

# Análise de sentimento em *feedbacks* discentes sobre a qualidade do ensino

Lucas Salvalaggio<sup>1</sup>, Fahad Kalil<sup>1</sup>

<sup>1</sup>Ciência da Computação – Atitus Educação  
Passo Fundo – RS – Brasil

lucassalvalaggio205@gmail.com, fahad.kalil@atitus.edu.br

**Abstract.** *This study compares the performance of different BERT-based models in sentiment analysis of student feedback written in Brazilian Portuguese. The main goal is to identify the most effective approach to automatically classify positive, negative, and neutral sentiments in academic evaluations. Four models were tested: Multilingual BERT (mBERT), BERTimbau Base, BERTimbau Large, and a hybrid BERTimbau + BiLSTM + Attention architecture. The experiments used a real dataset composed of manually labeled comments from Computer Science students. After supervised fine-tuning and cross-validation, monolingual models outperformed mBERT across all metrics. The hybrid model achieved the best overall performance, with an average F1-score of 0.8928, showing superior ability to capture semantic and emotional nuances in short, informal texts. The findings highlight the potential of Natural Language Processing as a tool to automate the interpretation of student feedbacks, supporting institutional decision-making and improving teaching quality.*

**Resumo.** *Este trabalho compara o desempenho de diferentes modelos baseados em BERT na análise de sentimentos de comentários discentes redigidos em português brasileiro. O objetivo é identificar a abordagem mais eficaz para classificar automaticamente sentimentos positivos, negativos e neutros em feedbacks acadêmicos. Foram avaliados quatro modelos: BERT Multilíngue (mBERT), BERTimbau Base, BERTimbau Large e uma arquitetura híbrida BERTimbau + BiLSTM + Attention. Utilizou-se um conjunto de dados real composto por comentários de alunos de um curso de Ciência da Computação, rotulados manualmente. Após o fine-tuning supervisionado e validação cruzada, observou-se que os modelos monolíngues superaram consistentemente o mBERT em todas as métricas. O modelo híbrido apresentou o melhor desempenho global, com F1-score médio de 0,8928, destacando-se pela capacidade de capturar nuances semânticas e emocionais em textos curtos e informais. Os resultados reforçam o potencial do Processamento de Linguagem Natural como ferramenta para automatizar a análise de feedbacks discentes, subsidiando decisões institucionais e a melhoria da qualidade do ensino.*

## 1. Introdução

Com o avanço das tecnologias digitais e o aumento da produção textual em plataformas educacionais, as instituições de ensino superior enfrentam novos desafios para compreender a percepção dos estudantes. Comentários dissertativos em avaliações institucionais

contêm informações valiosas sobre o ensino, o corpo docente e o ambiente acadêmico, mas sua análise manual é inviável em larga escala e sujeita a vieses interpretativos (Shaik *et al.*, 2023; Seemab *et al.*, 2024). Nesse cenário, o uso de técnicas de Processamento de Linguagem Natural (PLN) revela-se uma alternativa eficaz para extrair conhecimento estruturado de textos não estruturados. Entre essas técnicas, a Análise de Sentimentos destaca-se por identificar e classificar automaticamente a polaridade emocional (positiva, negativa ou neutra) presente nos comentários dos alunos (Zyout; Zyout, 2024; Shuqin; Raga, 2024). Essa abordagem permite compreender padrões de satisfação e insatisfação. Além disso oferece suporte à melhoria da qualidade do ensino (Wilbrod; Joshua; Avance International University, 2024).

No domínio educacional, a análise de sentimentos tem sido aplicada com sucesso para entender percepções estudantis sobre disciplinas, professores, metodologias e infraestrutura. Essa compreensão pode fornecer indicativos valiosos e relevantes para a melhoria da qualidade do ensino e para a tomada de decisões institucionais baseadas em evidências (Seemab *et al.*, 2024). No entanto, desafios persistem, como a detecção de ironia, ambiguidade de polaridade e diversidade linguística, especialmente em textos informais e curtos como os encontrados em plataformas educacionais.

Com o surgimento de modelos de aprendizado profundo, especialmente os baseados em arquiteturas *Transformer*, houve um salto significativo na precisão e na capacidade de generalização das aplicações de PLN. O *Bidirectional Encoder Representations from Transformers* (BERT) proposto por Devlin *et al.* (2019), consolidou-se como um dos principais modelos neste cenário, sendo amplamente utilizado em tarefas como classificação textual, extração de entidades e análise de sentimentos. No entanto, sua versão multilíngue (mBERT), embora compatível com o português, não foi treinada especificamente com dados do idioma, o que pode comprometer o desempenho do modelo em contextos linguísticos mais específicos.

Com o objetivo de superar essas limitações, foi desenvolvido o BERTimbau, um modelo BERT treinado exclusivamente com dados em português brasileiro (Souza; Nogueira; Lotufo, 2020). Em geral, modelos monolíngues, como o BERTimbau, apresentam melhor desempenho em tarefas específicas do idioma, especialmente em textos informais e curtos, características comuns aos *feedbacks* acadêmicos. Esse comportamento é amplamente observado nos estudos de Jazuli, Widowati e Kusumaningrum (2023) sobre o modelo monolíngue indonésio. Ainda assim, são escassos os trabalhos que comparam de forma sistemática o desempenho de diferentes variantes do BERT aplicadas ao contexto educacional em língua portuguesa.

Diante desse cenário, o presente trabalho compara o desempenho de diferentes modelos baseados em BERT na análise de sentimentos de comentários de feedback acadêmico redigidos em português brasileiro. São considerados o BERT Multilingue, o BERTimbau Base, o BERTimbau Large e uma arquitetura híbrida BERT + BiLSTM + Attention. A proposta inclui a preparação e rotulação manual de um conjunto textual real, a aplicação de técnicas de *fine-tuning* sobre os modelos selecionados e a comparação de seus desempenhos com base em métricas como acurácia, precisão, revocação, *F1-score* e matriz de confusão. Ao final, pretende-se identificar a abordagem mais eficaz para a tarefa, contribuindo para o aprimoramento da análise automatizada de sentimentos em ambientes educacionais e para a interpretação mais precisa das percepções estudantis em

contextos institucionais.

## 2. Referencial Teórico

Esta seção apresenta os fundamentos teóricos que embasam o desenvolvimento da pesquisa, os quais estão alinhados com o objetivo geral de identificar o modelo mais adequado para a análise de sentimentos em comentários de *feedback* acadêmico. São abordados conceitos centrais sobre Análise de Sentimentos, Modelos de Aprendizado Profundo aplicados à tarefa, o funcionamento e as aplicações do modelo BERT, a adaptação BERTimbau para o português brasileiro, bem como estudos sobre o uso dessas tecnologias em contextos educacionais.

O referencial teórico aqui apresentado busca fornecer o embasamento necessário para justificar a escolha dos métodos adotados e sustentar as comparações realizadas ao longo do trabalho.

