Revista Eletrônica de Iniciação Científica em Computação, 2026, 24:1
ISSN: 3085-8461 • doi: 10.5753/reic.2026.7222

Artigo de Pesquisa/Research Paper

SONBRA: um dataset público e anotado para pesquisas de classificação de gêneros musicais brasileiros

SONBRA: A Public and Annotated Dataset for Research on Brazilian Music Genre Classification

Anderson Vinicius Ribeiro [Universidade Federal do Amapá | *dev.anderson.vribeiro@gmail.com*]
João Marcos de Oliveira Santana [Universidade Federal do Amapá |*joaosantana1998@gmail.com*]
Marcos Abreu de Oliveira [Universidade Federal do Amapá |*marcosabr97@gmail.com*]
Cláudio Gomes [Universidade Federal de Pernambuco |*claudiorogerio@unifap.br*]

Bacharelado em Ciênciada Computação, Universidade Federal do Amapá, Rodovia Josmar Chaves Pinto - KM 02, Macapá, AP, 68903-419, Brasil.

Resumo. Este trabalho apresenta o SONBRA, um *dataset* público e anotado projetado para fomentar pesquisas na área de classificação automática de gêneros musicais brasileiros. O conjunto de dados é composto por oito gêneros nacionais de grande representatividade: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo. Para sua construção, foram coletadas faixas musicais das quais se extraíram fragmentos de 30 segundos em cinco regiões distintas de cada áudio (início, do início ao meio, meio, do meio ao fim e fim), totalizando 10.000 amostras. Cada amostra foi anotada com seu gênero correspondente e descrita por 785 parâmetros resultantes da extração de características acústicas, sendo eles *Chroma CENS*, *Chroma CQT*, *Chroma STFT*, *Fourier Tempogram*, *Mel-spectrogram*, *MFCC*, *RMS*, *Spectral Bandwidth*, *Spectral Centroid*, *Spectral Roll-Off*, *Tempogram*, *Tonnetz* e *ZCR*. Com o objetivo de validar a viabilidade do *dataset* para tarefas de aprendizado de máquina, realizou-se um estudo de caso de classificação. Foram avaliados os algoritmos Árvore de Decisão, K-Vizinhos Mais Próximos (KNN), Perceptron Multicamadas (MLP), *Random Forest*, Máquinas de Vetores de Suporte (SVM), *XGBoost* e um *ensemble* do tipo *Voting*. Os resultados demonstram a efetividade do conjunto de dados, com o modelo *Voting* atingindo a maior acurácia (83,2%) e as combinações de características que incluíram *Mel*, *MFCC*, *Tempogram* e *Chroma CENS* destacando-se como as mais informativas para a predição do gênero musical. Este artigo detalha a metodologia de criação e a estrutura do SONBRA, oferecendo-o como um recurso público para a comunidade científica, com o intuito de padronizar e avançar as pesquisas computacionais sobre a riqueza musical do Brasil.

Abstract. This paper introduces SONBRA, a publicly available annotated dataset designed to support and advance research in the automatic classification of Brazilian music genres. The dataset encompasses eight highly representative national genres: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo, and Sertanejo. For its creation, music tracks were collected and segmented into 30-second excerpts extracted from five distinct segments of each track (beginning, beginning-to-middle, middle, middle-to-end, and end), resulting in a total of 10,000 samples. Each sample is annotated with its corresponding genre label and is characterized by 785 features extracted from the audio, including *Chroma CENS*, *Chroma CQT*, *Chroma STFT*, *Fourier Tempogram*, *Mel-spectrogram*, *MFCC*, *RMS*, *Spectral Bandwidth*, *Spectral Centroid*, *Spectral Roll-Off*, *Tempogram*, *Tonnetz*, and *ZCR*. To validate the dataset's utility for machine learning tasks, we conducted a comprehensive classification benchmark. We evaluated the following algorithms: Decision Tree, K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), *Random Forest*, Support Vector Machines (SVM), *XGBoost*, and a *Voting* ensemble. The results demonstrate the dataset's effectiveness, with the *Voting* classifier achieving the highest accuracy (83.2%) and feature combinations involving *Mel*, *MFCC*, *Tempogram*, and *Chroma CENS* proving to be the most informative for genre prediction. This paper details the curation methodology and structure of the SONBRA dataset, which we release as a public resource to the scientific community to serve as a benchmark and to stimulate further computational research into the richness of Brazilian music.

Palavras-chave: Aprendizado de máquina, recuperação de informação musical, classificação de gêneros musicais

Keywords: Machine learning, music information retrieval, music genre classification

Recebido/Received: 18 December 2025 • **Aceito/Accepted:** 16 January 2026 • **Publicado/Published:** 30 January 2026

1 Introdução

Além do seu uso para o entretenimento, a música constitui uma forma de expressão, comunicação e mobilização, ligada diretamente a contextos socioculturais e ideológicos de uma determinada época. Sua estrutura é composta por harmonia, melodia, ritmo, timbre e textura, que expressam sua diversidade sonora. A categorização por gêneros pode ser definida por esses elementos, mas também por aspectos culturais e históricos, como origem geográfica, tradições e influências artísticas. Em um contexto nacional, a riqueza musical advém da interculturalidade de influências, como a indígena, africana e europeia, que foram a base para o desenvolvimento de diversos estilos musicais marcantes para a identidade do país Quadros Júnior [2019].

Aplicações e serviços utilizam dados extraídos de músi-

cas para classificar, organizar e fazer sugestões musicais aos usuários. Nesse contexto, as ferramentas de classificação de gêneros musicais tornam-se essenciais para encontrar similaridades de áudio. Essa área de pesquisa é amplamente estudada, com pesquisas resultando em bases de dados, sistemas e modelos de aprendizado de máquina Shirol and Kathiresan [2023]. No entanto, a maioria dessas é focada em música internacional e em gêneros mais populares, como pop e rock. A classificação de gêneros musicais brasileiros sofre uma carência de bases de dados balanceadas e com quantidade de amostras relevante de gêneros regionais, pois as que contêm gêneros nacionais possuem poucas amostras ou são mal balanceadas, não sendo efetivas para a criação de modelos de aprendizado de máquina eficientes Conceição *et al*. [2020].

Diante dessa lacuna, este trabalho propõe a criação de um conjunto de dados estruturado, balanceado, anotado e público, contendo amostras dos gêneros populares brasileiros: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo. A base será composta por características musicais extraídas de faixas musicais obtidas por meio de plataformas de *streaming*. Para validar a eficácia e a aplicabilidade em tarefas de classificação automática, realizou-se um estudo de caso em modelos de aprendizado de máquina, avaliando seis dos algoritmos mais comuns para a tarefa de classificação musical e também um sistema de *voting*. O *dataset* busca fortalecer a representatividade de gêneros nacionais em estudos de classificação e fomentar novas pesquisas e aplicações nessa área.

