Inteligência Artificial no Reconhecimento de Subgêneros Musicais Brasileiros

Beatriz Oliveira Lustosa[1], **Ayslan Trevizan Possebom**[1]

[1]Instituto Federal do Paraná (IFPR) – Campus Paranavaí/PR
Rua José Felipe Tequinha, 1400, Jardim das Nações – 87703-536 – Paranavaí – Brasil.

2023018845@estudantes.ifpr.edu.br, ayslan.possebom@ifpr.edu.br

Abstract. *This study investigates the use of machine learning techniques to classify Brazilian music subgenres—Sertanejo Raiz, Pagode, and Pop Rock Nacional—based on song lyrics. The lyrics were automatically collected via web scraping, followed by text preprocessing and feature extraction using TF-IDF. Several models were trained and evaluated with cross-validation, including Logistic Regression, Naive Bayes, Random Forest, SVM, and a Multilayer Perceptron (MLP) Neural Network. The results showed that the Multilayer Perceptron achieved the best performance with an average accuracy of 82.88%, while the SVM and Naive Bayes models also performed well on specific subgenres. The study confirms the feasibility of using machine learning for the classification of Brazilian music, paving the way for applications in music categorization and analysis systems.*

Resumo. *Este trabalho investiga o uso de técnicas de aprendizado de máquina para classificar subgêneros musicais brasileiros — Sertanejo Raiz, Pagode e Pop Rock Nacional — a partir das letras de músicas. Foi realizada a coleta automática das letras por meio de Raspagem de Dados, seguida de pré-processamento textual e extração de características utilizando TF-IDF. Diversos modelos foram treinados e avaliados com validação cruzada, incluindo Regressão Logística, Naive Bayes, Random Forest, SVM e Rede Neural Multicamada. Os resultados indicaram que a Rede Neural Multicamada obteve o melhor desempenho, com acurácia média de 82,88%, enquanto o SVM e o Naive Bayes também apresentaram bons resultados em subgêneros específicos. O estudo confirma a viabilidade do uso de aprendizado de máquina para a classificação de músicas brasileiras, abrindo caminho para aplicações em sistemas de categorização e análise musical.*

1. Introdução

A música desempenha um papel fundamental na cultura brasileira, refletindo a diversidade de estilos, histórias e identidades presentes no país [Trotta 2005]. Com o avanço da tecnologia, da inteligência artificial e a ampla disponibilidade de letras musicais em plataformas digitais, pode-se pensar em ferramentas automáticas capazes de organizar e classificar [Pereira 2009] as letras das músicas como adequadas para diferentes gêneros musicais de forma eficiente.

Nesse contexto, o uso de técnicas de aprendizado de máquina e processamento de linguagem natural (PLN) tem se destacado como uma alternativa promissora para a

análise e categorização automática de textos [Viana 2024, Silva 2024]. Embora existam estudos voltados à classificação de gêneros musicais internacionais [Borges et al. 2010, Malheiro et al. 2004], há uma carência de pesquisas que explorem especificamente os subgêneros brasileiros, como Sertanejo Raiz, Pagode e Pop Rock Nacional, utilizando exclusivamente as letras das músicas como fonte de dados.

Diante disso, este trabalho tem como objetivo investigar a aplicação de diferentes algoritmos de aprendizado de máquina na classificação automática desses subgêneros, a partir de um conjunto de letras obtido via Raspagem de Dados (*Web Scraping*) e processado com técnicas de PLN. Além de contribuir para a área acadêmica, este estudo pode servir de base para o desenvolvimento de sistemas de recomendação musical e análise de tendências culturais, reforçando a importância da inteligência artificial na área musical.

A estrutura deste trabalho está organizada da seguinte forma: a Seção 2 apresenta a fundamentação teórica, a Seção 3 descreve a metodologia empregada, a Seção 4 discute os resultados obtidos e a Seção 5 traz as conclusões e perspectivas futuras.

2. Fundamentação Teórica

A música é uma das formas mais antigas e universais de expressão humana. Ela está presente em todas as culturas conhecidas e cumpre diversas funções sociais, emocionais, artísticas e comunicativas. De acordo com [Ghirardi 2013], a música é uma manifestação cultural complexa, que envolve elementos estéticos, simbólicos e comunicativos, funcionando tanto como arte quanto como forma de interação social.

