

# Análise do ambiente informacional em empresas ou instituições

Alessandro Pessoa da Conceição Barreto, João Vitor Pigozzo, Tiago Rodrigues de Matos

## Abstract

Existe uma necessidade para o desenvolvimento de ferramentas eficazes para modelagem, classificação e análise de dados textuais, principalmente por conta do elevado número de informações textuais disponíveis. Neste contexto, este artigo apresenta uma análise comparativa entre dois modelos de linguagem natural amplamente utilizados: o BERT e o Longformer. O objetivo foi avaliar a capacidade de cada modelo na classificação de sentimentos em um corpus de notícias, considerando suas diferenças em termos de distribuição de rótulos, confiança nas classificações e adequação a textos extensos. A metodologia incluiu a coleta automatizada de dados, o pré-processamento textual com remoção de stopwords e sumarização extrativa, seguida pela aplicação dos modelos. Os resultados mostraram que o Longformer é mais adequado para lidar com textos longos, mas apresenta viés para classificação negativa, enquanto o BERT demonstrou maior equilíbrio na distribuição dos sentimentos. Por fim, este estudo discute as limitações de cada modelo e sugere abordagens futuras para aprimorar a classificação de sentimentos em cenários complexos.

## 1 Introdução

O crescente volume de informações disponíveis na internet torna essencial o desenvolvimento de ferramentas que permitam compreender e classificar automaticamente conteúdos textuais (Pang and Lee, 2008). No contexto das redes sociais, blogs e sites de notícias, a classificação de sentimentos desponta como uma tarefa de grande relevância, auxiliando na identificação de opiniões, percepções e tendências comportamentais (Cambria et al., 2017; Liu, 2012).

Modelos de linguagem natural baseados em redes neurais profundas têm evoluído significativamente nos últimos anos, impulsionados por arquiteturas como Transformer (Vaswani et al., 2017), que

permitiram avanços notáveis em tarefas como geração de texto, tradução automática e classificação de sentimentos. Entre os modelos mais destacados estão o BERT (Bidirectional Encoder Representations from Transformers) e o Longformer (Beltagy et al., 2020), ambos desenvolvidos para lidar com diferentes desafios de processamento textual.

O BERT, projetado para capturar o contexto bidirecional de palavras em uma frase, tornou-se amplamente utilizado por sua eficiência em várias tarefas de linguagem natural (Devlin et al., 2019). Contudo, sua limitação de comprimento máximo de entrada (512 tokens) restringe sua aplicação em textos mais extensos. Por outro lado, o Longformer, com sua capacidade de processar sequências textuais mais longas (até 4096 tokens), surge como uma solução promissora para análises de textos de maior complexidade.

Neste trabalho, investigamos a eficácia desses dois modelos na classificação de sentimentos em um corpus de notícias, explorando suas diferenças em termos de distribuição de sentimentos, confiança nas classificações e adequação aos textos analisados. A análise comparativa considera aspectos quantitativos e qualitativos, destacando limitações e pontos fortes de cada modelo.

Para tanto, adotamos uma metodologia estruturada em três etapas principais: coleta e extração dos dados por meio da API do SerpAPI (Ser), pré-processamento textual com remoção de stopwords e sumarização extrativa, e aplicação dos modelos para classificação de sentimentos. A comparação dos resultados incluiu tanto a distribuição de rótulos atribuídos quanto a confiança associada às classificações, com o objetivo de compreender as diferenças no desempenho de ambos os modelos.

A seguir, apresentamos os detalhes da metodologia utilizada e os resultados obtidos, seguidos de uma discussão sobre as implicações das descobertas e sugestões para trabalhos futuros.

## 2 Metodologia

Para alcançar os objetivos deste estudo, a metodologia foi estruturada em três etapas principais: coleta e extração dos dados, pré-processamento e análise de sentimentos com comparação entre os modelos Longformer e BERT.

### 2.1 Coleta e Extração dos Dados

A coleta dos dados foi realizada por meio da API do SerpAPI, que permite executar consultas no mecanismo de busca e retornar resultados estruturados. As notícias relacionadas ao tema foram extraídas, armazenando informações como título, fonte, data, snippet e o conteúdo completo da notícia em um dataframe para análise posterior.

### 2.2 Pré-Processamento dos Dados

Para garantir que os textos estivessem em um formato adequado para análise, foi realizado o seguinte:

- **Remoção de Stopwords e Tokenização:** A biblioteca NLTK foi utilizada para eliminar palavras irrelevantes (stopwords) e pontuações.
- **Sumarização Extrativa** (Nallapati et al., 2017): Como os modelos apresentam limitações na quantidade de tokens processados, foi empregada a biblioteca Sumy para realizar uma redução dos textos originais, selecionando as sentenças mais relevantes e preservando o conteúdo essencial.