Este artigo está organizado da seguinte forma: A Seção 2 faz uma revisão de trabalhos similares. Na Seção 3, apresenta-se a seleção das músicas e a extração das características para a criação da base de dados. A avaliação da base em modelos de aprendizado de máquina é detalhada na Seção 4. A Seção 5 apresenta os resultados obtidos dessa avaliação, e a Seção 6 trata das conclusões e trabalhos futuros.

2 Trabalhos relacionados

O trabalho de Farajzadeh *et al*. [2023] apresentou o *Persian Music Genre classification (PMG-Data)*, uma base de dados balanceada, anotada e disponível publicamente para a classificação de gêneros musicais persas. O *PMG-Data* é composto por 500 amostras de faixas musicais persas, distribuídas uniformemente entre os gêneros Rap, Traditional, Pop, Rock e Monody. O trabalho também propõe o *PMG-Net*, um modelo automatizado e eficiente baseado em redes neurais convolucionais (CNN). Foram realizados três experimentos utilizando o *PMG-Data*. Os dois primeiros empregaram o modelo *PMG-Net*. No primeiro, os sinais de áudio foram padronizados em segmentos de 32 segundos e foi adotada uma validação cruzada de 5 partições. A acurácia média nesse experimento foi de 76%, alcançando um máximo de 79%. O segundo experimento utilizou um método de fragmentação para aumento de dados. As faixas foram divididas em seis segmentos de oito segundos, obtendo 86% como a maior acurácia. O terceiro experimento consistiu em avaliar a base de dados com os algoritmos tradicionais *KNN, SVM, Random Forest* e *XGBoost*, utilizando as características *STFT* e *MFCC* como entrada. O *SVM* com *STFT* obteve 65% de acurácia, sendo o melhor resultado desse experimento, o que comprova a eficiência do *PMG-Net* em comparação às abordagens tradicionais.

O AMPOP, de Silva and Gomes [2022], é uma base de dados desenvolvida para o treinamento de um classificador automático de músicas populares da região amazônica. A base possui 125 amostras de cada um dos oito gêneros selecionados: Andino, Brega, Carimbó, Cúmbia, Merengue, Pasillo, Salsa e Vaqueirada. Foi utilizado um método de aumento de dados por fragmentação, no qual cada sinal de áudio foi segmentado em três partes de 10 segundos, extraídas do início, meio e fim de cada faixa. Dessas amostras, foram extraídas as características *Tempogram, Chroma STFT, Chroma CQT, RMS* e *ZCR*, totalizando 785 parâmetros. A base de dados foi avaliada com os algoritmos *KNN, MLP, SVC* e *XGBoost*, utilizando validação cruzada com 10 partições. Os fragmentos de início, meio e fim foram avaliados individualmente. Os melhores resultados foram obtidos com o fragmento do meio, com 61,33% de acurácia no *SVC* e 61,17% no *XGBoost*, valores que se mostraram bastante próximos.

A *Latin Music Database* (LMD), de Silla *et al*. [2018], contém 3.227 faixas, distribuídas em dez gêneros musicais latino-americanos: Axé, Bachata, Bolero, Forró, Gaúcha, Merengue, Pagode, Salsa, Sertanejo e Tango. Em um banco de dados relacional, estão registrados os metadados de artista, álbum e gênero de cada faixa, rotulados por dois especialistas da área musical. A base de dados está em formato MP3 e, para manter a qualidade de áudio, foram usadas apenas músicas comerciais. Por questões de direitos autorais, essa base não está disponível publicamente. A LMD suporta diversas tarefas de recuperação de informação musical, incluindo classificação de gêneros e análise de ritmos.

Os autores de Tzanetakis and Cook [2002] criaram a base de dados GTZAN com o objetivo de automatizar a tarefa de anotação de gênero musical. Ela é composta por 100 amostras de 30 segundos para cada um dos dez gêneros selecionados: Blues, Clássico, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae e Rock. Alguns desses gêneros possuem subclassificações em subgêneros, como Choir, Orchestra, Piano e String Quartet para o Clássico, e Big Band, Cool, Fusion, Piano, Quartet e Swing para o Jazz. A base contém 30 parâmetros, organizados em três categorias: Textura Tímbrica, Conteúdo Rítmico e Tonalidade. Para sua validação, foram utilizados os modelos Gaussian Mixture Model (GMM) e KNN, adotando-se validação cruzada com 10 partições. Com todas as características, o GMM obteve uma acurácia média de 61%, sendo o melhor resultado entre os modelos avaliados. Na avaliação de subconjuntos, os gêneros Jazz e Clássico alcançaram acurácias médias de 68% e 88%, respectivamente, também com o GMM.

O *Free Music Archive* (FMA), proposto em Deferrard *et al*. [2017], surge como uma alternativa à base de dados GTZAN, que é amplamente utilizada, mas apresenta diversas limitações. Seu acervo é constituído por arquivos sob a licença *Creative Commons (CC)*, o que permite sua disponibilização pública. A base é composta por 106.574 faixas, distribuídas em 16 gêneros musicais principais, totalizando 161 gêneros quando incluídos os subgêneros, e oferece 518 parâmetros extraídos de nove características principais, tais como *MFCC*, *Chroma*, *Tonnetz*, *Spectral Centroid*, *Spectral Contrast* e *ZCR*. Essa base de dados possui quatro subdivisões: **full**, que apresenta todos os dados; **large**, que limita

as faixas a 30 segundos extraídos do segmento médio das músicas; **medium**, que contempla 25.000 amostras apenas dos 16 gêneros principais, com segmentos de 30 segundos, porém é desbalanceada; e **small**, um conjunto balanceado que possui 1.000 amostras de cada um dos oito gêneros mais populares. A subdivisão **medium** foi avaliada com os algoritmos de aprendizado de máquina Regressão Linear (LR), KNN, SVM e MLP. Utilizando todos os parâmetros, a maior acurácia obtida nos testes foi de 63%, alcançada pelo SVM.

O *Bangla Music Dataset*, elaborado por Al Mamun *et al*. [2019], contém entre 250 e 350 músicas para cada um dos seis gêneros de origem bengali: *nazrulgeeti*, *rabindra sangeet*, *polligeeti*, Música de Banda, Hip-Hop e *adhunik gaan*, totalizando 1.742 faixas. As características extraídas das amostras de áudio foram: *ZCR*, *Spectral Centroid*, *MFCC*, *Spectral Roll-Off*, Frequência Cromática, *Spectral Bandwidth*, *Spectral Flux*, *Pitch* e *Tempo*. Com base nesse conjunto de dados, Hasib *et al*. [2022] implementou a BMNet-5, uma rede neural profunda para classificação de música bengali. A arquitetura da rede possui três camadas ocultas além das camadas de entrada e saída. Após o treinamento com 187 épocas, o modelo alcançou sua acurácia ótima de 90,32%.

