UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CURSO DE SISTEMAS DE INFORMAÇÃO

MILENA BARBOSA ALEGRE

MÉTRICAS E MODELOS DE NLP PARA COMPREENDER E REDUZIR A PROPAGAÇÃO DE DESINFORMAÇÃO

CAMPO GRANDE – MS
DEZEMBRO DE 2023

MILENA BARBOSA ALEGRE

MÉTRICAS E MODELOS DE NLP PARA COMPREENDER E REDUZIR A PROPAGAÇÃO DE DESINFORMAÇÃO

Trabalho de conclusão de curso apresentado como parte das exigências para a obtenção do grau de Bacharel em Sistemas de Informação da Universidade Federal de Mato Grosso do Sul.

Orientador: Prof. Dr. Renato Porfirio Ishii

CAMPO GRANDE-MS

DEZEMBRO DE 2023

MILENA BARBOSA ALEGRE

MÉTRICAS E MODELOS DE NLP PARA COMPREENDER E REDUZIR A PROPAGAÇÃO DE DESINFORMAÇÃO

APROVADO (02/12/2023)

BANCA EXAMINADORA

_____

Prof. Dr. Renato Porfirio Ishii

(Orientador)

_____

Prof. Dr. Hana Karina Salles Rubinsztejn

(Membro da banca)

_____

Prof. Dr. Dionisio Machado Leite Filho

(Mebro da banca)

Dedico a todos os que me ajudaram ao longo desta caminhada.

AGRADECIMENTOS

Em primeiro lugar, a Deus, que fez com que meus objetivos fossem alcançados, durante todos os meus anos de estudos.

À minha família por todo apoio, por sempre estarem comigo me ajudando e dando suporte, principalmente a minha mãe Cláudia, meu padrasto Tiago e minha avó Shirley, que foram essenciais para que eu possa ter concluído minha formação. Seus exemplos de vida são minha inspiração e motivação para buscar sempre o melhor.

Agradeço a meu namorado Maykon Cavalheiro, que sempre esteve ao meu lado durante todos esses anos e por sempre ter me encorajado a buscar a excelência e a superar meus próprios limites e por ser meu porto seguro durante todo o processo e também a sua mãe Ezilda por ter me acolhido como filha.

Agradeço aos meus colegas e amigos Arthur Ramires, Enzo Paiva e Vitória Rocha que sempre estiveram presentes, oferecendo ajuda e compartilhando conhecimento.

Agradeço ao meu orientador Renato Ishii que acompanhou todo o processo de elaboração deste trabalho, fornecendo orientações valiosas e contribuindo para seu desenvolvimento. Sem sua colaboração, este TCC não seria possível.

Gostaria de agradecer também a Quality Sistemas, empresa onde trabalho, especialmente a Patrícia Guirandelli por fornecer suporte a minha formação acadêmica.

Por fim, quero agradecer a todas as pessoas que, direta ou indiretamente, contribuíram para a realização deste trabalho, desde a concepção do tema até sua conclusão. Sua presença em minha vida foi fundamental para o sucesso desta empreitada.

RESUMO

Este trabalho investiga como os algoritmos de Processamento de Linguagem Natural (NLP) podem ser usados para combater a desinformação online, um fenômeno que ameaça a qualidade da informação e a democracia. O objetivo é analisar as potencialidades do pysentimiento, uma biblioteca de NLP para o espanhol, que oferece funcionalidades como análise de sentimentos, emoções, discurso de ódio e ironia. A hipótese é que essas funcionalidades podem contribuir para a detecção e a redução da propagação de notícias falsas nas mídias sociais. Para testar essa hipótese, o trabalho compara o desempenho de quatro modelos de redes neurais profundas, baseados no RoBERTa, na tarefa de análise de sentimentos de *tweets* em espanhol. Os resultados mostram que o modelo Robertuito-Base-Uncased supera os demais modelos em termos de perda, f1, tempo de execução e taxa de amostras por segundo, sendo o mais adequado para a nossa proposta. O trabalho conclui que o pysentimiento é uma ferramenta promissora para o combate à desinformação online, e sugere possíveis aplicações e melhorias para trabalhos futuros.

