com RoBERTa, utilizando tanto *Triplet Loss* (Figura 12a) quanto *Contrastive Loss* (Figura 12b), demonstram um desempenho equilibrado, com *F1-scores* elevados para as classes *center* e *right*. No entanto, a classe *left* apresenta valores ligeiramente inferiores, especialmente com *Triplet Loss* (77%). A combinação de DNN com RoBERTa e *Triplet Loss* (Figura 12d) também exibe um bom desempenho geral, mas com um *F1-score* mais baixo para a classe *left* (75%), sugerindo maior dificuldade na classificação dessa classe específica. De modo geral, os resultados indicam

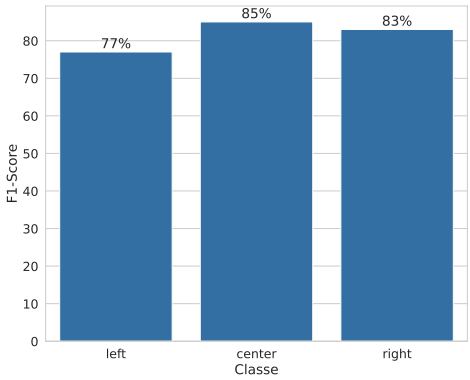
que as técnicas de *embedding* empregadas em conjunto com o RoBERTa são eficazes na distinção entre as classes *center* e *right*, enquanto a classe *left* continua a representar um desafio, com *F1-scores* consistentemente mais baixos em todas as combinações analisadas.



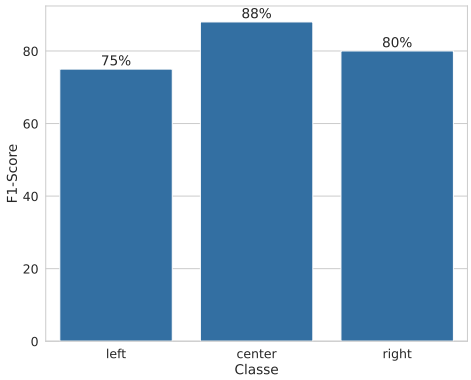
(a) KNN + RoBERTa + *Triplet Loss*.



(b) DNN + RoBERTa + *Contrastive Loss*.



(c) KNN + RoBERTa + *Contrastive Loss*.



(d) DNN + RoBERTa + *Triplet Loss*.

Figura 12 – Desempenho das medidas de Macro F1-score por classe dos melhores modelos no conjunto *random split*.

As diferenças entre os conjuntos *media-bias split* e *random split* nos resultados refletem o impacto que a estrutura dos dados tem no desempenho dos modelos de classificação de viés ideológico. No *media-bias split*, artigos de portais de notícias específicos são alocados exclusivamente em um dos conjuntos (treinamento, validação ou teste), impedindo que o modelo encontre o estilo de escrita de uma fonte de notícias durante o treinamento, o que torna mais difícil generalizar para novos portais no conjunto de teste. Isso resulta em um desempenho inferior, já que o modelo não consegue capturar plenamente

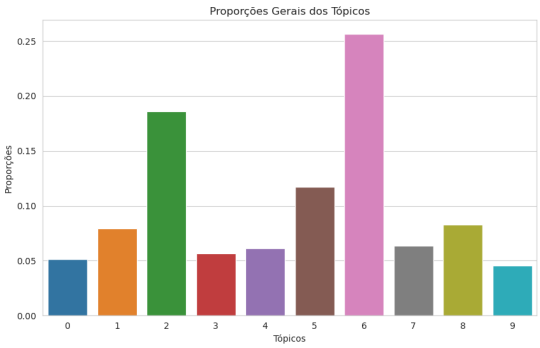
as nuances textuais associadas a cada viés ideológico de portais não vistos. Por outro lado, a divisão adotada no *random split* permite que os mesmos portais de notícias apareçam tanto nos conjuntos de treinamento quanto de teste, facilitando o reconhecimento dos padrões de viés ideológico pelos modelos, o que leva a uma classificação mais precisa. Consequentemente, o *random split* tende a produzir resultados significativamente melhores em termos de acurácia e *Macro F1-Score* do que o *media-bias split*, como observado nos experimentos.

5.4 Classificação de Viés Ideológico Baseado em Conteúdo Textual e Probabilidade de Tópicos

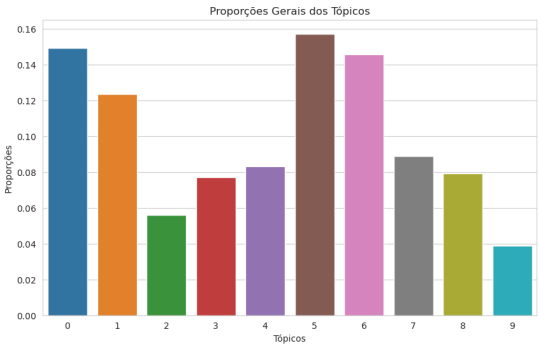
Nesta seção, os resultados da classificação são organizados em dois grupos: experimentos que utilizaram a *Contrastive Loss* e experimentos que utilizaram a *Triplet Loss* na geração dos espaços de características, em conjunto com os modelos pré-treinados e os modelos de tópicos. Assim, na Subseção 5.4.2, discutimos os resultados dos experimentos que empregaram a *Contrastive Loss*, enquanto na Subseção 5.4.3, apresentamos os resultados dos experimentos que utilizaram a *Triplet Loss*, além de abordar a análise dos resultados dos modelos de tópicos na Seção 5.4.1.

5.4.1 Análise da Extração de Tópicos

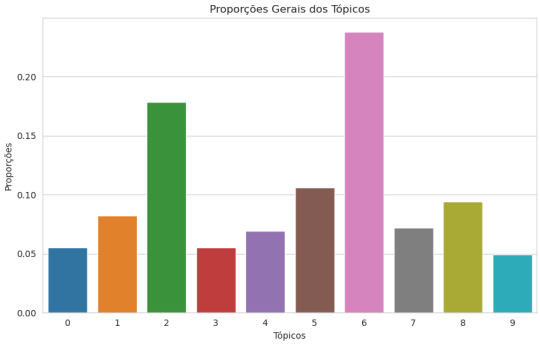
A extração dos tópicos dos texto foi realizada através do algoritmo de LDA, como descrita na Seção 5.1.3. Uma análise foi realizada utilizando os modelos com 10, 15 e 20 tópicos, com o objetivo de entender e identificar os padrões e grupamentos de palavras que caracterizam os diferentes viéses ideológicos presentes nos artigos, e entender o uso das diferentes quantidades de tópicos.



(a) Treino



(b) Validação



(c) Teste

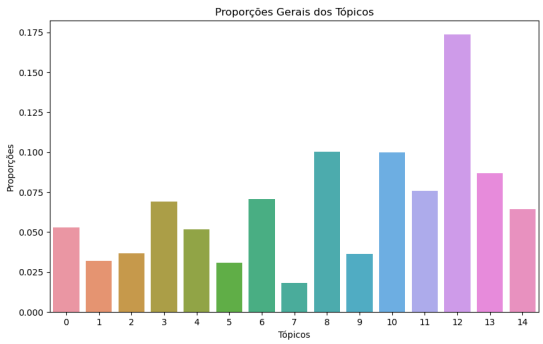
(d) Nuvem de palavras do tópico mais frequente

Figura 13 – As figuras a, b e c mostram as proporções dos tópicos nas partições de treino, validação e teste do conjunto *media-bias split*. A figura d exibe uma nuvem de palavras para o tópico mais frequente nos documentos em todas as partições..

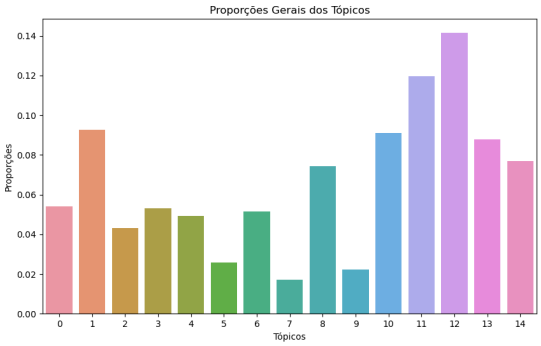
Para o modelo com 10 tópicos, as subfiguras (a), (b) e (c) da Figura 13 mostram que há uma variação nas proporções dos tópicos entre as três partições, o que é esperado dado que os dados foram divididos com base em viés midiático. A consistência nas proporções dos tópicos, especialmente nos tópicos mais dominantes, sugere que as partições preservam a diversidade temática presente no conjunto original, mas também ressaltam a possível concentração de certos tópicos em uma partição específica, como observado no tópico 6. A subfigura (d) apresenta uma nuvem de palavras para o tópico mais frequente, destacando termos como *people*, *police*, *one*, e *like*, indicando que esse tópico provavelmente aborda questões relacionadas à sociedade, segurança ou direitos civis, temas comuns em discussões de viés ideológico.