### 2.1. Análise de Sentimentos

A análise de sentimentos, também conhecida como mineração de opiniões, é uma subárea do Processamento de Linguagem Natural (PLN) e busca identificar, extrair e classificar informações subjetivas expressas em textos. Essa técnica busca determinar a polaridade emocional dos enunciados (positiva, negativa ou neutra) podendo também, em abordagens mais avançadas, identificar a intensidade e o alvo específico do sentimento (Shaik *et al.*, 2023).

Tradicionalmente, os métodos de análise de sentimentos se dividem em três níveis principais. No nível de documento, o sentimento geral de um texto completo é avaliado. No nível de sentença, cada frase é considerada individualmente. Já no nível de aspecto (*Aspect-Based Sentiment Analysis* – ABSA), busca-se associar sentimentos a elementos específicos mencionados no texto, como: “professor”, “infraestrutura” ou “método de ensino” no contexto educacional. Esta última abordagem tem ganhado destaque por sua capacidade de oferecer uma análise mais granular e útil para a tomada de decisão (Jazuli; Widowati; Kusumaningrum, 2023). Por exemplo, um comentário como “O professor explica muito bem, mas a sala é barulhenta” tende a ser classificado como neutro em uma análise tradicional, pois mistura elogios e críticas. Já o ABSA separa o texto por aspectos, identificando que o aluno está satisfeito com o professor (positivo) e insatisfeito com a infraestrutura (negativo). Essa capacidade de apontar exatamente \*sobre o que\* o sentimento se refere torna a abordagem mais útil para decisões institucionais, já que revela problemas ou qualidades que seriam ocultados pela polaridade geral.

Os algoritmos aplicados à análise de sentimentos variam desde métodos clássicos de aprendizado de máquina, como *Naive Bayes*, *SVM* e regressão logística, até abordagens mais recentes baseadas em aprendizado profundo (Wilbrod; Joshua; Avance International University, 2024). Modelos como *Long Short-Term Memory* (LSTM) e redes neurais convolucionais (CNN) já apresentam resultados expressivos em tarefas de classificação de texto. Mais recentemente, modelos pré-treinados como o BERT e suas variantes passaram a dominar o campo. Esses modelos obtêm resultados superiores especialmente em contextos onde o significado das palavras depende fortemente do entorno textual, como nos *feedbacks* acadêmicos (Zyout; Zyout, 2024).

## 2.2. Modelos de Aprendizado Profundo na Análise de Sentimentos

A aplicação de modelos de aprendizado profundo (*Deep Learning* – DL) na Análise de Sentimentos tem revolucionado essa área do Processamento de Linguagem Natural (PLN). Esses modelos são capazes de aprender representações hierárquicas e contextualmente ricas dos textos, dispensando a extração manual de atributos linguísticos, e alcançando resultados de estado da arte em diferentes tarefas de classificação de sentimentos (Zhang; Wang; Liu, 2018). Os modelos de aprendizado profundo aplicados à análise de sentimentos podem ser agrupados, principalmente, em três categorias: modelos não recursivos, modelos recursivos e modelos híbridos (Rojas-Barahona, 2016).

### 2.2.1. Modelos Não Recursivos

Modelos não recursivos incluem arquiteturas como Redes Neurais Convolucionais e Redes Neurais Recorrentes (RNNs). As CNNs são eficazes na detecção de padrões locais e mostraram alto desempenho especialmente em tarefas com textos curtos, como *tweets* e avaliações curtas (Zhang; Wang; Liu, 2018). Já as RNNs, e suas variantes como a LSTM, são projetadas para capturar dependências sequenciais de longo prazo, sendo mais apropriadas para textos mais extensos ou contextualmente complexos (Rojas-Barahona, 2016). As CNNs têm como vantagem a eficiência computacional e a habilidade de capturar n-gramas relevantes independentemente de sua posição no texto. Kim (2014) propôs diversas variantes de CNNs, incluindo versões com *embeddings* treináveis e não treináveis, obtendo resultados competitivos em tarefas de classificação de sentimento.

### 2.2.2. Modelos Recursivos

Os modelos recursivos (*Recursive Neural Networks* – RecNNs) exploram a estrutura gramatical do texto, representada geralmente por árvores sintáticas. Nesses modelos, as representações das palavras são combinadas em nós da árvore para construir significados mais complexos em um processo bottom-up. O *Recursive Neural Tensor Network* (RNTN), por exemplo, permite uma modelagem mais sofisticada das interações entre palavras e frases, sendo útil na identificação de fenômenos linguísticos como negação ou intensificadores (Rojas-Barahona, 2016; Zhang; Wang; Liu, 2018). Modelos como o Tree-LSTM estendem o LSTM tradicional para estruturas em árvore, permitindo que a memória e o aprendizado contextual sejam organizados pela estrutura sintática da frase, o que resulta em desempenho superior em tarefas com granularidade fina, como a análise de sentimento em frases (Zhang; Wang; Liu, 2018).

### 2.2.3. Modelos Híbridos e Mecanismos de Atenção

Modelos híbridos combinam as vantagens das redes convolucionais e recorrentes com estruturas de árvore, frequentemente incorporando mecanismos de atenção. O uso de atenção permite ao modelo “focar” nas partes mais relevantes do texto para a tarefa em questão. Isso é particularmente útil na análise de sentimentos por aspecto, onde é necessário correlacionar palavras contextuais com o alvo da opinião. Autores destacam também a

importância de modelos como redes com memória (*Memory Networks*) e redes com atenção recorrente (*Recurrent Attention Networks*), que contribuem para uma melhor extração de sentimentos em contextos mais complexos e longos (Zhang; Wang; Liu, 2018; Shuqin; Raga, 2024).

### 2.3. Modelo BERT

Utilizando a arquitetura Transformer, BERT (*Bidirectional Encoder Representations from Transformers*) é um modelo de linguagem pré-treinado sobre grandes volumes de texto não rotulado. Seu diferencial está na capacidade de capturar representações contextuais profundas por meio de um treinamento bidirecional, o que o torna mais eficiente que modelos anteriores unidirecionais. Para isso, utiliza duas tarefas durante o pré-treinamento: a Modelagem de Linguagem Mascarada (Masked Language Modeling - MLM) e a Predição da Próxima Frase (Next Sentence Prediction - NSP). No MLM, determinadas palavras da frase são intencionalmente substituídas pelo token [MASK], e o modelo deve prever qual termo original ocupa aquela posição com base no contexto bidirecional. Esse mecanismo permite que o BERT aprenda relações semânticas profundas entre palavras e desenvolva representações contextuais mais robustas. Complementarmente, o NSP treina o modelo a identificar se duas sentenças aparecem de forma consecutiva em um texto, favorecendo a captura de relações de coerência e continuidade entre enunciados (Devlin *et al.*, 2019).

A combinação do aprendizado bidirecional com os objetivos de pré-treinamento MLM e NSP resulta em *embeddings* contextualizados altamente informativos, capazes de capturar nuances semânticas e relações de coerência textual. Essas representações tornam o BERT particularmente eficaz em tarefas posteriores, como classificação de sentimentos, inferência textual e extração de entidades nomeadas, superando abordagens anteriores ao lidar com dependências contextuais complexas (Souza; Nogueira; Lotufo, 2020).

