UNIVERSIDADE FEDERAL DO AMAPÁ
DEPARTAMENTO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

CIÊNCIA DA COMPUTAÇÃO

SONBRA: UMA BASE DE DADOS PARA A CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS BRASILEIROS

ANDERSON VINÍCIUS RIBEIRO AMORIM
JOÃO MARCOS DE OLIVEIRA SANTANA
MARCOS ABREU DE OLIVEIRA

Orientador(a): Me. Claudio Rogerio Gomes da Silva

MACAPÁ
ABRIL DE 2025

ANDERSON VINÍCIUS RIBEIRO AMORIM
JOÃO MARCOS DE OLIVEIRA SANTANA
MARCOS ABREU DE OLIVEIRA

SONBRA: UMA BASE DE DADOS PARA A CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS BRASILEIROS

Trabalho de Conclusão de Curso apresentado à Universidade Federal do Amapá como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador(a): Me. Claudio Rogerio Gomes da Silva

MACAPÁ
ABRIL DE 2025

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Central/UNIFAP-Macapá-AP
Elaborado por Cristina Fernandes – CRB-2 / 1569

A524s Amorim, Anderson Vinícius Ribeiro.
SONBRA: uma base de dados para a classificação automática de gêneros musicais brasileiros / Anderson Vinícius Ribeiro Amorim, João Marcos de Oliveira Santana, Marcos Abreu de Oliveira. - Macapá, 2025.
1 recurso eletrônico. 55 folhas.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal do Amapá, Coordenação do Curso de Ciência da Computação, Macapá, 2025.
Orientador: Claudio Rogerio Gomes da Silva.
Coorientador: .

Modo de acesso: World Wide Web.
Formato de arquivo: Portable Document Format (PDF).

1. Base de dados. 2. Aprendizado de máquina. 3. Recuperação de informação musical. I. Silva, Claudio Rogerio Gomes da orientador. II. Universidade Federal do Amapá. III. Título.

CDD 23. ed. – 004

AMORIM, Anderson Vinícius Ribeiro; SANTANA, João Marcos de Oliveira; OLIVEIRA, Marcos Abreu de. **SONBRA**: uma base de dados para a classificação automática de gêneros musicais brasileiros. Orientador: Claudio Rogerio Gomes da Silva. 2025. 55 f. Trabalho de Conclusão de Curso (Graduação) - Ciência da Computação. Universidade Federal do Amapá, Macapá, 2025.

UNIVERSIDADE FEDERAL DO AMAPÁ
DEPARTAMENTO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
COORDENAÇÃO DO CURSO DE CIÊNCIA DA COMPUTAÇÃO

ATA DE DEFESA DE TCC

Realizou-se, no dia 10 de abril de 2025, às 16h, na Universidade Federal do Amapá, a defesa do TCC intitulado: **“SONBRA: UMA BASE DE DADOS PARA A CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS BRASILEIROS”**, dos discentes Anderson Vinícius Ribeiro Amorim matrícula 201612200040, João Marcos de Oliveira Santana matrícula 2017002935, Marcos Abreu de Oliveira matrícula 2017011237. A Banca Examinadora foi composta pelo Prof. Me. Cláudio Rogério Gomes da Silva, orientador; Prof. Me. Marco Antonio Leal da Silva, presidente da banca, e Prof. Me. Thiago Pinheiro do Nascimento, examinadores. Concluída a defesa, foram realizadas as arguições e comentários. Em seguida, procedeu-se o julgamento pelos membros da Banca Examinadora, tendo o trabalho sido APROVADO.

E, para constar, eu, Prof. Me. Marco Antonio Leal da Silva, presidente da Banca Examinadora, lavrei a presente ata que, após lida e achada conforme, foi assinada por mim e demais membros da Banca Examinadora.

Macapá-AP, 10 de abril de 2025.

gov.br Documento assinado digitalmente
CLAUDIO ROGERIO GOMES DA SILVA
Data: 10/04/2025 18:00:47-0300
Verifique em https://validar.iti.gov.br

_____
Prof. Me. Cláudio Rogério Gomes da Silva
Orientador do Projeto de TCC (UNIFAP)

gov.br Documento assinado digitalmente
MARCO ANTONIO LEAL DA SILVA
Data: 14/04/2025 15:00:33-0300
Verifique em https://validar.iti.gov.br

_____
Prof. Me. Marco Antonio Leal da Silva
Examinador (UNIFAP)

gov.br Documento assinado digitalmente
THIAGO PINHEIRO DO NASCIMENTO
Data: 14/04/2025 21:54:39-0300
Verifique em https://validar.iti.gov.br