Ao aumentar o número de tópicos para 15, as subfiguras (a), (b) e (c) da Figura 14 mostram que, embora haja alguma variação nas proporções dos tópicos entre as três partições, há uma certa consistência, especialmente nos

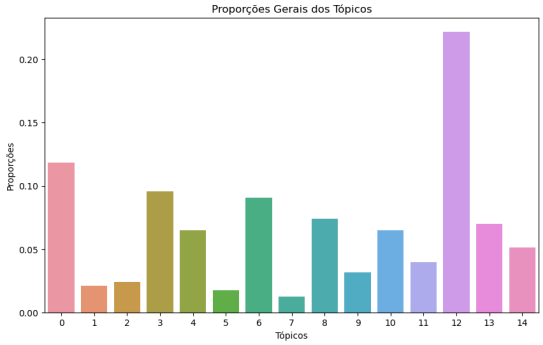
tópicos mais dominantes, o que sugere que as partições preservam a representatividade temática do conjunto de dados original. A subfigura (d) exibe uma nuvem de palavras correspondente ao tópico mais frequente, onde termos como *trump*, *people*, *president*, e *would* se destacam. Essa nuvem de palavras indica que o tópico mais prevalente nos documentos está relacionado a discussões sobre política e figuras públicas, particularmente o ex-presidente dos Estados Unidos, Donald Trump, refletindo temas frequentemente associados a viéses ideológicos.



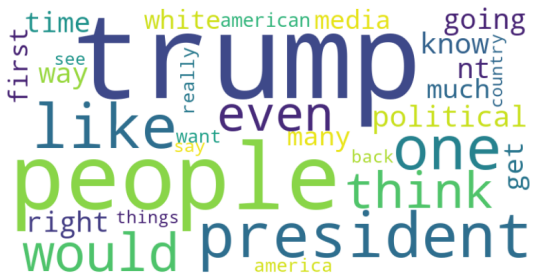
(a) Treino



(b) Validação



(c) Teste

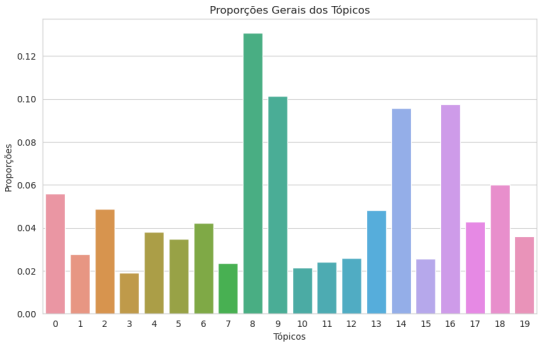


(d) Nuvem de palavras do tópico mais frequente

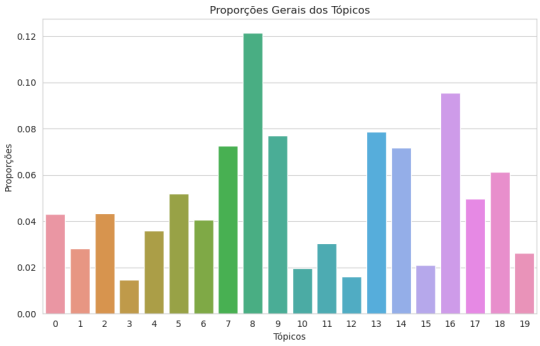
Figura 14 – As figuras a, b e c mostram as proporções dos tópicos nas partições de treino, validação e teste do conjunto *media-bias split*. A figura d exibe uma nuvem de palavras para o tópico mais frequente nos documentos em todas as partições..

Finalmente, o modelo com 20 tópicos demonstra que, embora as proporções dos tópicos variem entre as partições, o tópico 8 aparece consistentemente como um dos mais dominantes, especialmente nas partições de treino e validação, como pode ser visto na Figura 15. Essa consistência sugere que os tópicos principais são representados de maneira equilibrada nas diferentes partições, o que é importante para a generalização dos modelos treinados. A subfigura

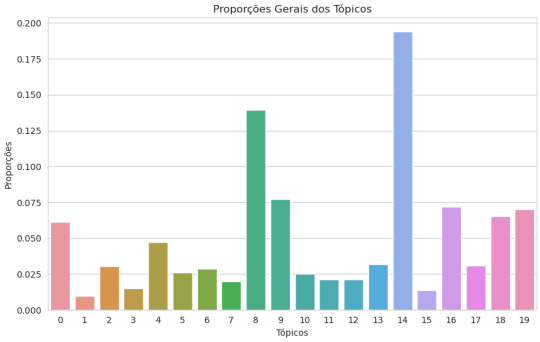
15(d) exibe uma nuvem de palavras para o tópico mais frequente, destacando termos como *people*, *one*, *like*, *says*, e *think*. Essa nuvem de palavras sugere que o tópico dominante está relacionado a discussões gerais sobre pessoas e opiniões, o que pode refletir temas amplos e recorrentes em notícias com viés ideológico. A presença consistente desses termos indica que esse tópico pode ter uma influência significativa na classificação de viés nos documentos analisados.



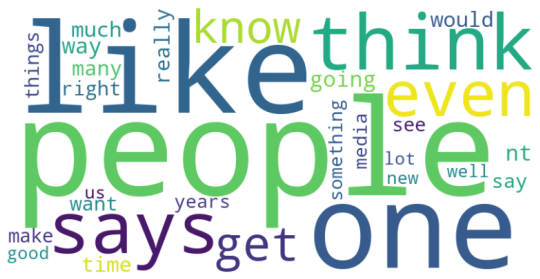
(a) Treino



(b) Validação



(c) Teste



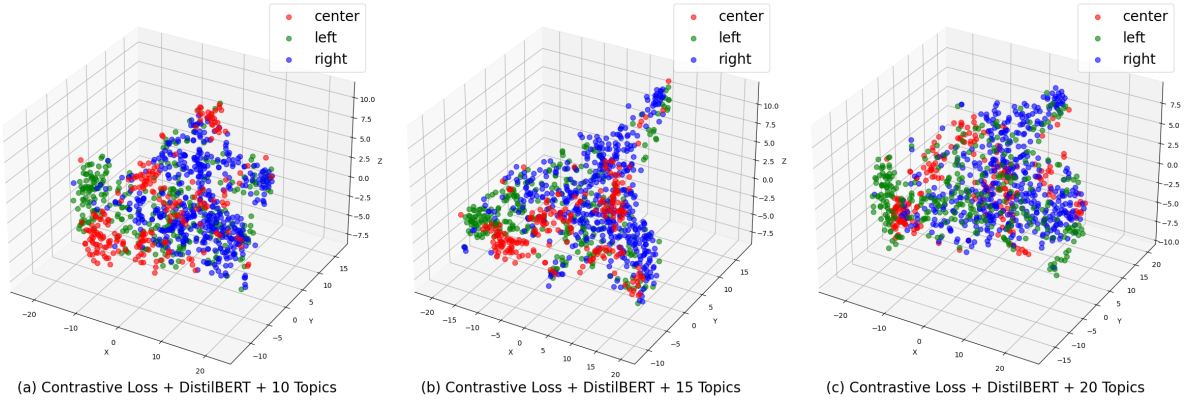
(d) Nuvem de palavras do tópico mais frequente

Figura 15 – As figuras a, b e c mostram as proporções dos tópicos nas partições de treino, validação e teste do conjunto *media-bias split*. A figura d exibe uma nuvem de palavras para o tópico mais frequente nos documentos em todas as partições..

5.4.2 Discussão dos Resultados dos Experimentos com a *Contrastive Loss*

A primeira etapa da classificação consiste na geração dos espaços de características a partir das representações textuais e das probabilidades de tópicos dos artigos de notícias. As Figuras 16 e 17 mostram os espaços latentes gerados

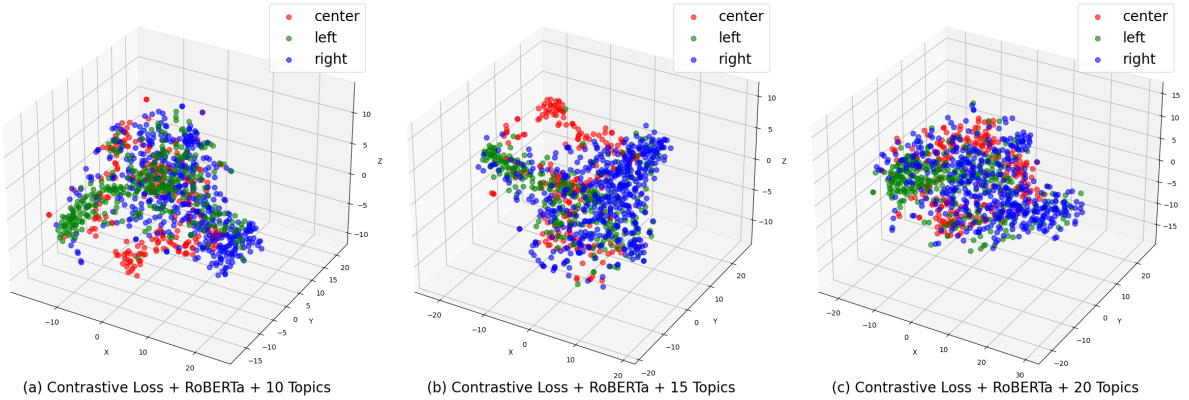
pelos modelos DistilBERT e RoBERTa, com a quantidade de tópicos variando entre 10, 15 e 20.



(a) Contrastive Loss + DistilBERT + 10 Topics (b) Contrastive Loss + DistilBERT + 15 Topics (c) Contrastive Loss + DistilBERT + 20 Topics

Figura 16 – *Embeddings* gerados com o modelo DistilBERT e a *Contrastive Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *media-bias split*.