Devido a essa robustez, o BERT rapidamente se consolidou como um dos modelos mais adotados no Processamento de Linguagem Natural (PLN), superando modelos clássicos em diversos *benchmarks*. Em especial, sua aplicação em análise de sentimentos tem-se mostrado altamente eficaz, já que consegue capturar nuances semânticas complexas e padrões subjetivos nos textos (Zyout; Zyout, 2024).

Com o objetivo de expandir essas capacidades, Shuqin e Raga (2024) propuseram a arquitetura BERT + BiLSTM + Attention, que combina as representações contextuais do BERT com a modelagem sequencial do BiLSTM e mecanismos de atenção. Essa combinação permite destacar palavras-chave com maior relevância emocional no texto. Os experimentos realizados pelos autores em avaliações de cursos online demonstraram que o modelo híbrido alcançou melhor desempenho preditivo em relação a abordagens isoladas.

### 2.4. BERTimbau: Um Modelo para o Português Brasileiro

A crescente popularização dos modelos baseados em BERT trouxe significativas melhorias para diversas tarefas de PLN. No entanto, esses avanços ocorreram majoritariamente em inglês e em alguns poucos idiomas com ampla disponibilidade de dados anotados. Diante desse cenário, surgiu a necessidade de desenvolver modelos específicos para outras línguas, como o português brasileiro. Foi nesse contexto que se desenvolveu o BERTimbau, o primeiro modelo BERT treinado exclusivamente com dados em português brasileiro (Souza; Nogueira; Lotufo, 2020).

Assim como o BERT original, BERTimbau utiliza a arquitetura Transformer, sendo pré-treinado com dois objetivos: MLM e NSP. O BERTimbau foi treinado com dados do corpus brWaC (*Brazilian Web as Corpus*), composto por mais de 2,68 bilhões de *tokens* provenientes de páginas web brasileiras, garantindo diversidade temática e linguística no pré-treinamento. A construção do vocabulário foi realizada com 2 milhões de sentenças retiradas da Wikipedia em português<sup>1</sup>, por meio do algoritmo Byte-Pair Encoding (BPE) com a biblioteca SentencePiece, resultando em um vocabulário de 30.000 subpalavras (Souza; Nogueira; Lotufo, 2023).

A avaliação do BERTimbau foi conduzida em três tarefas clássicas de PLN em português: (i) Similaridade textual de sentenças (STS); (ii) Reconhecimento de inferência textual (RTE) e (iii) Reconhecimento de entidades nomeadas (NER). Nos três casos, o BERTimbau superou o desempenho do modelo Multilingual BERT (mBERT), demonstrando que modelos monolíngues treinados com dados específicos apresentam melhor desempenho em tarefas no próprio idioma (Souza; Nogueira; Lotufo, 2023).

Além disso, os autores realizaram uma análise detalhada da tokenização, demonstrando que a segmentação em subpalavras tem impacto direto no desempenho do modelo nas tarefas finais. BERTimbau apresentou segmentações mais consistentes do que o mBERT, o que contribuiu para seu melhor desempenho geral (Souza; Nogueira; Lotufo, 2023).

O modelo BERTimbau está disponível em versões Base e Large, sendo a versão Base inicializada a partir do mBERT, enquanto a versão Large utilizou pesos do BERT em inglês. Ambas as versões alcançaram estado da arte em seus respectivos *benchmarks*, e estão publicamente acessíveis via plataforma GitHub<sup>2</sup>, contribuindo para o avanço da pesquisa em PLN para a língua portuguesa.

Portanto, o BERTimbau representa um marco no desenvolvimento de recursos linguísticos avançados para o português brasileiro, abrindo espaço para aplicações mais precisas em tarefas como análise de sentimentos, sumarização, classificação textual, e diversas outras.

## 2.5. Aplicações em Ambientes Educacionais

A crescente digitalização do ensino superior impulsionou o uso de tecnologias de Processamento de Linguagem Natural (PLN) em ambientes educacionais, especialmente para o tratamento de grandes volumes de *feedback* textual gerados por estudantes. A Análise de Sentimentos (*Sentiment Analysis* – SA) desponta como uma ferramenta fundamental para interpretar automaticamente avaliações dissertativas de alunos sobre disciplinas, docentes, infraestrutura e serviços acadêmicos.

Para Wilbrod, Joshua e Avance International University (2024), técnicas de SA aplicadas ao *feedback* estudantil permitem categorizar automaticamente comentários como positivos, negativos ou neutros, auxiliando instituições no diagnóstico de problemas pedagógicos e no aprimoramento do processo de ensino-aprendizagem. No estudo realizado, os autores desenvolveram um classificador baseado em algoritmos de aprendizado de máquina (como *Naive Bayes*) para analisar sentimentos expressos por alunos,

---

<sup>1</sup> Acesso em: <http://pt.wikipedia.com>

<sup>2</sup> Acesso em: <https://github.com/neuralmind-ai/portuguese-bert>

demonstrando que tal abordagem pode melhorar a qualidade educacional.

Em contextos multilíngues, a aplicação de modelos de linguagem pré-treinados como o BERT e suas variantes tem-se mostrado eficaz para lidar com nuances sintáticas e semânticas em comentários estudantis. Jazuli, Widowati e Kusumaningrum (2023) exploraram a técnica de Análise de Sentimentos Baseada em Aspectos (*Aspect-Based Sentiment Analysis* – ABSA), utilizando o modelo IndoBERT para processar 10 mil comentários de alunos de universidades da Indonésia. O modelo demonstrou precisão superior (*F1-score* de até 0,974) na identificação de sentimentos atrelados a aspectos específicos como ensino, infraestrutura, serviços e currículo, reforçando o papel do ABSA como ferramenta para avaliação educacional detalhada e segmentada.

Adicionalmente, Zyout e Zyout (2024) propuseram um modelo híbrido que combina redes neurais recorrentes com camadas de atenção e *embeddings* do BERT. Aplicado ao *feedback* estudantil, o modelo alcançou F-scores acima de 88% e se destacou pela habilidade de capturar emoções sutis e construções linguísticas complexas. Os autores argumentam que a integração de mecanismos de atenção e *embeddings* contextuais aprimora significativamente a detecção da polaridade emocional nos textos, mesmo quando estes são curtos ou subjetivos. A aplicação prática desses modelos é ampla. Instituições podem usar SA para:

- Monitorar, em tempo real, a satisfação estudantil com disciplinas e professores;
- Identificar padrões de insatisfação em aspectos específicos, como metodologia, infraestrutura ou apoio administrativo;
- Antecipar riscos de evasão escolar a partir da análise emocional de interações em Ambientes Virtuais de Aprendizagem (AVA);
- Personalizar intervenções pedagógicas com base no perfil emocional dos estudantes.