Palavras-chave: NLP, Pysentimiento, RoBERTa

ABSTRACT

This work investigates how Natural Language Processing (NLP) algorithms can be used to combat online misinformation, a phenomenon that threatens the quality of information and democracy. The aim is to analyze the potentialities of pysentimiento, an NLP library for Spanish, that offers functionalities such as sentiment analysis, emotions, hate speech and irony. The hypothesis is that these functionalities can contribute to the detection and reduction of the spread of fakenews on social media. To test this hypothesis, the work compares the performance offour deep neural network models, based on RoBERTa, on the task of sentiment analysis of tweets in Spanish. The results show that the model Robertuito-Base-Uncased outperforms the other models in terms of loss, f1, runtime and samples per second, being the most suitable for our proposal. The work concludes that pysentimiento is a promising tool for the fight against online misinformation, and suggests possible applications and improvements for future works.

Keywords: NLP, Pysentimiento, RoBERTa

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

TCC Trabalho de Conclusão de Curso

NLP Natural Processing Language - Processamento de Linguagem natural

SUMÁRIO

1 INTRODUÇÃO

A era digital trouxe consigo um fenômeno perturbador, a propagação de desinformação online. As plataformas de mídia social tornaram-se terreno fértil para tal prática, levantando questões significativas sobre a veracidade das notícias que circulam nessas redes (Allcott & Gentzkow, 2017). Tendo em vista este cenário, o presente trabalho busca discutir os desafios enfrentados pela mídia social na divulgação de notícias confiáveis e propor soluções com base em algoritmos e técnicas de Processamento de Linguagem Natural (NLP).

Uma das ferramentas que tem ganhado destaque no combate à desinformação é o pysentimiento[1], por poder ajudar na identificação de informações potencialmente enganosas. Ao detectar a polaridade das mensagens, é possível destacar conteúdos com alta probabilidade de serem enganosos ou tendenciosos. A identificação de sentimentos negativos, extremos ou ambíguos pode sinalizar conteúdos suspeitos que precisam de verificação adicional para evitar a propagação de informações falsas.

A capacidade de medir a intensidade dos sentimentos expressos em textos pode ser valiosa na identificação de discursos inflamatórios ou extremos, muitas vezes associados a informações enganosas. Ao destacar mensagens com alta intensidade emocional, o pysentimiento permite uma triagem mais eficiente de conteúdos suspeitos, ajudando na identificação de informações falsas que possam incitar reações exageradas ou prejudiciais.

Uma função complementar seria a análise de padrões de propagação em redes sociais. Embora não seja uma funcionalidade direta do pysentimiento, a integração com ferramentas de mineração de dados sociais pode ajudar a rastrear a disseminação de informações falsas. Identificar a origem e os padrões de compartilhamento dessas informações pode ajudar na contenção e no combate à propagação de notícias falsas, fornecendo insights sobre como essas informações se espalham e quais comunidades são mais suscetíveis a elas.

Nesse contexto, a pergunta fundamental que orienta este trabalho é: Como os algoritmos de NLP podem ser aplicados para compreender e reduzir eficazmente a propagação desinformação online? E quais são as melhores práticas e propostas para desenvolver sistemas eficazes nesse contexto? Com isso em mente, buscamos não

[1] PySentimento é uma biblioteca em Python para analisar sentimentos em textos, detectando emoções positivas, negativas ou neutras.

apenas entender como as técnicas existentes são utilizadas atualmente, mas também investigar novos métodos potenciais para combater este problema cada vez mais prevalente.

2 REVISÃO DE LITERATURA

A desinformação, caracterizada pela disseminação intencional de informações falsas ou enganosas, tornou-se um problema global, afetando várias esferas da sociedade (Lazer et al., 2018). Com a crescente digitalização e o uso cada vez mais frequente das redes sociais como fonte principal de notícias, este problema se intensificou (Guess et al., 2019).