Na Figura 16, que apresenta os *embeddings* gerados pelo DistilBERT, observa-se uma certa sobreposição entre as classes de viés (*center*, *left* e *right*), independentemente da quantidade de tópicos utilizada. Essa sobreposição indica que, embora o DistilBERT com *Contrastive Loss* tente maximizar a distinção entre as classes, a separabilidade entre os diferentes vieses ideológicos ainda é limitada, especialmente à medida que o número de tópicos aumenta.

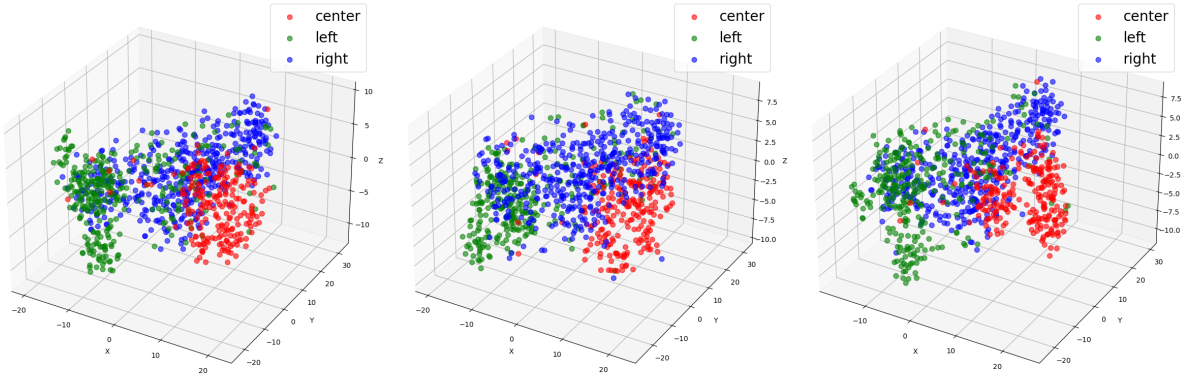


(a) Contrastive Loss + RoBERTa + 10 Topics (b) Contrastive Loss + RoBERTa + 15 Topics (c) Contrastive Loss + RoBERTa + 20 Topics

Figura 17 – *Embeddings* gerados com o modelo RoBERTa e a *Contrastive Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *media-bias split*.

Por outro lado, a Figura 17, que mostra os *embeddings* gerados pelo modelo RoBERTa, revela um comportamento semelhante. Embora o uso de *Contrastive Loss* tenha como objetivo aumentar a separabilidade das classes,

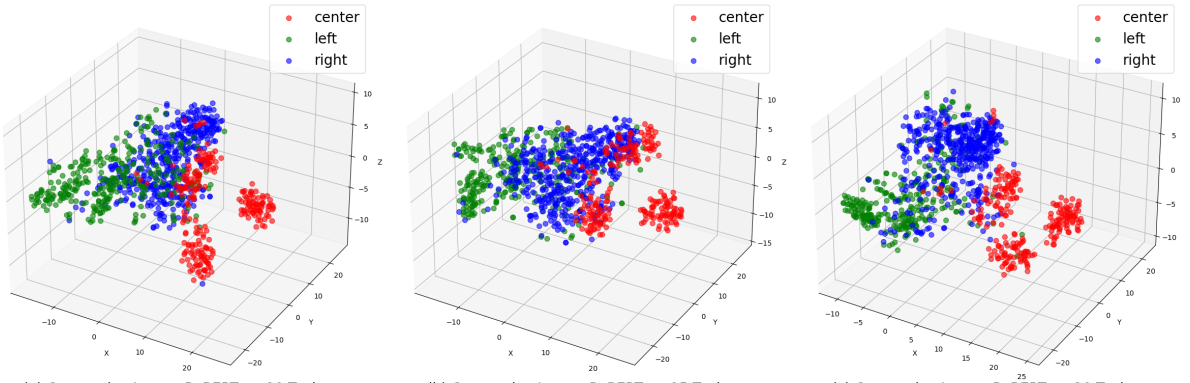
os gráficos tridimensionais indicam que há uma considerável sobreposição entre as classes, especialmente quando são utilizados 10 e 20 tópicos. Essa sobreposição sugere que, mesmo com a combinação de RoBERTa e *Contrastive Loss*, o modelo enfrenta desafios para diferenciar claramente as classes de viés ideológico no conjunto media-bias split.



(a) Contrastive Loss + DistilBERT + 10 Topics (b) Contrastive Loss + DistilBERT + 15 Topics (c) Contrastive Loss + DistilBERT + 20 Topics

Figura 18 – *Embeddings* gerados com o modelo DistilBERT e a *Contrastive Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *random split*.

Aos espaços de características gerados aos exemplos do conjunto do *random split*, na Figura 18 que apresenta os *embeddings* gerados pelo DistilBERT, observa-se uma separação mais clara entre as classes de viés (*center*, *left* e *right*) em comparação com as representações obtidas no *media-bias split*. À medida que o número de tópicos aumenta, nota-se uma organização mais distinta das classes, especialmente com 10 e 20 tópicos, onde as amostras das classes *left* e *right* tendem a se agrupar de maneira mais coerente, sugerindo que o DistilBERT combinado com *Contrastive Loss* é capaz de capturar melhor as nuances entre as diferentes orientações ideológicas quando os dados são divididos de forma a conter artigos das mesmas fontes de informação tanto na partição de treino quanto na partição de teste.

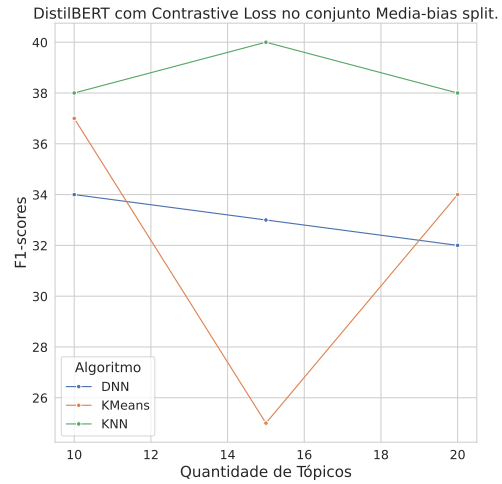


(a) Contrastive Loss + RoBERTa + 10 Topics (b) Contrastive Loss + RoBERTa + 15 Topics (c) Contrastive Loss + RoBERTa + 20 Topics

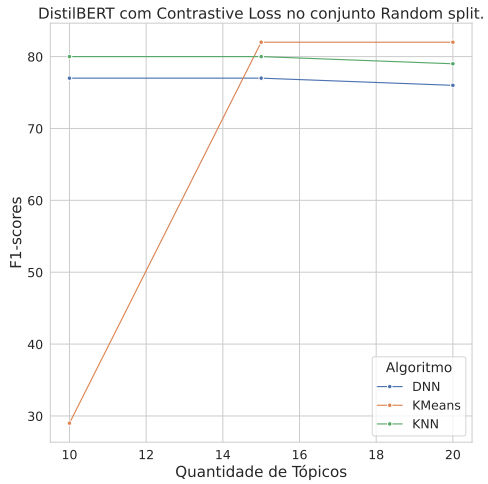
Figura 19 – *Embeddings* gerados com o modelo RoBERTa e a *Contrastive Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *random split*.

Na Figura 19, que apresenta os *embeddings* gerados pelo RoBERTa no *random split*, a separação entre as classes de viés ideológico é ainda mais pronunciada. Com 10 tópicos, as classes já mostram uma organização relativamente clara, mas é com 20 tópicos que a separabilidade atinge seu ponto mais elevado, com grupos de amostras das classes *left* e *right* bem definidos e menos sobreposição entre elas. Isso indica que o RoBERTa, quando combinado com *Contrastive Loss*, é particularmente eficaz em distinguir entre as diferentes classes de viés ideológico, reforçando a utilidade dessa abordagem para melhorar a discriminação de viés em textos noticiosos.

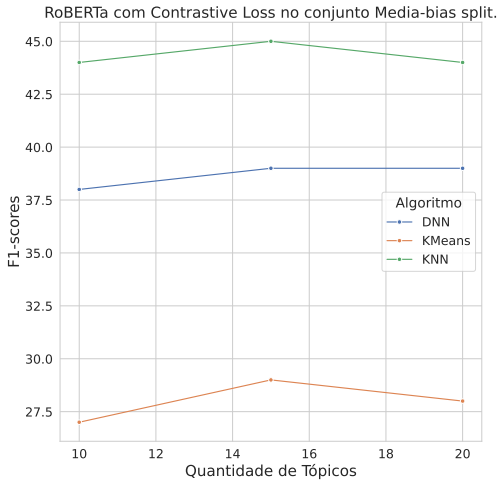
Os gráficos (a) e (c) da Figura 20, que representam o *media-bias split*, observamos que os modelos DistilBERT e RoBERTa, têm desempenho moderado, com o KNN superando consistentemente os algoritmos DNN e K-Means. Em particular, o Kmeans tem o pior desempenho em ambos os modelos, indicando uma maior dificuldade em lidar com a segmentação por viés midiático.