A maioria dos conjuntos de dados disponíveis na literatura contemporânea concentra-se em gêneros musicais internacionais, como rock e pop. Os poucos que incluem músicas brasileiras possuem uma quantidade limitada de amostras ou classificam-nas genericamente como "ritmos latinos". Portanto, a SONBRA apresenta-se como uma base de dados estratificada, balanceada, disponível publicamente e especificamente focada na diversidade e na complexidade dos gêneros musicais populares brasileiros. A Tabela 1 apresenta uma comparação entre os trabalhos relacionados e a proposta.

Tabela 1. Comparação de Datasets

Base de dados	Gêneros	Disponível	Amostras	Estratificada
PMG-Data	5	Sim	3000	Sim
Ampop	8	Sim	3000	Sim
FMA	16	Sim	25000	Não
GTZAN	10	Sim	1000	Sim
LMD	10	Não	3227	Não
Hasib	6	Sim	1742	Não
SONBRA	8	Sim	10000	Sim

3 Base de dados SONBRA

A criação da base de dados foi iniciada com a seleção dos gêneros musicais brasileiros. Em seguida, procedeu-se à seleção e coleta das faixas para, logo após, realizar a extração das características musicais. Estas etapas estão descritas a seguir.

3.1 Seleção dos gêneros musicais

A música brasileira contém diversos ritmos populares e, para este trabalho, foram escolhidos: Samba, Samba-Enredo, Pagode, Bossa Nova, Funk, Forró, Forró Piseiro e Sertanejo. O **Samba** é de origem afro-brasileira, com forte uso de instrumentos percussivos e uma síncope característica. Dele surgiram outros ritmos, como o **Samba-Enredo**, que é uma variação usada nos populares desfiles de escola de samba, e o **Pagode**, que nos anos 1990 tornou-se mais popular com a comercialização musical mais moderna Quadros Júnior [2019]. A **Bossa Nova** surgiu como uma tentativa de modernizar a música brasileira; para isso, usou-se de elementos característicos dos ritmos brasileiros, mas também de influências de ritmos estrangeiros, como *Jazz* e *Blues*. Outro gênero marcado pela influência estrangeira foi o **Funk**, que se tornou popular nas periferias do Rio de Janeiro por meio de festas em que o público criava suas próprias letras sobre as músicas do *Funky* americano e, posteriormente, do *Miami Bass* Bezerra [2017]; Viana [2010]. Por sua vez, o **Forró** surgiu como nome de uma festa em que se tocavam vários ritmos. Dessa maneira, tornou-se o nome de um conjunto de ritmos e, mais adiante, o nome de um gênero, sendo característico o uso da zabumba, do triângulo e da sanfona. O **Forró Piseiro** é uma variante do Forró que utiliza instrumentos eletrônicos, como teclados e sintetizadores, guitarras, além da sanfona e de outros instrumentos Dias and Dupan [2022]. Por último, o **Sertanejo** é uma modernização da música caipira, originária do interior de São Paulo e de outros estados do Centro-Sul, sendo característico o uso de duplas e trios para sua execução e da viola (violão) Quadros Júnior [2019]; Alonso [2011]; Caldas [1987].

3.1.1 Criação da base de dados

A plataforma de *streaming* de vídeos *YouTube* foi utilizada para a seleção das músicas. Para cada um dos oito gêneros, foram selecionadas 250 faixas, totalizando 2.000 músicas. Visando à melhor qualidade sonora e consistência, priorizaram-se gravações de estúdio, o que possibilita uma análise mais precisa e comparável entre os ritmos. Além disso, foram consideradas apenas faixas com duração entre dois e dez minutos, garantindo assim uma variação sonora adequada dos segmentos após a fragmentação.

Após a seleção, as músicas foram coletadas utilizando a ferramenta *Pytube*[1] no formato MP4. Em seguida, procedeu-se à conversão para o formato WAV por meio do *ffmpeg* Tomar [2006], com o objetivo de minimizar eventuais interferências e ruídos, assegurando a integridade dos dados para a posterior extração de características.

Uma faixa de áudio completa pode apresentar alta variabilidade estrutural ao longo de sua duração. Processar seu conteúdo integral pode ocultar padrões locais que poderiam ser identificados por um modelo Silla *et al*. [2007]. Além disso, do ponto de vista computacional, esse processamento é custoso em termos de memória e tempo de execução, especialmente quando aplicado a grandes bases de dados Costa *et al*. [2004].

Diante desse cenário, adotou-se uma estratégia de amostragem multi-segmento. As faixas musicais foram fragmentadas em segmentos de 30 segundos, extraídos de cinco regiões distintas: início (*begin*), meio (*middle*), transição do início para o meio (*middle_1*), transição do meio para o final (*middle_2*) e final (*end*). A Tabela 2 detalha o cálculo utilizado para a segmentação das músicas, enquanto a Figura 1 ilustra as etapas do fluxo de criação da base de dados.

Dessa forma, as 250 faixas de cada gênero geraram 1.250 amostras após a fragmentação (250 × 5 tipos de fragmentos). Consequentemente, a base de dados musical resultante possui, no total, 10.000 fragmentos de áudio de 30 segundos (1.250 ×

[1] https://pytube.io/en/latest/

Tabela 2. Intervalos de extração dos segmentos de áudio de 30 segundos.

Região	Rótulo	Intervalo (T = duração total)
Início	`begin`	$[0,\ 30]$
Início-Meio	`middle_1`	$[T/2-30,\ T/2]$
Meio	`middle`	$[T/2-15,\ T/2+15]$
Meio-Final	`middle_2`	$[T/2,\ T/2+30]$
Final	`end`	$[T-30,\ T]$

8 gêneros), que serão utilizados exclusivamente para fins acadêmicos. O conjunto de dados está disponível para consulta em repositório Zenodo Ribeiro *et al*. [2025].

