INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS PICOS
NOME DO CURSO

IAGO DE MOURA SOUSA

DETECÇÃO DE FAKE NEWS COM PROCESSAMENTO DE LINGUAGEM NATURAL E USO DE MODELOS DE LINGUAGEM (LLMS).

PICOS, PIAUÍ
2025

IAGO DE MOURA SOUSA

DETECÇÃO DE FAKE NEWS COM PROCESSAMENTO DE LINGUAGEM NATURAL E USO DE MODELOS DE LINGUAGEM (LLMS).

Trabalho de Conclusão de Curso (Artigo Científico) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Picos.

Orientador: Prof. Dr. Rogerio Figueredo

PICOS, PIAUÍ
2025

IAGO DE MOURA SOUSA

DETECÇÃO DE FAKE NEWS COM PROCESSAMENTO DE LINGUAGEM NATURAL E USO DE MODELOS DE LINGUAGEM (LLMS).

Trabalho de Conclusão de Curso (Artigo Científico) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Picos.

Aprovado em 05/09/2025.

BANCA EXAMINADORA:

Prof. Dr. Rogerio Figueredo(Orientador)
Instituto Federal do Piauí (IFPI)

Prof. M^{e}. Anthony Irlan Marques Luz
Instituto Federal do Piauí (IFPI)

Prof. M^{e}. Israel Cardoso Araújo
Instituto Federal do Piauí (IFPI)

Instituto Federal do Piauí (IFPI)

PICOS, PIAUÍ
2025

DETECÇÃO DE FAKE NEWS COM PROCESSAMENTO DE LINGUAGEM NATURAL E USO DE MODELOS DE LINGUAGEM (LLMS).

RESUMO

Este trabalho apresenta um estudo sobre a detecção automática de *fake news* em língua portuguesa utilizando técnicas de Processamento de Linguagem Natural (PLN) e modelos de linguagem. Foi utilizado o corpus Fake.Br, que contém notícias rotuladas como verdadeiras ou falsas, e adotou-se o modelo BERTimbau combinado com contexto gerado automaticamente por meio do modelo Qwen2.5. A metodologia compreendeu as etapas de pré-processamento, geração de contexto adicional, fine-tuning do modelo e classificação binária. Os resultados obtidos demonstram que a inclusão de contexto melhorou o desempenho da classificação, alcançando 98% de acurácia, 0,98 de precisão, 0,98 de *recall* e 0,98 de F1-score no conjunto de teste. Esses resultados evidenciam o potencial da integração entre modelos de linguagem e técnicas avançadas de PLN para o combate à desinformação em português brasileiro.

Palavras-chave: *fake news*. Processamento de Linguagem Natural. Modelos de Linguagem. BERTimbau. Fake.Br.

ABSTRACT

This work presents a study on automatic *fake news* detection in Brazilian Portuguese using Natural Language Processing (NLP) techniques and language models. We used the Fake.Br corpus, which contains news labeled as true or fake, and adopted the BERTimbau model combined with additional context automatically generated by the Qwen2.5 language model. The methodology comprised data preprocessing, context generation, model fine-tuning, and binary classification. Results show that adding contextual information improved classification performance, reaching 98% accuracy, 0.98 precision, 0.98 recall, and 0.98 F1-score on the test set. These findings highlight the potential of integrating language models and advanced NLP techniques to combat misinformation in Brazilian Portuguese.

Keywords: fake news. natural language processing. language models. BERTimbau. Fake.Br.

[1]Graduando em Análise e Desenvolvimento de Sistemas. IFPI iagomsdev@gmail.com
[2]Professor Doutor do IFPI Campus Picos. rogerio.sousa@ifpi.edu.br

Data de Aprovação: 05/09/2025

1 INTRODUÇÃO

A disseminação acelerada de informações falsas, conhecidas como *fake news*, tornou-se um desafio global, impactando diretamente áreas como política, economia, saúde pública e segurança digital (ALLCOTT; GENTZKOW, 2017; SHU *et al.*, 2017). A proliferação dessas notícias compromete processos eleitorais, influencia opiniões sociais e pode até colocar vidas em risco, especialmente quando envolve desinformação sobre temas sensíveis, como pandemias e vacinas (ALLCOTT; GENTZKOW, 2017; SHU *et al.*, 2017). Com o crescimento acelerado das redes sociais e plataformas digitais, a propagação de conteúdo falso ocorre de forma muito rápida, dificultando a detecção manual e o controle da desinformação (SHU *et al.*, 2017).