Contudo, a aplicação da análise de sentimentos em contextos educacionais ainda enfrenta desafios relevantes, como a detecção de ironia, sentimentos mistos e críticas implícitas, além da escassez de corpora anotados em português. Comentários vagos ou enviesados podem comprometer a acurácia dos modelos, que, apesar de avançados, como o BERT, ainda operam com baixa interpretabilidade. Além disso, aspectos emocionais nem sempre capturam sugestões qualitativas importantes, e mudanças no ambiente educacional, como o ensino híbrido, exigem constante adaptação dos modelos (Seemab *et al.*, 2024).

### 3. Trabalhos Relacionados

Neste item serão apresentados os principais trabalhos que possuem uma relação com o assunto definido neste estudo, que trata da análise de sentimentos em comentários de *feedback* acadêmico utilizando modelos BERT e suas variantes.

- **Sentiment Analysis of Student Feedback in Higher Education Using Natural Language Processing (Seemab *et al.*, 2024)**

O artigo aborda a importância da automatização da análise de *feedback* estudantil como ferramenta para melhorar a qualidade do ensino superior. Destaca as limitações dos métodos tradicionais, como o viés subjetivo e a morosidade na avaliação manual.

O estudo teve como objetivo explorar técnicas de PLN, como SVM, LSTM e BERT, para classificar sentimentos em categorias positivas, negativas e neutras. Foi desenvolvido um *framework* que utilizou esses modelos para identificar padrões temáticos e classificar sentimentos em *feedbacks* educacionais. A pesquisa também apresentou uma metodologia robusta de pré-processamento e avaliação de modelos.

Os autores concluíram que as técnicas de PLN, especialmente com modelos BERT, são eficazes para gerar indicadores relevantes que promovem melhorias contínuas na educação. Difere-se do presente trabalho por não focar especificamente na língua portuguesa nem em dados de instituições brasileiras.

- **Sentiment Analysis of Student Feedback Using Attention-Based RNN and Transformer Embedding (Zyout; Zyout, 2024)**

Neste artigo, os autores propuseram um modelo de análise de sentimentos utilizando RNNs com camadas de atenção, combinadas com *embeddings* como GloVe e BERT.

O estudo utilizou um conjunto híbrido de dados, incluindo comentários reais e sintéticos, para balancear sentimentos positivos e negativos, com foco em melhorar o desempenho dos modelos na classificação de sentimentos educacionais.

Foi observado que a arquitetura Bi-LSTM combinada com BERT e atenção obteve os melhores resultados, com *F1-scores* superiores a 88%.

O trabalho contribui para o uso de técnicas avançadas de PLN na educação, mas difere do presente estudo por empregar dados parcialmente gerados por IA e não ser aplicado ao português brasileiro.

- **Sentiment Analysis of Student Feedback. An Implementation of a Natural Language Processing (Wilbrod; Joshua; Avance International University, 2024)**

Este artigo discute o uso do classificador *Naive Bayes* para análise de sentimentos em *feedbacks* estudantis, com foco em simplificar e automatizar a categorização de sentimentos em positivo, negativo ou neutro.

O objetivo principal foi construir um modelo de classificação eficaz com base em uma base de dados rotulada, destacando o impacto do pré-processamento de dados na acurácia do modelo.

A metodologia incluiu limpeza textual, divisão dos dados em treino e teste e aplicação do modelo de *Naive Bayes* com avaliação por métricas como acurácia, precisão e *recall*.

O estudo demonstrou que é possível atingir resultados consistentes mesmo com modelos simples, sendo eficaz para identificar percepções estudantis. Porém, seu enfoque em técnicas clássicas e ausência de BERT marcam uma diferença importante em relação ao trabalho proposto.

- **A Deep Learning Model for Student Sentiment Analysis on Course Reviews (Shuqin; Raga, 2024)**

Este trabalho propõe um modelo de deep learning denominado BERT + BiLSTM + Attention (BBA), voltado à análise de sentimentos em avaliações de cursos online (MOOCs).

O estudo visou aprimorar a extração de sentimentos com base em dados reais, utilizando a combinação de BERT para codificação contextual, BiLSTM para de-



pendências sequenciais e mecanismos de atenção para priorização de termos relevantes.

Os autores realizaram experimentos comparativos com diversos modelos de aprendizado profundo, comprovando a superioridade do BBA na precisão e interpretação de sentimentos.

A principal contribuição foi a proposta de uma arquitetura híbrida avançada, que melhora a representatividade semântica do texto. O presente trabalho se diferencia por focar na língua portuguesa e por considerar também variações do BERT específicas para o português.

- **Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback (Jazuli; Widowati; Kusumaningrum, 2023)**

Este estudo introduz um modelo de análise de sentimentos baseado em aspectos (ABSA) utilizando o IndoBERT, uma versão do BERT ajustada para a língua indonésia.

O objetivo foi identificar sentimentos relacionados a aspectos específicos, como qualidade do ensino, currículo, infraestrutura e serviços, com alto grau de precisão.

O experimento utilizou 10 mil comentários estudantis e atingiu métricas robustas (*F1-score* acima de 0,97), demonstrando a eficácia da abordagem ABSA com BERT otimizado.

Embora compartilhe a proposta metodológica de aplicar BERT adaptado à língua local e de realizar análise aspectual, o estudo é centrado no idioma indonésio. O presente trabalho, por sua vez, busca adaptar essas ideias para o português brasileiro utilizando o BERTimbau, avaliando seu desempenho em dados reais de comentários de discentes de um curso superior em Ciência da Computação no Brasil.

## 4. Materiais e Métodos

Esta seção apresenta os recursos, ferramentas e procedimentos empregados na condução do presente estudo, cuja proposta consiste na comparação de modelos baseados em BERT aplicados à análise de sentimentos em comentários de *feedback* acadêmico. O delineamento metodológico contempla a definição do tipo de pesquisa, as ferramentas e bibliotecas utilizadas, a base de dados, o processo de preparação textual, a seleção de modelos, a configuração dos experimentos e os critérios de avaliação.

### 4.1. Ferramentas e Ambiente de Desenvolvimento

Os experimentos foram conduzidos em ambiente virtual, utilizando a linguagem de programação Python, amplamente adotada em tarefas de Processamento de Linguagem Natural (PLN) por sua simplicidade e diversidade de bibliotecas especializadas. O treinamento foi realizado no ambiente Google Colab<sup>3</sup>, utilizando GPU NVIDIA T4 (16 GB), essencial para reduzir o tempo de treinamento dos modelos.

As principais bibliotecas utilizadas incluem:

- Transformers: para carregamento de modelos BERT, tokenização e *fine-tuning*;

---

<sup>3</sup> Acesso em: <https://colab.research.google.com>

- Scikit-learn: para cálculo de métricas e aplicação de validação cruzada;
- Pandas e NumPy: para manipulação da base de dados e operações vetoriais;
- Matplotlib e Seaborn: para visualização dos resultados.