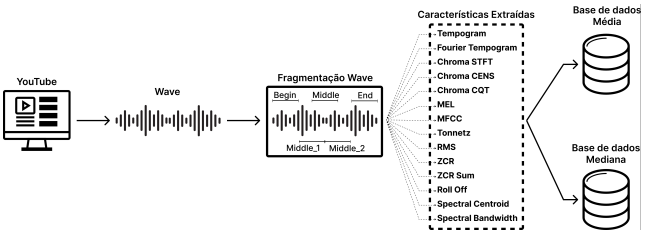

Figura 1. Visão geral das etapas para a construção da base de dados

3.1.2 Extratores musicais

De acordo com a Figura 1, utilizando a biblioteca *Librosa* McFee *et al*. [2015] para a construção da base de dados, aplicaram-se, nas amostras de 30 segundos, os extratores musicais de *Tempogram, Mel-espectrogram, Fourier Tempogram, Chroma STFT, Chroma CENS, Chroma CQT, MFCC, Tonnetz, RMS, ZCR, Spectral Roll-Off, Spectral Centroid* e *Spectral Bandwidth*.

O *tempogram* é uma representação que evidencia os tempos mais relevantes em batidas por minuto (BPM) ao longo da música e pode ser extraído por diferentes métodos, a partir da detecção de mudanças significativas no sinal de áudio por meio de uma função chamada *novelty*, que identifica os momentos em que notas são tocadas. Entre as técnicas utilizadas estão o Fourier Tempogram, que enfatiza o que se chama de harmônicos de tempo, e o Tempogram Autocorrelacionado, que enfatiza os sub-harmônicos Müller [2015].

O *mel-espectrogram (Mel)* utiliza-se da escala mel para representar os sons conforme a percepção humana, enquanto o *Mel-Frequency Cepstrum Coeficient (MFCC)* utiliza o mesmo conceito, porém usa coeficientes cepstrais, aplicando uma Transformada Discreta do Cosseno ao espectro mel Klapuri and Davy [2006].

O *Chroma*, baseado nas doze classes de altura da escala temperada, representa notas que compartilham identidade sonora mesmo em diferentes oitavas. É extraído por técnicas como *Chroma STFT*, *Chroma CQT* e *Chroma CENS*, esta última aplicando normalização e suavização para reduzir ruídos e enfatizar estruturas relevantes Müller and Balke [2018]; Müller [2015].

As características espectrais complementam a análise dos sinais musicais. O *Spectral Centroid* indica a concentração média das frequências Sharma *et al*. [2020], enquanto o *Spectral Bandwidth* e o *Spectral Roll-Off* descrevem a dispersão e a distribuição de energia no espectro Klapuri and Davy [2006].

Extratores de domínio temporal fornecem informações sobre as propriedades temporais dos sinais de áudio ao longo do tempo. O *Zero Crossing Rate (ZCR)* mede as transições de sinal em um *frame*, associadas a ruído, enquanto o *Root Mean Square (RMS)* calcula a energia geral das amplitudes, auxiliando na detecção da presença e intensidade do som Klapuri and Davy [2006].

O *Tonnetz* captura relações harmônicas entre notas e acordes em um contexto tonal, principalmente das quintas perfeitas e terças menores e maiores, por meio de um vetor Tonal Centroid de seis dimensões Harte *et al*. [2006].

As características *Tempogram*, *Fourier Tempogram*, *Chroma STFT*, *Chroma CENS*, *Chroma CQT*, *Tonnetz*, *RMS* e *Spectral Centroid* foram extraídas utilizando os parâmetros padrão da biblioteca *Librosa*. Para o *Mel-spectrogram*, o argumento *n_mels* foi configurado como 128, determinando o número de bandas de frequência geradas. Para os *MFCC*, o argumento *n_mfcc* foi definido como 20, extraindo-se os vinte primeiros coeficientes cepstrais.

Para o *ZCR*, além da extração padrão, calcularam-se as estatísticas de média e mediana ao longo do tempo. Adicionalmente, criou-se uma variante denominada *ZCR sum*, correspondente à soma total das taxas de cruzamento por zero do sinal.

No caso do *Spectral Roll-Off*, realizaram-se três extrações distintas, cada uma com um valor diferente do parâmetro *roll_percent*: 0,01; 0,85 e 0,99. Procedimento análogo foi aplicado ao *Spectral Bandwidth*, no qual o argumento *p* foi variado de 2 a 12, resultando em 11 extrações que capturam a largura de banda espectral sob diferentes normas.

A extração das características gera uma representação numérica, que pode assumir a forma de escalares, vetores ou matrizes. Com o intuito de reduzir o uso de recursos computacionais, foram realizadas análises estatísticas considerando a média e a mediana como valores de uma determinada característica. Isso resultou em duas bases de dados, como ilustra a Figura 1. Este processo impossibilita que os dados sejam retornados à sua originalidade, ou seja, é impossível reconvertê-los para áudio. A Tabela 3 mostra quantos parâmetros a base de dados possui para cada extrator.

Tabela 3. Quantidade de parametros por extrator

Extrator	Quantidade de Parametros
Fourier Tempogram	193
Tempogram	384
Chroma STFT	12
Chroma CENS	12
Chroma CQT	12
Mel-spectogram	128
MFCC	20
Tonnetz	6
RMS	1
ZCR	1
ZCR Sum	1
Spectral Roll-Off	3
Spectral Centroid	1
Spectral Bandwidth	11
Total	**785**

4 Avaliação

A base de dados completa, com 10.000 amostras, foi dividida para fins de avaliação. Inicialmente, foi separada em dois conjuntos: desenvolvimento (70%) e teste (30%). A divisão do conjunto de desenvolvimento em treinamento (80%) e validação (20%) seguiu o método de reamostragem (*resampling*) sem reposição, também conhecido como *hold-out*. Nesta técnica, uma amostra é reservada para formar o conjunto de validação, enquanto o restante compõe o conjunto de aprendizado. O primeiro é utilizado para treinar os modelos, e o segundo, para estimar o erro de generalização Oneto [2018]. O esquema completo de divisão da base de dados está ilustrado na Figura 2. Essas partições foram empregadas no treinamento dos modelos de aprendizado de máquina, visando identificar aquele que alcançasse o melhor desempenho conforme a métrica de acurácia.

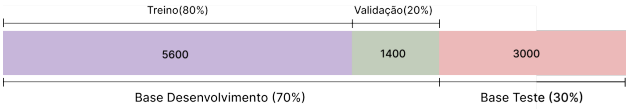

Figura 2. Divisão da Base de Dados

4.1 Modelos de aprendizado de máquina

A SONBRA é uma base de dados anotada, na qual todos os registros possuem o rótulo de sua classe correspondente. Por esse motivo, optou-se pela abordagem de aprendizado supervisionado, que, com base em registros rotulados, cria um modelo generalizado capaz de prever o rótulo de novos registros. Dentre os algoritmos supervisionados disponíveis, foram selecionados aqueles mais frequentemente utilizados nos trabalhos relacionados: Árvores de Decisão, *K Vizinhos Mais Próximos (KNN)*, *Perceptron Multicamadas (MLP)*, *Random Forest*, *Máquinas de Vetores de Suporte (SVM)* e *eXtreme Gradient Boosting (XGBoost)*.