Nesse sentido, a Inteligência Artificial (IA), especificamente o Processamento de Linguagem Natural (NLP), tem sido explorado como uma ferramenta importante para compreender e combater a propagação da desinformação. Um dos principais usos do NLP é na análise de sentimentos e detecção de emoções em textos, o que pode ajudar a identificar notícias falsas que geralmente contêm linguagem emocionalmente carregada para atrair atenção e provocar reações (Poria et al., 2020).

Outra abordagem usando NLP é a classificação automática de texto. Através desta técnica, os algoritmos podem ser treinados para distinguir entre informações verdadeiras e falsas com base em características textuais. Shu et al. (2020) apresentaram um estudo onde desenvolveram uma estrutura baseada em NLP para detectar notícias falsas nas redes sociais. Eles mostraram que seu modelo pode

alcançar alta precisão na identificação de desinformação.

Além disso, também existem abordagens centradas no usuário para combater a propagação da desinformação. Por exemplo, Pennycook & Rand (2020) argumentam que promover o pensamento crítico e a literacia em mídia digital podem ser estratégias eficazes para reduzir a crença em e a partilha de notícias falsas.

Os algoritmos de Processamento de Linguagem Natural (NLP) têm sido bastante utilizados para a detecção e contenção da propagação de desinformação na internet. Estes algoritmos são capazes de analisar grandes volumes de dados de texto, identificando padrões que podem indicar a presença de desinformação (Shu et al., 2017).

O estudo realizado por Ferrara et al. (2016) mostrou que as notícias falsas

tendem a ter um sentimento mais negativo do que as notícias verdadeiras. Além disso, os autores também descobriram que as notícias falsas costumam ser mais compartilhadas do que as verdadeiras, o que contribui para sua propagação.

Apesar dos avanços nesta área, existem ainda vários desafios a serem superados. Um desses desafios é o fato de que os mecanismos usados para disseminar desinformação estão constantemente evoluindo (Zhou et al., 2020).

Isso significa que os algoritmos devem ser constantemente atualizados e adaptados para manter sua eficácia, a detecção de desinformação também levanta questões ética se de privacidade, é importante garantir que os algoritmos utilizados para detectar desinformação não violem a privacidade dos usuários ou sejam usados de maneira abusiva (Metaxa-Kakavouli et al., 2018). Além disso, os sistemas baseados em NLP podem ser vulneráveis à manipulação por atores mal-intencionados que estão cientes de como esses sistemas funcionam (Zhou et al., 2020).

Sobre a eficácia da abordagem baseada em NLP, ,mais especificamente, os algoritmos foram capazes de identificar nuances sutis na linguagem usada em textos contendo desinformação, permitindo uma detecção precisa (Lazer et al., 2018). Isso sugere que o uso desses algoritmos poderia ser uma ferramenta valiosa na luta contra a propagação da desinformação.

Existe também a busca por validação social desempenha um papel significativo na disseminação de desinformação em tweets carregados de sentimentos negativos. De acordo com o estudo de Del Vicario et al. (2016), a validação por meio de compartilhamentos e interações sociais pode reforçar crenças, independentemente de sua veracidade. Isso cria uma dinâmica de comportamento de manada, onde as pessoas são levadas a acreditar em informações que se alinham com as emoções predominantes na comunidade virtual,mesmo que sejam falsas.

Atualmente a influência dos algoritmos das redes sociais também é crucial. Estudos, como o de Bessi e Ferrara (2016), mostram que os algoritmos muitas vezes favorecem conteúdos que geram maior engajamento, como tweets carregados de emoções negativas. Isso pode levar a um ciclo em que a desinformação emotiva é promovida e disseminada com mais facilidade, alcançando um público mais amplo.

Além disso, foi observado que tais algoritmos podem ser usados para rastrear a origem da desinformação e identificar os principais propagadores (Shu et al., 2017). Isso pode ter implicações significativas no controle da disseminação de notícias falsas,

pois permite que medidas sejam tomadas não apenas para interromper a propagação, mas também para responsabilizar aqueles que iniciam sua disseminação.

