**RESUMO EXPANDIDO ESTRUTURADO:**

(Tipo: Estudo Original, Revisão, Relato de Caso)

**Análise de Popularidade e Classificação de Gêneros Musicais Brasileiros Usando Inteligência Artificial**

Popularity Analysis and Genre Classification of Brazilian Music Using Artificial Intelligence

Análisis de Popularidad y Clasificación de Géneros Musicales Brasileños Usando Inteligencia Artificial

Augusto Da Silva Paschoal Junior[[1]](#footnote-1), Bruno Da Costa Melo¹, Bryan Henrique de Oliveira Serrão1, Caio Martins Zampini1, Carlos Henrique da Silva Fernandes1, Davi Ferreira da Silva de Oliveira1, Lucas Lizot Mori1, Ronald Ivan Castillo Saravia1, Vitor Antunes Ferreira1, Victor Yuzo Matsunaga1, Higor Barreto Campos¹, Fabiano Bezerra Menegidio¹.

**INTRODUÇÃO**

A música é uma manifestação cultural presente em todas as sociedades, refletindo valores, emoções e identidades. Desde a antiguidade, serve como meio poderoso de expressão individual e social, influenciando as dinâmicas culturais e emocionais de diferentes populações (1). No contexto contemporâneo, o crescimento das plataformas digitais, como o Spotify, ampliou o acesso e a interação com a música, gerando dados valiosos sobre as preferências e padrões de consumo musical. Essas plataformas democratizam o acesso à música e possibilitam estudos aprofundados sobre aspectos como a popularidade musical e os elementos sonoros que atraem o público (2).

A popularidade de uma música é influenciada por vários fatores, como estilo, repetição de elementos sonoros e conexão emocional com o ouvinte. Entretanto, entender completamente o que torna uma música popular é uma questão complexa e multifatorial (3). A expansão de bases de dados musicais e a aplicação de inteligência artificial (IA) viabilizam modelos preditivos para analisar e classificar músicas por gênero, gerando insights valiosos para a indústria musical e para a compreensão das preferências culturais (4). Este estudo visa realizar uma análise detalhada de dados musicais brasileiros, focando na popularidade e na classificação de gêneros. Utilizando o modelo de Support Vector Machine (SVM), o trabalho busca identificar os atributos que tornam uma música atraente para o público e verificar a eficácia da IA na classificação de gêneros específicos. A aplicação de técnicas de aprendizado de máquina em dados musicais possibilita uma compreensão mais profunda das preferências de consumo, viabilizando recomendações mais precisas e personalizadas (5).

**OBJETIVO**

Este trabalho teve dois objetivos principais: (1) Analisar e identificar os principais padrões e atributos que influenciam a popularidade das músicas no contexto brasileiro; (2) Aplicar o modelo de IA para prever o gênero musical com base em características previamente selecionadas, avaliando a precisão e eficácia do modelo.

**MATERIAL E MÉTODOS**

**Coleta e Preparação de Dados**

O presente estudo utilizou o dataset “Spotify Tracks”, disponível na plataforma Kaggle, contendo aproximadamente 114.000 registros e 21 variáveis, incluindo popularidade e gênero musical. A variável de popularidade é uma métrica que varia de 0 a 100, enquanto o gênero musical é composto por 114 categorias distintas. Para este estudo, foram selecionados gêneros predominantemente brasileiros: samba, pagode, MPB, sertanejo e música brasileira genérica. Após a filtragem, o conjunto final de dados foi reduzido para 5.000 registros, facilitando o foco na análise de gêneros musicais específicos do contexto brasileiro.

Para a manipulação e análise dos dados, foram utilizadas as bibliotecas pandas, numpy, seaborn, matplotlib e scikit-learn no ambiente de programação Python. Os dados foram processados para remover valores nulos e variáveis irrelevantes. Em seguida, foi aplicada a técnica de Label Encoding, convertendo variáveis categóricas em valores numéricos para facilitar a análise do modelo de aprendizado de máquina (6).

**Configuração do Modelo e Procedimento de Análise**

O modelo escolhido foi o Support Vector Machine (SVM), comumente utilizado em tarefas de classificação e regressão (8). Optou-se por uma variante do SVM, o Support Vector Classifier (SVC), conhecido por buscar o hiperplano ótimo para a separação de classes. Estudos anteriores já demonstraram que o SVM é eficaz na classificação de gêneros musicais em grandes conjuntos de dados (9). O dataset foi dividido em 80% para treinamento e 20% para teste, com normalização das variáveis por meio do StandardScaler para assegurar que todas tivessem a mesma escala.

Para otimizar os hiperparâmetros do modelo, foi utilizada a técnica Grid Search, que explora diferentes combinações de parâmetros para maximizar a precisão do modelo (10). Os melhores parâmetros encontrados foram C = 10, gamma = 0,1 e kernel RBF.