## 4.2. Fonte de Dados

A base textual utilizada é composta por comentários anônimos de estudantes de um curso superior de Ciência da Computação em uma instituição privada do Brasil, extraídos de formulários institucionais de avaliação de disciplinas. Os dados foram previamente anonimizados e, em seguida, rotulados manualmente pelo pesquisador em três classes: positivo (*Positive*), negativo (*Negative*) e neutro (*Neutral*). O conteúdo textual apresenta-se predominantemente curto, informal e redigido em português brasileiro, o que impõe desafios à tarefa de análise de sentimentos, especialmente na interpretação semântica de enunciados breves e ambíguos.

**Tabela 1. Amostra de comentários e classificação de sentimento presentes na base de estudo**

Comentário (text)	Classe (label_text)	Rótulo (label)
A internet não funcionou muito bem, não consegui usar o computador na aula.	Negative	0
Ótima aula e didática, porém é difícil de visualizar o que tem escrito nos slides.	Neutral	1
Professor muito didático, abrangente. Dinâmica da aula muito boa.	Positive	2
Ótima aula!	Positive	2

A Tabela 2 apresenta a distribuição das classes no conjunto de dados utilizado para o treinamento e avaliação dos modelos. Observa-se que a categoria positiva representa a maior proporção das instâncias, com 569 exemplos, seguida pelas classes neutra (353 instâncias) e negativa (307 instâncias).

**Tabela 2. Distribuição das classes na base de dados.**

Classe	Rótulo (label)	Quantidade
Positivos	2	569
Neutros	1	353
Negativos	0	307

## 4.3. Pré-processamento dos Dados

Antes dos dados serem submetidos aos modelos foi necessário realizar a etapa de pré-processamento, composta pelas seguintes tarefas:

- Conversão para letras minúsculas;
- Remoção de acentuação, pontuação, URLs, emojis e caracteres especiais;

- Exclusão de entradas vazias ou duplicadas;
- Tokenização com os tokenizers específicos de cada modelo (por exemplo, BERTimbauTokenizer e BertTokenizer);
- Segmentação dos textos com limite máximo de 128 *tokens* por entrada, conforme o padrão da arquitetura BERT.

#### 4.4. Modelos de treinamento

Foram comparados quatro modelos distintos que compartilham a arquitetura Transformer, mas diferem em profundidade, escopo linguístico e complexidade. Enquanto BERTimbau (Base e Large) permitem avaliar o impacto de modelos monolíngues treinados em português, o mBERT fornece uma perspectiva multilíngue. Já a arquitetura híbrida BERT + BiLSTM + Attention incorpora camadas sequenciais e de atenção, buscando maior sensibilidade contextual. Essa comparação visa identificar a abordagem mais eficaz para análise de sentimentos em comentários acadêmicos curtos e informais. Os modelos foram ajustados via **fine-tuning** supervisionado, com base nos comentários previamente rotulados.

1. BERTimbau Base: modelo monolíngue treinado exclusivamente em português brasileiro (Souza; Nogueira; Lotufo, 2020);
2. BERTimbau Large: versão mais profunda do BERTimbau, com maior capacidade de representação;
3. BERT Multilingue (mBERT): modelo treinado em 104 idiomas, incluindo o português (Devlin *et al.*, 2019);
4. Arquitetura híbrida BERTimbau + BiLSTM + Attention: estrutura que combina as representações do BERTimbau com uma camada bidirecional de LSTM e mecanismo de atenção, inspirada no modelo proposto por Shuqin e Raga (2024).

#### 4.5. Configuração de Treinamento

A configuração experimental seguiu os seguintes parâmetros:

- **Épocas:** 5. Número de vezes que o modelo percorre todo o conjunto de treinamento.
- **Batch size:** 16. Quantidade de amostras processadas simultaneamente em cada atualização dos pesos.
- **Taxa de aprendizado:**  $2e-5$ . Controla o tamanho das atualizações dos parâmetros do modelo.
- **Otimizador:** AdamW. Algoritmo de otimização utilizado para ajustar os pesos durante o treinamento.
- **Função de perda:** *CrossEntropyLoss*. Mede o erro em tarefas de classificação multiclasse.
- **Dropout:** 0.4. Taxa de desativação aleatória de neurônios para reduzir sobreajuste.
- **Tamanho máximo de sequência:** 128 *tokens*. Limita o número de *tokens* considerados em cada entrada.

- Balanceamento de classes: aplicado por meio da função `compute_class_weight` da biblioteca `scikit-learn`, atribuindo pesos inversamente proporcionais à frequência de cada classe. Os pesos foram integrados à função de perda, reduzindo o viés em favor das classes majoritárias e melhorando o desempenho nas categorias minoritárias;

Os dados foram divididos na proporção de 80% para conjunto de treinamento e 20% para validação. Além disso, foi aplicada validação cruzada estratificada com cinco dobras (*5-folds*) para reduzir o viés de avaliação e aumentar a robustez estatística dos resultados. Tais configurações seguiram diretrizes adotadas em múltiplos estudos da literatura. A revisão sistemática de Fatwanto, Zamakhsyari e Ndungi (2024) destaca que *batch sizes* 16 e 32, além da taxa de aprendizado  $2e-5$ , estão entre os valores mais utilizados em experimentos com BERT. Souza, Nogueira e Lotufo (2023) aplicaram *5-folds* com empilhamento (*stacking*) de modelos BERT, atingindo *F1-score* de 88,3 na tarefa de inferência textual. Além disso, o estudo de Habbat *et al.* (2023) empregou validação cruzada com 10 dobras (*10-folds*) para melhorar a robustez da análise de sentimentos em conjuntos de dados desbalanceados, obtendo desempenho elevado com *ensemble* de BERT e BiLSTM.

#### 4.6. Avaliação de Desempenho

A avaliação dos modelos foi realizada por meio das seguintes métricas:

- Acurácia: porcentagem de classificações corretas;
- Precisão (*Precision*): proporção de verdadeiros positivos entre os positivos previstos;
- Revocação (*Recall*): proporção de verdadeiros positivos entre os exemplos da classe real;
- *F1-score*: média harmônica entre precisão e revocação, calculada na versão macro e ponderada.
- Matriz de Confusão: representação tabular que mostra as classificações corretas e incorretas por classe, permitindo identificar os principais erros do modelo.
- Curva ROC (*Receiver Operating Characteristic*): representação gráfica da taxa de verdadeiros positivos (*TPR*) em função da taxa de falsos positivos (*FPR*), possibilitando avaliar a capacidade discriminativa do modelo.

As métricas foram extraídas com auxílio da biblioteca `scikit-learn` e os resultados são apresentados em tabelas e gráficos de comparação, com suporte das bibliotecas `Matplotlib` e `Seaborn`.

### 5. Resultados e Discussão

Esta seção apresenta e analisa os resultados obtidos com os modelos de deep learning e a arquitetura híbrida descritos na Seção 4, visando avaliar o desempenho na classificação multiclasse de sentimentos em feedbacks discentes. Primeiramente, são expostos os resultados quantitativos com base nas métricas de acurácia, precisão, revocação e *F1-score*. Em seguida, realiza-se uma análise qualitativa das matrizes de confusão e curvas ROC para identificar padrões de acertos e erros. Por fim, apresenta-se uma discussão comparativa com trabalhos relacionados e as limitações observadas no estudo.