Conforme explica [Deckert 2014], os elementos musicais podem ser comparados a ingredientes utilizados na culinária. Assim como um cozinheiro escolhe e combina ingredientes para criar um prato, o compositor seleciona e organiza elementos como ritmo, melodia, timbre, intensidade, harmonia e forma musical para construir uma obra sonora. Esses componentes são essenciais para a estruturação da música, influenciando diretamente sua identidade e expressão.

2.1. Gênero e Subgênero Musical

Os gêneros musicais vão além de simples classificações, pois possuem significados sociais e simbólicos que influenciam tanto a criação quanto a recepção das músicas. Nesse sentido, [Janotti Jr. and Sá 2019] ressaltam que os gêneros funcionam como referência para disputas de gosto e como forma de consolidar a identidade dos artistas. Os subgêneros, por sua vez, surgem quando um gênero se expande e passa a englobar diferentes estilos. Para [Holt 2007], os subgêneros refletem transformações estilísticas e culturais, além do surgimento de novas cenas musicais. Assim, a análise de subgêneros como Sertanejo Raiz, Pagode e Pop Rock Nacional é importante para compreender como diferentes estilos se relacionam com aspectos culturais e sociais do Brasil:

- Sertanejo Raiz: também conhecido como canção caipira, se desenvolveu como um subgênero do gênero Sertanejo com características próprias e bem definidas dentro da música popular brasileira. Algumas características são: melancolia nas letras; ritmos lentos, palavras ou expressões tipicamente rurais que abordam o cotidiano do campo. Tende a se distinguir com facilidade, por abordar temas tão específicos.

- Pagode: é um subgênero do samba, trazendo temas mais centrados no cotidiano e no entretenimento. Enquanto o samba é associado a questões sociais e políticas, o pagode (em especial a partir dos anos 1980) foca mais em temas de amor, relações pessoais e experiências do dia a dia.
- Pop Rock Nacional: é um subgênero do rock que se distingue por incorporar elementos da identidade e da cultura brasileiras. Ao longo das décadas, foi influenciado e influenciou diferentes contextos históricos.

2.2. Aprendizagem de Máquinas

A Inteligência Artificial (IA) é uma área do conhecimento que busca simular as capacidades cognitivas humanas. Segundo [Kaufman 2018], a IA "refere-se a um campo de conhecimento associado à linguagem e à inteligência, ao raciocínio, à aprendizagem e à resolução de problemas". Esse conceito é fundamental para entender as diversas aplicações da IA, incluindo sua influência na música, onde pode ser usada para criar, analisar e até predizer tendências musicais. A área de IA cresceu muito nos últimos anos, fazendo com que diversas subáreas evoluíssem em paralelo, tais como formas de representação do conhecimento, aprendizagem de máquinas e, principalmente, o desenvolvimento de sistemas multiagentes. A ideia principal é fazer com que sistemas informatizados possam agir, atuar ou tomar decisões semelhantes àquelas tomadas por seres humanos.

Uma das subáreas da IA é a Aprendizagem de Máquinas. O conceito fundamental é que os sistemas são capazes de melhorar automaticamente por meio da experiência, sem a necessidade de intervenção constante por parte do programador. Isso significa que, em vez de programar explicitamente cada tarefa, o objetivo do aprendizado de máquina é criar algoritmos que possam aprender e evoluir com base em dados de entrada e saídas previamente definidas [Faceli et al. 2021].

Dentro dessa definição de aprendizado por meio da experiência, o aprendizado supervisionado é um dos processos mais explorado. Este tipo de aprendizado consiste em um processo de aprender com exemplos previamente rotulados, ou seja, exemplos nos quais a saída esperada é fornecida durante o treinamento, permitindo que os algoritmos de aprendizagem supervisionada melhorem seu desempenho com base na experiência já conhecidas. Estes exemplos rotulados são geralmente lidos a partir de conjuntos de dados chamados de *dataset* que contém os atributos (conhecidos como *features*) independentes e o atributo dependente (conhecido como *class* ou *target*). Com isso, os algoritmos podem realizar a classificação de novos itens (ex: determinar a classe para estes novos itens) utilizando os padrões aprendidos anteriormente no treinamento dos algoritmos.