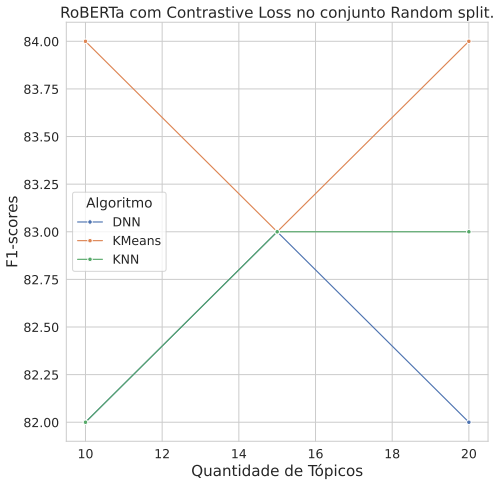
(a) DistilBert + *Contrastive Loss* no conjunto *media-bias* split.



(b) DistilBERT + *Contrastive Loss* no conjunto *random* split.



(c) RoBERTa + *Contrastive Loss* no conjunto *media-bias* split.



(d) RoBERTa + *Contrastive Loss* no conjunto *random* split.

Figura 20 – Resultados do *Macro F1-score* obtidos pelos modelos KNN, K-means e DNN na tarefa de classificação de viés ideológico. Para a geração de *embeddings*, foram utilizados os modelos pré-treinados DistilBERT e RoBERTa, em conjunto com a *Contrastive Loss*. Além disso, os modelos foram aplicados a representações de probabilidade de tópicos geradas pelo LDA, variando o número de tópicos entre 10, 15 e 20.

Nos gráficos (b) e (d), correspondentes ao *random* split, há uma clara melhoria no desempenho geral, especialmente para o Kmeans, que atinge *F1-scores* significativamente mais altos, próximos ou acima de 83%, enquanto os outros algoritmos apresentam resultados variáveis. O KNN, em especial,

mostra um aumento de desempenho com o aumento da quantidade de tópicos no modelo RoBERTa. Esses resultados sugerem que a segmentação aleatória (*random* split) favorece a performance dos modelos, especialmente para o KNN e o Kmeans, enquanto a segmentação no conjunto do *media-bias* impõe desafios adicionais, refletidos em *F1-scores* mais baixos e maior variabilidade entre os algoritmos.

Tabela 7 – Performance das métricas no conjunto de teste na tarefa de classificação de viés ideológico.

Modelo	**Split**	**Representação**	**Função de Perda**	**Número de Tópicos**	**Classificador**	***Macro F1-score***	**Acurácia**
Baly et al. (2020)		GloVe - embeddings	-	*0*	LSTM	31,51	32,30%
Baly et al. (2020)	*Media-bias split*	BERT - embeddings	-	*0*	BERT	35, 53	36,75%
4		RoBERTa - embeddings	*Contrastive Loss*	*0*	KNN	45,44	48,89%
25		RoBERTa - embeddings	*Contrastive Loss*	*15*	**KNN**	**45,87**	**49,73**%
Baly et al. (2020)		GloVe - embeddings	-	*0*	LSTM	65, 50	66,17%
Baly et al. (2020)	*Random split*	BERT - embeddings	-	*0*	BERT	80,19	79,83%
18		RoBERTa - embeddings	*Contrastive Loss*	*0*	DNN	83,90	**83,89%**
26		RoBERTa - embeddings	*Contrastive Loss*	*20*	**Kmeans**	**84,00**	83,85%

Na Tabela 7, observa-se que, no particionamento *media-bias split*, o modelo KNN combinado com *embeddings* gerados pelo RoBERTa e utilizando a função de perda *Contrastive Loss* alcançou o melhor desempenho, com um *Macro F1-score* de $45,87\%$ e uma acurácia de $49,73\%$. No particionamento *random split*, o KMeans utilizando também *embeddings* do RoBERTa com *Contrastive Loss* e 20 tópicos superou todos os outros, com um *Macro F1-score* de $84,00\%$ e uma acurácia de $83,85\%$. Estes resultados sugerem que, enquanto o *media-bias split* representa um desafio maior para os modelos, o *random split* permite uma melhor performance, possivelmente devido à presença de fontes de informação compartilhadas entre os conjuntos de treino e teste, o que facilita a tarefa de classificação.

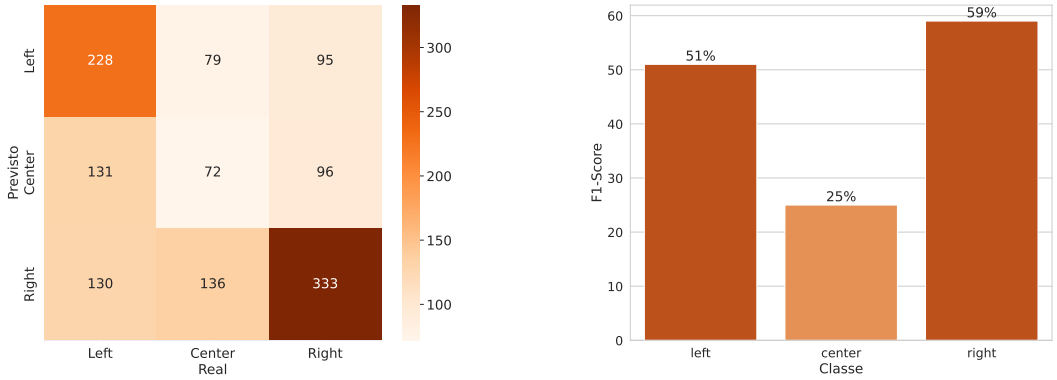


Figura 21 – Matriz de confusão e gráfico de barra com as medidas de macro F1-score por classe do KNN + DistilBERT + *Contrastive Loss* + 15 Tópicos, no conjunto *media-bias split*.

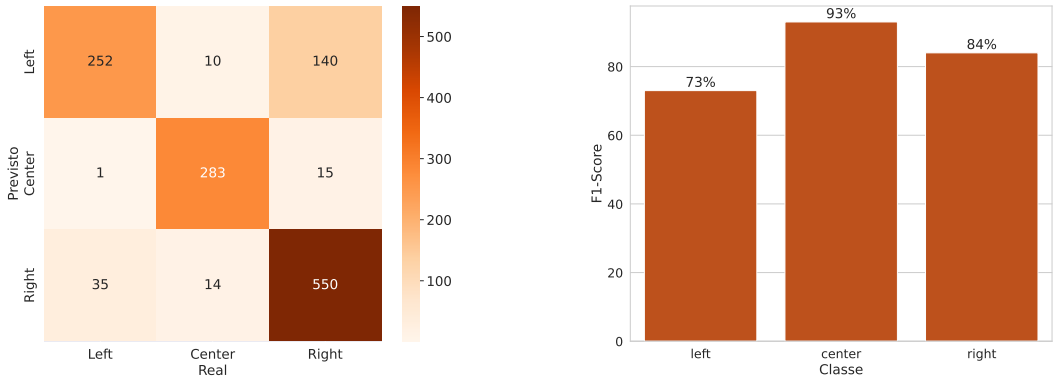


Figura 22 – Matriz de confusão e gráfico de barra com as medidas de macro F1-score por classe do Kmeans + RoBERTa + *Contrastive Loss* + 20 Tópicos, no conjunto *random split*.

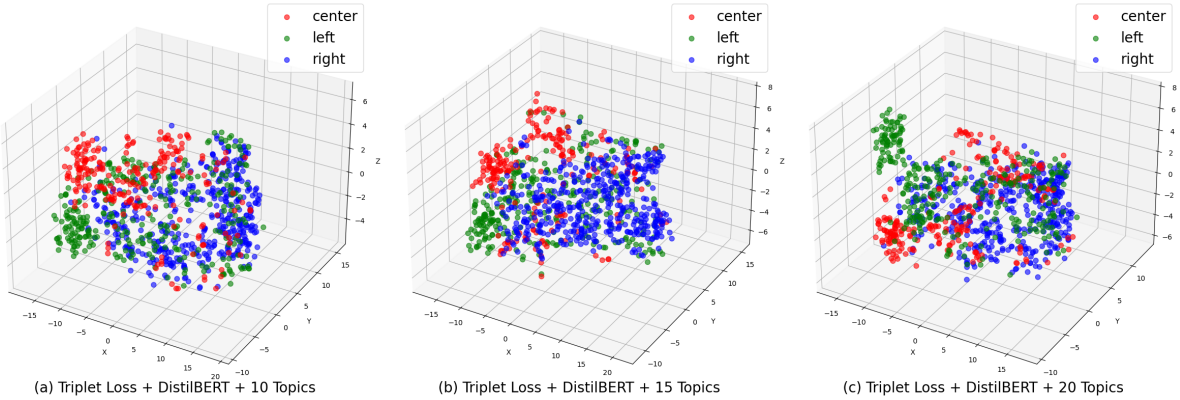
Além disso, a Figura 21 mostra a matriz de confusão e o Macro *F1-score* por classe para o melhor modelo aplicado ao conjunto *media-bias*. Observa-se que o KNN tem maior dificuldade em classificar corretamente as amostras da classe *center*, com um *F1-score* de apenas 25%. As classes *left* e *right* apresentam *F1-scores* ligeiramente melhores, com 51% e 59%, respectivamente, mas ainda assim, abaixo do ideal. A matriz de confusão indica que muitas amostras da classe *center* foram erroneamente classificadas como pertencentes a outras classes, especialmente a classe *right*.