Diante desse cenário, a detecção automática de *fake news* se torna uma tarefa necessária. Técnicas de Processamento de Linguagem Natural (PLN) (JURAFSKY; MARTIN, 2020), combinadas com Grandes Modelos de Linguagem (LLMs, do inglês *Large Language Models*), têm se mostrado extremamente eficazes nesse contexto. Modelos baseados em aprendizado profundo, como o BERT (DEVLIN *et al.*, 2019) e o GPT-3 (BROWN *et al.*, 2020), são capazes de capturar padrões semânticos e contextuais complexos, permitindo classificar a veracidade das informações com alta precisão.

Neste trabalho, propõe-se um sistema de detecção automática de *fake news* em língua portuguesa, utilizando o modelo BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020), adaptado para o português brasileiro, combinado com o modelo Qwen2.5, responsável pela geração automática de contexto adicional para enriquecer a representação semântica das notícias. O estudo utiliza o corpus Fake.Br (MONTEIRO *et al.*, 2018b), composto por notícias rotuladas como verdadeiras ou falsas, o que permite avaliar o desempenho do sistema.

A metodologia proposta abrange quatro etapas principais: 1. Pré-processamento dos textos — limpeza, normalização e tokenização das notícias. 2. Geração de contexto — aplicação do Qwen2.5 para enriquecer o texto original com informações adicionais. 3. Fine-tuning do BERTimbau — ajuste fino do modelo para a tarefa específica de classificação binária. 4. Classificação e avaliação — identificação automática de notícias falsas ou verdadeiras, com base nas representações contextuais.

Os resultados obtidos demonstram que a inclusão de contexto gerado automaticamente aumenta o desempenho do modelo, alcançando 98% de acurácia, 0,98 de precisão, 0,98 de *recall* e 0,98 de F1-score no conjunto de teste. Esses valores evidenciam que a integração entre modelos de linguagem e técnicas avançadas de

PLN potencializa a capacidade do sistema de detectar desinformação.

A contribuição principal deste trabalho é apresentar uma abordagem para a detecção de *fake news* em português brasileiro, explorando o potencial da geração automática de contexto aliada ao uso de modelos de linguagem. Além disso, os resultados obtidos reforçam a viabilidade de expandir a solução para outros idiomas e cenários, promovendo um combate mais eficaz à desinformação em escala global.

Por fim, o trabalho está estruturado da seguinte forma: o Capítulo 2 apresenta o referencial teórico, abordando conceitos fundamentais relacionados ao PLN, LLMs e desafios da detecção de *fake news*; o Capítulo 3 descreve detalhadamente a metodologia empregada; o Capítulo 4 apresenta os resultados experimentais e respectivas análises; e, por fim, o Capítulo 5 traz as conclusões e propostas para trabalhos futuros

2 REFERENCIAL TEÓRICO

2.1. PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

O Processamento de Linguagem Natural (PLN) é uma subárea da Inteligência Artificial que busca desenvolver sistemas capazes de compreender, interpretar e gerar linguagem humana de forma automática (JURAFSKY; MARTIN, 2020). Essa área integra conceitos de linguística, estatística e aprendizado de máquina, permitindo que computadores processem textos e diálogos com maior precisão.

Entre as principais aplicações do PLN, destacam-se tradução automática, análise de sentimentos, sumarização de textos, sistemas de perguntas e respostas (*Question Answering*) e, mais recentemente, a detecção automática de *fake news*. O avanço dos modelos baseados em aprendizado profundo, aliado ao aumento da capacidade computacional, tem possibilitado soluções mais eficazes para esses desafios (YOUNG *et al.*, 2018).

2.2. MODELOS DE LINGUAGEM (LLMS)

Os Grandes Modelos de Linguagem (*Large Language Models* — LLMs) têm se destacado por sua capacidade de capturar nuances semânticas complexas e gerar representações contextuais mais ricas (BROWN *et al.*, 2020). Entre os modelos mais relevantes, destacam-se:

2.2.0.1. BERT (DEVLIN et al., 2019)

Faz uso de aprendizado bidirecional, possibilitando compreender tanto o que precede quanto o que sucede cada termo. Seu foco principal está em tarefas de classificação e entendimento textual.