Os algoritmos de aprendizagem de máquinas supervisionados utilizados neste trabalho são:

- Regressão Logística: utiliza uma função logística para estimar a probabilidade de uma instância pertencer a uma determinada classe, sendo bastante eficaz em problemas binários;
- Multinomial Naive Bayes: aplica o teorema de Bayes assumindo independência entre as variáveis, sendo especialmente usado em classificação de textos e análise de frequência de palavras;
- Random Forest: combina diversas árvores de decisão, agregando seus resultados para melhorar a precisão;

- Máquina de Vetores de Suporte (SVM – *Support Vector Machine*): busca encontrar um hiperplano ótimo que separe as classes com a maior margem possível, sendo apropriado em dados de alta dimensionalidade;
- Rede Neural Multicamada (MLP – *Multi-Layer Perceptron*): tenta imitar o funcionamento do cérebro humano por meio de camadas de neurônios artificiais conectados, sendo capaz de capturar relações complexas nos dados.

Para se criar os *datasets* contendo letras de músicas dos subgêneros Sertanejo Raiz, Pagode e Pop Rock Nacional a serem analisadas pelos algoritmos, foi utilizado a técnica de Raspagem de Dados. Segundo [Mitchell 2015], a raspagem de dados pode ser compreendido sob duas perspectivas: teórica e prática. Na abordagem teórica, consiste na coleta de dados realizada manualmente por um usuário por meio de um navegador web. Já na prática, envolve um conjunto amplo de técnicas e tecnologias de programação, abrangendo desde a análise de dados até aspectos relacionados à segurança da informação. Essa metodologia mostra-se especialmente útil para a extração automatizada de grandes volumes de informações disponibilizadas em páginas web, como, por exemplo, letras de músicas publicadas em sites especializados. Esta técnica foi uma das partes fundamentais para a etapa de coleta das letras de música para a construção do *dataset* utilizado.

2.3. Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é uma área interdisciplinar que combina conhecimentos da Inteligência Artificial, da Linguística Computacional e da Ciência da Computação, com o objetivo de permitir que máquinas compreendam, interpretem e processem a linguagem humana de forma automatizada [Jurafsky and Martin 2025]. Essa tecnologia atua na transformação de dados textuais não estruturados em representações estruturadas, compreensíveis para algoritmos.

O PLN envolve diversas etapas, como:

- Tokenização: segmentação do texto em unidades menores, como palavras ou sentenças;
- Remoção de *stopwords*: eliminação de palavras de baixa relevância semântica, como artigos e preposições;
- *Stemming* e lematização: redução das palavras à sua forma base, diminuindo a variação morfológica;
- Vetorização: conversão do texto em uma representação numérica para processamento computacional.

Segundo [Aggarwal and Zhai 2012], a natureza esparsa e de alta dimensionalidade dos textos os torna especialmente adequados para classificadores lineares, que conseguem explorar relações entre atributos e redundâncias para separar as classes de forma eficiente.

Para que os algoritmos realizem essa tarefa de forma eficaz, é necessário transformar os documentos em representações numéricas. Uma das técnicas mais utilizadas para esse fim é o TF-IDF (*Term Frequency-Inverse Document Frequency*). O TF-IDF é uma técnica de ponderação de termos amplamente utilizada em mineração de texto e recuperação de informação. A técnica mede a importância de uma palavra dentro de um documento em relação a um conjunto de documentos (corpus).

O cálculo da importância das palavras envolve dois componentes: TF (*Term Frequency*) que mede quantas vezes um termo aparece em um documento, refletindo sua

relevância local; e, IDF (*Inverse Document Frequency*) que avalia a raridade do termo no corpus, atribuindo menor peso a palavras comuns e maior peso a termos pouco frequentes.

De acordo com [Manning et al. 2008], o produto dessas duas medidas gera um valor que destaca palavras distintivas de um documento, reduzindo a influência de termos genéricos. Essa abordagem é particularmente eficaz para tarefas de classificação de texto, pois enfatiza o vocabulário específico de cada categoria.