Os resultados obtidos após o treinamento e validação dos modelos são apresentados na Tabela 3. As métricas correspondem à média dos *5-folds* de validação cruzada, calculadas para acurácia, precisão, revocação e *F1-score*. Todas as avaliações consideram o balanceamento de classes por meio da ponderação automática dos pesos, assegurando maior equidade entre as categorias positivas, neutras e negativas.

**Tabela 3. Resultados médios de desempenho dos modelos avaliados**

Modelo	Acurácia	Precisão	Revocação	<i>F1-score</i>
mBERT	0,8226	0,8259	0,8226	0,8228
BERTimbau Base	0,8771	0,8788	0,8771	0,8768
BERTimbau Large	0,8771	0,8792	0,8771	0,8771
<b>BERTimbau + BiLSTM + Attention</b>	<b>0,8926</b>	<b>0,8949</b>	<b>0,8926</b>	<b>0,8928</b>

Os resultados evidenciam um desempenho superior dos modelos monolíngues em comparação ao mBERT, confirmando observações anteriores da literatura (Souza; Nogueira; Lotufo, 2020, 2023). O modelo BERTimbau Base já apresenta melhora significativa em relação ao mBERT, com ganho médio de aproximadamente 5,4% na acurácia, o que demonstra a relevância de um pré-treinamento específico em português brasileiro para a análise de sentimentos.

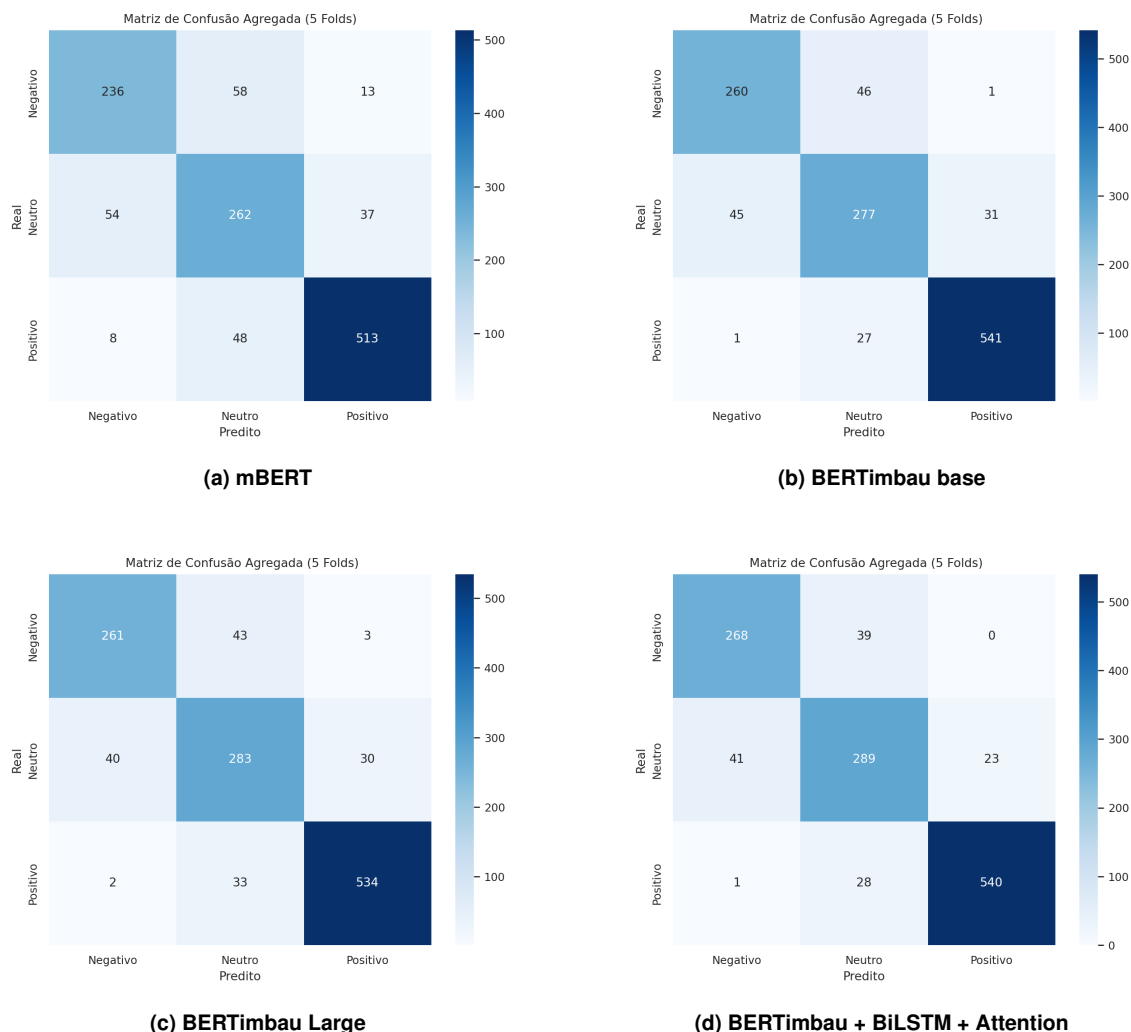
A diferença entre BERTimbau Base e BERTimbau Large foi marginal, indicando saturação de desempenho em conjuntos moderados. Esse resultado indica que o aumento da profundidade do modelo nem sempre se traduz em ganhos expressivos de desempenho em conjuntos de dados de tamanho moderado, possivelmente devido à saturação da capacidade de generalização ou à limitação do volume de exemplos rotulados disponíveis para o *fine-tuning*.

O melhor desempenho geral foi obtido pela arquitetura híbrida BERTimbau + BiLSTM + Attention, que atingiu acurácia média de 0,8926 e *F1-score* médio de 0,8928. Essa combinação entre codificações contextuais do BERT e camadas sequenciais com mecanismos de atenção mostrou-se eficaz na identificação de padrões semânticos e nuances afetivas nos textos curtos e informais característicos de feedbacks estudantis. Os resultados corroboram o estudo de Shuqin e Raga (2024), que reportaram desempenho superior de arquiteturas híbridas em cenários educacionais semelhantes, com melhorias notáveis na detecção de sentimentos ambíguos ou implícitos.

### 5.1. Análise das Matrizes de Confusão e Curva ROC

A Figura 1 apresenta as matrizes de confusão agregadas para *5-folds* da validação cruzada, enquanto a Figura 2 exibe a curva ROC multiclasse comparando o desempenho dos quatro modelos avaliados. As análises complementam as métricas quantitativas apresentadas na Tabela 3, permitindo uma interpretação mais detalhada sobre os padrões de acertos e erros entre as classes de sentimento.