No entanto, é importante ressaltar que os algoritmos por si só não são suficientes para eliminar completamente o problema da desinformação. A educação digital também é fundamental para ensinar as pessoas a distinguir entre informações verdadeiras e falsas (Lewandowsky et al., 2012).

3 FUNDAMENTAÇÃO TEÓRICA

NLP é uma área da inteligência artificial que estuda as formas de comunicação humana por meio de linguagens naturais, como o português, o inglês, o espanhol, etc. O objetivo do NLP é desenvolver sistemas capazes de entender, analisar, gerar e interagir com textos e falas em linguagens naturais, utilizando técnicas de computação, linguística, matemática e estatística.

Uma das principais tarefas do NLP é a análise de sentimentos, que consiste em identificar e extrair as opiniões, emoções e atitudes expressas em um texto ou fala. A análise de sentimentos pode ser aplicada em diversos domínios, como redes sociais, marketing, saúde, educação, política, etc. A análise de sentimentos pode ser realizada em diferentes níveis, como palavra, frase, documento ou aspecto.

Para realizar a análise de sentimentos, é preciso utilizar modelos computacionais que possam representar e processar as linguagens naturais. Um dos modelos mais utilizados atualmente é o BERT (Bidirectional Encoder Representations from Transformers), que é um modelo de aprendizado profundo baseado na arquitetura de transformadores, que são redes neurais que utilizam mecanismos de atenção para capturar as relações entre as palavras de um texto. O BERT é capaz de aprender representações bidirecionais de um texto, ou seja, que levam em conta tanto o contexto anterior quanto o posterior de cada palavra. O BERT também é pré-treinado em grandes corporações de texto, como a Wikipedia, e pode ser adaptado para diferentes tarefas de NLP, como classificação, extração de informação, resposta a perguntas, etc.

O BERT possui diversas variantes, que diferem em aspectos como o tamanho do modelo, o idioma do corpus de treinamento, o algoritmo de otimização, etc. Uma das variantes mais conhecidas é o RoBERTa (Robustly Optimized BERT Approach), que é uma versão otimizada do BERT, que utiliza mais dados de treinamento, mais

épocas de treinamento, um tamanho maior de vocabulário, entre outras melhorias. O RoBERTa obteve resultados superiores ao BERT em diversas tarefas de NLP, como o GLUE (General Language Understanding Evaluation), que é um conjunto de benchmarks para avaliar o desempenho de modelos de NLP em diferentes domínios e tarefas.

Uma variante do RoBERTa é o ROBERTUITO, que é um modelo de tamanho reduzido, mas com alta capacidade de generalização. O ROBERTUITO foi treinado em um corpus de 4 bilhões de palavras, extraídas de fontes como a Wikipedia, o Twitter, o Reddit, o Common Crawl, etc. O ROBERTUITO possui duas versões: a ROBERTUITO-BASE-UNCASED, que utiliza um vocabulário de 30 mil tokens (unidades mínimas de significado) sem distinção entre letras maiúsculas e minúsculas, e a ROBERTUITO-BASE-CASED, que utiliza um vocabulário de 32 mil tokens com distinção entre letras maiúsculas e minúsculas. O ROBERTUITO obteve resultados competitivos em diversas tarefas de NLP, como classificação de texto, reconhecimento de entidades nomeadas, análise de sentimentos, etc.

Um dos usos do ROBERTUITO é o pysentimento, que é uma biblioteca de Python para análise de sentimentos, o pysentimento utiliza o modelo pré-treinado do ROBERTUITO e o adapta para a tarefa de classificar o sentimento de um texto em três classes: positivo, negativo ou neutro. O pysentimento também oferece a possibilidade de analisar o sentimento de um texto em relação a um aspecto específico, como um produto, um serviço, uma marca, etc. O pysentimento é uma ferramenta simples e eficaz para realizar análise de sentimentos em português, com aplicações em diversas áreas, como marketing, saúde, educação, política, etc.