2.4. Bibliotecas em Python

No desenvolvimento deste trabalho, foram utilizadas bibliotecas em Python amplamente reconhecidas pela comunidade científica, tais como:

- Requests: realizar requisições HTTP de forma simples para a coleta das letras das músicas;
- BeautifulSoup: extrair dados estruturados a partir de páginas da web;
- Pandas: manipular e organizar textos (manipulação de datasets);
- NLTK (Natural Language Toolkit): realizar o pré-processamento textual;
- Scikit-learn: aplicar os algoritmos de aprendizado de máquina na classificação dos subgêneros musicais.

3. Metodologia

Este trabalho se baseia em parte nas abordagens apresentadas por [Pimenta and Pugliesi 2022], que propõem uma metodologia eficaz para a classificação de gêneros musicais utilizando algoritmos de aprendizagem supervisionada. As técnicas descritas pelos autores foram fundamentais para o desenvolvimento da metodologia empregada neste estudo, especialmente no que diz respeito à extração de características.

O processo adotado para o reconhecimento de subgêneros musicais brasileiros (Sertanejo Raiz, Pagode e Pop Rock Nacional) com base nas letras das músicas foi dividido em quatro fases principais, conforme apresentada na Figura 1.

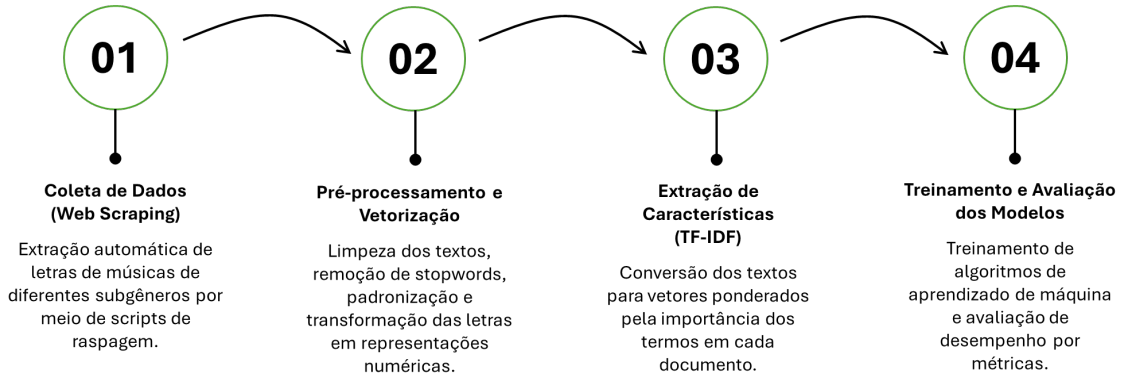

Figura 1. Etapas para desenvolvimento do classificador.

Para a construção do *dataset* utilizado neste trabalho, foi realizada a coleta automática de letras de músicas dos três subgêneros brasileiros (Sertanejo Raiz, Pagode e Pop Rock Nacional). Foi utilizada a linguagem Python e as bibliotecas Requests e BeautifulSoup. A escolha dos artistas para cada subgênero considerou a relevância histórica e

a representatividade no cenário musical brasileiro, de forma a garantir variedade e autenticidade nas composições.

A coleta dos dados foi realizada por meio da técnica de Raspagem de Dados. O processo ocorreu em duas etapas principais:

- Extração de URLs das músicas: para cada artista, foi acessada sua página no site `https://www.letras.mus.br` e coletados os links das 20 músicas mais populares, utilizando o seletor HTML que identifica a lista de faixas.
- Coleta das letras: para cada URL obtida, foi acessada a página individual da música, e o conteúdo da letra foi extraído diretamente da estrutura HTML, com a remoção de elementos irrelevantes, como anúncios e *tags*.

Além disso, para evitar sobrecarga no servidor do site, foi implementado um intervalo de 1,5 segundos entre as requisições, seguindo boas práticas de raspagem de dados. Ao final, foi construído um *dataset* contendo 1.500 músicas, com 500 músicas para cada subgênero. Foi escolhido 25 artistas representativos de cada subgênero e inserido as 20 músicas mais famosas de cada um deles. Os dados foram organizados em um arquivo CSV com as seguintes colunas:

- gênero: Subgênero musical da música;
- artista: Nome do artista ou banda;
- música: Título da música;
- letra: Letra completa da música.