As Árvores de Decisão são modelos hierárquicos capazes de lidar com dados complexos. A estrutura é representada como uma árvore binária, na qual os parâmetros de entrada são recursivamente divididos em subdomínios, denominados nós. O nó final, chamado de folha, representa uma classe à qual um registro pode pertencer. A classificação é realizada com base na ordenação dos valores das características, buscando-se o particionamento que define o caminho ótimo até uma folha, conforme critérios de decisão predefinidos Kotsiantis *et al*. [2006].

O *Random Forest* emprega a técnica de *ensemble*, que combina múltiplos modelos para melhorar a previsão final Bühlmann [2012], para aprimorar o desempenho das árvores de decisão. Diversas árvores são construídas a partir de reamostragens dos dados iniciais, utilizando o método de *bootstrap* (amostragem com reposição), que cria subconjuntos de treinamento que podem conter instâncias repetidas Bühlmann [2012]. Posteriormente, uma função de agregação, como a votação majoritária, é aplicada para determinar o resultado final, combinando as predições individuais de todas as árvores Suthaharan [2016].

O XGBoost foi desenvolvido para aprimorar o desempenho do *gradient boosting*, uma evolução da técnica de *boosting* na qual os modelos são combinados sequencialmente, com cada etapa usando o resultado da anterior para minimizar o viés (*bias*) da função Bühlmann [2012]. Especificamente, o *gradient boosting* aplica uma função de custo diferenciável às previsões entre cada etapa, visando minimizá-la para aumentar a precisão Ramraj *et al*. [2016]. O XGBoost otimiza esse processo, tornando-o mais rápido e eficiente computacionalmente. O modelo alcança esse objetivo por meio do armazenamento de dados em memória, o que permite reutilizar cálculos já realizados anteriormente por meio de *cache*. Além disso, o XGBoost executa os estimadores em paralelo, utilizando múltiplas *threads* do processador. Essas otimizações tornam o modelo significativamente mais rápido em comparação a métodos similares, especialmente em bases de dados de grande volume Chen and Guestrin [2016].

O KNN é um modelo não paramétrico, o que significa que não possui uma quantidade fixa ou predefinida de parâmetros a serem aprendidos. O algoritmo opera selecionando os k vizinhos mais próximos de uma determinada instância, utilizando para isso uma medida de distância, como a Euclidiana. A premissa fundamental do KNN é a de que instâncias próximas no espaço de características tendem a pertencer à mesma classe. Assim, após identificar os vizinhos mais próximos, o algoritmo atribui à instância analisada a classe mais frequente entre aquelas k amostras selecionadas Scaringella *et al*. [2006].

As SVMs são algoritmos robustos para tarefas de classificação, tanto linear quanto não linear, apresentando alta eficácia principalmente em conjuntos de dados complexos de pequeno a médio porte. O objetivo central do método é identificar o hiperplano ótimo que maximize a margem de separação entre as classes. Para definir esse hiperplano, o algoritmo identifica os pontos de dados mais críticos, conhecidos como vetores de suporte, que são as amostras mais próximas do hiperplano e delineiam os limites das margens. Em cenários onde os dados não são linearmente separáveis no espaço original, as SVMs empregam uma função *kernel*. Essa função realiza um mapeamento não linear dos dados para um espaço de características de dimensionalidade superior, onde se torna possível encontrar um hiperplano que separe as classes de forma linear Kotsiantis *et al*. [2006].

O MLP é uma arquitetura de rede neural artificial composta por três elementos fundamentais: uma camada de entrada, uma ou mais camadas ocultas com neurônios interconectados e uma camada de saída responsável por gerar as probabilidades associadas a cada classe. Nas camadas ocultas, cada neurônio processa os dados de entrada aplicando uma função de ativação não linear e transmite o resultado para a camada subsequente. A aprendizagem ocorre por meio da minimização de uma função de custo (ou perda), que quantifica a discrepância entre as previsões do modelo e os valores reais. Para efetuar essa minimização, emprega-se o algoritmo de retropropagação (*backpropagation*), que calcula o gradiente do erro em relação a todos os pesos da rede e os ajusta iterativamente, direcionando o modelo para uma configuração que reduza progressivamente o erro. Esse processo de ajuste é repetido ao longo de múltiplas épocas (*epochs*) até que o modelo atinja um nível satisfatório de desempenho Géron [2019]; Kotsiantis *et al*. [2006].

4.2 Otimização de hiperparâmetros e seleção de características

A construção de modelos de aprendizado de máquina para a classificação de gêneros musicais envolve a definição de diversos parâmetros ajustáveis, conhecidos como hiperparâmetros. A otimização de hiperparâmetros consiste em buscar, dentro de um conjunto pré-definido, os valores que minimizem os erros do modelo, maximizando assim sua acurácia Yu and Zhu [2020]. Para esse fim, foi empregada a técnica de busca em grade (*grid search*), na qual se define um conjunto de valores candidatos para cada hiperparâmetro e, a partir disso, realiza-se uma varredura exaustiva de combinações, treinando e avaliando o modelo para cada uma delas. Ao final do processo, são selecionados os valores que proporcionaram o melhor desempenho. Uma limitação dessa abordagem é que, dependendo do número de hiperparâmetros e dos valores testados, o espaço de busca pode tornar-se computacionalmente proibitivo, não sendo recomendada para cenários com grandes conjuntos de dados ou alta complexidade Yu and Zhu [2020].

A seleção de características foi realizada em duas etapas principais. Primeiramente, os modelos foram avaliados individualmente para cada característica musical extraída, registrando-se as cinco melhores para cada algoritmo. Posteriormente, foram conduzidos experimentos combinando pares (combinações 2 a 2) e trios (combinações 3 a 3) de características, com o objetivo de identificar sinergias que pudessem proporcionar uma representação mais rica e discriminativa dos dados. Exemplos dessas combinações incluem: *Tempogram*; *Tempogram* com *MFCC*; e *Tempogram*, *MFCC* e *Mel*. Paralelamente a essa etapa de combinação, procedeu-se à otimização dos hiperparâmetros, descartando ou ajustando os valores que não produziam resultados minimamente satisfatórios. As seções seguintes descrevem em detalhe os métodos de treinamento e os critérios de seleção aplicados em cada fase.