_____
Prof. Me. Thiago Pinheiro do Nascimento
Examinador (UNIFAP)

UNIVERSIDADE FEDERAL DO AMAPÁ
DEPARTAMENTO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
COORDENACÃO DO CURSO DE CIÊNCIA DA COMPUTACÃO

FICHA DE AVALIAÇÃO DE TCC

Discentes: **Anderson Vinícius Ribeiro Amorim,** matrícula 201612200040
João Marcos de Oliveira Santana, matrícula 2017002935
Marcos Abreu de Oliveira, matrícula 2017011237

Título: **"SONBRA: UMA BASE DE DADOS PARA A CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS BRASILEIROS".**

Orientador (a): Prof. Me. Cláudio Rogério Gomes da Silva
Examinador(a) 1 da Banca Avaliadora: Prof. Me. Marco Antonio Leal da Silva
Examinador(a) 2 da Banca Avaliadora: Prof. Me. Thiago Pinheiro do Nascimento

	Avaliador 1	Avaliador 2
Trabalho Escrito (0 a 5)	10	10
Apresentação Oral (0 a 5)	10	10
Nota Final	NF1 = 10	NF2 = 10

No item TRABALHO ESCRITO, a banca examinadora deverá avaliar: organização sequencial, argumentação, profundidade do tema, relevância e contribuição acadêmica da pesquisa, correção gramatical, clareza, apresentação estética, adequação aos aspectos formais às normas da ABNT.

No item APRESENTAÇÃO ORAL, a banca examinadora deverá avaliar: domínio do conteúdo, organização da apresentação, habilidades de comunicação e expressão, capacidade de argumentação, uso dos recursos audiovisuais, correção gramatical e apresentação estética do trabalho.

MÉDIA FINAL: A média final será calculada pela soma das duas notas finais (NF1 e NF2) dividida por dois.

$$\text{MÉDIA FINAL} = \frac{NF1 + NF2}{2} = \frac{10}{2}.$$

Documento assinado digitalmente
gov.br CLAUDIO ROGERIO GOMES DA SILVA
Data: 10/04/2025 18:00:47-0300
Verifique em https://validar.iti.gov.br

Macapá-Ap., 10 de abril de 2025.

Prof. Me. Cláudio Rogério Gomes da Silva
Orientador do Projeto de TCC (UNIFAP)

Documento assinado digitalmente
gov.br MARCO ANTONIO LEAL DA SILVA
Data: 14/04/2025 15:02:19-0300
Verifique em https://validar.iti.gov.br

Prof. Me. Marco Antonio Leal da Silva
Examinador (UNIFAP)

Documento assinado digitalmente
gov.br THIAGO PINHEIRO DO NASCIMENTO
Data: 15/04/2025 10:17:00-0300
Verifique em https://validar.iti.gov.br

Prof. Me. Thiago Pinheiro do Nascimento
Examinador (UNIFAP)

Agradecimentos

Gostaria de expressar minha profunda gratidão a todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

Agradeço ao meu orientador, Prof. Me. Cláudio Rogério, pela oportunidade, discussões enriquecedoras e valiosas orientações ao longo desta jornada. Sua expertise foi fundamental para a conclusão desta pesquisa.

À minha família, em especial à minha mãe, Antônia, e à minha irmã, Ataíres, pelo amor incondicional, apoio nos momentos desafiadores e por sempre acreditarem em mim. Vocês são minha base e minha maior motivação.

Aos meus colegas de curso João Santana e Marcos Oliveira, que compartilharam os desafios, conquistas e aprendizados deste trabalho.

Aos meus amigos de longa data, que estiveram ao meu lado nos momentos de estresse e me apoiaram incondicionalmente, apesar da minha ausência e as vezes falta de paciência.

Aos professores do Colegiado de Ciência da Computação, que, com seu conhecimento e comprometimento, enriqueceram minha formação acadêmica e profissional.

Por fim, a todos que, de alguma forma, contribuíram para este trabalho e para minha trajetória universitária. Cada palavra de incentivo, cada ajuda e cada lição foram essenciais para que eu chegasse até aqui.

– Anderson Ribeiro

Agradecimentos

Aos meus colegas, expresso minha gratidão pelo empenho e dedicação ao longo desta jornada. Ao meu orientador, agradeço imensamente pela paciência e pelo apoio diante dos meus questionamentos infindáveis, sempre contribuindo para o meu crescimento acadêmico.

Aos meus pais, à minha irmã e aos meus padrinhos, sou profundamente grato por todo suporte, pelos conselhos, pela presença constante e, principalmente, por jamais permitirem que eu desistisse. Aos meus familiares, que sempre estiveram presentes na minha formação, deixo meu reconhecimento e carinho.