Após a coleta das letras, realizou-se o pré-processamento dos textos para garantir que os dados estivessem limpos e padronizados antes da extração de características. Esse processo é essencial para reduzir ruídos e melhorar a qualidade das informações utilizadas pelos modelos de aprendizado de máquina.

O pré-processamento foi implementado em Python e bibliotecas pandas, re (expressões regulares), NLTK e unicodedata. Todas as etapas foram encapsuladas na classe PreVetorizacaoTexto, que realiza o carregamento do *dataset*, a limpeza das letras e a geração de um novo arquivo CSV com os textos já processados.

As principais etapas e recursos utilizados foram:

- Conversão para letras minúsculas: O método .lower() do Python foi usado para padronizar o texto em letras minúsculas, evitando diferenças entre palavras iguais que variam apenas em maiúsculas/minúsculas.
- Remoção de acentuação e caracteres especiais: A biblioteca unicodedata foi utilizada para remover acentos e caracteres especiais, garantindo que palavras equivalentes, como “coração” e “coracao”, fossem tratadas de forma idêntica e evitando inconsistências no modelo.
- Remoção de pontuações, números e normalização de espaços: A biblioteca re foi usada para aplicar expressões regulares que mantêm apenas caracteres alfabéticos, removendo pontuações, números e espaços duplicados, além de padronizar a separação entre palavras (exemplo: re.sub(r’[ˆa-z\s]’, ”, texto)).
- Tokenização e remoção de *stopwords*: Com a biblioteca NLTK, foi utilizada a lista de *stopwords* em português para remover termos muito frequentes e pouco informativos (exemplo: de, a, em), evitando interferências na análise semântica do texto.

Ao final dessas etapas, o texto pré-processado foi salvo em um novo arquivo CSV (dataset_pre_vetorizado.csv), que servirá como base para a próxima fase de extração de características. Nesta etapa, foi aplicado a técnica TF-IDF para calcular a importância de cada palavra em uma letra de música em relação a todo o corpus do *dataset*. Nesse processo, as letras limpas são convertidas em um formato numérico interpretável pelos modelos de aprendizado de máquina, por meio da biblioteca scikit-learn em Python. Para a vetorização, foram definidos os seguintes parâmetros: min_df=5, para ignorar termos presentes em menos de cinco letras e reduzir ruídos; max_df=0.8, para desconsiderar termos muito frequentes (em mais de 80% das letras); e ngram_range=(1,2), para incluir unigramas e bigramas, permitindo capturar não apenas palavras isoladas, mas também expressões relevantes.

Para classificar os subgêneros musicais com base nas letras, foram comparados cinco algoritmos de aprendizado de máquina supervisionado: Regressão Logística, Multinomial Naive Bayes, Random Forest, SVM e uma MLP. A avaliação do desempenho de cada modelo foi realizada utilizando a técnica de Validação Cruzada K-Fold com 5 dobras (k=5). Esse método garante uma avaliação mais robusta e menos enviesada, pois o *dataset* é dividido em 5 partes, e O modelo é treinado e testado 5 vezes, usando uma parte diferente como conjunto de teste a cada iteração, o que corresponde a 20% do total das amostras em cada iteração.

Os modelos foram treinados e avaliados da seguinte forma:

- Regressão Logística: Configurada com max_iter=1000 e solver='liblinear' para garantir a convergência;
- Multinomial Naive Bayes: Usado com sua configuração padrão;
- Random Forest: Utilizou-se n_estimators=100, um número suficiente de árvores para uma performance robusta;
- SVM: O kernel linear foi escolhido por sua eficácia com dados esparsos como os gerados pelo TF-IDF.
- MLP: Implementada com camadas densas e Dropout para evitar overfitting. O treinamento utilizou *Early Stopping* para interromper o processo se o desempenho parasse de melhorar, otimizando o tempo de execução e a performance.