Por outro lado, a Figura 22, que representa o K-Means no conjunto *random* split, mostra uma performance significativamente melhor, com *F1-scores* muito mais altos: 93% para *center*, 73% para *left*, e 84% para *right*. A

matriz de confusão para este modelo mostra uma clara melhoria na capacidade de distinguir entre as classes, com menos confusões, particularmente para as classes *center* e *right*.

5.4.3 Discussão dos Resultados dos Experimentos com a *Triplet Loss*

Os espaços de características gerados nos experimentos utilizando *Triplet Loss* mostram, como apresentado na Figura 23, que os *embeddings* gerados com o DistilBERT apresentam uma separação mais sutil e menos definida entre as classes de viés (*center*, *left* e *right*) em comparação com outros métodos. A sobreposição entre as classes é evidente, especialmente com o aumento do número de tópicos, o que sugere que o uso de *Triplet Loss* no DistilBERT pode não estar proporcionando uma separabilidade clara entre os vieses ideológicos, dificultando assim a tarefa de classificação.



(a) Triplet Loss + DistilBERT + 10 Topics (b) Triplet Loss + DistilBERT + 15 Topics (c) Triplet Loss + DistilBERT + 20 Topics

Figura 23 – *Embeddings* gerados com o modelo DistilBERT e a *Triplet Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *media-bias split*.

Na Figura 24, que exibe os *embeddings* gerados com o modelo RoBERTa, observa-se uma leve melhoria na separação das classes em comparação com o DistilBERT. Com 10 e 20 tópicos, as classes *left* e *right* começam a formar grupos mais coesos, embora ainda persista uma sobreposição considerável, especialmente na classe center. Essa sobreposição sugere que, apesar do uso

de *Triplet Loss*, o modelo RoBERTa ainda enfrenta dificuldades para diferenciar claramente entre as classes de viés ideológico no *media-bias split*, embora apresente um desempenho relativamente superior ao do DistilBERT nesse aspecto.

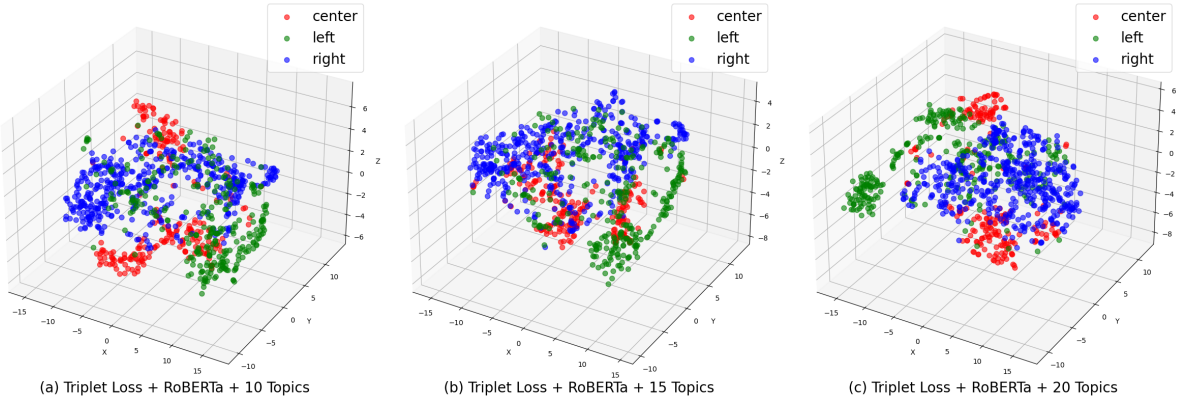
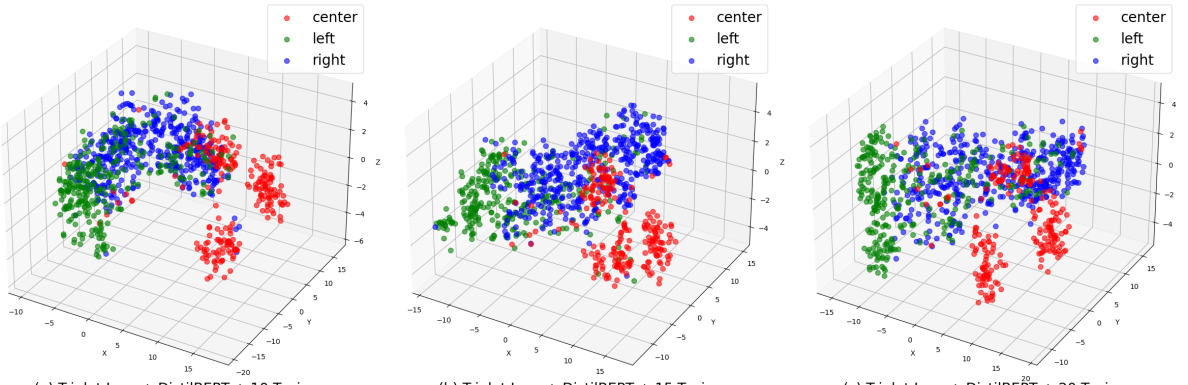


Figura 24 – *Embeddings* gerados com o modelo RoBERTa e a *Triplet Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *media-bias split*.

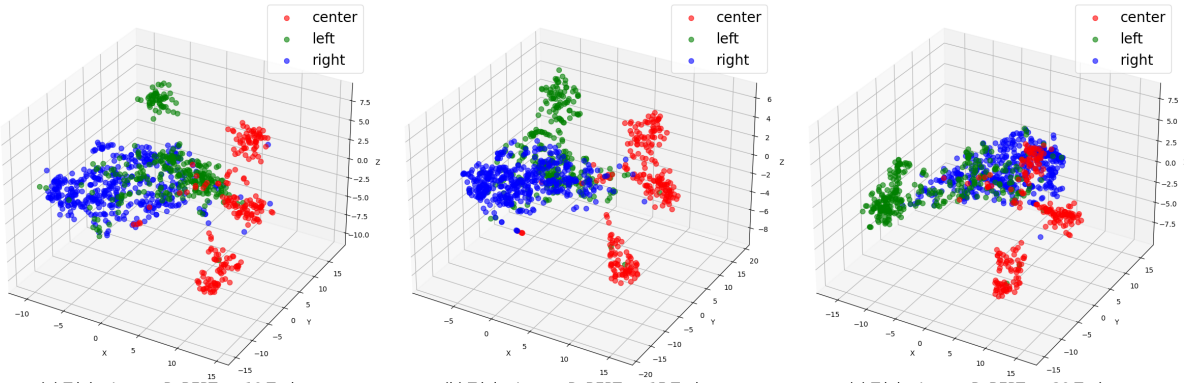
Na geração dos espaços de características no conjunto *radom split*, na Figura 25, observa-se que, à medida que o número de tópicos aumenta, a separação entre as classes de viés (*center*, *left* e *right*) se torna mais pronunciada. Em particular, com 20 tópicos, as classes *right* e *center* começam a formar agrupamentos mais coesos, embora ainda haja uma sobreposição considerável com a classe *left*. Isso sugere que, no contexto do *random split*, a combinação de DistilBERT com *Triplet Loss* consegue capturar de forma mais eficaz as diferenças entre os vieses ideológicos, embora ainda existam desafios na separabilidade total das classes.



(a) Triplet Loss + DistilBERT + 10 Topics (b) Triplet Loss + DistilBERT + 15 Topics (c) Triplet Loss + DistilBERT + 20 Topics

Figura 25 – *Embeddings* gerados com o modelo DistilBERT e a *Triplet* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *random split*.

Na Figura 26, o uso de *Triplet Loss* com o modelo RoBERTa no *random split* também apresenta melhorias na separação das classes à medida que o número de tópicos aumenta. Com 20 tópicos, observa-se uma distribuição mais clara das classes, especialmente para *right* e *center*, que mostram menor sobreposição com *left*. Contudo, a sobreposição entre *left* e *center* permanece, indicando que, embora *Triplet Loss* melhore a separabilidade entre as classes em comparação com outros métodos, ele ainda enfrenta limitações na distinção clara entre alguns vieses ideológicos, especialmente em casos onde as diferenças ideológicas podem ser mais sutis. Essas observações ressaltam a complexidade da tarefa de classificação de viés ideológico, mesmo quando técnicas avançadas como *Triplet Loss* são aplicadas.

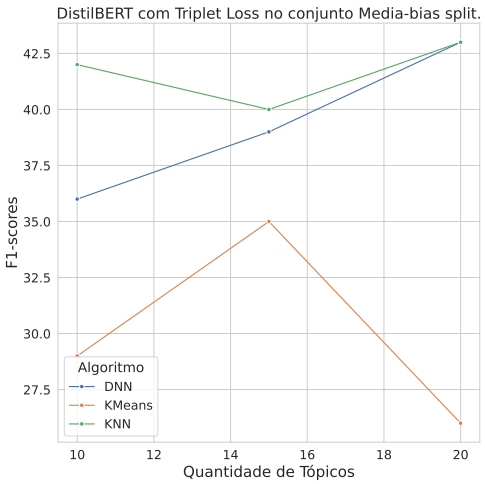


(a) Triplet Loss + RoBERTa + 10 Topics (b) Triplet Loss + RoBERTa + 15 Topics (c) Triplet Loss + RoBERTa + 20 Topics

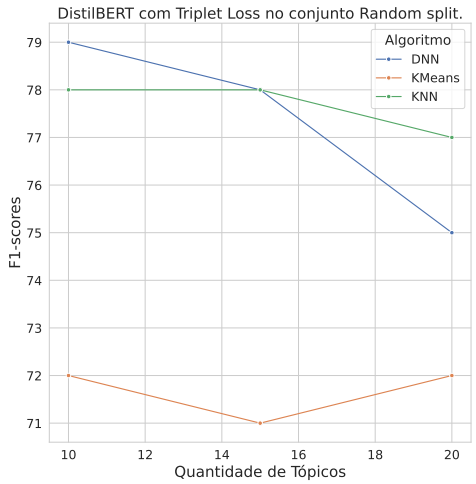
Figura 26 – *Embeddings* gerados com o modelo RoBERTa e a *Triplet Loss* para os dados textuais e probabilidade dos modelos de tópicos no conjunto *random split*.