Aos professores que marcaram essa trajetória com seus ensinamentos e orientações, meu sincero agradecimento. Aos amigos que compartilharam comigo não apenas os desafios, mas também os momentos de descontração, obrigado por ouvirem meus devaneios e reclamações sem jamais me abandonarem.

Por fim, dedico este trabalho aos meus avôs, Amiraldo Braga e Eulálio Oliveira Filho, que partiram antes que eu pudesse concluir esta etapa, mas que sempre demonstraram orgulho pela minha caminhada. A eles, minha eterna gratidão.

– João Santana

Agradecimentos

Agradeço a todos que, de alguma forma, contribuíram para que isto se tornasse possível.

– Marcos Oliveira

"E de guerra em paz
De paz em guerra
Todo o povo dessa terra
Quando pode cantar
Canta de dor"
– Canto das Três Raças, Clara Nunes

Resumo

Considerando as bases de dados musicais mais relevantes disponíveis para estudos de músicas e gêneros musicais na área de Recuperação de Informação Musical, observou-se a escassez de bases que contenham gêneros musicais brasileiros. Sendo assim, este trabalho teve como objetivo criar a SONBRA, uma base de dados estratificada com os seguintes gêneros musicais: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo. Com o intuito de aumentar o número de amostras para cada gênero, dividiu-se cada faixa em cinco fragmentos e, posteriormente, foram extraídas, de cada um, as seguintes características: *Fourier Tempograma, Tempograma, Mel, MFCC, Chroma STFT, Chroma CQT, Chroma CENS, Tonnetz, ZCR, Spectral Centroid, Spectral Roll Off, Spectral Bandwith* e *RMS*. Para avaliar se a base de dados está qualificada para uso na classificação de gêneros musicais, optou-se por avaliá-la com modelos de classificação: KNN, MLP, XGBoost, *Decision Tree*, *Random Forest*, SVM e uma proposta de *ensemble*, mais especificamente um sistema de *Voting*. O melhor resultado obtido foi com o *Voting*, que obteve 83,2% de acurácia. Dentre os modelos individuais, o XGBoost obteve o melhor desempenho com 82,5% de acurácia e o modelo *Decision Tree* obteve a menor performance com 62,7% de acurácia. No geral, os gêneros que os modelos conseguiram classificar com maior facilidade foram: Forró Piseiro, Funk e Samba-Enredo, destacando-se o Funk. Em contrapartida, o gênero que apresentou maior dificuldade de ser classificado foi o Sertanejo. Quanto às características extraídas, os modelos foram mais eficientes com combinações entre *Mel*, *MFCC*, *Tempograma* e *Chroma CENS*. Portanto, mostrou-se que a base de dados e as características extraídas podem ser usadas para trabalhos de classificação de gêneros.

Palavras-chave: Base de dados. Aprendizado de máquina. Recuperação de informação musical. Classificação de gêneros musicais.

Abstract

After reviewing the most relevant musical databases available for studying music and musical genres in the field of Music Information Retrieval, a scarcity of databases containing Brazilian musical genres was observed. Therefore, this study aimed to create the SONBRA, a stratified database with the following musical genres: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo and Sertanejo. In order to increase the number of samples for each genre, each track was divided into five fragments, and the following features were extracted from each fragment: *Fourier Tempogram, Tempogram, Mel, MFCC, Chroma STFT, Chroma CQT, Chroma CENS, Tonnetz, ZCR, Spectral Centroid, Spectral Roll Off, Spectral Bandwidth* and *RMS*. To assess whether the database is suitable for genre classification, it was evaluated using classification models: KNN, MLP, XGBoost, Decision Tree, Random Forest, SVM, and an ensemble approach, specifically a Voting system. The best result was achieved with the Voting model, which obtained 83.2% accuracy. Among the individual models, XGBoost performed the best with 82.5% accuracy, while the Decision Tree model had the lowest performance, achieving 62.7% accuracy. Overall, the genres that the models classified most easily were Forró Piseiro, Funk, and Samba-Enredo, with Funk standing out. In contrast, the genre that exhibited the greatest difficulty in classification was Sertanejo. Regarding the extracted features, the models performed best with combinations of *Mel, MFCC, Tempogram*, and *Chroma CENS*. Therefore, it was demonstrated that the database and the extracted features can be used for genre classification tasks.

Keywords: Dataset. Machine learning. Music information retrieval. Music genre classification.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

Além de um recurso de manipulação de sons, ruídos e silêncios organizados em uma sequência lógica, a música é uma forma de expressão, comunicação, gratificação, mobilização e auto-realização; repertoriada em um contexto social, cultural e ideológico; definida por um tempo e uma época (ZAMPRONHA, 2002).