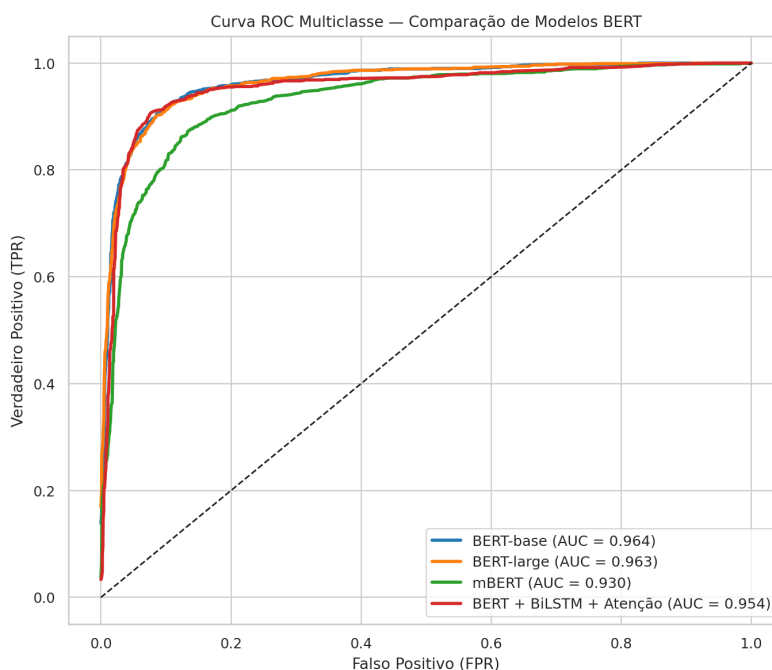
**Figura 1. Matriz de confusão para 5-folds**



As matrizes de confusão evidenciam que todos os modelos alcançam desempenho consistente na identificação da classe positiva, com altos índices de verdadeiros positivos e baixo número de falsos negativos. Esse comportamento sugere que as expressões linguísticas positivas tendem a apresentar padrões lexicais e sintáticos mais regulares, o que facilita a detecção automática por parte dos modelos baseados em representações contextuais.

A classe neutra, por outro lado, concentrou a maioria dos erros de classificação. Observa-se confusão frequente entre sentenças neutras e negativas, fenômeno que pode ser explicado pela ambiguidade semântica de comentários que expressam avaliações descritivas, sugestivas ou moderadamente críticas, sem o uso explícito de marcadores afetivos. Esse padrão também pôde ser observado nos resultados experimentais, indicando que as fronteiras entre neutralidade e negatividade nem sempre são claramente distinguíveis pelo modelo. O modelo híbrido BERTimbau + BiLSTM + Attention apresentou leve melhora nesse aspecto, reduzindo o número de falsos negativos e demonstrando maior sensibilidade à neutralidade contextual.

A curva ROC multiclasse (Figura 2) confirma o bom desempenho global dos modelos, com valores de *AUC* superiores a 0,93 para todos os casos. O mBERT apresentou o menor *AUC* (0,930), novamente indicando limitações decorrentes de seu treinamento multilíngue. As versões BERTimbau Base e Large obtiveram resultados muito próximos (*AUC* = 0,964 e 0,963, respectivamente), sugerindo que o aumento da complexidade do modelo não implicou ganhos expressivos. O modelo BERTimbau + BiLSTM + Attention atingiu *AUC* = 0,954, resultado ligeiramente inferior ao Base, porém com melhor equilíbrio entre as classes nas matrizes de confusão, o que reforça sua robustez em cenários reais de dados heterogêneos.



**Figura 2. Curva ROC multiclasse para os modelos avaliados**

Em síntese, a análise combinada das métricas, matrizes de confusão e curvas ROC evidencia que os modelos monolíngues, especialmente o BERTimbau, apresentaram desempenho superior ao mBERT, demonstrando a importância de utilizar conjuntos de dados linguísticos alinhados ao português brasileiro e ao contexto educacional. Entre as abordagens testadas, a arquitetura híbrida BERTimbau + BiLSTM + Attention destacou-se pela maior consistência geral, conciliando alta acurácia, menor viés entre classes e estabilidade nas métricas de desempenho. Esses resultados reforçam o potencial das abordagens baseadas em aprendizado profundo para a interpretação automatizada de percepções discentes, contribuindo para análises mais precisas sobre a qualidade do ensino em instituições de educação superior.

Ao confrontar os resultados deste estudo com os de Zyout e Zyout (2024) e Shuqin e Raga (2024), observa-se que todos alcançaram desempenhos expressivos com arquiteturas baseadas em BERT, confirmando sua eficácia na análise de sentimentos educacionais. O modelo BERT + BiLSTM + Attention de Zyout e Zyout (2024) obteve *F1-score* de 0,89 em uma classificação binária (positiva e negativa) de comentários estudantis em inglês, utilizando um conjunto de 3.820 amostras, incluindo dados sintéticos gerados por prompt

engineering para balanceamento das classes. Já o de Shuqin e Raga (2024), também com BERT + BiLSTM + Attention, atingiu *F1-score* de 0,966 em avaliações de cursos MOOCs, igualmente em uma tarefa binária. Em contraste, o presente trabalho aplicou uma classificação multiclasse (negativo, neutro e positivo) em português brasileiro, idioma de estrutura morfosintática mais complexa e com ampla variação regional, o que naturalmente eleva a dificuldade da tarefa e tende a reduzir ligeiramente os índices globais. Ainda assim, o modelo híbrido manteve resultados consistentes e equilibrados entre as classes (*F1-score* de 0,8928), demonstrando sensibilidade a nuances intermediárias. Embora o desempenho absoluto seja um pouco inferior, o modelo proposto oferece análise mais detalhada e interpretativamente rica, reforçando a versatilidade e a capacidade de generalização da arquitetura BERT + BiLSTM + Attention em diferentes idiomas e contextos educacionais.

Algumas limitações foram observadas ao longo deste estudo. A principal refere-se ao tamanho reduzido da base de comentários discentes, o que pode restringir a capacidade de generalização dos modelos e a representatividade das diferentes polaridades de sentimento. Apesar da aplicação de validação cruzada e balanceamento de classes, o número limitado de exemplos, sobretudo nas categorias neutra e negativa, ainda representa um desafio para uma classificação totalmente equilibrada.

Outra limitação está relacionada à rotulação manual dos textos, que envolve certo grau de subjetividade, especialmente em comentários ambíguos ou com linguagem indireta. Mesmo com critérios definidos, a interpretação humana pode variar, afetando a consistência das etiquetas e, consequentemente, o desempenho dos modelos supervisionados.

Cabe ressaltar que os comentários estudantis tendem a conter expressões coloquiais, abreviações e erros ortográficos. Esses fenômenos afetam diretamente a etapa de tokenização, especialmente em modelos pré-treinados que não lidam bem com ortografia irregular. Além disso, os modelos avaliados não capturam informações pragmáticas ou contextuais externas (como disciplina, semestre ou professor), que poderiam refinar a interpretação dos sentimentos.

Adicionalmente, a metodologia empregada identifica apenas a polaridade global do comentário, sem indicar a que aspecto o sentimento se refere. Assim, elogios e críticas presentes em um mesmo texto são reduzidos a uma única classificação, o que limita a interpretação sobre quais elementos do processo educacional motivam satisfação ou insatisfação. Por isso, este modelo deve ser entendido como parte de um conjunto maior de ferramentas analíticas, não sendo recomendado seu uso isolado para decisões institucionais.