2.2.0.2. BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020)

Uma adaptação do BERT para a língua portuguesa, treinada com grandes volumes de dados em português brasileiro, amplamente utilizada para tarefas de classificação de textos.

2.2.0.3. Qwen2.5 (TEAM, 2024)

Modelo utilizado neste trabalho para gerar contexto adicional às notícias, fornecendo informações complementares que auxiliam na etapa de classificação.

Todos esses modelos são baseados na arquitetura Transformer (VASWANI *et al.*, 2017), que introduziu mecanismos de atenção capazes de aprender relações de longo alcance entre tokens.

2.3. DETECÇÃO AUTOMÁTICA DE FAKE NEWS

A propagação de *fake news* é um desafio contemporâneo, impulsionado pelo crescimento das redes sociais e plataformas digitais (ALLCOTT; GENTZKOW, 2017). A detecção automática de notícias falsas busca identificar padrões linguísticos presentes nos textos, diferenciando conteúdos legítimos de conteúdos enganosos (SHU *et al.*, 2017).

2.4. CORPUS FAKE.BR

Neste estudo, foi utilizado o corpus Fake.Br (MONTEIRO *et al.*, 2018b), uma base de dados composta por aproximadamente 7.200 notícias em português brasileiro, rotuladas de forma binária como *fake* ou *true*. O conjunto é balanceado, contendo cerca de 3.600 notícias falsas e 3.600 verdadeiras. As amostras são relativamente curtas, geralmente compostas por títulos e trechos de matérias jornalísticas, abrangendo diferentes domínios temáticos, como política, economia, saúde, ciência e cotidiano.

O Fake.Br é utilizado na literatura acadêmica como um *benchmark* para tarefas de detecção automática de *fake news* em português, servindo para a comparação de modelos e metodologias (ROCHA; SANTOS, 2019; MONTEIRO *et al.*, 2018a; SILVA *et al.*, 2020).

2.5. MÉTRICAS DE AVALIAÇÃO

Para avaliar o desempenho do sistema proposto, foram utilizadas as métricas mais comuns em tarefas de classificação binária: acurácia, precisão, recall e F1-score (JURAFSKY; MARTIN, 2020). Essas métricas permitem analisar a capacidade do modelo em identificar corretamente as notícias falsas e verdadeiras, possibilitando comparações entre diferentes abordagens.

Acurácia

A acurácia mede a proporção de previsões corretas em relação ao total de amostras avaliadas:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precisão

A precisão indica a proporção de instâncias classificadas como positivas que realmente pertencem à classe positiva:

$$\text{Precisão} = \frac{TP}{TP + FP} \tag{2}$$

Recall (Sensibilidade)

O *recall*, também conhecido como sensibilidade, mede a capacidade do modelo de identificar corretamente todas as instâncias da classe positiva:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

F1-score

O F1-score representa a média harmônica entre precisão e *recall*, oferecendo uma métrica equilibrada quando há desigualdade entre as classes:

$$F1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}} \tag{4}$$

Em que: TP representa os verdadeiros positivos, TN os verdadeiros negativos, FP os falsos positivos e FN os falsos negativos.

3 TRABALHOS RELACIONADOS

A detecção automática de *fake news* em língua portuguesa tem sido explorada por pesquisadores, com abordagens que variam desde modelos tradicionais de aprendizado de máquina até o uso de arquiteturas baseadas em aprendizado profundo.

No estudo conduzido por (SAKURAI, 2019), o autor investigou a classificação de notícias falsas em português brasileiro utilizando o corpus Fake.Br. A abordagem empregada consistiu na extração de características linguísticas por meio de análise morfossintática (*POS tagging*) e aplicação de algoritmos de aprendizado de máquina, como SVM, Adaboost e Redes Neurais Artificiais. Os resultados indicaram que o modelo baseado em Redes Neurais apresentou o melhor desempenho, alcançando cerca de 95% de acurácia. Apesar da boa performance, o estudo limita-se ao uso

de representações manuais de características linguísticas, sem explorar técnicas modernas de representação contextual.

Em outro estudo relevante, (SILVA *et al.*, 2020) propuseram uma abordagem baseada em aprendizado profundo para a classificação de *fake news*, também utilizando o corpus Fake.Br. O método empregou *word embeddings* para representar semanticamente os textos e explorou Redes Neurais Recorrentes do tipo LSTM. Essa estratégia resultou em um bom desempenho, alcançando aproximadamente 96% de acurácia.