A música pode ser organizada através de elementos. A melodia refere-se à sequência de notas de diferentes sons. A harmonia ocorre quando duas ou mais notas diferentes são tocadas simultaneamente, o que constitui um acorde. O ritmo é um modo de organizar o tempo e a acentuação de sons. O timbre é o som particular que cada instrumento emite. A forma se refere à estrutura ou organização das seções de uma composição. Finalmente, a textura musical é para expressar sonoridades, que podem produzir um efeito denso ou penetrante e agressivo (BENNET, 1986).

Uma maneira comum de categorizar músicas é através de gêneros, que podem ser definidos com base na manipulação de seus elementos musicais, bem como em aspectos culturais e históricos, como local de origem, época, tradições e influências culturais (JANNOTI JÚNIOR, 2020). No entanto, a definição de um gênero musical pode variar dependendo da perspectiva, havendo sobreposição ou interseção entre diferentes gêneros.

No Brasil, a música é um importante elemento cultural que reflete a diversidade e riqueza da história do país. Segundo (QUADROS JÚNIOR, 2019), a partir da mistura de diferentes influências, como a música indígena, africana e europeia, desenvolveram-se vários estilos musicais que se tornaram símbolos da identidade do país.

Muitas aplicações e serviços atualmente utilizam dados e características musicais para classificar, organizar e sugerir novas músicas aos usuários. Nesse contexto, os gêneros musicais servem como uma ferramenta essencial para classificar músicas e personalizar recomendações.

Encontramos na literatura muitos estudos relacionados à classificação automática de gêneros musicais (STURM, 2012; SHIROL; KATHIRESAN, 2023). No entanto, a maioria desses estudos foca na classificação de gêneros populares internacionais, como o rock e o pop, e a classificação automática de ritmos brasileiros ainda é escassa e enfrenta desafios relacionados à limitação de bases de dados com uma quantidade equilibrada de músicas regionais.

Este trabalho apresenta uma base de dados que poderá ser usada para classificar automaticamente os gêneros populares brasileiros de Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo. A base de dados contém 250 músicas

de cada gênero, extraídas da plataforma de *streaming* YouTube [1].

Este capítulo apresenta aspectos introdutórios relativos ao desenvolvimento da presente pesquisa, tais como problemática, justificativa e objetivos. Finalmente, a organização do presente trabalho é discutida e apresentada.

1.1 Problema e Justificativa

A classificação automática de gêneros musicais tem sido amplamente estudada, com muitos sistemas e modelos que foram desenvolvidos a partir de bases de dados com gêneros musicais internacionais (STURM, 2012; SHIROL; KATHIRESAN, 2023). No entanto, quando se trata dos gêneros musicais brasileiros, como Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo, há uma lacuna na disponibilidade de bases de dados robustas e balanceadas. A maioria dos métodos existentes para classificação de gêneros musicais internacionais não considera as especificidades rítmicas e culturais das músicas brasileiras, resultando em classificações imprecisas ou pouco representativas quando se trata desses gêneros.

Além disso, os poucos estudos que se debruçam sobre a classificação de músicas brasileiras enfrentam problemas relacionados à escassez de dados adequados. As bases de dados disponíveis são geralmente limitadas, mal balanceadas ou não contêm uma quantidade suficiente de amostras representativas de cada gênero, dificultando o treinamento de modelos de aprendizado de máquina eficientes (CONCEIÇÃO *et al.*, 2020).

Portanto, o problema central que este trabalho busca resolver é a falta de uma base de dados ampla e bem estruturada que contemple adequadamente a diversidade e a complexidade dos gêneros musicais populares brasileiros. A ausência dessa base impede o desenvolvimento e a utilização eficaz de modelos de classificação automática, o que limita o reconhecimento e a valorização da riqueza musical do Brasil em aplicações tecnológicas.

1.2 Objetivos

Os objetivos do presente trabalho encontram-se estruturados em objetivos geral e objetivos específicos, os quais são detalhados nas subseções, por conseguinte.

[1] https://www.youtube.com

1.2.1 Objetivo Geral

Desenvolver uma base de dados estruturada de músicas de gêneros populares brasileiros, a partir de amostras extraídas de plataformas digitais, com o intuito de facilitar e promover o estudo e a classificação automática de gêneros musicais através de métodos computacionais.