4. Resultados

A acurácia e a matriz de confusão de cada modelo foram calculadas a partir da média dos resultados obtidos nas cinco partições do processo, o que proporciona uma estimativa de desempenho mais robusta e confiável. Após a aplicação da validação cruzada K-Fold em cada um dos modelos de classificação, as acurácias médias obtidas foram:

- Random Forest: 0.7565
- Multinomial Naive Bayes: 0.7766
- Regressão Logística:0.7920
- SVM: 0.8000
- Rede Neural Multicamada: 0.8288

A acurácia média mais alta foi alcançada pelo modelo de Rede Neural Multicamada, indicando sua superioridade na tarefa de classificação. As matrizes de confusão obtidas foram as apresentadas na Figura 2.

Na figura matriz de confusão da Rede Neural Multicamada, observa-se que:

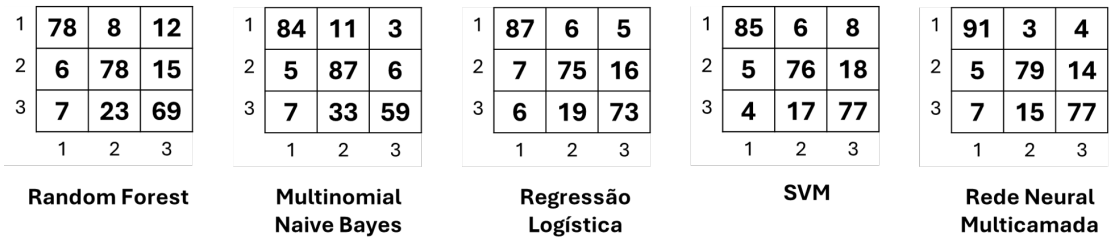

Figura 2. Matriz de Confusão: classificação dos subgêneros Sertanejo Raiz (1), Pagode (2) e Pop Rock Nacional (3).

- Sertanejo Raiz foi classificado corretamente em 91 ocasiões, demonstrando uma alta capacidade de identificação desse subgênero. A confusão com Pagode (3) e Pop Rock Nacional (4) foi mínima;
- Pagode foi corretamente identificado em 79 casos. A maior parte das confusões ocorreu com Pop Rock Nacional (14), indicando uma possível sobreposição de características textuais entre esses dois gêneros;
- Pop Rock Nacional foi classificado corretamente em 77 vezes, com a principal confusão ocorrendo com Pagode (15), o que reforça a observação anterior sobre a proximidade textual entre esses subgêneros.

De forma geral, o modelo de Rede Neural Multicamada apresentou o melhor desempenho geral entre todos os classificadores testados. A Tabela 1 resume as métricas de precisão, *recall* e *F1-score*.

Tabela 1. Precisão, *Recall* e *f1-Score* do classificador Rede Neural Multicamada.

Gênero	Precisão	Recall	F1-Store
Sertanejo Raiz	**0.88**	**0.93**	**0.91**
Pagode	**0.81**	**0.81**	**0.81**
Pop Rock Nacional	**0.81**	**0.78**	**0.79**

As métricas obtidas complementam a análise dos valores obtidos na matriz de confusão. O Sertanejo Raiz obteve o melhor desempenho, com um *f1-Score* de 0.91, impulsionado por um alto *Recall* (0.93). Isso indica que a maioria das músicas do gênero foi corretamente identificada, e o modelo teve pouquíssimos falsos negativos. Para Pagode e Pop Rock Nacional, os valores de Precisão e *Recall* são mais equilibrados e ligeiramente inferiores, resultando em um *f1-Score* de 0.81 e 0.79, respectivamente. Isso reforça a observação da matriz de confusão: a Rede Neural teve um desempenho robusto, mas com uma pequena dificuldade em diferenciar esses dois subgêneros.

A validação cruzada K-Fold resultou uma acurácia média de 0.8288 (± 0.0191). Como a variação entre as dobras foi pequena (+/- 0.0191), pode-se concluir que o desempenho do modelo é consistente, sugerindo ausência de sobreajuste a uma única partição dos dados.