4.2.1 Etapa 1 - validação cruzada

A validação cruzada de k-folds é uma técnica na qual o conjunto de dados é dividido em k partições (folds). O modelo é treinado k vezes, utilizando $k-1$ partições para treino e a partição restante para validação em cada iteração Hastie *et al.* [2009]. Para a seleção dos melhores modelos nesta etapa, foi empregada uma validação cruzada estratificada com 10 partições, utilizando exclusivamente o subconjunto de Treino da Base de Desenvolvimento. A estratificação garante que a proporção das classes (gêneros musicais) seja preservada em cada partição, prevenindo desequilíbrios que poderiam enviesar a avaliação do desempenho Sammut and Webb [2011]. O desempenho de cada configuração de modelo foi resumido pela acurácia média ao longo das partições, e a configuração com o maior valor foi selecionada para representar cada abordagem de treinamento.

4.2.2 Etapa 2 - erro de generalização

A segunda etapa consistiu no treinamento dos modelos utilizando a base de treino e na subsequente avaliação de seu desempenho no conjunto de validação. A acurácia obtida neste conjunto fornece uma estimativa pontual do erro de generalização do modelo, ou seja, uma medida de seu desempenho esperado em dados não vistos durante o treinamento Oneto [2018]. Para cada configuração de modelo, calculou-se a diferença absoluta entre essa acurácia de validação e a acurácia média obtida na etapa de validação cruzada. O modelo considerado ideal foi aquele que apresentou a menor diferença, um critério que seleciona a configuração com o equilíbrio mais favorável entre ajuste aos dados de treino e capacidade de generalização, minimizando assim evidências de *overfitting*.

4.2.3 Sistema voting

Adicionalmente aos modelos individuais, foi implementado um sistema de votação (*Voting*) Géron [2019]. Para compor este *ensemble*, foi selecionada a melhor configuração de cada algoritmo (KNN, MLP, SVM, Random Forest, XGBoost e Decision Tree) identificada na validação cruzada da Etapa 1. O sistema de votação foi avaliado sob dois critérios de decisão:

- **Votação *Hard***: a classe final é determinada pelo voto majoritário entre as previsões dos modelos individuais.
- **Votação *Soft***: a classe final é determinada pela média das probabilidades preditas por cada modelo, escolhendo-se a classe com maior probabilidade agregada.

O treinamento e a avaliação do *ensemble* seguiram o mesmo fluxo descrito nas subseções anteriores, com uma exceção na etapa de seleção de características. Para o sistema de votação, foram utilizadas as combinações de três características que proporcionaram a maior acurácia individual para cada um dos modelos componentes.

4.3 Avaliação final

Ao final das etapas de seleção de características e otimização de hiperparâmetros, foram selecionados 14 modelos: seis provenientes da etapa de validação cruzada, seis da etapa de cálculo do erro de generalização e dois do sistema de votação. Cada um desses modelos foi então treinado no conjunto de Treino completo e sua acurácia final foi calculada no conjunto de Teste, que fornece a estimativa mais robusta do desempenho em dados não vistos. Com base nessas acurácias finais, identificou-se o modelo com o melhor desempenho como o mais adequado para o contexto deste estudo. Por fim, selecionaram-se sete modelos computacionais que se destacaram pelo seu desempenho, consolidando os resultados da avaliação.

5 Resultados

Durante a etapa de validação cruzada, a combinação das características *Mel, MFCC* e *Tempogram* apresentou os melhores resultados em cinco dos seis modelos analisados, enquanto o sexto modelo obteve seu desempenho ótimo com a combinação *Chroma CENS, Mel* e *Tempogram*. Consequentemente, essas duas configurações de características foram adotadas como parâmetros de entrada para a avaliação do modelo de *voting*.

Os resultados da Etapa Final de Avaliação (subseção **??**) são apresentados na Tabela 4, que contempla sete modelos: KNN, Árvore de Decisão, MLP, Random Forest, SVM, XGBoost e Voting. Os três que apresentaram a maior acurácia

foram: Voting (83,1%), XGBoost (82,5%) e Random Forest (81,3%). Diante desses resultados, o modelo de Voting foi selecionado como o de melhor desempenho global.

Tabela 4. Acurácias da avaliação final

Modelo	Acurácia
Árvore de Decisão	0,627
KNN	0,771
MLP	0,749
Random Forest	0,813
SVM	0,799
XGBoost	0,825
Voting	**0,832**

É possível identificar que os gêneros com maior dificuldade de classificação foram Bossa Nova, Forró, Pagode, Samba e Sertanejo, cujos desempenhos por modelo estão resumidos na Tabela 5. Em negrito, destacam-se os piores desempenhos por modelo, e são apresentadas a média, a mediana e o desvio padrão de cada gênero. O Sertanejo obteve o pior desempenho em três modelos (MLP, SVM e XGBoost) e registrou as menores média e mediana. O Forró apresentou o maior desvio padrão, indicando a classificação mais instável entre os modelos. Desse modo, o gênero mais difícil de classificar no geral foi o Sertanejo, seguido por Samba, Bossa Nova e Forró.

Tabela 5. Gêneros com menor desempenho geral

Modelo	Bossa Nova	Forró	Pagode	Samba	Sertanejo
Árvore de Decisão	52,3%	**43,7%**	59,7%	45,9%	53,6%
KNN	**57,3%**	76,3%	78,7%	72,0%	68,8%
MLP	74,1%	50,4%	91,2%	54,7%	**48,5%**
Random Forest	72,3%	73,6%	84,8%	**66,9%**	74,1%
SVM	66,4%	87,2%	65,1%	84,8%	**57,1%**
XGBoost	75,2%	84,8%	79,2%	73,1%	**63,2%**
Voting	73,1%	77,9%	88,5%	**59,7%**	86,1%
Média	67,24%	70,56%	78,17%	65,30%	**64,49%**
Mediana	72,3%	76,3%	79,2%	66,9%	**63,2%**
Desvio Padrão	8,93%	**15,6%**	10,91%	12,00%	12,00%

Em contrapartida, os gêneros com melhor desempenho foram Forró Piseiro, Funk e Samba-Enredo. As métricas resumidas de cada um estão descritas na Tabela 6. Nota-se que o Funk foi o gênero com a melhor performance geral, porém, em comparação com os demais, observa-se um equilíbrio mais acentuado entre eles, sendo o Samba-Enredo o que apresentou o menor desvio padrão.

Tabela 6. Gêneros com melhor desempenho geral

Modelo	Forró Piseiro	Funk	Samba-Enredo
Árvore de Decisão	**84,8%**	80,0%	81,6%
KNN	80,3%	86,7%	**96,5%**
MLP	93,9%	**96,00%**	90,1%
Random Forest	90,9%	**94,9%**	93,1%
SVM	89,9%	**95,7%**	93,3%
XGBoost	94,9%	**95,7%**	93,6%
Voting	**94,4%**	92,8%	92,8%
Média	89,87%	**91,69%**	91,57%
Mediana	90,9%	**94,9%**	93,1%
Desvio Padrão	5,08%	5,65%	**4,42%**

A Tabela 7 exibe a acurácia, a precisão média, a revocação média e o *F1-score* médio de cada modelo. Conforme citado, os melhores resultados foram obtidos por Voting, XGBoost e Random Forest, enquanto a Árvore de Decisão foi o de desempenho mais baixo. Nos modelos KNN, MLP e SVM, a precisão média supera a revocação média, o que sugere uma tendência a um maior número de falsos positivos para certos gêneros, indicando uma sensibilidade diferenciada por classe. Os modelos com maior equilíbrio entre precisão e revocação, refletido por um *F1-score* mais elevado, foram também aqueles com maior acurácia global.