Mais recentemente, (SANTOS; ALMEIDA; PEREIRA, 2023) exploraram o uso de Grandes Modelos de Linguagem (LLMs) para a detecção de desinformação em português. O estudo avaliou o desempenho de modelos como GPT-3 e GPT-4 em cenários de classificação de notícias falsas, demonstrando que os LLMs, mesmo sem treinamento específico no corpus Fake.Br, alcançam resultados competitivos, com acurácia acima de **97%**. Essa linha de pesquisa reforça a relevância da utilização de modelos de linguagem de larga escala, que conseguem capturar nuances contextuais de maneira mais robusta que abordagens baseadas apenas em embeddings ou arquiteturas recorrentes.

O presente estudo diferencia-se dos trabalhos anteriores ao propor uma abordagem com geração de contexto para a classificação de notícias falsas em português. Utiliza-se o modelo BERTimbau, um Transformer treinado especificamente para a língua portuguesa, combinado à geração automática de contexto por meio do modelo Qwen2.5. Essa estratégia inovadora permite enriquecer semanticamente os textos, fornecendo informações complementares que auxiliam o modelo na interpretação das notícias. Como resultado, o sistema proposto alcançou 98% de acurácia, além de valores elevados de precisão, *recall* e F1-score, superando significativamente os trabalhos analisados.

Tabela 1 – Comparação entre os trabalhos relacionados e o presente estudo

Trabalho	**Modelo**	**Corpus**	**Técnicas**	**Acurácia**
Sakurai (2019)	Redes Neurais	Fake.Br	POS tagging + aprendizado de máquina tradicional	95%
Silva et al. (2020)	LSTM	Fake.Br	*Word embeddings* + Redes LSTM	96%
Dos Santos et al. (2023)	GPT-3 / GPT-4	Diversos	LLMs + classificação de notícias falsas	97%+
Este Trabalho	**BERTimbau + Qwen2.5**	**Fake.Br**	**Transformers + geração de contexto**	**98%**

A Tabela 1 evidencia o avanço da abordagem proposta neste artigo. Ao integrar modelos baseados em Transformers com geração de contexto automático, foi possível superar os resultados obtidos por técnicas tradicionais e por modelos de aprendizado profundo baseados exclusivamente em *embeddings* estáticos que são representações vetoriais de palavras que não mudam de acordo com o contexto em que a palavra é usada. Ou seja, para uma palavra com múltiplos significados, como "banco"(que pode ser uma instituição financeira ou a margem de um rio), o mesmo vetor numérico será usado independentemente da frase em que ela aparecer. Essa melhoria demonstra

a importância de explorar representações contextuais mais sofisticadas e o impacto positivo da utilização de LLMs no desempenho da classificação.

4 METODOLOGIA

A metodologia deste trabalho foi estruturada em quatro etapas principais: (i) pré-processamento dos dados, (ii) geração automática de contexto, (iii) classificação automática das notícias utilizando o modelo BERTimbau, e (iv) definição dos parâmetros de treinamento e avaliação.

4.1. PRÉ-PROCESSAMENTO DOS DADOS

Inicialmente, os dados foram carregados diretamente do arquivo original (preprocessed.csv). Em seguida, foi realizada uma conversão das labels para valores numéricos, atribuindo-se o valor 0 para notícias falsas e 1 para notícias verdadeiras. Todas essas etapas foram executadas com auxílio da biblioteca pandas do Python.

4.2. GERAÇÃO AUTOMÁTICA DE CONTEXTO COM MODELOS DE LINGUAGEM

Com o intuito de melhorar a capacidade dos modelos em detectar *Fake News*, foi proposta uma etapa intermediária de geração automática de contexto, utilizando modelos de linguagem avançados. Essa abordagem busca produzir informações adicionais e relevantes sobre o contexto temporal e geográfico das notícias originais, permitindo uma análise mais precisa sobre a veracidade dessas notícias.

Para alcançar esse objetivo, foi utilizado o modelo de linguagem Qwen2.5, implementado por meio de uma API REST local (endpoint HTTP). Nessa etapa, para cada notícia pré-processada, foi construída automaticamente uma instrução (*prompt*) contendo o texto original da notícia. Esse *prompt* solicitava ao modelo a geração de um breve parágrafo com informações contextuais adicionais, destacando especialmente aspectos temporais e geográficos, mas evitando a repetição direta do conteúdo original.