## **6. Considerações Finais**

O presente trabalho teve como propósito comparar o desempenho de diferentes modelos baseados em BERT na tarefa de análise de sentimentos de comentários discentes em português brasileiro. Foram avaliadas quatro abordagens: mBERT, BERTimbau Base, BERTimbau Large e a arquitetura híbrida BERTimbau + BiLSTM + Attention, aplicadas a um conjunto de dados composto por textos curtos e informais provenientes de avaliações institucionais. Os resultados indicaram que os modelos monolíngues superaram o mBERT em todas as métricas, confirmando a importância do pré-treinamento em dados específicos



da língua portuguesa. O BERTimbau Base e o BERTimbau Large apresentaram desempenhos semelhantes, enquanto a arquitetura BERTimbau + BiLSTM + Attention obteve o melhor resultado global, com *F1-score* médio de 0,8928 e melhor equilíbrio entre as classes.

Em termos institucionais, os achados deste estudo possuem implicações diretas na gestão da qualidade do ensino. A implementação dos modelos propostos pode viabilizar sistemas automatizados de monitoramento da satisfação discente, auxiliando na detecção precoce de problemas pedagógicos, avaliação contínua de cursos e identificação de boas práticas docentes. Dessa forma, a instituição pode fundamentar decisões estratégicas em dados empíricos e em análises linguísticas precisas, promovendo uma cultura de avaliação baseada em evidências e de melhoria contínua do processo educativo.

Esses achados evidenciam que a combinação entre representações contextuais do BERT e mecanismos de atenção aprimora a identificação de padrões semânticos sutis, especialmente em textos informais e curtos. Em termos aplicados, os resultados reforçam o potencial do Processamento de Linguagem Natural (PLN) como ferramenta de apoio para a análise automatizada de feedbacks estudantis, contribuindo para diagnósticos mais rápidos e consistentes sobre a qualidade do ensino e a percepção discente.

Os resultados obtidos neste estudo revelam oportunidades de aprofundamento e diversificação de abordagens para o avanço da análise de sentimentos em contextos educacionais. Um primeiro caminho consiste na expansão do escopo analítico para além da polaridade de sentimento. Futuras pesquisas podem investigar a detecção de emoções específicas, como satisfação, frustração, engajamento ou desmotivação, utilizando modelos de classificação *multilabel* ou técnicas de *emotion recognition*. Essa ampliação permitiria compreender de forma mais granular o impacto emocional das experiências de ensino. Outra linha de investigação envolve o monitoramento temporal de sentimentos, analisando a evolução das percepções discentes ao longo dos semestres. Essa perspectiva longitudinal possibilitaria identificar tendências, efeitos de mudanças pedagógicas e variações sazonais, oferecendo subsídios estratégicos para a tomada de decisão institucional.

Sugere-se, ainda, explorar análises comparativas entre cursos ou áreas do conhecimento, aplicando o modelo a bases de dados de diferentes contextos acadêmicos. Tal abordagem permitiria avaliar se os padrões linguísticos e emocionais variam conforme a natureza da disciplina, o formato das aulas ou o perfil dos estudantes.

## Referências

DEVLIN, J. *et al.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: BURSTEIN, J.; DORAN, C.; SOLORIO, T. (ed.). **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**. Minneapolis, Minnesota: Association for Computational Linguistics, jun. 2019. p. 4171–4186. DOI: 10.18653/v1/N19-1423. Disponível em: <https://aclanthology.org/N19-1423/>.

FATWANTO, A.; ZAMAKHSYARI, F.; NDUNGI, R. A Systematic Literature Review of BERT-based Models for Natural Language Processing Tasks. **JURNAL INFOTEL**, v. 16, p. 713–728, dez. 2024.

HABBAT, N. *et al.* Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning. **Engineering Applications of Artificial Intelligence**, v. 126, p. 106999, 2023. ISSN 0952-1976. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0952197623011831>.

JAZULI, A.; WIDOWATI, W.; KUSUMANINGRUM, R. Aspect-based sentiment analysis on student reviews using the Indo-Bert base model. **E3S Web of Conferences**, v. 448, nov. 2023. Disponível em: <https://doi.org/10.1051/e3sconf/202344802004>. Acesso em: 25 abr. 2025.

KIM, Y. Convolutional Neural Networks for Sentence Classification. *In*: MOSCHITTI, A.; PANG, B.; DAELEMANS, W. (ed.). **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Doha, Qatar: Association for Computational Linguistics, out. 2014. p. 1746–1751. DOI: 10.3115/v1/D14-1181. Disponível em: <https://aclanthology.org/D14-1181/>.

ROJAS-BARAHONA, L. M. Deep learning for sentiment analysis. **Language and Linguistics Compass**, v. 10, n. 12, p. 701–719, 2016. Disponível em: <https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/lnc3.12228>.

SEEMAB, S. *et al.* Sentiment Analysis of Student Feedback in Higher Education Using Natural Language Processing (NLP). **Contemporary Journal Of Social Science Review**, v. 2, 4 2024. Disponível em: <https://contemporaryjournal.com/index.php/14/article/view/256>.

SHAIK, T. *et al.* Sentiment analysis and opinion mining on educational data: A survey. **Natural Language Processing Journal**, v. 2, p. 100003, 2023. ISSN 2949-7191. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2949719122000036>.

SHUQIN, H.; RAGA, R. C. A Deep Learning Model for Student Sentiment Analysis on Course Reviews. **IEEE Access**, v. 12, p. 136747–136758, 2024. Disponível em: <https://doi.org/10.1109/ACCESS.2024.3463793>.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERT models for Brazilian Portuguese: Pretraining, evaluation and tokenization analysis. **Applied Soft Computing**, v. 149, p. 110901, 2023. ISSN 1568-4946. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1568494623009195>.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. *In*: INTELLIGENT Systems. Cham: Springer International Publishing, out. 2020. p. 403–417. Disponível em: [https://doi.org/10.1007/978-3-030-61377-8\\_28](https://doi.org/10.1007/978-3-030-61377-8_28). Acesso em: 20 abr. 2025.

WILBROD, M.; JOSHUA, T.; AVANCE INTERNATIONAL UNIVERSITY, A. Sentiment Analysis of Student Feedback. An Implementation of a Natural Language

Processing (NLP) Algorithm. **Metropolitan Journal Of Science And Technology**, v. 3, p. 78–86, 10 out. 2024.

ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. **WIREs Data Mining and Knowledge Discovery**, v. 8, n. 4, e1253, 2018. Disponível em: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1253>.

ZYOUT, I.; ZYOUT, M. Sentiment analysis of student feedback using attention-based RNN and transformer embedding. **IAES International Journal of Artificial Intelligence (IJ-AI)**, v. 13, p. 2173–2184, jun. 2024. Disponível em: <https://doi.org/10.11591/ijai.v13.i2.pp2173-2184>.