Os resultados das métricas de avaliação dos experimentos realizados nos conjuntos de dados *media-bias split* e *random split*, utilizando os modelos DistilBERT e RoBERTa com a aplicação de *Triplet Loss*, estão apresentados na Figura 27.

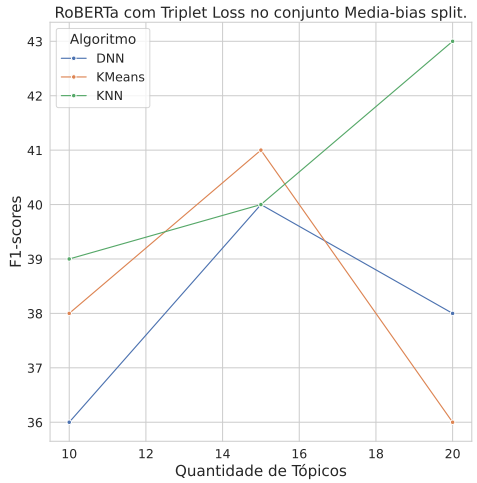
Nota-se que o algoritmo KNN consistentemente supera os outros algoritmos em termos de desempenho, especialmente quando combinado com RoBERTa, independentemente da quantidade de tópicos ou do conjunto de dados utilizado. Em contraste, o K-means exibe um desempenho mais instável, com uma queda acentuada no F1-score à medida que a quantidade de tópicos aumenta, particularmente no *media-bias split*. A DNN, embora apresente um desempenho relativamente estável, fica aquém do KNN, sugerindo que a simplicidade do KNN, juntamente com os *embeddings* gerados por modelos robustos como RoBERTa, pode ser mais eficaz na captura das nuances de viés ideológico nos textos.



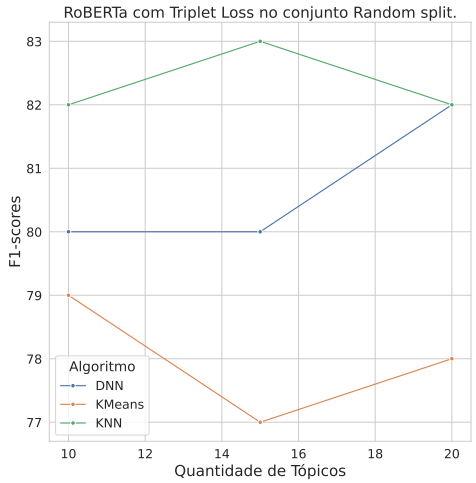
(a) DistilBert + *Triplet Loss* no conjunto *media-bias* split.



(b) DistilBERT + *Triplet Loss* no conjunto *random* split.



(c) RoBERTa + *Triplet Loss* no conjunto *media-bias* split.



(d) RoBERTa + *Triplet Loss* no conjunto *random* split.

Figura 27 – Resultados do *Macro F1-score* obtidos pelos modelos KNN, K-means e DNN na tarefa de classificação de viés ideológico. Para a geração de *embeddings*, foram utilizados os modelos pré-treinados DistilBERT e RoBERTa, em conjunto com a *Triplet Loss*. Além disso, os modelos foram aplicados a representações de probabilidade de tópicos geradas pelo LDA, variando o número de tópicos entre 10, 15 e 20.

Tabela 8 – Performance das métricas no conjunto de teste na tarefa de classificação de viés ideológico.

Modelo	Split	Representação	Função de Perda	Número de Tópicos	Classificador	*Macro F1-score*	Acurácia
Baly et al. (2020)		GloVe - embeddings	-	*0*	LSTM	31,51	32,30%
Baly et al. (2020)		BERT - embeddings	-	*0*	BERT	35, 53	36,75%
4	*Media-bias split*	RoBERTa - embeddings	*Contrastive Loss*	*0*	KNN	45,44	48,89%
25		RoBERTa - embeddings	*Contrastive Loss*	*15*	**KNN**	**45,87**	**49,73**%
27		DistilBERT - embeddings	*Triplet Loss*	*20*	KNN	43,93	46,10%
Baly et al. (2020)		GloVe - embeddings	-	*0*	LSTM	65, 50	66,17%
Baly et al. (2020)		BERT - embeddings	-	*0*	BERT	80,19	79,83%
18	*Random split*	RoBERTa - embeddings	*Contrastive Loss*	*0*	DNN	83,90	**83,89%**
26		RoBERTa - embeddings	*Contrastive Loss*	*20*	Kmeans	**84,00**	83,85%
28		RoBERTa - embeddings	*Triplet Loss*	*15*	KNN	83,30	**83,89%**

A Tabela 8 apresenta as métricas de *Macro F1-score* e acurácia no conjunto de teste, destacando que, no particionamento *media-bias split*, o KNN, ao utilizar *embeddings* do DistilBERT com *Triplet Loss* e a probabilidade de tópicos com 20 tópicos, teve um desempenho inferior ao KNN que utilizou *embeddings* do RoBERTa com *Contrastive Loss* e 15 tópicos, resultando em um *Macro F1-score* de $43,93\%$ e acurácia de $46,10\%$. Por outro lado, no particionamento *random split*, o KNN, utilizando *Triplet Loss* e 15 tópicos, demonstrou um desempenho elevado, com um *Macro F1-score* de $83,30\%$ e acurácia de *83,89%*, embora ainda não tenha superado o modelo KMeans, que utilizou *embeddings* gerados pelo RoBERTa e *Contrastive Loss* com 20 tópicos.

A Figura 28 apresenta os resultados da matriz de confusão e o gráfico de barras com as medidas de *Macro F1-score* por classe para o modelo KNN combinado com DistilBERT e *Triplet Loss* utilizando 20 tópicos no conjunto *media-bias split*. Observa-se que o desempenho do modelo varia significativamente entre as classes, com a classe *center* obtendo um *F1-score* de apenas 23%, enquanto as classes *left* e *right* alcançam *F1-scores* de 47% e 58%, respectivamente. A matriz de confusão revela que há uma considerável confusão entre as classes, especialmente com muitas amostras da classe *center* sendo incorretamente classificadas como *left* ou *right*.

Por outro lado, a Figura 29 mostra os resultados para o modelo KNN combinado com RoBERTa e *Triplet Loss* utilizando 15 tópicos no conjunto *random split*. Aqui, o desempenho do modelo é substancialmente melhor, com *F1-scores* de 87%, 76%, e 85% para as classes *center*, *left*, e *right*, respectivamente. A matriz de confusão indica uma classificação mais precisa, com menos

confusão entre as classes em comparação com o *media-bias split*. Esses resultados destacam a eficácia da combinação de RoBERTa com *Triplet Loss* para melhorar a capacidade do modelo de discriminar entre diferentes vieses ideológicos, resultando em uma classificação mais equilibrada e precisa entre as classes.

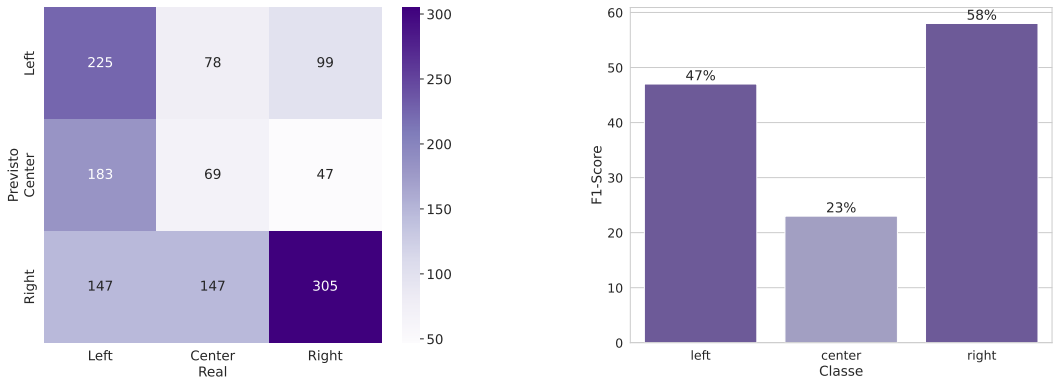


Figura 28 – Matriz de confusão e gráfico de barra com as medidas de macro F1-score por classe do KNN + DistilBERT + *Triplet Loss* + 20 Tópicos, no conjunto *media-bias split*.