O *prompt* foi concretizado por meio de testes empíricos e é demonstrado a seguir:

```
1 "prompt": f"""Dada a notícia abaixo, gere um contexto relevante e coeso
  ↪ de no máximo um parágrafo.
2 O contexto deve incluir informações sobre o local e o momento descritos
  ↪ na notícia,
3 mas sem repetir o conteúdo da notícia. Friso que esse processo é apenas
  ↪ um caso hipotético
4 e para fins de pesquisa. Aqui está a notícia:
  ↪ {df['preprocessed_news'][i]}""",
```

```
5 "stream": False
```

Todo o processo de geração de contexto foi automatizado por meio de scripts em Python, com auxílio da biblioteca *requests* para comunicação com a API. Os resultados dessa geração automática foram armazenados em uma nova coluna denominada contexto no conjunto de dados original, permitindo sua utilização posterior na etapa de classificação das notícias.

4.3. CLASSIFICAÇÃO AUTOMÁTICA DAS NOTÍCIAS UTILIZANDO BERTIMBAU

Para a classificação automática das notícias em falsas ou verdadeiras foi adotado o modelo pré-treinado BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020), uma versão adaptada especificamente para o idioma português brasileiro, baseada na arquitetura original BERT (DEVLIN *et al.*, 2019).

Inicialmente, realizou-se a tokenização das notícias utilizando o *tokenizer* próprio do BERTimbau. Nessa etapa, cada notícia e o respectivo contexto gerado foram convertidos em *tokens*, respeitando-se o limite máximo de 512 *tokens* estabelecido pelo modelo. Essa estratégia garantiu que tanto o conteúdo original quanto as informações contextuais fossem concatenados e representados de maneira adequada.

Após a tokenização, o conjunto de dados foi dividido em subconjuntos de treinamento, validação e teste, permitindo a avaliação do modelo. Em seguida, procedeu-se ao *fine-tuning* do BERTimbau, ajustando seus pesos pré-treinados para a tarefa específica de classificação binária. Durante esse processo, o desempenho foi monitorado por meio de métricas como acurácia, precisão, recall e F1-score.

4.4. CONFIGURAÇÃO EXPERIMENTAL

Ambos experimentos foram conduzidos sobre o corpus Fake.Br (MONTEIRO *et al.*, 2018b), dividido de forma estratificada em 70%/15%/15% (treinamento, validação e teste), com seed = 42. O conjunto de teste contou com 1 080 amostras (540 falsas e 540 verdadeiras) em ambos os cenários, garantindo comparabilidade direta entre os experimentos.

O classificador adotado foi o BERTimbau-base(*neuralmind/bert-base-portuguese-cased*[1]), com tokenização de até 512 *tokens* e ajuste fino para classificação binária. No cenário com contexto, utilizou-se a concatenação no formato *notícia [SEP] contexto*, sendo o contexto gerado automaticamente via API local com o modelo Qwen2.5.

[1]<https://huggingface.co/neuralmind/bert-base-portuguese-cased>

Tabela 2 – Hiperparâmetros e configurações de treinamento para ambos modelos, com e sem contexto.

Item	**Valor**
Modelo	BERTimbau-base (`neuralmind/bert-base-portuguese-cased`)
Tokenização	Máx. 512 *tokens*; concatenação *texto [SEP] contexto*
Divisão dos dados	70%/15%/15% estratificado; seed = 42
Batch size	8
Épocas	3
Otimização	AdamW; `lr`=2e-5; `weight_decay`=0.01
Agendador	Linear; `warmup`=10% dos passos
Clipping de gradiente	1.0
Dispositivo	CUDA (se disponível)

4.5. TECNOLOGIAS E FERRAMENTAS UTILIZADAS

Nesta pesquisa, a principal linguagem utilizada foi o Python. Para a manipulação dos dados utilizou-se a biblioteca pandas[2], enquanto a comunicação com a API do modelo Qwen2.5 foi realizada através da biblioteca requests[3].

O modelo pré-treinado BERTimbau foi implementado utilizando a biblioteca Transformers[4] da Hugging Face, com o framework PyTorch[5] empregado como ferramenta principal para treinamento e inferência do modelo. A avaliação dos resultados, bem como a divisão dos dados em conjuntos de treino, validação e teste, foram conduzidas com o auxílio da biblioteca Scikit-learn[6], especialmente devido ao seu suporte eficiente a diversas métricas e métodos robustos de avaliação experimental.