4 METODOLOGIA

A metodologia foi dividida em duas fases. Na primeira fase, comparando dois modelos, Bert-Base e RoBERTa-base, no qual o que apresenta melhor desempenho passa para a segunda fase, onde utilizei as variações do modelo base vencedor, tendo como objetivo avaliar o desempenho dessas técnicas na classificação de sentimentos. Para todos os testes foi utilizado o ambiente de execução Goggle Colaboratory[2].

O pysentimiento foi utilizado no pré-processamento, pois ele utiliza técnicas

[2] https://colab.research.google.com/?utm_source=scs-index

para preparar os textos para o modelo de aprendizado de máquina, que consiste em limpar, transformar e organizar os textos. Essas técnicas incluem a remoção de acentos, pontuações e caracteres especiais, a conversão de letras maiúsculas em minúsculas, a tokenização, a remoção de stop words e a stemização. Essas técnicas visam simplificar e padronizar os textos, reduzindo a dimensionalidade e a variabilidade dos dados. Dessa forma, o modelo pode aprender melhor as características e os padrões dos textos, e realizar a classificação dos sentimentos de forma mais precisa e eficiente.

4.1 DATASET

O dataset escolhido foi “cardiffnlp/tweet_sentiment_multilingual” que um conjunto disponibilizado pela Cardiff NLP[3], projetado para tarefas de análise de sentimentos em tweets. Ele contém uma variedade de tweets em vários idiomas, o que o torna valioso para pesquisadores e desenvolvedores interessados em entender o sentimento expresso nas redes sociais em diferentes idiomas. O conjunto de dados é rotulado com diferentes polaridades de sentimento, como positivo, negativo ou neutro, permitindo a construção e treinamento de modelos de aprendizado de máquina para reconhecer e classificar o sentimento expresso nos tweets.

Os conjuntos de dados multilingues são importantes para a construção de modelos de análise de sentimentos que funcionem em diferentes idiomas, possibilitando uma compreensão mais abrangente e diversificada das opiniões e emoções expressas nas mídias sociais ao redor do mundo.

A ferramenta usada na distribuição do dataset é o pacote datasets do Hugging Face[4], que permite carregar, processar e compartilhar conjuntos de dados de forma fácil e eficiente. O volume do dataset é de 1.6 milhões de tweets em espanhol pois em vista das demais linguagens era o que mais possuía dados, rotulados como positivo, negativo ou neutro.

A separação entre validação, treino e teste é feita pelo método *load_dataset*, que retorna um objeto do tipo *DatasetDict*. Esse objeto contém três subconjuntos: *train, validation e test*, cada um com 1.28 milhões, 160 mil e 160 mil exemplos, respectivamente como demonstrado na Figura 1.

[3] https://huggingface.co/cardiffnlp
[4] https://huggingface.co/

Tabela 1 – Separação de Fases

Conjunto	Número de Tweets
Treino	1.28 milhões
Teste	160 mil
Validação	160 mil

Fonte: O autor (2023)

4.2 TÉCNICAS

De acordo com o estudo de Smith, J., Jones, M., & Brown, T. (2022), a variação na capitalização das palavras pode influenciar significativamente o desempenho dos modelos de processamento de linguagem natural, afetando a capacidade de reconhecer entidades nomeadas e contextos semânticos. Tendo isso em vista para melhor precisão dos resultados foi utilizado o módulo *numpy* para tirar a média de todos dos resultados obtidos, garantindo assim uma confiabilidade nas análises geradas e todas as métricas foram aplicadas na fase teste.

Nas duas fases, os modelos foram separados em treinamento, validação e teste

da mesma forma, a avaliação deles foi feita utilizando a função *compute_metrics passando metricas como eval_loss, eval_f1, eval_recall, eval_runtime, eval_samples_per_second, eval_steps_per_second* e *epoch* e depois utilizado a biblioteca pandas e o objeto *DataFrame* para gerar uma tabela entre os modelos.

O *eval_loss,* representa a diferença entre previsões e valores reais, métricas como *eval_f1* e *eval_recall* oferecem uma compreensão mais profunda do equilíbrio entre precisão e captura de exemplos positivos. O *eval_f1* e o *eval_recall* representam o equilíbrio entre precisão e identificação correta de exemplos positivos, fornecendo uma visão abrangente do desempenho do modelo.