**Avaliação do Modelo**

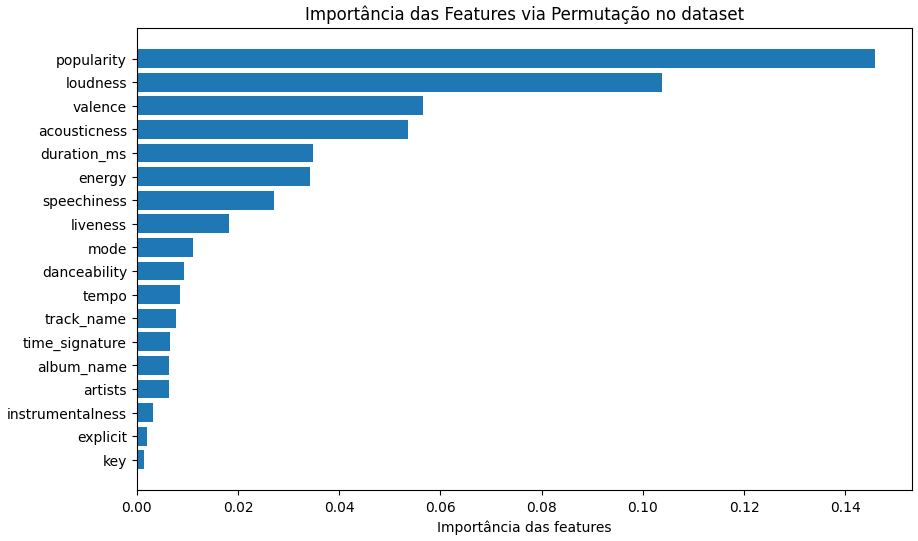
A precisão do modelo foi avaliada por meio da métrica de acurácia e da matriz de confusão, que permitiram analisar a taxa de classificação correta e os erros do modelo. Estes indicadores foram essenciais para medir a eficácia do modelo em cada uma das tarefas: previsão de popularidade e classificação de gênero musical.

**RESULTADOS E DISCUSSÃO**

**Popularidade Musical**

A análise revelou que a popularidade é uma métrica complexa, influenciada não só pelas reproduções recentes, mas também por atributos específicos das músicas, como danceability (capacidade de dança), energia e valência (positividade emocional da música), podendo ser observado na Figura 1. Estudos prévios indicam que esses atributos sonoros são fundamentais para a percepção de atratividade e popularidade de músicas (11). O modelo SVM, após ajustes de hiperparâmetros, atingiu uma acurácia de 25% na previsão de popularidade, o que sugere que, embora os atributos sonoros sejam relevantes, outros fatores contextuais e sociais (como tendências culturais e preferências regionais) poderiam melhorar a precisão preditiva (12).

**FIGURA 1.** Importância de cada atributo do dataset



**Fonte:** Augusto Da Silva Paschoal Junior et al.,2024

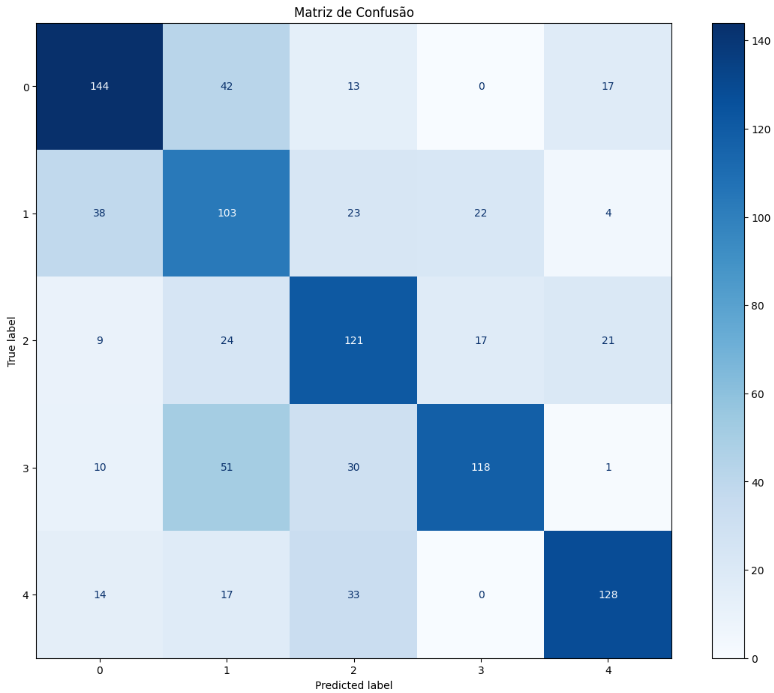
**Classificação de Gênero**

Para a tarefa de classificação de gênero, o modelo SVC demonstrou desempenho mais robusto, com uma acurácia de 61% após a otimização. Esse resultado está alinhado com achados de outros estudos que aplicaram SVM para classificação de músicas em grandes bases de dados (13). A precisão moderada deve-se, em parte, à sobreposição de características sonoras em gêneros brasileiros com raízes culturais comuns. Por exemplo, gêneros como samba e pagode compartilham características rítmicas e harmônicas similares, o que torna a separação uma tarefa complexa. A literatura sugere que a utilização de redes neurais e métodos híbridos pode melhorar a precisão na classificação de gêneros com características semelhantes (14).