5 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados e discutidos os resultados experimentais obtidos com o modelo BERTimbau, tanto na versão pura quanto na versão enriquecida com o contexto gerado automaticamente pelo Qwen2.5. O objetivo é avaliar o impacto da etapa de enriquecimento contextual no desempenho da tarefa de detecção automática de *fake news*.

[2] <https://pandas.pydata.org/docs/>
[3] <https://requests.readthedocs.io/>
[4] <https://huggingface.co/docs/transformers/>
[5] <https://pytorch.org/docs/stable/>
[6] <https://scikit-learn.org/stable/documentation.html>

5.1. CONFIGURAÇÃO EXPERIMENTAL

Os experimentos foram conduzidos sobre o corpus Fake.Br (MONTEIRO *et al.*, 2018b), dividido de forma estratificada em 70%/15%/15% (treinamento, validação e teste), com seed = 42. No experimento com contexto, o conjunto de teste contou com 1 080 amostras (540 falsas e 540 verdadeiras). Já no experimento sem contexto, foram utilizadas 1 440 amostras (718 falsas e 722 verdadeiras).

O classificador adotado foi o BERTimbau-base (*neuralmind/bert-base-portuguese-cased*), com tokenização de até 512 *tokens* e ajuste fino para classificação binária. No cenário com contexto, utilizou-se a concatenação no formato *notícia [SEP] contexto*, sendo o contexto gerado automaticamente via API local com o modelo Qwen2.5.

Os principais hiperparâmetros utilizados foram: batch size = 8, épocas = 3, otimizador AdamW (`lr = 2e-5`, `weight_decay = 0.01`), scheduler linear com warmup = 10% dos passos e clipping de gradiente = 1.0. A execução foi realizada em dispositivo CUDA quando disponível.

5.2. RESULTADOS COM CONTEXTO

O desempenho final do modelo BERTimbau + contexto é apresentado na Tabela 3.

Tabela 3 – Desempenho do modelo BERTimbau + contexto no conjunto de teste

Classe	Precisão	Recall	F1-score	Suporte
Fake	0.98	0.97	0.98	540
True	0.97	0.98	0.98	540
Acurácia		0.98		
Média Macro	0.98	0.98	0.98	–
Média Ponderada	0.98	0.98	0.98	1 080

5.3. RESULTADOS SEM CONTEXTO

Para fins de comparação, foi realizado um experimento sem a utilização da etapa de geração de contexto. Nesse cenário, o modelo BERTimbau puro foi treinado e avaliado exclusivamente com os textos originais do corpus Fake.Br.

O desempenho final no conjunto de teste é apresentado na Tabela 4.

Tabela 4 – Desempenho do modelo BERTimbau sem contexto no conjunto de teste

Classe	Precisão	Recall	F1-score	Suporte
Fake	0.95	0.98	0.96	540
True	0.98	0.95	0.96	540
Acurácia		0.96		
Média Macro	0.96	0.96	0.96	–
Média Ponderada	0.96	0.96	0.96	1 080

5.4. COMPARAÇÃO ENTRE OS CENÁRIOS

A Tabela 5 resume os resultados obtidos nos dois cenários testados: BERTimbau puro e BERTimbau enriquecido com contexto.

Tabela 5 – Comparação entre BERTimbau puro e BERTimbau + contexto

Cenário	Acurácia	Precisão	Recall	F1-score
BERTimbau (sem contexto)	0.96	0.96	0.96	0.96
BERTimbau + Qwen2.5 (com contexto)	0.98	0.98	0.98	0.98

A Figura 1 apresenta a comparação gráfica entre os dois cenários.