1.2.2 Objetivos Específicos

- Coletar e organizar amostras musicais de gêneros brasileiros de Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo, a partir de plataformas de *streaming*;
- Aplicar técnicas de extração de características musicais para a descrição quantitativa de elementos sonoros das amostras;
- Garantir que a base de dados resultante seja balanceada e representativa dos diferentes gêneros, de forma a possibilitar o uso futuro em sistemas de classificação automática de gêneros musicais;
- Disponibilizar a base de dados finalizada em um formato acessível para a comunidade acadêmica e para desenvolvedores de sistemas de reconhecimento musical.

2 Referencial Teórico

Neste capítulo são apresentadas as abordagens utilizadas na classificação de gêneros musicais. A seção 2.1 aborda conceitos e técnicas fundamentais da recuperação de informação musical. Na seção 2.2 é descrito o processo de construção de um classificador de gêneros musicais, sendo dividido em quatro etapas principais: a subseção 2.2.1 trata da construção da base de dados, detalhando as estratégias de coleta, padronização e extração de características musicais, enquanto a subseção 2.2.2 apresenta a construção do modelo de aprendizado de máquina, incluindo a escolha dos algoritmos, treinamento e avaliação do desempenho do classificador. Por fim, na seção 2.3 são discutidos trabalhos relacionados ao tema, destacando pesquisas e avanços na área.

2.1 Recuperação de informação musical

Segundo Schedl (2008), a recuperação de informação musical (em inglês: Music Information Retrieval - MIR) é um campo de pesquisa que estuda a manipulação de dados musicais através da extração, análise e uso de informações que podem estar representadas em diferentes formatos, como sinais de áudio e letras de músicas. A MIR é essencial para muitas aplicações, incluindo indexação, organização de conteúdo e sistemas de recomendação de música, bastante utilizados em serviços de *streaming* de músicas como Spotify[1] e Deezer[2].

Devido à sua heterogeneidade, o campo de pesquisa em MIR é extremamente amplo. No entanto, quando o foco está na classificação automática de músicas por meio de aprendizado de máquina, o subcampo de extração de características dos sinais de áudio se torna um elemento central. Isso ocorre porque os atributos extraídos dos sinais permitem estabelecer relações de similaridade entre diferentes instrumentos musicais e vozes, utilizando a forma de onda de um sinal de áudio (SCHEDL, 2008).

Um sinal de áudio é representado em termos de amplitude ao longo do tempo. Para analisar suas componentes espectrais, é necessário convertê-lo para o domínio tempo-frequência. Esse processo é realizado por meio da Transformada Rápida de Fourier de Curto Tempo (em inglês: *Short-Time Fourier Transform* – STFT), que segmenta o sinal em pequenas janelas de tempo. Para cada intervalo, a STFT calcula o espectro de frequências correspondente e computa um coeficiente para cada uma, o que resulta em uma representação bidimensional tempo-frequência com seus respectivos coeficientes. Isso permite realizar a análise da variação espectral ao longo do tempo. Ao calcular o

[1] https://www.spotify.com
[2] https://www.deezer.com

quadrado da magnitude da STFT, cria-se o espectrograma, que é uma representação bidimensional tempo-frequência. Quando usado como imagem, a magnitude é retratada por cores. Quanto mais intensa uma cor, maior a magnitude (MÜLLER, 2015).

Quando uma STFT é realizada e seu espectrograma calculado, isso é feito de forma linear, ou seja, todas as frequências estão igualmente espaçadas pelo espectro. Porém, quando se trata da percepção sonora do som pelos humanos, a percepção é diferente: os sons de frequência mais baixas são mais facilmente diferenciados do que frequências mais altas. Isto é, um som com frequência de 50Hz é, para os ouvidos humanos, bem diferente de um som com frequência 150Hz. Isso não é verdade para um som com 20.000Hz e outro de 20.0150Hz. Para isso, foi proposta a escala mel, que permite uma representação mais próxima à percepção humana ao som. O *mel-spectrograma* utiliza essa relação, dada pela equação 1, para representar o espectro do sinal. O *Mel-Frequency Cepstrum Coeficient* (MFCC) utiliza o mesmo conceito da escala mel, que permite uma representação mais próxima à percepção humana ao som. Porém, usa os coeficientes cepstrais, que são obtidos, de forma simplificada, aplicando uma Transformada Discreta do Cosseno sobre o espectro mapeado para a escala mel (KLAPURI; DAVY, 2006).