Tabela 7. Métricas de desempenho dos modelos

Modelo	Acurácia	Precisão média	Revocação média	F1-score médio
Árvore de Decisão	0,627	0,628	0,627	0,627
KNN	0,771	0,792	0,771	0,771
MLP	0,749	0,783	0,749	0,745
Random Forest	0,813	0,816	0,813	0,814
SVM	0,799	0,833	0,799	0,801
XGBoost	0,825	0,832	0,825	0,824
Voting	0,832	0,831	0,832	0,83

Considerando a acurácia, o melhor modelo foi o Voting. A Tabela 8 apresenta as métricas do Voting para cada gênero musical. Como o número de amostras é balanceado entre os gêneros, a revocação pode ser interpretada como acurácia por classe. Observa-se que o Samba foi o gênero com a menor acurácia, apresentando, portanto, o pior desempenho, enquanto Forró Piseiro, Funk e Samba-Enredo obtiveram as maiores acurácias. A Figura 3 apresenta a matriz de confusão para o modelo.

Tabela 8. Métricas para o modelo *Voting*

Gênero	Precisão	Revocação	F1-Score	Amostras
Bossa Nova	0,749	0,731	0,74	375
Forró	0,8	0,779	0,789	375
Forró Piseiro	0,915	0,944	0,929	375
Funk	0,98	0,928	0,953	375
Pagode	0,798	0,885	0,839	375
Samba	0,744	0,597	0,663	375
Samba-Enredo	0,906	0,928	0,917	375
Sertanejo	0,758	0,861	0,806	375

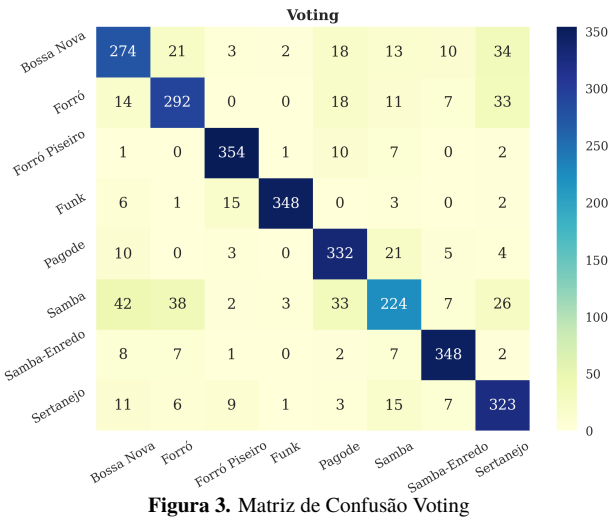

Figura 3. Matriz de Confusão Voting

Destaca-se, por fim, que as características que mais se sobressaíram nos sete modelos da etapa final foram *Mel, MFCC, Tempogram* e *Chroma CENS*. Enquanto *Mel* e *MFCC* capturam informações espectrais relevantes, o *Tempogram* repre-

senta padrões rítmicos e estruturas temporais características de cada gênero, e o *Chroma CENS* codifica a progressão harmônica e as propriedades tonais das músicas.

6 Conclusão e trabalhos futuros

Este trabalho teve como objetivo principal a criação e disponibilização do SONBRA, um dataset público e anotado contendo gêneros musicais populares brasileiros, destinado ao estudo e à classificação automática desses estilos. Para validar a viabilidade e a utilidade da base, diversos modelos de classificação foram aplicados, permitindo avaliar sua adequação para a tarefa proposta.

Os resultados demonstram que a base de dados é efetivamente viável para a classificação de gêneros musicais. Dentre todos os modelos avaliados, o sistema de votação (*Voting*) obteve o melhor desempenho, alcançando 83,2% de acurácia. Os gêneros Forró Piseiro, Funk e Samba-Enredo foram os mais facilmente classificados, enquanto o modelo apresentou maior dificuldade para prever o gênero Samba. Entre os modelos individuais, o XGBoost obteve o melhor resultado, com 82,5% de acurácia. Em contraste, a Árvore de Decisão registrou o pior desempenho, com 62,7% de acurácia. De modo geral, três gêneros destacaram-se pela facilidade de reconhecimento pelos modelos: Forró Piseiro, Funk e Samba-Enredo. Por outro lado, o gênero Sertanejo apresentou a maior dificuldade de classificação, seguido por Samba, Bossa Nova e Forró.

Este estudo contribui significativamente para o campo da classificação de gêneros musicais ao disponibilizar uma base de dados abrangente que contempla os seguintes estilos populares brasileiros: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo. Cada faixa musical na base é descrita por um conjunto abrangente de características extraídas, incluindo *Fourier Tempogram*, *Tempogram*, *Mel*, *MFCC*, *Chroma STFT*, *Chroma CQT*, *Chroma CENS*, *Tonnetz*, *ZCR*, *Spectral Centroid*, *Spectral Roll-Off*, *Spectral Bandwidth* e *RMS*. Esses atributos permitem uma análise detalhada das faixas, abrangendo aspectos espectrais, temporais, harmônicos e de timbre.

Entretanto, algumas limitações devem ser consideradas. A categorização manual das músicas pode conter ruídos de anotação que impactam a precisão dos modelos. Além disso, não foi implementado um controle rigoroso para evitar que faixas de um mesmo artista ou trechos distintos de uma mesma faixa pudessem aparecer simultaneamente nos conjuntos de treino e teste durante a validação cruzada, o que pode ter introduzido vazamento de dados (*data leakage*) e otimismo artificial nos resultados. Outro ponto é a possível alta repetição de artistas em determinados gêneros, o que pode ter facilitado a classificação por meio de assinaturas artísticas específicas, em vez de padrões musicais genuínos do gênero.

Para trabalhos futuros, sugere-se:

- A inclusão de outros gêneros musicais brasileiros para ampliar a abrangência e representatividade da base de dados.
- A aplicação de arquiteturas mais avançadas, como redes neurais convolucionais (CNNs) ou transformadores, para capturar padrões mais complexos e aprimorar a classificação.
- A realização de uma análise de importância de características mais rigorosa, removendo atributos redundantes ou pouco informativos.
- A exploração de técnicas de aprendizado não supervisionado ou semi-supervisionado para identificar de forma mais precisa os limites e a estrutura interna dos gêneros.
- A adoção de métodos mais sofisticados de seleção e validação de modelos, como validação cruzada aninhada ou critérios de penalização baseados em complexidade, complementando as técnicas de *hold-out* utilizadas.