4.1. Comparação entre modelos

A análise dos outros algoritmos revelou as seguintes características:

- Máquina de Vetores de Suporte (SVM): Com a segunda melhor acurácia (0.8000), o SVM demonstrou ser um classificador robusto para essa tarefa. Sua matriz de confusão mostrou uma performance sólida para todos os gêneros, com poucas confusões, e um desempenho no Pop Rock Nacional (77 acertos) que se igualou ao da Rede Neural;
- Regressão Logística: Alcançou uma acurácia de 0.7920. Este resultado sugere que a representação por características extraídas pelo TF-IDF captura um grau de separabilidade linear relevante para a tarefa de classificação.
- Multinomial Naive Bayes: Com 0.7766 de acurácia, este modelo obteve um bom desempenho, especialmente na classificação do Pagode, onde obteve o maior número de acertos (87) entre todos os modelos. No entanto, sua maior limitação foi a alta confusão na classificação do Pop Rock Nacional;
- Random Forest: Obtendo a acurácia mais baixa (0.7565), o Random Forest não foi tão eficaz quanto os outros modelos. Sua principal limitação foi a grande dificuldade em diferenciar o Pop Rock Nacional do Pagode.

4.2. Desempenho Específico por Gênero

Embora a Rede Neural Multicamada tenha alcançado a maior acurácia média geral, uma análise mais aprofundada das matrizes de confusão revela que a performance dos modelos variou significativamente entre os subgêneros. Essa observação é crucial, pois sugere que a escolha do algoritmo ideal pode depender do objetivo de classificação:

- Sertanejo Raiz: O modelo de Rede Neural Multicamada se mostrou o mais eficaz na identificação das letras deste gênero, com acurácia de 93% (91 acertos). Esse desempenho superior sugere que a complexidade do modelo foi capaz de capturar as nuances e padrões textuais específicos que distinguem o Sertanejo Raiz, com seu vocabulário e temáticas recorrentes;
- Pagode: O modelo Multinomial Naive Bayes apresentou o maior número de classificações corretas para a classe 'Pagode', com acurácia de 89% (87 acertos). O desempenho do Naive Bayes, um classificador probabilístico mais simples, indica que as características textuais do Pagode são mais distintivas e menos ambíguas;
- Pop Rock Nacional: A classificação deste subgênero apresentou um desafio maior para a maioria dos modelos, devido à sua provável sobreposição textual com os outros gêneros. No entanto, houve um empate técnico entre a Rede Neural Multicamada e a Máquina de Vetores de Suporte, ambas com acurácia de 78% (77 acertos).

Em resumo, a análise detalhada por gênero musical demonstrou que o desempenho dos modelos varia significativamente entre as diferentes classes, não havendo um único algoritmo ideal para todas as tarefas. O melhor desempenho da Rede Neural na acurácia geral sugere o seu potencial de generalização, enquanto a performance pontual do Naive Bayes no Pagode e o comportamento estável do SVM no Pop Rock representam informações valiosas sobre a adequação de cada algoritmo para tarefas específicas. Esta avaliação reforça a utilidade da análise combinada da acurácia geral e da performance pontual (e.g., f1-score por classe) para guiar a seleção do algoritmo, dependendo do foco da aplicação.

5. Conclusões

Este trabalho teve como objetivo aplicar técnicas de aprendizado de máquina para o reconhecimento automático de subgêneros musicais brasileiros — Sertanejo Raiz, Pagode e Pop Rock Nacional — a partir das letras de músicas. Por meio da coleta sistemática de dados via Raspagem de Dados, pré-processamento rigoroso dos textos, extração de características com TF-IDF e treinamento de diversos modelos de classificação, foi possível avaliar o potencial dessas técnicas na identificação de padrões linguísticos característicos de cada subgênero. Os resultados obtidos a partir das técnicas de aprendizado de máquina utilizadas demonstram que a Rede Neural Multicamada alcançou o melhor desempenho geral, com acurácia média de 82,88%, evidenciando sua capacidade de generalização. A análise das matrizes de confusão revelou que este modelo apresentou excelente desempenho na classificação do Sertanejo Raiz, enquanto o Multinomial Naive Bayes se destacou na identificação do Pagode, e o SVM obteve resultados robustos no Pop Rock Nacional, empatando com a Rede Neural neste subgênero.