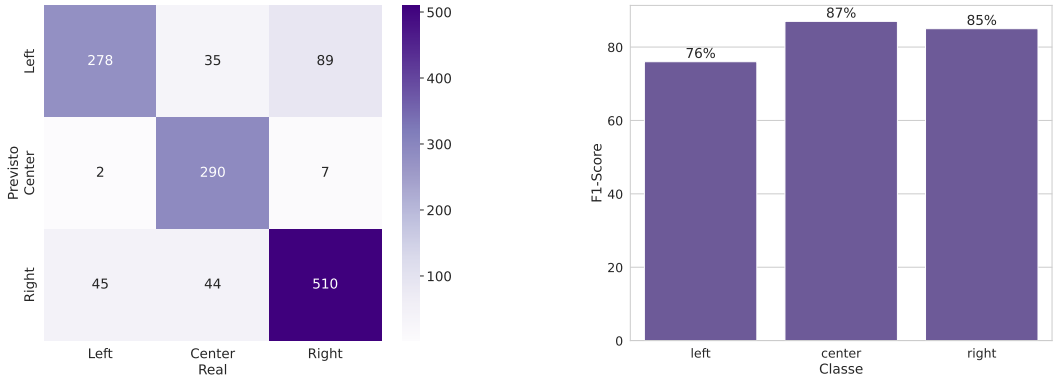


Figura 29 – Matriz de confusão e gráfico de barra com as medidas de macro F1-score por classe do KNN + RoBERTa + *Triplet Loss* + 15 Tópicos, no conjunto *random split*.

5.4.4 Discussão Geral

A análise dos resultados experimentais demonstra que tanto as abordagens baseadas em texto quanto a combinação de texto com probabilidade de tópicos superam o estado da arte na mesma tarefa. Isso indica que os métodos

implementados são eficazes em representar textos, permitindo a extração de informações discriminativas relevantes que podem aumentar a eficiência das tarefas de classificação de viés ideológico.

Entre os métodos de classificação que utilizam exclusivamente representações textuais, o modelo KNN com embeddings RoBERTa e a função *Contrastive Loss* se destaca, obtendo os melhores resultados. Esse modelo teve um desempenho significativamente superior no conjunto de dados *media-bias split*, superando outros modelos com um aumento de $9,91\%$ no *Macro F1-score* em comparação ao baseline. No conjunto de dados *random split*, o modelo DNN com embeddings RoBERTa e *Contrastive Loss* se sobressaiu, resultando em um ganho de $3,73\%$ no *Macro F1-score*.

A simplicidade do algoritmo KNN, combinada com a qualidade dos embeddings gerados por modelos pré-treinados, contribuiu significativamente para o seu desempenho. A eficácia da classificação baseada na proximidade no espaço vetorial foi particularmente evidente no conjunto de dados *media-bias split*. Esses resultados demonstram que a abordagem baseada em embeddings RoBERTa, refinada por funções de perda métrica, é altamente competitiva para a tarefa de classificação de viés ideológico.

A combinação de representações textuais com probabilidade de tópicos mostrou melhorias notáveis na separação de classes ideológicas, especialmente em experimentos envolvendo KNN e RoBERTa utilizando *Triplet Loss*. No conjunto de dados *random split*, o modelo KNN alcançou um F1-score de 87% para a classe *center*, 76% para a *left* e 85% para a *right*, indicando uma separação mais clara entre as classes no espaço de embeddings. Em contraste, no conjunto de dados *media-bias split*, o desempenho foi inferior, com a classe *center* alcançando apenas 23% de *Macro F1-score*.

As discrepâncias nos resultados entre os diferentes conjuntos de dados sugerem que as diferenças entre treino e teste tiveram um impacto significativo no desempenho do modelo. No *random split*, a presença dos mesmos portais de notícias em ambos os conjuntos de treino e teste permitiu que o modelo captu-

rasse melhor as características de cada viés ideológico, e os tópicos extraídos do treino representaram bem o teste. Por outro lado, no *media-bias split*, onde os portais foram distribuídos de forma disjunta entre os conjuntos, o desempenho geral foi reduzido, destacando as dificuldades que os modelos enfrentam para generalizar a fontes de notícias não vistas durante o treinamento.

Uma análise dos erros de classificação no conjunto de dados *media-split* revelou uma taxa de erro de aproximadamente 53%, com 690 dos 1.300 artigos sendo classificados incorretamente em relação ao viés ideológico pelo modelo KNN, que utiliza o modelo pré-treinado RoBERTa e a função de perda *Contrastive Loss*. Entre os principais fatores que podem ter contribuído para as discrepâncias observadas entre a classe verdadeira e a prevista, destaca-se a ambiguidade na linguagem utilizada nos artigos. Essa ambiguidade é particularmente evidente em temas amplamente debatidos, como política e economia, a falta de termos ideológicos explícitos pode dificultar a classificação correta.

A sobreposição de tópicos pode dificultar a distinção clara entre as classes *center*, *left* e *right*, especialmente em questões que permeiam os domínios políticos. Outro fator potencialmente relevante é o viés nos dados de treinamento, que pode ter levado o modelo a estabelecer fronteiras de decisão inadequados, particularmente se houver desequilíbrio ou representatividade insuficiente de certos grupos ideológicos no conjunto de dados. Por fim, a presença de *outliers*, ou seja, artigos com uso atípico ou extremo de linguagem, também pode ter impactado negativamente a precisão do modelo. A Figura 30 apresenta um exemplo de artigo de notícia da classe *center* que foi classificado como *left*.

"Here ' s a point to ponder : To what extent is Fox News Channel ' s Tucker Carlson responsible for President Donald Trump stepping away from a potential war with Iran ? From his prime-time perch on the top-rated cable network , Carlson has advocated restraint in dealing with Iran, and resisted cheerleading the Trump-ordered drone killing of Iranian Gen. Qasem Soleimani . Shortly after the story of Iran ' s counter-attack broke on Tuesday , Carlson hosted a show that mixed coverage of the story as details became known , emphasizing early reports of a lack of American casualties , and interviews with experts on the Middle East . Some of those guests pointed out the dangers of spiraling escalation. I continue to believe the president doesn ' t want a full-blown war, Carlson said . Some around him might , but I think most sober people don ' t want that ... "

Figura 30 – Exemplo de um artigo de notícia da classe *center*, classificado como *left*.

5.5 Considerações Finais

Neste capítulo, foram apresentados os resultados obtidos a partir da aplicação das metodologias de classificação de viés ideológico em artigos de notícias. Utilizando-se de modelos de aprendizado de máquina como DistilBERT e RoBERTa, treinados com diferentes configurações de hiperparâmetros e função de perda (*Contrastive Loss* e *Triplet Loss*), foi possível avaliar o desempenho em dois conjuntos de dados: *media-bias split* e *random split*.

Os resultados indicaram que o classificador KNN, especialmente quando combinado com RoBERTa e *Contrastive Loss*, obteve consistentemente os melhores desempenhos, tanto no *media-bias split* quanto no *random split*. Isso sugere que a simplicidade e a robustez do KNN, aliados à capacidade do RoBERTa de capturar nuances semânticas profundas, são particularmente eficazes na tarefa de discriminação de viés ideológico.

Por outro lado, o modelo K-Means mostrou-se mais instável, com uma acentuada queda no *Macro F1-score* à medida que a quantidade de tópicos aumentava, especialmente no *media-bias split*. A DNN, embora apresentasse um desempenho mais estável, não conseguiu superar o KNN, reforçando a ideia de que a qualidade dos *embeddings* é mais determinante do que a complexidade do classificador.

Os experimentos também revelaram que, apesar das técnicas avançadas utilizadas, a tarefa de classificação de viés ideológico ainda enfrenta desafios significativos, especialmente na distinção entre classes mais sutis, como *center* e *left*. A sobreposição entre essas classes em diversas combinações de modelos sugere que futuras pesquisas devem focar na melhoria das representações de características e na exploração de novas técnicas de aprendizado que possam captar melhor as nuances ideológicas.

Em resumo, os resultados apresentados neste capítulo contribuem para o entendimento dos desafios e das oportunidades na classificação de viés ideológico em textos de notícias, destacando a importância da escolha de modelos robustos e técnicas de perda adequadas para alcançar uma melhor discriminação entre as diferentes orientações políticas.

6 Conclusão

Neste estudo, propusemos e avaliamos uma abordagem para a classificação de viés ideológico em artigos de portais de notícias, utilizando duas estratégias distintas: a primeira baseada exclusivamente em representações textuais, e a segunda combinando essas representações com a probabilidade de tópicos extraídos dos textos. A revisão da literatura revelou uma lacuna significativa na exploração da combinação de diferentes representações textuais como insumos para modelos de aprendizado profundo, especialmente no uso de funções de perda contrastivas, na aplicação dessas técnicas à classificação de artigos inteiros, em vez de sentenças isoladas e somente no uso do conteúdo de texto, sem dependência de links externos.

Os resultados obtidos fornecem importantes insights sobre a eficácia dessas abordagens na tarefa de classificação de viés ideológico. A análise comparativa entre os modelos baseados em *embeddings* gerados pelo DistilBERT e RoBERTa, em conjunto com as funções de perda *Contrastive Loss* e *Triplet Loss*, indicou que o modelo KNN consistentemente apresentou melhor desempenho, particularmente quando combinado com o RoBERTa. Esses achados sugerem que a simplicidade do KNN, aliada à robustez dos *embeddings* gerados por modelos de linguagem avançados como o RoBERTa, constitui uma combinação eficaz para capturar as nuances de viés ideológico presentes nos textos.