Essas propostas visam aprimorar a robustez e a generalidade do SONBRA, consolidando sua utilidade como recurso aberto para o estudo computacional da música brasileira e de seus diversos estilos.

Declarações complementares

Contribuições dos autores

Claudio Rogerio: Conceitualização, Supervisão, Metodologia, Validação. Anderson Ribeiro: Curadoria de Dados, Metodologia, Investigação, Redação – Rascunho original, Redação – Revisão e Edição. João Santana: Curadoria de Dados, Metodologia, Investigação, Revisão Bibliográfica, Redação – Rascunho original, Redação – Revisão e Edição. Marcos Oliveira: Curadoria de Dados, Investigação, Revisão Bibliográfica, Redação – Rascunho original, Redação – Revisão e Edição. Todos os autores leram e aprovaram o manuscrito final.

Conflitos de interesse

Os autores declaram que não têm nenhum conflito de interesses

Disponibilidade de dados e materiais

Os conjuntos de dados gerados e análisados durante o estudo atual estão disponíveis em `https://doi.org/10.5281/zenodo.17958966`

Referências

Al Mamun, M. A., Kadir, I., Rabby, A. S. A., and Al Azmi, A. (2019). Bangla music genre classification using neural network. In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages 397–403. DOI: 10.1109/SMART46866.2019.9117400.

Alonso, G. (2011). *Cowboys do Asfalto*. Doutorado em história, Programa de Pós-Graduação em História, Universidade Federal Fluminense, Niterói.

Bezerra, J. (2017). *Funk: a batida dos bailes cariocas que contagiou o Brasil*. Panda Books, São Paulo, 1 edition.

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. *Handbook of Computational Statistics*, page 39. DOI: 10.1007/978-3-642-21551-3_33.

Caldas, W. (1987). *O que é a música sertaneja?* Brasiliense, São Paulo.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 785–794, Nova York. Association for Computing Machinery. DOI: 10.1145/2939672.2939785.

Conceição, J., Freitas, R., Gadelha, B., Kienen, J., Anders, S., and Cavalcante, B. (2020). Applying supervised learning techniques to brazilian music genre classification. *2020*

XLVI Latin American Computing Conference (CLEI), pages 102–107. DOI: 10.1109/CLEI52000.2020.00019.

Costa, C., Valle, J., and Koerich, A. (2004). Automatic classification of audio data. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 1, pages 562–567 vol.1. DOI: 10.1109/ICSMC.2004.1398359.

Deferrard, M., Benzi, K., Vandergheynst, P., and Bresson, X. (2017). Fma: A dataset for music analysis.

Dias, I. and Dupan, S. (2022). *O que é o Forró?* Meroveu, Campina Grande, 3 edition.

Farajzadeh, N., Sadeghzadeh, N., and Hashemzadeh, M. (2023). Pmg-net: Persian music genre classification using deep neural networks. *Entertainment Computing*, 44:100518. DOI: 10.1016/j.entcom.2022.100518.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly, Sebastopol, 2 edition.

Harte, C., Sandler, M., and Gasser, M. (2006). Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, page 21–26, New York. Association for Computing Machinery.

Hasib, K. M., Tanzim, A., Shin, J., Faruk, K. O., Mahmud, J. A., and Mridha, M. F. (2022). Bmnet-5: A novel approach of neural network to classify the genre of bengali music based on audio features. *IEEE Access*, 10:108545–108563. DOI: 10.1109/ACCESS.2022.3213818.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2 edition.

Klapuri, A. and Davy, M., editors (2006). *Signal Processing Methods for Music Transcription*. Springer, New York.

Kotsiantis, S., Zaharakis, I., and Pintelas, P. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26:159–190. DOI: 10.1007/s10462-007-9052-3.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8.

Müller, M. (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, Cham, 1 edition.

Müller, M. and Balke, S. (2018). *Short-Time Fourier Transform and Chroma Features*. International Audio Laboratories Erlangen.

Oneto, L. (2018). Model selection and error estimation without the agonizing pain. *WIREs Data Mining and Knowledge Discovery*, 8(4):e1252. DOI: https://doi.org/10.1002/widm.1252.

Quadros Júnior, J. F. S. d. (2019). *Música Brasileira*. PPGCOM/UFMG, Belo Horizonte.

Ramraj, S., Uzir, N., Sunil, R., and Banerjee, S. (2016). Experimenting xgboost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40):651–662.

Ribeiro, A. V., SANTANA, J. M. d. O., and OLIVEIRA, M. A. d. (2025). Sonbra dataset. `https://doi.org/10.5281/zenodo.17958966`. Acessado em: 17 de dezembro de 2025.

Sammut, C. and Webb, G. I., editors (2011). *Encyclopedia of Machine Learning*. Springer, Boston.

Scaringella, N., Zoia, G., and Mlynek, D. (2006). Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23(2):133–141. DOI: 10.1109/MSP.2006.1598089.

Sharma, G., Umapathy, K., and Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, 161:107201. DOI: 10.1016/j.apacoust.2019.107201.

Shirol, S. and Kathiresan, R. S. (2023). A comprehensive survey of music genre classification using audio files. *International Journal of Enhanced Research in Science, Technology & Engineering*, 12:183–192.

Silla, Jr., C. N., Kaestner, C. A. A., and Koerich, A. L. (2007). Automatic music genre classification using ensemble of classifiers. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 1687–1692. DOI: 10.1109/ICSMC.2007.4414136.

Silla, Jr., C. N., Koerich, A. L., and Kaestner, C. A. A. (2018). The latin music database. In *Proceedings of the 9th International Conference on Music Information Retrieval*, pages 451–456, Philadelphia. ISMIR. DOI: 10.5281/zenodo.1416282.

Silva, D. and Gomes, C. (2022). Modelo de aprendizado de máquina para classificação de gêneros musicais populares da região amazônica legal internacional. *Revista Eletrônica de Iniciação Científica em Computação*, 20(4).

Suthaharan, S. (2016). *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Springer US.

Tomar, S. (2006). Converting video formats with ffmpeg. *Linux Journal*, 2006(146):10.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302. DOI: 10.1109/TSA.2002.800560.

Viana, L. R. (2010). O funk no brasil: Música desintermediada na cibercultura. In *Revista Sonora, Unicamp*, volume 3.

Yu, T. and Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *ArXiv*, abs/2003.05689.