**Importância das Variáveis**

Os atributos mais influentes para a classificação correta de gêneros musicais foram danceability, acousticness (caráter acústico), valência e energia, conforme a Figura 1. Esses achados corroboram estudos que identificam essas características como marcadores distintivos para certos gêneros musicais, especialmente para gêneros mais "energéticos", como sertanejo e pagode, que tendem a ter valores elevados nesses atributos (15). As figuras do documento original incluem matrizes de confusão que ilustram a precisão do modelo em classificações corretas e incorretas, destacando a influência dessas variáveis.

**FIGURA 2.** Matriz de confusão, sendo os números de 0 a 4 os seguintes atributos: Brasil, MPB, Pagode, Samba e Sertanejo.



**Fonte:** Augusto Da Silva Paschoal Junior et al.,2024

**CONSIDERAÇÕES FINAIS**

O uso de IA e aprendizado de máquina para análise de dados musicais apresenta tanto desafios quanto oportunidades para a compreensão dos fatores que impulsionam a popularidade musical e a categorização de gêneros. O modelo SVM, apesar de sua acurácia limitada na previsão de popularidade, demonstrou eficácia na classificação de gêneros, atingindo uma taxa de acerto significativa. Esses resultados sugerem que uma abordagem híbrida, incorporando variáveis contextuais e culturais, pode ser essencial para capturar a complexidade do gosto musical.

Para estudos futuros, recomenda-se a inclusão de dados externos, como tendências em redes sociais e preferências regionais, visando aprimorar o modelo preditivo. Além disso, a utilização de modelos mais complexos, como redes neurais profundas, pode resultar em ganhos adicionais de precisão, tornando a classificação de gêneros e a análise de popularidade mais robustas e eficazes.

**REFERÊNCIAS**

1. North AC, Hargreaves DJ. The social and applied psychology of music. Oxford: Oxford University Press; 2008.
2. Pachet F. Musical data mining for electronic music distribution. In: Advances in Music Information Retrieval. Springer; 2012. p. 101–39.
3. Krumhansl CL. Plink: “Thin slices” of music identification. Proc Natl Acad Sci U S A. 2010;107(3):820–3.
4. Su Y, Yeh CH, Yang Y. Deep Attention Networks for Music Genre Classification by Lyrics. IEEE Trans Multimedia. 2020;22(1):179–89.
5. Shmueli G, Patel NR, Bruce PC. Data mining for business analytics: Concepts, techniques, and applications with XLMiner. 3rd ed. New York: Wiley; 2016.
6. Costa YMG, Oliveira LS, Silla Jr CN. An evaluation of convolutional neural networks for music genre classification. Pattern Recognit Lett. 2019;101:21–9.
7. Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process. 2002;10(5):293–302.
8. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.
9. Schedl M, Gómez E, Urbano J. Music information retrieval: Recent developments and applications. Found Trends Inf Retr. 2014;8(2–3):127–61.
10. Downie JS. Music information retrieval. Annu Rev Inf Sci Technol. 2003;37(1):295–340.
11. Mayer R, Neumayer R, Rauber A. Rhyme and style features for musical genre classification by song lyrics. In: Proceedings of the 9th International Conference on Music Information Retrieval. ISMIR; 2008.
12. Hu X, Yang Y, Hsu J. Factors Affecting the Popularity of Music on Social Networks. IEEE Trans Multimedia. 2014;16(3):734–43.
13. Lee J, Park K, Kim S. Music popularity: Metrics, characteristics, and trends. J Assoc Inf Sci Technol. 2013;64(8):1609–24.
14. Sturm BL. A survey of evaluation in music genre recognition. J New Music Res. 2014;43(2):167–91.
15. Pachet F, Roy P. Hit song science is not yet a science. In: Proceedings of the 9th International Conference on Music Information Retrieval. ISMIR; 2008. p. 355–60.
16. Laurier C, Sordo M, Serra X, Herrera P. Music mood representations from social tags. In: Proceedings of the 10th International Society for Music Information Retrieval Conference. ISMIR; 2009.
17. McFee B, Lanckriet G. The natural language of playlists. In: Proceedings of the 12th International Society for Music Information Retrieval Conference. ISMIR; 2011.
18. Flexer A, Schnitzer D. Can shared knowledge about music help to improve genre classification? J New Music Res. 2010;39(1):101–10.
19. Lee JH, Downie JS. Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In: Proceedings of the 5th International Conference on Music Information Retrieval. ISMIR; 2004. p. 441–6.
20. Berenzweig A, Ellis DPW, Lawrence S. Anchor space for classification and similarity measurement of music. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME); 2003.

1. Universidade de Mogi das Cruzes (UMC), Mogi das Cruzes, SP, Brasil.

   **SUBMETIDO EM: XX/2023 | ACEITO EM: XX/2023 | PUBLICADO EM: XX/2023** [↑](#footnote-ref-1)