$$mel(f) = 2595 \log_{10} \left[\frac{f}{700} + 1 \right] \tag{1}$$

A música é organizada em torno de unidades temporais chamadas batidas (pulsos). A duração de cada pulso é dada de acordo com o *tempo* (andamento), medido em Batidas por Minuto (BPM). Um tempograma é uma representação em que em uma unidade de tempo da música são destacados os *tempos*, em BPM, mais relevantes. Tempogramas podem ser extraídos com diferentes técnicas, dentre elas: Fourier Tempograma e Tempograma Autocorrelacionado. Para essa extração, primeiramente é necessário detectar os momentos que as notas são inicialmente tocadas. Para isso, faz-se uma função chamada *novelty*, em que os picos demonstram locais que tiveram grandes mudanças na música, podendo ser um grande indicativo de início de uma batida. Exposto isso, a primeira técnica, Fourier Tempograma, utiliza STFT nesta função *novelty* para criar um tempograma. Nesse processo, a função *novelty* é comparada com sinusoides que representam cada *tempo* em uma determinada janela (segundos). Um coeficiente é dado de acordo com a semelhança entre a sinusoide e a região analisada. Quanto mais parecidas, maior é o valor atribuído a esse coeficiente que, numa imagem de tempograma, é representado com uma cor de maior intensidade, semelhante ao espectrograma. Já o Tempograma Autocorrelacionado usa a autocorrelação, em que compara-se uma função com si mesma, mas deslocada no tempo. Da mesma forma que na outra técnica, essa análise é realizada em janelas (segundos). Em cada deslocamento, em segundos, compara-se a função *novelty* deslocada com sua versão não deslocada, e

quanto maior a semelhança, maior o coeficiente de correlação. Esses coeficientes são usados para destacar no tempograma os *tempos* mais relevantes daquele instante. Nota-se que o Fourier Tempograma enfatiza o que se chama de harmônicos de *tempo*, enquanto a segunda técnica enfatiza os sub-harmônicos. A partir daqui, o termo Tempograma refere-se à técnica Tempograma Autocorrelacionado (MÜLLER, 2015).

Outra técnica de conversão para o domínio do tempo-frequência é a *Constant Q Transform* (CQT), que organiza as frequências em uma escala logarítmica, diferente da STFT que as organiza linearmente. Essa escala permite um alinhamento com as notas musicais, já que as frequências fundamentais são espaçadas semelhantemente. Isso é possível ao garantir que todas as bandas tenham o mesmo fator Q (relação entre frequência central e largura de banda - definida como faixa de frequências que cada filtro abrange). Portanto, a CQT oferece melhor resolução para baixas frequências e melhor precisão no tempo para altas frequências, semelhante ao que os ouvidos percebem (BROWN, 1991; SCHÖRKHUBER; KLAPURI, 2010).

O *Chroma* é um conceito que deriva do termo *chromatic*, originário do grego *chroma*, que significa "cor". Em contexto musical, esse termo está intimamente associado às doze classes de altura presentes na escala temperada de doze tons: C, C#, D, D#, E, F, F#, G, G#, A, A# e B. Cada classe de altura representa um conjunto de notas que, embora possam pertencer a oitavas diferentes, compartilham a mesma identidade sonora. Assim, notas como C2 e C5, por exemplo, possuem o mesmo valor de *chroma*, pois ambas correspondem à classe de altura C, evidenciando uma similaridade perceptual, mesmo que apresentem diferentes frequências fundamentais devido à variação de oitava. Essa propriedade de invariância em relação à oitava permite que o valor de *chroma* seja interpretado como uma "cor sonora", realçando a ideia de que notas pertencentes a diferentes classes de altura são percebidas como distintas em termos de timbre e qualidade sonora.

Dessa maneira, um cromagrama representa, em cada unidade de tempo de uma música, a distribuição de energia das músicas nas doze classes de altura, permitindo uma análise da progressão harmônica e melódica da música. Para extrair essas informações, diversas técnicas podem ser aplicadas: a primeira é o que se chama de *Chroma* STFT que, como o nome diz, utiliza STFT para transformar o sinal da relação amplitude-tempo para frequência-tempo. Assim, destacam-se as frequências com maior energia em um determinado intervalo. Em seguida, as frequências são convertidas para se adequarem de acordo com as classes de altura, resultando em uma representação em que o eixo vertical mostra as classes de altura e o eixo horizontal o tempo e, usando o mesmo conceito do espectrograma, uma cor mais forte implica que aquela nota teve maior energia. A segunda forma é o *Chroma* CQT, que usa a CQT para converter o sinal para o domínio tempo-frequência. Por fim, o *Chroma Energy distribution Energy Statistics*,

referido neste trabalho como *Chroma* CENS, é obtido com a aplicação das seguintes etapas adicionais: na primeira, é realizada a normalização e quantização dos vetores de chroma, levando em consideração uma escala logarítmica de percepção humana do som. Após isso, o cromagrama passa por um processo de suavização, que tem como objetivo diminuir a sensibilidade a pequenas mudanças sonoras, deixando visíveis somente as mais relevantes, diminuindo o ruído geral (MÜLLER; BALKE, 2018; MÜLLER, 2015).