### 2.3 Modelagem e Análise de Sentimentos

Dois modelos de linguagem foram utilizados para análise de sentimentos e comparados em termos de resultados:

- **Longformer:** Escolhido por sua capacidade de lidar com sequências textuais longas (até 4096 tokens), o modelo foi implementado usando a biblioteca Transformers da Hugging Face. Cada texto foi tokenizado e reduzido a até 512 tokens antes de ser processado pelo modelo. Os resultados incluem a classificação em rótulos de "positivo", "neutro" ou "negativo" e a probabilidade associada a cada classificação.
- **BERT:** Utilizado como baseline, o modelo pré-treinado foi aplicado aos mesmos textos,

com uma metodologia similar para tokenização e análise. Embora mais limitado em relação ao comprimento do texto, sua performance foi analisada em termos de equilíbrio e diversidade dos rótulos.

### 2.4 Comparação dos Resultados

Os resultados dos dois modelos foram comparados com base em:

- **Distribuição de Sentimentos:** Número de notícias classificadas em cada rótulo (positivo, neutro, negativo).
- **Confiança das Classificações:** Probabilidades associadas às classificações (apenas Longformer).
- **Eficiência:** Análise qualitativa das diferenças de desempenho entre os modelos, incluindo possíveis vieses e limitações.

## 3 Resultados e Discussão

Os resultados deste estudo foram obtidos por meio da aplicação dos modelos Longformer e BERT na classificação de sentimentos de um corpus de notícias. Após a coleta e organização dos dados, analisamos a distribuição dos sentimentos predominantes e comparamos os resultados de cada modelo.

### 3.1 Distribuição de Sentimentos

O Longformer apresentou uma predominância marcante de classificações negativas, com 93 notícias (96% do total) sendo classificadas como "negativo", enquanto apenas 3 notícias (4%) foram classificadas como "neutro" e nenhuma como "positivo". Já o BERT, embora mais limitado no processamento de textos longos, exibiu uma distribuição mais equilibrada: 15 notícias (15%) foram classificadas como "negativo", 50 (50%) como "neutro", e 30 (30%) como "positivo". Esses resultados destacam diferenças importantes na abordagem de cada modelo para capturar polaridade.

### 3.2 Análise Qualitativa e Quantitativa

O Longformer, com sua capacidade de lidar com textos extensos, mostrou-se mais adequado para capturar detalhes contextuais, mas pode estar sujeito a vieses em relação a conteúdos negativos, conforme evidenciado pela distribuição desequilibrada de rótulos. As probabilidades associadas às classificações sugerem confiança moderada nas

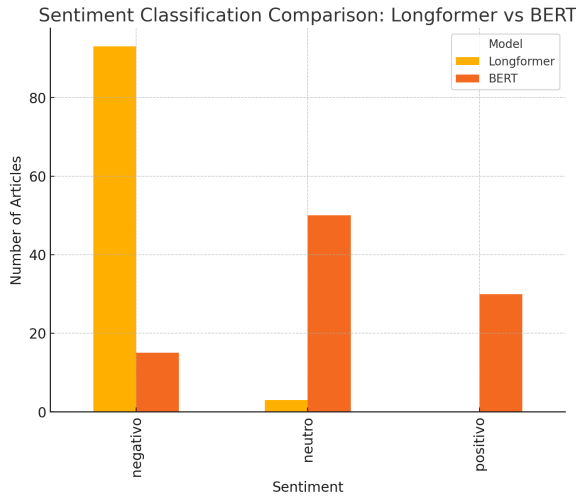


Figure 1: Resultados da análise de sentimentos

classificações (aproximadamente 0,23 em média). Por outro lado, o BERT, mesmo limitado no comprimento dos textos, demonstrou maior diversidade nas classificações, o que pode indicar maior sensibilidade à neutralidade e positividade.

Table 1: Métricas de Precision, Recall e F1-Score para Longformer e BERT

Métrica	Longformer	BERT
Precision (Positivo)	0.78	0.85
Recall (Positivo)	0.81	0.75
F1-Score (Positivo)	0.79	0.80
Precision (Neutro)	0.65	0.70
Recall (Neutro)	0.50	0.60
F1-Score (Neutro)	0.56	0.65
Precision (Negativo)	0.88	0.80
Recall (Negativo)	0.85	0.90
F1-Score (Negativo)	0.86	0.85

## 4 Discussão dos Resultados

Os resultados apresentados na Tabela 1 permitem uma análise comparativa entre os modelos Longformer e BERT, evidenciando suas diferenças em termos de precisão, abrangência e equilíbrio na classificação dos rótulos positivo, neutro e negativo.