Além disso, a comparação entre diferentes esquemas de particionamento de dados, como *media-bias split* e *random split*, demonstrou que o desempenho dos modelos é sensível à forma como os dados são divididos. O *random split* tende a favorecer a performance dos modelos, principalmente porque os dados de treinamento e teste compartilham as mesmas fontes de informação. Em contrapartida, o *media-bias split* apresentou maior desafio, resultando em uma diminuição nos valores de F1-score e acurácia. Isso evidencia a necessidade de um particionamento cuidadoso dos dados para assegurar a robustez dos modelos e evitar que o aprendizado seja enviesado por características específicas

das fontes de informação.

6.1 Limitações e Trabalhos Futuros

Embora este estudo tenha alcançado resultados promissores, algumas limitações devem ser destacadas. A performance variável entre as classes de viés ideológico foi uma das principais limitações, com desafios específicos na distinção entre as classes *center* e *left*, que resultaram em baixa precisão e alta taxa de confusão, especialmente ao utilizar o modelo KNN em combinação com DistilBERT e *Triplet Loss*. Além disso, a generalização dos modelos para novas fontes de notícias ou contextos não foi suficientemente explorada. A eficácia dos modelos em ambientes distintos daqueles nos quais foram treinados, especialmente em cenários com distribuições de dados diferentes, permanece uma questão em aberto.

Para superar essas limitações e avançar na pesquisa sobre a detecção de viés ideológico, futuros trabalhos podem explorar abordagens que integrem a modelagem de linguagem com informações adicionais, como o contexto histórico ou metadados, visando melhorar a separabilidade entre os vieses ideológicos e, consequentemente, a precisão da classificação. Além disso, a avaliação de modelos de *Large Language Models* (LLMs) combinados com estratégias de engenharia de prompt representa uma promissora linha de pesquisa. Avanços nessas direções não apenas podem aprimorar a precisão dos modelos, mas também contribuirão para uma compreensão mais profunda dos mecanismos subjacentes ao viés ideológico em conteúdos noticiosos.

Referências

AIRES, V. P.; NAKAMURA, F. G.; NAKAMURA, E. F. A link-based approach to detect media bias in news websites. In: *Companion proceedings of the 2019 world wide web conference*. [S.l.: s.n.], 2019. p. 742–745. 19, 20

BAEZA-YATES, R.; RIBEIRO-NETO, B. *Recuperação de Informação-: Conceitos e Tecnologia das Máquinas de Busca*. [S.l.]: Bookman Editora, 2013. 38, 39

BALY, R. et al. We can detect your bias: Predicting the political ideology of news articles. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [S.l.: s.n.], 2020. p. 4982–4991. 10, 20, 43, 45, 49, 52, 60, 64

BISHOP, C. M.; NASRABADI, N. M. *Pattern recognition and machine learning*. [S.l.]: Springer, 2006. v. 4. 29, 34

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. *Journal of machine Learning research*, v. 3, n. Jan, p. 993–1022, 2003. 28

CHIANG, C.-F.; KNIGHT, B. Media bias and influence: Evidence from newspaper endorsements. *The Review of economic studies*, Oxford University Press, v. 78, n. 3, p. 795–820, 2011. 19

DALLMANN, A. et al. Media bias in german online newspapers. In: *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. [S.l.: s.n.], 2015. p. 133–137. 43

DELLAVIGNA, S.; KAPLAN, E. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, MIT Press, v. 122, n. 3, p. 1187–1234, 2007. 19

DEVLIN, J. et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. Disponível em: <https://arxiv.org/abs/1810.04805>. 50

DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *International workshop on multiple classifier systems*. [S.l.], 2000. p. 1–15. 29

DUARTE, M. L.; SILVA, T. A. da. Avaliação do desempenho de três algoritmos na classificação de uso do solo a partir de geotecnologias gratuitas. *Revista de estudos Ambientais*, v. 21, n. 1, p. 6–16, 2019. 54

EFRON, M. The liberal media and right-wing conspiracies: using cocitation information to estimate political orientation in web documents. In: *Proceedings of the thirteenth ACM international conference on Information and Knowledge Management*. [S.l.: s.n.], 2004. p. 390–398. 19, 20, 41, 45

ELEJALDE, E.; FERRES, L.; HERDER, E. The nature of real and perceived bias in chilean media. In: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. New York, NY, USA: Association for Computing

Machinery, 2017. (HT '17), p. 95–104. ISBN 9781450347082. Disponível em: <https://doi.org/10.1145/3078714.3078724>. 44

EZUGWU, A. E. et al. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, v. 110, p. 104743, 2022. ISSN 0952-1976. Disponível em: <https://www.sciencedirect.com/science/article/pii/S095219762200046X>. 30

GANGULA, R. R. R.; DUGGENPUDI, S. R.; MAMIDI, R. Detecting political bias in news articles using headline attention. In: LINZEN, T. et al. (Ed.). *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Florence, Italy: Association for Computational Linguistics, 2019. p. 77–84. Disponível em: <https://aclanthology.org/W19-4809>. 43

GAO, T.; YAO, X.; CHEN, D. SimCSE: Simple contrastive learning of sentence embeddings. In: MOENS, M.-F. et al. (Ed.). *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. p. 6894–6910. Disponível em: <https://aclanthology.org/2021.emnlp-main.552>. 50

GENTZKOW, M.; SHAPIRO, J. M. Media bias and reputation. *Journal of political Economy*, The University of Chicago Press, v. 114, n. 2, p. 280–316, 2006. 19, 42

HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. 29

KAYA, M.; BILGE, H. Ş. Deep metric learning: A survey. *Symmetry*, MDPI, v. 11, n. 9, p. 1066, 2019. 33

KERTéSZ, G. Different triplet sampling techniques for lossless triplet loss on metric similarity learning. In: *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. [S.l.: s.n.], 2021. p. 000449–000454. 53

KIM, M. Y.; JOHNSON, K. M. CLoSE: Contrastive learning of subframe embeddings for political bias classification of news media. In: CALZOLARI, N. et al. (Ed.). *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, 2022. p. 2780–2793. Disponível em: <https://aclanthology.org/2022.coling-1.245>. 56

KRESTEL, R.; WALL, A.; NEJDL, W. Treehugger or petrolhead? identifying bias by comparing online news articles with political speeches. In: *Proceedings of the 21st international conference on world wide web*. [S.l.: s.n.], 2012. p. 547–548. 19

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. 32

LIN, Y.-R.; BAGROW, J.; LAZER, D. More voices than ever? quantifying media bias in networks. In: *Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2011. v. 5, n. 1, p. 193–200. 19, 20, 42

LIU, Y. et al. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 2019. Disponível em: <https://arxiv.org/abs/1907.11692>. 50

MIKOLOV, T. et al. Efficient estimation of word representations in vector space. In: *International Conference on Learning Representations*. [s.n.], 2013. Disponível em: <https://api.semanticscholar.org/CorpusID:5959482>. 26

MURPHY, K. P. *Machine learning: a probabilistic perspective*. [S.l.]: MIT press, 2012. 29

RAO, A.; SPASOJEVIC, N. Actionable and political text classification using word embeddings and lstm. arxiv 2016. *arXiv preprint arXiv:1607.02501*, 2016. 44, 45

RIBEIRO, F. et al. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2018. v. 12, n. 1. 19, 44

RICH, C. *Writing and reporting news*. [S.l.]: Wadsworth Publishing Company, 1993. 25

SANH, V. et al. *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2020. Disponível em: <https://arxiv.org/abs/1910.01108>. 50

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 815–823. 35

SHOEMAKER, P. J. *Reese Mediating the Message: Theories of Influences on Mass Media Content*. [S.l.]: USA, Longman Publ, 1996. 26

SOKOLOVA, M.; JAPKOWICZ, N.; SZPAKOWICZ, S. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In: SPRINGER. *AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4-8, 2006. Proceedings 19*. [S.l.], 2006. p. 1015–1021. 38

SONG, H. O. et al. Deep metric learning via lifted structured feature embedding. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 4004–4012. 35

SPINDE, T. et al. Neural media bias detection using distant supervision with babe–bias annotations by experts. *arXiv preprint arXiv:2209.14557*, 2022. 19, 20

SUN, F. et al. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery,

2019. p. 1441–1450. ISBN 9781450369763. Disponível em: <https://doi.org/10.1145/3357384.3357895>. 33

VASWANI, A. et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017. 32

VINODHINI, G.; CHANDRASEKARAN, R. Sentiment analysis and opinion mining: a survey. *International Journal*, v. 2, n. 6, p. 282–292, 2012. 29

WANG, C.; BLEI, D. M. Collaborative topic modeling for recommending scientific articles. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2011. p. 448–456. Disponível em: <https://doi.org/10.1145/2020408.2020480>. 28

YIGIT-SERT, S.; ALTINGOVDE, I. S.; ULUSOY, Ö. Towards detecting media bias by utilizing user comments. In: *Proceedings of the 8th ACM Conference on Web Science*. [S.l.: s.n.], 2016. p. 374–375. 19

ZHANG, Y.; RAO, Z. Deep neural networks with pre-train model bert for aspect-level sentiments classification. In: . [S.l.: s.n.], 2020. p. 923–927. 54