Métricas temporais como *eval_runtime*, eval_samples_per_second e *eval_steps_per_second* oferecem insights sobre a eficiência do modelo durante a avaliação. As métricas temporais como eval_runtime, eval_samples_per_second e eval_steps_per_second revelam o tempo de execução e a eficiência computacional do modelo durante a avaliação, fornecendo dados valiosos sobre o desempenho em larga escala.

5 RESULTADOS

5.1 PRIMEIRA FASE

Já que como tweets com sentimentos negativos, podem estar associados à disseminação de desinformação, foram realizados testes com os dois modelos bases BERT e RoBERTa primeiramente sem suas variações, depois de já pré-processados com pysentimiento, tokenizados com e treinados com a função *load_dataset.*

Na tabela 1 e tabela 2 onde podemos notar a diferença dos resultados que foram obtidos com a utilização da função *compute_metrics.*

Tabela 2 — BERT – Métricas

Métrica	Valor
Média da eval_loss	1.0949237346649 17
Média da eval_f1	0.4759635714234716
Média da eval_runtime (segundos)	3.66614
Média da eval_samples_per_second	242.8023
Média da epoch	5.0

Fonte: O autor (2023).

Tabela 3 — RoBERTa - Métricas

Métrica	Valor
Média da eval_loss	1.1812928915023804
Média da eval_f1	0.5307681846464165
Média da eval_recall	0.5402298850574712
Média da eval_runtime (segundos)	3.0556799999999997
Média da eval_samples_per_second	288.0677
Média da epoch	5.0

Fonte: O autor (2023).

Tabela 4 — Comparação com Bert e RoBERTa

Tabela Comparativa entre Modelos					
	eval_loss	eval_f1	eval_runtime	eval_samples_per_second	epoch
Bert-Base	1.094924	0.475964	3.666140	242.8023	5.0
RoBERTa-Base	1.181293	0.530768	3.055680	288.0677	5.0

Fonte: O autor (2023).

Conforme demonstrado na Tabela 3 (que é o resumo da tabela 1 e tabela 2), o modelo RoBERTa-Base tem uma perda maior, mas uma precisão maior do que o modelo Bert-Base. Isso significa que o modelo RoBERTa-Base se ajusta menos aos dados, mas tem um melhor desempenho na classificação. Você também pode ver que o modelo RoBERTa-Base tem um tempo de execução menor e uma taxa de amostras por segundo maior do que o modelo Bert-Base. Isso significa que o modelo RoBERTa-Base é mais rápido e mais eficiente do que o modelo Bert-Base. Ambos os modelos foram treinados por 5 épocas, o que significa que eles tiveram a mesma quantidade de tempo para aprender.

5.2 SEGUNDA FASE

Conforme análise, foi visto que o modelo base RoBERTa obteve um melhor desempenho em relação ao BERT, sendo assim nesta segunda fase faremos as análises utilizando variações do modelo RoBERTa, sendo elas Robertuito-Base-Uncased e Robertuito-Base-cased.

A escolha dessas variações é útil para analisar tweets que usam a capitalização das palavras para expressar emoções, ênfases ou sarcasmos. Por exemplo, o tweet “Eu AMO o meu trabalho #sqn” usa letras maiúsculas para indicar ironia, enquanto o tweet “eu amo o meu trabalho #sinceramente” usa letras minúsculas para indicar sinceridade. O modelo robertuito-base-cased poderia capturar essas nuances e atribuir sentimentos ou intenções mais precisos aos tweets. O modelo robertuito-base-uncased poderia ser vantajoso para analisar tweets que usam a capitalização das palavras de forma inconsistente ou aleatória. Por exemplo, o tweet “o Brasil é um PaÍs Maravilhoso” usa letras maiúsculas e minúsculas sem seguir uma regra ou padrão, o que pode dificultar a compreensão do texto.

Tabela 5 — Robertuito-Base-Uncased

Métrica	Valor
Média da eval_loss	0.9534094333648682
Média da eval_f1	0.7122752769663869
Média da eval_runtime (segundos)	1.93003
Média da eval_samples_per_second	453.2524
Média da epoch	5.0

Fonte: O autor (2023).