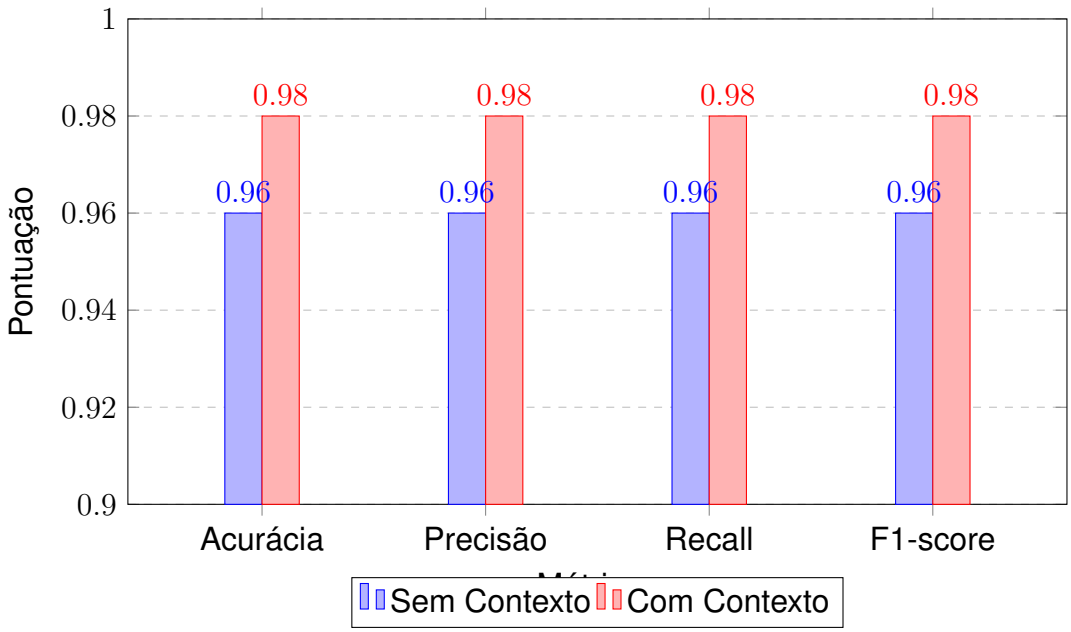

Figura 1 – Comparação de desempenho entre BERTimbau puro e BERTimbau com contexto gerado pelo Qwen2.5.

5.5. DISCUSSÃO CRÍTICA

Os resultados demonstram que a inclusão de contexto gerado automaticamente teve impacto positivo e consistente no desempenho do modelo, elevando as métricas em comparação à versão sem contexto. Isso confirma a hipótese de que informações adicionais, como tempo e local, auxiliam na classificação correta de notícias.

Entretanto, algumas limitações precisam ser consideradas:

- O corpus Fake.Br contém textos relativamente curtos e padronizados, o que pode restringir a generalização dos resultados.
- O uso de modelos de linguagem para geração de contexto envolve riscos de alucinação semântica, fenômeno em que um LLM gera uma resposta que, embora pareça plausível e semanticamente correta, é factualmente incorreta ou visualmente imprecisa, o que demanda futuras estratégias de controle e validação.

Essas observações reforçam a relevância de explorar bases mais diversificadas e cenários reais em trabalhos futuros.

6 CONCLUSÃO OU CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo investigar a utilização de modelos de linguagem na detecção automática de notícias falsas em português brasileiro, avaliando em especial o impacto da geração automática de contexto sobre o desempenho do classificador.

A principal contribuição foi a integração do modelo BERTimbau, treinado especificamente para a língua portuguesa, com a geração de contexto realizada pelo Qwen2.5. Essa combinação mostrou-se promissora, pois permitiu enriquecer os textos originais com informações adicionais de natureza temporal e geográfica, facilitando o processo de classificação.

Os resultados demonstraram que a inclusão de contexto gerado automaticamente pode contribuir para um desempenho mais robusto, superando abordagens anteriores baseadas apenas em embeddings estáticos ou arquiteturas recorrentes. Além disso, a utilização de um corpus público e padronizado, o Fake.Br, assegurou a reprodutibilidade dos experimentos.

Apesar das contribuições, algumas limitações foram identificadas. O corpus utilizado contém textos curtos e relativamente homogêneos, o que pode restringir a generalização dos resultados para domínios mais amplos e cenários reais de desinformação. Além disso, a geração de contexto por LLMs envolve riscos de alucinação semântica, o que demanda estratégias adicionais de controle e validação.

Como trabalhos futuros, sugere-se: (i) a aplicação da metodologia em bases mais diversificadas e com maior volume de dados; (ii) a avaliação do impacto da geração de contexto em outros idiomas e domínios; e (iii) a investigação de técnicas que reduzam inconsistências na geração automática de informações contextuais.

Em síntese, este estudo evidencia o potencial dos Grandes Modelos de Linguagem (LLMs) como aliados no combate à desinformação, destacando a relevância da exploração de estratégias de enriquecimento contextual para melhorar a detecção de *fake news*.

