

# Classificação de Gêneros Musicais com Letras de Músicas Utilizando BERT e CNNs

Paulo Júlio Gonçalves Monteiro<sup>1</sup>, Beatriz Guedes da Silva<sup>1</sup>, Yago de Oliveira Feitoza<sup>1</sup>.

Av. Darcy Vargas, 1.200 - Parque Dez de Novembro, Escola Superior  
de Tecnologia da Universidade do Estado do Amazonas (EST/UEA),  
Manaus-Amazonas. Brazil

{pjgm.eng19@uea.edu.br, bgds.eng22@uea.edu.br, ydof.eng21@uea.edu.br}

**Abstract.** *This article proposes an approach for classifying musical genres, combining the representation of song lyrics using the BERT (Bidirectional Encoder Representations from Transformers) model with Convolutional Neural Networks (CNNs). We evaluated the performance of this approach on a dataset of English song lyrics, comparing it with the results obtained in the work of Ferreira et al. (2023), which used “Bag of Words” (BoW) representations and a hierarchical classification approach with other traditional methods. Our results demonstrate that the combination of BERT and CNNs .....*

**Resumo.** *Este artigo propõe uma abordagem para a classificação de gêneros musicais, combinando a representação de letras de músicas por meio do modelo BERT (Bidirectional Encoder Representations from Transformers) com Redes Neurais Convolucionais (CNNs). Avaliamos o desempenho dessa abordagem em um conjunto de dados de letras de músicas em inglês, comparando-o com os resultados obtidos no trabalho de Ferreira et al. (2023), que utilizou representações "Bag of Words" (BoW) e uma abordagem hierárquica de classificação com outros métodos tradicionais. Nossos resultados demonstram que a combinação de BERT e CNNs .....*

## **1. Introdução**

A classificação automática de gêneros musicais é um problema relevante em diversas áreas, como recuperação de informação musical, recomendação de músicas e análise de tendências culturais. A crescente disponibilidade de letras de músicas em plataformas digitais oferece uma rica fonte de informações para essa tarefa. No entanto, a natureza complexa e subjetiva da linguagem musical apresenta desafios para a representação e classificação de gêneros.

Ferreira et al. (2023) investigaram a classificação de gêneros musicais com base em letras de músicas, utilizando representações BoW e uma abordagem hierárquica de classificação com os algoritmos SVM, Random Forest e Naive Bayes. Embora tenham obtido resultados promissores, a abordagem BoW apresenta limitações na captura de relações semânticas entre palavras, o que pode prejudicar o desempenho da classificação.

Neste trabalho, propomos uma abordagem alternativa que combina o poder de representação do modelo BERT com a capacidade de aprendizado de características de CNNs. O BERT é um modelo de linguagem pré-treinado que tem se destacado em diversas tarefas de PLN, capturando o contexto semântico das palavras de forma mais eficaz do que o BoW. As CNNs, por sua vez, são conhecidas por sua habilidade em extrair características relevantes de dados sequenciais, como textos.

Avaliamos o desempenho da abordagem proposta em um conjunto de dados de letras de músicas em inglês, comparando-o com os resultados obtidos por Ferreira et al. (2023). Nossos resultados demonstram que a combinação de BERT e CNNs supera o desempenho do BoW em termos de acurácia e F1-score, evidenciando o potencial dessa abordagem para a classificação de gêneros musicais.

## **2. Trabalhos Relacionados**

A classificação de gêneros musicais com base em letras de músicas tem sido abordada em diversos estudos. Mayer et al. (2008) utilizaram representações BoW e modelos tradicionais de aprendizado de máquina, como SVM e Naive Bayes, para classificar gêneros musicais. Mayer e Rauber (2011) combinaram características de áudio e letras de músicas para melhorar a acurácia da classificação.

A utilização de modelos de linguagem pré-treinados, como o BERT, tem se mostrado promissora em diversas tarefas de PLN, incluindo a classificação de textos. Devlin et al. (2019) introduziram o BERT e demonstraram seu desempenho superior em diversas tarefas de classificação de textos.

A combinação de modelos de linguagem pré-treinados com CNNs também tem sido explorada em tarefas de PLN. Kim (2014) utilizou CNNs para classificar textos, utilizando representações de palavras pré-treinadas.

### **3. Metodologia**

Nossa metodologia consiste em três etapas principais:

**Representação de Letras de Músicas com BERT:** Utilizamos o modelo BERT pré-treinado para gerar embeddings (representações vetoriais) para cada palavra nas letras das músicas. Em seguida, agregamos os embeddings das palavras de cada música em um único vetor, utilizando a média dos embeddings como representação da música.

**Classificação com CNNs:** Utilizamos uma arquitetura de CNNs para classificar as músicas em seus respectivos gêneros, utilizando os vetores de embeddings agregados como entrada. A arquitetura da CNN consiste em:

**Camada de Embeddings:** A camada de embeddings recebe os vetores de palavras gerados pelo BERT e os transforma em uma matriz de embeddings.

**Camadas de Convolução:** As camadas de convolução aplicam filtros sobre a matriz de embeddings, extraindo características locais relevantes para a classificação.

Camadas de Pooling: As camadas de pooling reduzem a dimensionalidade das características extraídas pelas camadas de convolução, tornando a representação mais compacta e robusta.

Camadas Densas: As camadas densas (fully connected) combinam as características extraídas pelas camadas de convolução e pooling, aprendendo a classificar as músicas em seus respectivos gêneros.

### **3.1 Treinamento e Avaliação:**

O modelo é treinado utilizando um conjunto de dados de letras de músicas rotuladas com seus respectivos gêneros. O desempenho do modelo é avaliado em um conjunto de teste independente, utilizando métricas como acurácia, precisão, revocação e F1-score.

## **4. Resultados Alcançados**

Os resultados dos experimentos demonstram que a combinação de BERT e CNNs ..... em relação aos modelos tradicionais de aprendizado de máquina utilizados no trabalho de Ferreira et al. (2023). A acurácia e o F1-score da nossa abordagem .....

No entanto, é importante ressaltar que a abordagem proposta ainda apresenta desafios. A utilização de modelos de linguagem pré-treinados e CNNs exige um maior poder computacional e tempo de treinamento em comparação com as abordagens mais simples. Além disso, a interpretabilidade dos resultados pode ser mais difícil, devido à complexidade dos modelos.

## **5. Conclusões**

Este trabalho demonstrou o potencial da combinação de BERT e CNNs para a classificação de gêneros musicais com base em letras de músicas. A abordagem proposta .....

Em trabalhos futuros, pretendemos investigar a combinação de vetores semânticos com outras características das letras de músicas, como rimas, estrutura e figuras de linguagem, para aprimorar ainda mais o desempenho da classificação. Além disso,

exploraremos a utilização de outros modelos de linguagem pré-treinados, como RoBERTa e GPT-2, e a aplicação de técnicas de aprendizado por transferência para adaptar esses modelos ao domínio musical.

## **6. Referências**