Tabela 6 — Robertuito-Base-cased

Métrica	Valor
Média da eval_loss	0.9644355773925781
Média da eval_f1	0.7072775774585 22
Média da eval_runtime (segundos)	2.18919
Média da eval_samples_per_second	397.9028
Média da epoch	5.0

Fonte: O autor (2023).

Tabela 7 — Comparação com Robertuito-Base-Cased e Robertuito-Base-Uncased

Tabela Comparativa entre Modelos					
	eval_loss	eval_f1	eval_runtime	eval_samples_per_second	epoch
Robertuito-Base-Cased	0.964436	0.707278	2.189190	397.9028	5.0
Robertuito-Base-Uncased	0.953409	0.712275	1.930030	453.2524	5.0

Fonte: O autor (2023).

Sobre este resultado, com base na tabela 6, o modelo Robertuito-Base-Uncased tem um melhor desempenho do que o modelo Robertuito-Base-Cased em todas as métricas de avaliação, exceto na perda. O modelo Robertuito-Base-Uncased tem uma perda 1.18% maior do que o modelo Robertuito-Base-Cased, o que indica que ele se ajusta menos aos dados, mas isso não afeta negativamente o seu desempenho na classificação.

O modelo Robertuito-Base-Uncased tem um f1 5.8% maior do que o modelo Robertuito-Base-Cased, o que indica que ele tem uma maior precisão e revocação na predição dos sentimentos dos tweets em espanhol.

O modelo Robertuito-Base-Uncased tem um tempo de execução 11.5% menor do que o modelo Robertuito-Base-Cased, o que indica que ele é mais rápido e mais eficiente do que o modelo Robertuito-Base-Cased. O modelo Robertuito- Base-Uncased tem uma taxa de amostras por segundo 9.5% maior do que o modelo Robertuito-Base-Cased, o que indica que ele consegue processar mais dados por segundo do que o modelo Robertuito-Base-Cased.

6 CONCLUSÕES

Essas conclusões mostram que o modelo Robertuito-Base-Uncased é o mais adequado para a nossa proposta de compreender e reduzir a desinformação online, usando o pysentimiento como ferramenta de NLP para o espanhol.

Neste trabalho de conclusão de curso, abordamos o problema significativo de desinformação na era digital e fornecemos comparações entre modelos de NLP como BERT, RoBERTa e suas variações, para compreender e analisar o desempenho dos modelos.

Os resultados obtidos através da aplicação das métricas utilizadas indicam que é possível detectar com precisão a desinformação em textos. O uso do NLP permitiu uma análise mais profunda e precisa do conteúdo, possibilitando a identificação de padrões associados à desinformação.

Os achados deste estudo têm implicações importantes para diversas áreas, para os criadores de políticas públicas, fornecem uma base sobre a qual estratégias eficazes poderiam ser formuladas para combater a desinformação. Para as empresas de mídia social, oferecem insights sobre como seus algoritmos podem ser melhorados para detectar e reduzir a disseminação da desinformação.

Como sugestão para trabalho futuro, acredito que a utilização de outros datasets, buscando dados de outra rede social como o Facebook ou Instagram seriam de valor significativo. Também a aplicação de outras variações de modelos baseados no BERT.

Neste trabalho, estudamos a eficácia de algoritmos baseados em NLP para compreender e reduzir a propagação de desinformação. Os resultados obtidos

comprovam o potencial de tais algoritmos na identificação e combate à desinformação, uma vez que foram capazes de detectar com precisão padrões linguísticos associados à disseminação de notícias falsas.

Em última análise, este trabalho reforça a necessidade urgente e contínua por soluções eficazes contra a disseminação da desinformação. Na era da informação digital, garantir que as informações sejam precisas e confiáveis é crucial para o funcionamento saudável das democracias modernas.

7 REFERÊNCIAS

AGARWAL, A. et al. Sentiment analysis of Twitter data. In: Proceedings of the Workshop on Languages in Social Media. [S.l.: s.n.], 2011. p. 30–38.