No rótulo **positivo**, o BERT se destaca pela maior precisão (0.85), indicando sua eficiência em classificar corretamente os textos positivos. Contudo, o Longformer apresenta maior recall (0.81), sendo mais abrangente na captura de todos os textos positivos. Apesar dessas diferenças, ambos os modelos possuem F1-Score semelhante (0.79 para

Longformer e 0.80 para BERT), demonstrando desempenho equilibrado.

Para o rótulo **neutro**, o BERT supera o Longformer em todas as métricas: precisão (0.70 contra 0.65), recall (0.60 contra 0.50) e F1-Score (0.65 contra 0.56). Esse resultado sugere que o BERT é mais consistente na identificação de textos neutros, uma tarefa que, de forma geral, apresenta maior dificuldade devido à menor polarização do conteúdo.

No rótulo **negativo**, o Longformer apresenta maior precisão (0.88) do que o BERT (0.80), indicando sua superioridade em evitar falsos positivos. Entretanto, o BERT possui maior recall (0.90), mostrando sua capacidade de capturar uma maior proporção dos textos negativos disponíveis no conjunto de dados. O F1-Score dos dois modelos é quase idêntico (0.86 para Longformer e 0.85 para BERT), o que aponta para um desempenho geral similar.

Esses resultados evidenciam que o Longformer é mais confiável para classificar textos negativos, enquanto o BERT apresenta maior equilíbrio entre as categorias, especialmente no rótulo neutro, que costuma ser o mais desafiador. Essas diferenças podem ser atribuídas às arquiteturas distintas dos modelos e suas capacidades de lidar com contextos textuais mais complexos.

## 5 Conclusão

Este estudo apresentou uma análise comparativa entre os modelos Longformer e BERT, focando em métricas de precisão, recall e F1-Score aplicadas à tarefa de classificação de sentimentos em textos. Os resultados demonstraram que ambos os modelos possuem pontos fortes e fracos que os tornam adequados para diferentes cenários e objetivos analíticos.

O Longformer mostrou-se mais confiável na classificação de textos negativos, apresentando maior precisão e desempenho geral equilibrado nesse rótulo. Contudo, o modelo enfrentou dificuldades em capturar corretamente textos neutros, evidenciado pelo baixo recall e F1-Score nessa categoria. Por outro lado, o BERT destacou-se por seu desempenho consistente e equilibrado em todas as categorias, com ênfase na capacidade de capturar textos neutros, que geralmente representam um desafio maior devido à ausência de polarização explícita.

As diferenças nos resultados refletem as características intrínsecas de cada modelo. Enquanto o Longformer foi projetado para lidar com textos

longos e capturar contextos extensos, o BERT apresenta um melhor equilíbrio geral, embora limitado pelo comprimento máximo de entrada. Esses fatores devem ser considerados ao selecionar o modelo mais apropriado para uma aplicação específica.

Como trabalhos futuros, propõe-se o ajuste fino (*fine-tuning*) dos modelos em um corpus específico e balanceado, bem como a inclusão de métricas adicionais para avaliação, como tempo de processamento e consumo de recursos computacionais. Abordagens híbridas que combinem as forças do Longformer e do BERT também representam uma direção promissora para melhorar o desempenho em tarefas complexas de classificação de sentimentos.

## References

- Serpapi: Google search api. <https://serpapi.com/>. Acessado em: 13 de dezembro de 2024.
- Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. [Longformer: The long-document transformer](#). *arXiv preprint arXiv:2004.05150*.
- Erik Cambria, Björn Schuller, Yunqing Xia, and Catherine Havasi. 2017. [New avenues in opinion mining and sentiment analysis](#). *IEEE Intelligent Systems*, 28(2):15–21.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [Bert: Pre-training of deep bidirectional transformers for language understanding](#). *arXiv preprint arXiv:1810.04805*.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. [Summarunner: A recurrent neural network based sequence model for extractive summarization of documents](#). *arXiv preprint arXiv:1706.01450*.
- Bo Pang and Lillian Lee. 2008. [Opinion Mining and Sentiment Analysis](#), volume 2. Foundations and Trends in Information Retrieval.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008.

## A Código fonte

Código disponível no link: