

# Avaliação Automática de Redações no Modelo do ENEM por meio do *fine-tuning* do BERTimbau

## Introdução

O surgimento de modelos de linguagem como o BERT (Devlin et al., 2018), baseados em *transformers*, superou técnicas tradicionais de classificação de texto ao utilizar mecanismos que consideram o significado e a relação entre palavras de forma eficiente. A capacidade de ajuste fino de tais modelos pode, ainda, destacá-los na tarefa de avaliação automática de redações, já que permite a eles um treinamento especializado para a assimilação de padrões. Tal abordagem pode ser aplicada a exames como o ENEM, que possui uma cadeia logística complexa e lenta para a realização das correções. Este trabalho visa utilizar conjuntos de dados anotados e abertos para realizar o treinamento de cinco sistemas especialistas em avaliações de redações no modelo do ENEM, cada um com foco em uma competência distinta do exame.

## Objetivos

- ▶ Desenvolver modelos de correção automática de redações baseados no BERT.
- ▶ Comparar desempenho de técnicas de treinamento e arquitetura de sistemas especialistas.
- ▶ Avaliar precisão da avaliação automática em relação à correção humana.
- ▶ Discutir vantagens ou desvantagens da abordagem automatizada e documentar resultados obtidos.

## Fundamentação Teórica

O BERT, um modelo de linguagem baseado em transformadores, supera limitações recorrentes ao utilizar o mecanismo de atenção para modelar dependências bidirecionais. Ele atua principalmente na codificação, usando o *encoder* dos transformadores.

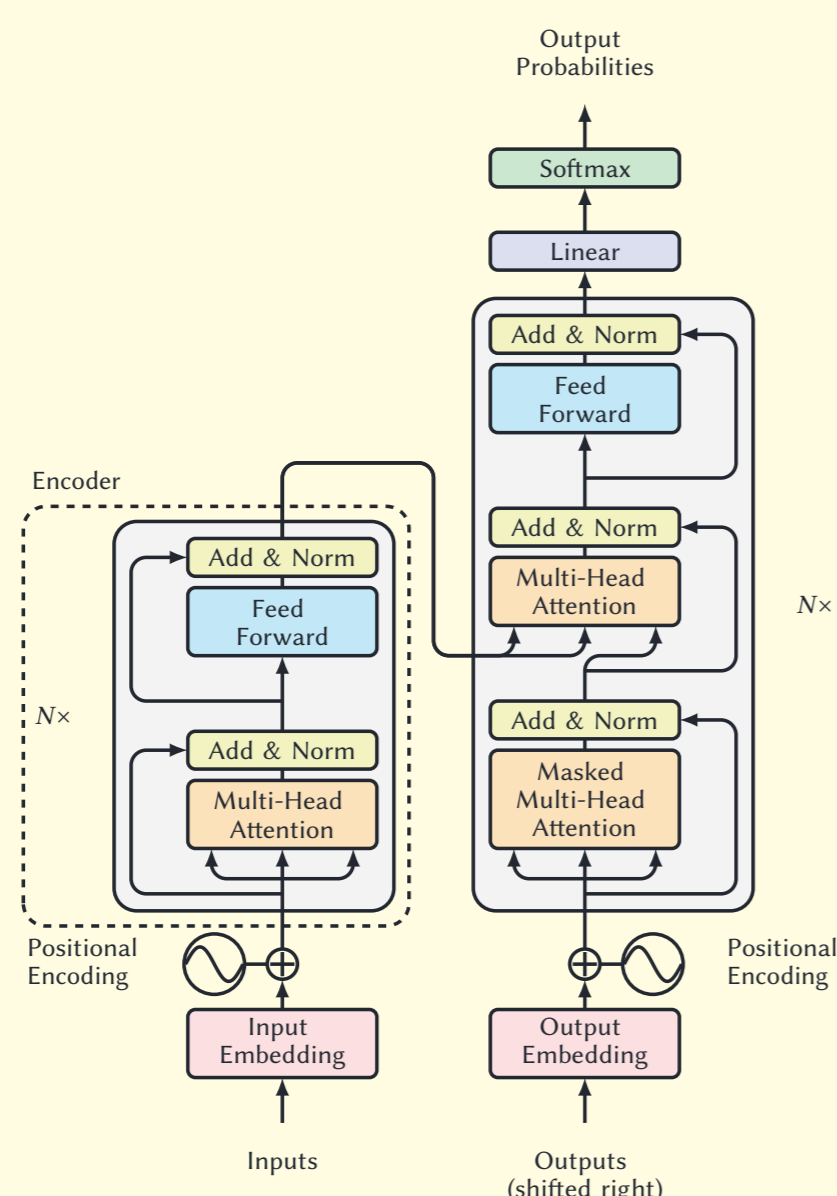


Figura 1: Arquitetura de um transformador (Vaswani et al., 2017).

## Metodologia

Utilizando a base de dados Essay-BR (Marinho et al., 2022), em suas versões simples e estendida, foram treinados modelos especializados em atribuir notas às cinco competências do ENEM (Brasil, 2023). A arquitetura de cada sistema foi elaborada com base na técnica de *fine-tuning* do BERT, usando o modelo pré-treinado BERTimbau para capturar nuances linguísticas em textos em língua portuguesa. Após o treinamento, os modelos foram rigorosamente avaliados com métricas de desempenho de aprendizado de máquina, como o *Quadratic Weighted Kappa* (QWK) (Cohen, 1968), e métricas do ENEM.

## Arquitetura dos Modelos

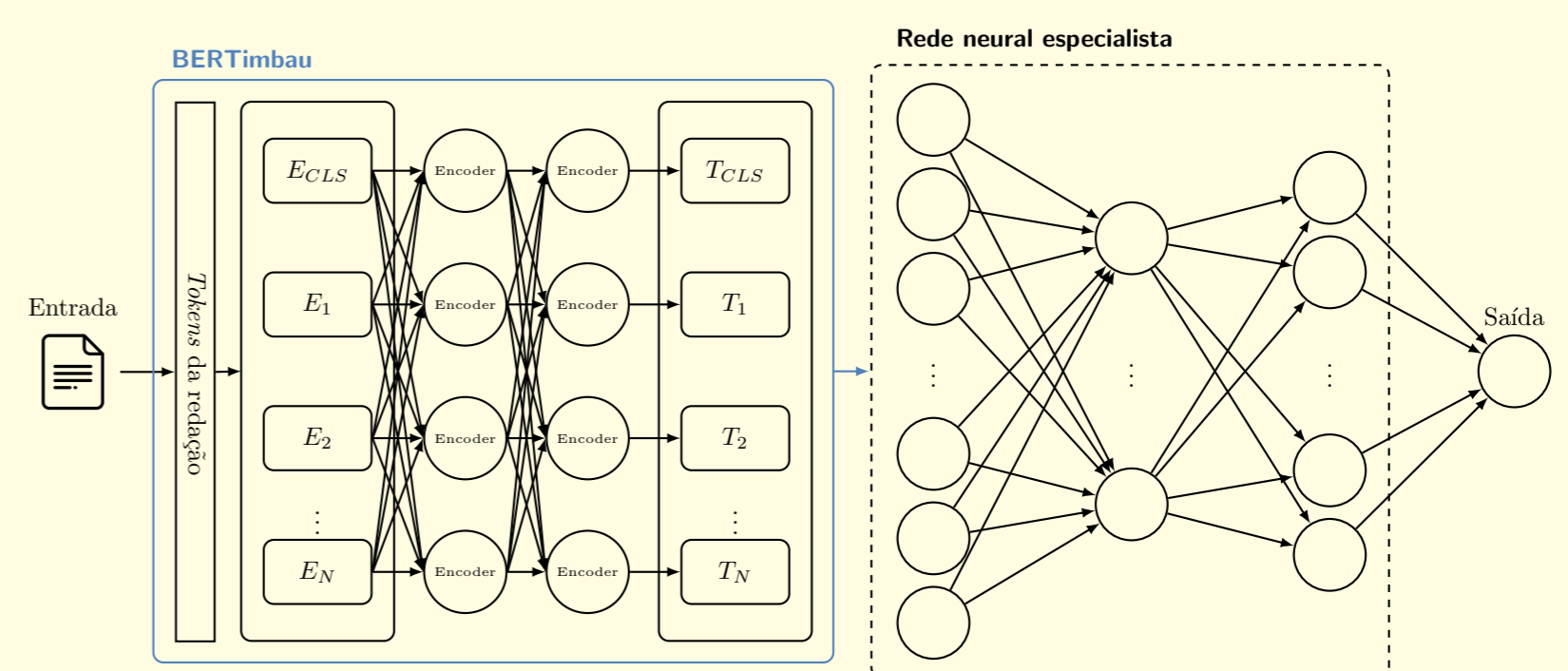


Figura 2: Arquitetura dos modelos de correção automática.

## Resultados

Sendo PCE a proporção de correspondência exata e MSE o erro quadrático médio, os melhores resultados dos sistemas avaliadores são mostrados pela tabela abaixo.

Modelos	QWK	PCE	MSE
<b>Competência I</b>	0,595	68,3%	0,358
<b>Competência II</b>	0,562	57,2%	0,640
<b>Competência III</b>	0,539	59,3%	0,554
<b>Competência IV</b>	0,621	50,4%	0,701
<b>Competência V</b>	0,548	49,4%	0,932

## Bibliografia

- ▶ Brasil (2023). *A Redação do Enem 2023: cartilha do participante*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
- ▶ Cohen, Jacob (1968). «Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.». *Psychological Bulletin* 70, pp. 213–220. URL: <https://api.semanticscholar.org/CorpusID:29694079>.
- ▶ Devlin, Jacob et al. (2018). «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding». *CoRR* abs/1810.04805. arXiv: 1810.04805. URL: <http://arxiv.org/abs/1810.04805>.
- ▶ Marinho, Jeziel et al. (2022). «Essay-BR: a Brazilian Corpus to Automatic Essay Scoring Task». *Journal of Information and Data Management* 13.1, pp. 65–76. DOI: 10.5753/jidm.2022.2340. URL: <https://sol.sbc.org.br/journals/index.php/jidm/article/view/2340>.
- ▶ Vaswani, Ashish et al. (2017). «Attention Is All You Need». *CoRR* abs/1706.03762. arXiv: 1706.03762. URL: <http://arxiv.org/abs/1706.03762>.