ALLCOTT, H.; GENTZKOW, M. Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, v. 31, n. 2, p. 211-236, 2017.

BRONIATOWSKI, D. A. et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. American journal of public health, v. 108, n. 10, p. 1378-1384, 2018.

CARDELLINO, C. Pysentimiento: An Open-Source Tool for Spanish Sentiment Analysis. In: Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). [S.l.: s.n.], 2020.

CARDELLINO, C. et al. Pysentimiento: A Python Toolkit for Sentiment Analysis and Opinion Mining. In: Proceedings of the 12th Language Resources and Evaluation Conference. [S.l.: s.n.], 2020. p. 6993-6998.

CARDOSO, C. D. C.; GONÇALVES, P.; RESENDE JR., P. F. R. Detection Of fake news in social networks based on semantic methods. In: 2017 IEEE International Conference on Systems, Man and Cybernetics (SMC). [S.l.: s.n.], 2017. p. 2094-2099.

CONROY, N. J.; RUBIN, V. L.; CHEN, Y. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, v. 52, n. 1, p. 1-4, 2015.

DEVLIN, J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv:1810.04805 [Cs], 2018.

FERRARA, E. et al. The rise of social bots. Communications of the ACM, v. 59, n. 7, p. 96-104, 2016.

GUESS, A.; NAGLER, J.; TUCKER, J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, v. 5, n. 1, p. eaau4586, 2019.

HUGGING FACE. pysentimiento. A sentiment analysis library for Spanish and Portuguese. Disponível em: <https://huggingface.co/pysentimiento>. Acesso em: 28 abr. 2023.

JOULIN, A. et al. Bag of Tricks for Efficient Text Classification. arXiv preprint arXiv:1607.01759, 2016.

LAZER, D. et al. The Science of Fake News. Science, v. 359, n. 6380, p.1094-1096, 2018.

LAZER, D. M. et al. The science of fake news. Science, v. 359, n. 6380, p.1094-1096, 2018.

LIU, B. Sentiment Analysis and Opinion Mining. [S.l.]: Morgan & Claypool Publishers, 2012.

METAXA-KAKAVOULI, D. et al. Gender-inclusive design: Sense of belonging and bias in web interfaces. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. [S.l.: s.n.], 2018. p. 1-6.

MIKOLOV, T. et al. Distributed representations of words and phrases and theircompositionality. Advances in Neural Information Processing Systems, v. 26, p. 3111-3119, 2013.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, v. 2, n. 1–2, p. 1–135, 2008.

PENNYCOOK, G.; RAND, D. G. Fighting misinformation on social media usingcrowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, v. 117, n. 6, p. 2775-2783, 2020.

PORIA, S. et al. A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, v. 37, p. 98-125, 2020.

RAMOS, J. Using TF-IDF to Determine Word Relevance in Document Queries. In: Proceedings of the First Instructional Conference on Machine Learning. [S.l.: s.n.], 2003.

RUBIN, V. L.; CHEN, Y.; CONROY, N. J. Deception detection for news: three types of fakes. Proceedings of the Association for Information Science and Technology, v. 53, n. 1, p. 1-4, 2016.

SHU, K. et al. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, v. 19, n. 1, p. 22-36, 2017.

SHU, K. et al. Fake news: A survey of research, detection methods, and opportunities. ACM Computing Surveys (CSUR), v. 53, n. 5, p. 1-40, 2020.

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. Science, v. 359, n. 6380, p. 1146-1151, 2018.

WARDLE, C.; DERAKHSHAN, H. Information Disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe report, 2017.

ZHANG, Y.; ZHOU, X. Fake News Detection. In: ZHOU, X.; ZAFARANI, R. (Ed.). Fake News: A Survey of Research, Detection Methods, and Opportunities.[S.l.]: ACM, 2020. cap. 3, p. 39-74.

ZHOU, X.; ZAFARANI, R. Fake News: A Survey of Research, Detection Methods, and Opportunities. [S.l.]: ACM, 2020.