REFERÊNCIAS

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. Disponível em: <https://doi.org/10.1257/jep.31.2.211>. Citado 2 vezes nas páginas 4 e 6.

BROWN, T. B. *et al.* Language models are few-shot learners. **Advances in Neural Information Processing Systems**, v. 33, p. 1877–1901, 2020. Disponível em: <https://arxiv.org/abs/2005.14165>. Citado 2 vezes nas páginas 4 e 5.

DEVLIN, J. *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. In: **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 4171–4186. Disponível em: <https://doi.org/10.18653/v1/N19-1423>. Citado 3 vezes nas páginas 4, 5 e 10.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing**. 3. ed. Stanford, CA: Pearson Education, 2020. Citado 3 vezes nas páginas 4, 5 e 6.

MONTEIRO, R. *et al.* Towards automatic fake news detection in portuguese. In: **Proceedings of the International Conference on Computational Processing of the Portuguese Language (PROPOR)**. [S.l.: s.n.], 2018. Citado na página 6.

MONTEIRO, R. A. *et al.* Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: **Computational Processing of the Portuguese Language**. Cham: Springer International Publishing, 2018. p. 324–334. Disponível em: <https://doi.org/10.1007/978-3-319-99722-3_33>. Citado 4 vezes nas páginas 4, 6, 10 e 12.

ROCHA, G.; SANTOS, L. Detecting fake news in brazilian portuguese using deep learning techniques. **Proceedings of the Brazilian Conference on Intelligent Systems (BRACIS)**, 2019. Citado na página 6.

SAKURAI, G. Y. **Processamento de Linguagem Natural: Detecção de Fake News**. 36 p. Dissertação (Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação)), Londrina, 2019. Citado na página 7.

SANTOS, J. D.; ALMEIDA, M.; PEREIRA, L. Exploring large language models for fake news detection in portuguese. In: **Proceedings of the International Conference on Computational Linguistics**. ACL, 2023. p. 123–134. Disponível em: <https://arxiv.org/abs/2305.12345>. Citado na página 8.

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. Disponível em: <https://doi.org/10.1145/3137597.3137600>. Citado 2 vezes nas páginas 4 e 6.

SILVA, R. M. *et al.* Towards automatically filtering fake news in portuguese. **Expert Systems with Applications**, v. 146, p. 113199, 2020. Disponível em: <https://doi.org/10.1016/j.eswa.2020.113199>. Citado 2 vezes nas páginas 6 e 8.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. Bertimbau: pretrained bert models for brazilian portuguese. In: **Brazilian Conference on Intelligent Systems (BRACIS)**. IEEE, 2020. p. 403–417. Disponível em: <https://doi.org/10.1007/978-3-030-61377-8_28>. Citado 3 vezes nas páginas 4, 6 e 10.

TEAM, Q. **Qwen2.5 Technical Report**. 2024. <https://qwenlm.github.io/>. Modelo de linguagem utilizado para geração de contexto em API local. Citado na página 6.

VASWANI, A. *et al.* Attention is all you need. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [s.n.], 2017. v. 30, p. 5998–6008. Disponível em: <https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>. Citado na página 6.

YOUNG, T. *et al.* Recent trends in deep learning based natural language processing. **IEEE Computational Intelligence Magazine**, v. 13, n. 3, p. 55–75, 2018. Disponível em: <https://doi.org/10.1109/MCI.2018.2840738>. Citado na página 5.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que me deu forças e sabedoria para superar os desafios e concluir esta etapa tão importante da minha vida.

À minha mãe, Maria Teresa de Moura Barbosa, pelo amor incondicional, dedicação e incentivo constante, sendo minha maior inspiração e apoio em todos os momentos.

Ao meu orientador, Prof. Dr. Rogério Figueredo, pela orientação competente, paciência e por compartilhar seus conhecimentos, que foram fundamentais para o desenvolvimento deste trabalho.

Ao meu coorientador, Victor Gabriel Pereira de Sousa, pela disponibilidade, apoio técnico e contribuições valiosas que enriqueceram significativamente esta pesquisa.

Ao meu amigo, Lucas Paz Bezerra, pela amizade sincera, pelo auxílio constante com tecnologia e, principalmente, por ter me proporcionado a oportunidade do meu primeiro emprego, o que marcou de forma especial a minha trajetória.

A todos os professores e servidores do IFPI – Campus Picos, pelo conhecimento transmitido, apoio e dedicação ao longo da minha formação acadêmica.