Esses achados sugerem que, embora modelos mais complexos, como as Redes Neurais, apresentem desempenho global superior, classificadores mais simples, como o Naive Bayes, podem ser mais adequados para subgêneros específicos, dependendo da natureza textual das letras. Além disso, a dificuldade observada na separação entre Pagode e Pop Rock Nacional indica uma possível sobreposição temática e lexical, o que representa um desafio adicional para os modelos.

Portanto, o estudo confirma a viabilidade do uso de aprendizado de máquina na classificação de subgêneros musicais brasileiros a partir de letras, oferecendo subsídios para pesquisas futuras. Como perspectivas de continuidade, sugerem-se:a ampliação do conjunto de dados para incluir mais subgêneros e artistas, o uso de técnicas de processamento de linguagem natural mais avançadas e a investigação de arquiteturas de redes neurais mais profundas, visando melhorar a diferenciação entre gêneros com maior similaridade textual.

Com isso, abre-se espaço para aplicações práticas, como sistemas automáticos de categorização por meio das letras das músicas e análises semânticas mais sofisticadas no contexto da música brasileira.

Referências

Aggarwal, C. C. and Zhai, C., editors (2012). *Mining Text Data*. Springer, New York, NY, USA.

Borges, E., Simas Filho, E., Farias, C., Ribeiro, I., and Lopes, D. (2010). Classificação do gênero musical utilizando redes neurais artificiais. In *X Congresso Norte-Nordeste de Pesquisa e Inovação*, pages 1–8.

Deckert, M. (2014). *Educação Musical: da teoria à prática na sala de aula*. Moderna, São Paulo, SP. 1ª edição.

Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A., and Carvalho, A. C. P. d. L. F. (2021). *Inteligência Artificial: uma abordagem de aprendizado de máquina*. LTC, Rio de Janeiro, RJ, 2ª edition.

Ghirardi, R. (2013). *A música como linguagem: uma abordagem interdisciplinar*. Editora Perspectiva, São Paulo, SP. 1ª edição.

Holt, F. (2007). *Genre in Popular Music*. University of Chicago Press, Chicago and London, illustrated edition.

Janotti Jr., J. and Sá, S. P. d. (2019). Revisitando a noção de gênero musical em tempos de cultura musical digital. *Galáxia: Revista Interdisciplinar de Comunicação e Cultura*, 41(mai-ago):128–139.

Jurafsky, D. and Martin, J. H. (2025). Speech and language processing: An introduction to natural language processing. `https://web.stanford.edu/~jurafsky/slp3/`. 3º edição (online draft). Acessado em 18/09/2025.

Kaufman, D. (2018). *A inteligência artificial irá suplantar a inteligência humana?* Estação das Letras e Cores, Barueri, SP, 1ª edition.

Malheiro, R., Paiva, R. P., Mendes, A. J., Mendes, T., and Cardoso, A. (2004). Sistemas de classificação musical com redes neuronais.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK.

Mitchell, R. (2015). *Web Scraping with Python: Collecting Data from the Modern Web*. O'Reilly Media, Sebastopol, CA, 1º edition.

Pereira, E. M. (2009). *Estudos sobre uma ferramenta de classificação musical*. Dissertação (mestrado em engenharia elétrica), Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP.

Pimenta, M. F. and Pugliesi, J. B. (2022). Reconhecimento de gêneros musicais com técnicas de aprendizagem de máquina supervisionada. *Revista Eletrônica de Computação Aplicada*, 3(1).

Silva, L. T. d. (2024). Bamport: construção de uma base de dados multimodal para análise de músicas em português. Trabalho de conclusão de curso (graduação em ciência da computação), Universidade Federal do Rio de Janeiro, Instituto de Computação, Rio de Janeiro, RJ.

Trotta, F. (2005). Música e mercado: a força das classificações. *Revista Contemporânea*, 3(2):181–196.

Viana, R. T. C. (2024). Análise de padrões temáticos e emocionais em letras de músicas periféricas brasileiras utilizando ciência de dados. Dissertação (mestrado em ciência da computação), Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação, Belo Horizonte, MG.