O *Spectral Centroid* indica em que região do espectro as frequências de um sinal estão mais concentradas. É possível obtê-la ao considerar o espectro como uma distribuição, sendo as frequências os valores e as amplitudes normalizadas são as probabilidades de serem observadas (SHARMA; UMAPATHY; KRISHNAN, 2020). Portanto, essa característica nada mais é que uma média ponderada do espectro analisado, em que evidencia qual a tendência do sinal: se a energia está mais concentrada nas frequências baixas ou altas. Disso vem a noção de brilho, pois quanto maior o *spectral centroid*, mais brilho o espectro contém, isto é, a energia do sinal está mais distribuída nas frequências mais altas.

Por sua vez, o *Spectral Bandwith* é definido por como uma medida de dispersão das frequências em torno do *Spectral Centroid*, logo define a largura de banda (do inglês: *bandwith*) do sinal de áudio no espectro definido (SHARMA; UMAPATHY; KRISHNAN, 2020). Por fim, o *Spectral Roll Off* retorna a frequência em que um percentual definido, geralmente 85% ou 95%, de energia está abaixo dela. Se for um valor baixo, isso significa que a maior parte da energia do som está concentrada nas frequências baixas (KLAPURI; DAVY, 2006).

O *Zero Crossing Rate* (ZCR) define quantas vezes o sinal de áudio mudou de positivo para negativo e vice-versa em um determinado *frame*. Pode ser usado para verificar os níveis de ruído de um sinal. Essa interpretação diz que quanto mais ruidoso um sinal é, menor é o valor de ZCR (SHARMA; UMAPATHY; KRISHNAN, 2020). Ao passo que o *Root Mean Square* (RMS) é usado para indicar o volume do sinal de áudio (KLAPURI; DAVY, 2006).

O *Tonnetz* é uma representação das relações harmônicas entre as notas musicais, mais especificamente das quintas perfeitas, terças menores e terças maiores. Computacionalmente, pode-se usar o *tonnetz* para verificar mudanças de acordes e registrar progressões harmônicas dos sinais de áudio, usando um vetor *Tonal Centroid* de seis dimensões, em que se indica a proximidade das notas e acordes com cada relação citada anteriormente (HARTE; SANDLER; GASSER, 2006).

2.2 Construção de classificador de gêneros musicais

A classificação de gêneros musicais é um dos principais desafios da área de MIR. O avanço das técnicas de aprendizado de máquina permitiu o desenvolvimento de modelos capazes de categorizar músicas de forma precisa e eficiente; entretanto, a qualidade da base de dados é determinante para a performance desses modelos.

2.2.1 Construção da base de dados

Sturm (2012) destaca que mais de 79% das pesquisas sobre reconhecimento de gênero musical utilizam dados ou recursos de áudio. A construção de uma base de dados eficiente a partir desse tipo de informação envolve diversas etapas, incluindo a coleta, padronização dos arquivos de áudio e a extração de características relevantes para a modelagem dos dados.

Além de bases de dados privadas, existem várias bases de dados disponíveis publicamente que atendem a diversos requisitos para construção de um modelo, como GTZAN (TZANETAKIS; COOK, 2002) e Latin Music Database (SILLA JR.; KOERICH; KAESTNER, 2018). Esses conjuntos oferecem uma ampla variedade de amostras musicais e têm sido frequentemente empregados em pesquisas acadêmicas na área (STURM, 2012). No entanto, essas bases podem não abranger todos os gêneros musicais de interesse, tornando necessária a criação de uma base de dados personalizada.

Para a construção de uma base própria, é essencial selecionar e coletar músicas de fontes diversas, como plataformas de *streaming* de músicas ou de vídeos. A curadoria do conjunto de dados deve garantir uma distribuição equilibrada entre os gêneros, evitando viés nos modelos e assegurando a generalização dos resultados.

As músicas coletadas podem estar disponíveis em diferentes formatos e qualidades, tornando necessária sua conversão para um padrão unificado. A escolha desse padrão depende dos requisitos e objetivos do projeto, considerando fatores como compatibilidade com ferramentas de processamento, preservação da qualidade do áudio e otimização do armazenamento e processamento dos dados.

Após a coleta das faixas musicais, é feita a extração de características dos sinais de áudio. Tempograma, MFCC e recursos baseados em *Chroma* são algumas das características amplamente analisadas para a diferenciação entre gêneros musicais (SHIROL; KATHIRESAN, 2023). Tais recursos podem ser extraídos com ferramentas especializadas, como Essentia[3] e Librosa[4].

[3] https://essentia.upf.edu
[4] https://librosa.org

2.2.2 Construção do modelo de aprendizado de máquina

A construção de um modelo de aprendizado de máquina eficiente para tarefas de classificação requer a preparação adequada da base de dados. Isso envolve uma etapa de pré-processamento dos dados, que inclui a remoção ou tratamento de valores faltantes e eliminação de ruídos (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006). A escolha do algoritmo de aprendizado é uma etapa crítica na construção do modelo, pois a seleção da melhor abordagem depende dos requerimentos da tarefa, das características da base de dados e dos níveis de predição desejados (SHIROL; KATHIRESAN, 2023).

A aprendizagem supervisionada é uma categoria de algoritmos que consiste no treinamento do modelo em uma base de dados que contém os rótulos das classes de cada instância. Com base nesses dados rotulados, o modelo aprende a criar uma generalização, permitindo que consiga classificar novas instâncias desconhecidas (SUTHAHARAN, 2016; SCARINGELLA; ZOIA; MLYNEK, 2006).

Combinar diferentes algoritmos em um único modelo pode maximizar a capacidade preditiva de um sistema. Essa técnica, conhecida como *ensemble*, é baseada na previsão feita a partir do resultado de diferentes modelos. Dentre as abordagens de *ensemble*, as mais comuns são *bagging* e *boosting*. O objetivo do *bagging* é reduzir a variância nas funções de decisão. Por meio da reamostragem com reposição (*bootstrap*), são selecionadas amostras dos dados iniciais, podendo conter repetição. Em seguida, combina paralelamente os resultados de diferentes algoritmos para obter um melhor desempenho nas previsões. Já o *boosting* realiza uma combinação sequencial, em que cada etapa usa o resultado da anterior para aprimorar o seu treinamento, com o propósito de minimizar o viés (*bias*) da função (BÜHLMANN, 2012). Como uma evolução ao *boosting*, o *gradient boosting* aplica uma função de custo diferenciável ao resultado das previsões entre cada etapa. O objetivo é minimizar essa função de custo, tornando o modelo mais preciso (RAMRAJ *et al.*, 2016). Essas técnicas visam aumentar a precisão e robustez do modelo final .

A classificação de gêneros musicais por meio de aprendizado supervisionado pode ser realizada utilizando diversas técnicas, variando desde métodos tradicionais até arquiteturas mais complexas de aprendizado profundo. Dentre esses modelos estão: Árvores de Decisão (em inglês: *Decision Tree*), *Random Forests*, *eXtreme Gradient boosting* (XGBoost), K-Vizinhos mais Próximos (em inglês: *K-Nearest Neighbors* - KNN), Máquinas de Vetores de Suporte (em inglês: *Support-Vector Machine* - SVM) e *Multilayer Perceptron* (MLP).

Árvores de decisão são algoritmos hierárquicos amplamente utilizados para tarefas de classificação, capazes de lidar com dados complexos. Sua estrutura é baseada em uma árvore binária, onde as características são divididas em subdomínios (também

chamados de nós). Cada nó final, chamado de folha, representa uma classe à qual a instância pode pertencer. A classificação é feita com base na ordenação dos valores das características, para encontrar o particionamento que define o melhor caminho até uma folha, de acordo com critérios de decisão pré-definidos (SUTHAHARAN, 2016; KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

Random Forests utilizam a técnica de *ensemble* para otimizar o desempenho das árvores de decisão. Várias árvores são criadas a partir do *bootstrap* dos dados iniciais. Em seguida, uma função de decisão é aplicada para determinar a classificação final, combinando o resultado das árvores (SUTHAHARAN, 2016).

O XGBoost foi desenvolvido para aprimorar o desempenho do *gradient boosting*, tornando-o mais rápido e eficiente computacionalmente. O modelo alcança esse objetivo através do armazenamento de dados em memória, que permite reutilizar cálculos já feitos 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 comparado a métodos similares, especialmente em bases de dados muito grandes (CHEN; GUESTRIN, 2016).

KNN é um modelo não paramétrico, ou seja, não possui uma quantidade de parâmetros pré-definidos. O algoritmo seleciona os k vizinhos mais próximos de uma instância com base em uma medida de distância. Em seguida, a classe mais frequente entre as escolhidas é atribuída à instância analisada (SCARINGELLA; ZOIA; MLYNEK, 2006). O KNN parte do pressuposto de que instâncias próximas no espaço de características tendem a pertencer à mesma classe.

As SVMs são algoritmos poderosos para classificação, tanto linear quanto não linear, sendo eficientes principalmente em bases de dados complexas de pequeno e médio porte. A ideia central desse método é encontrar um hiperplano que represente a melhor separação entre as classes. A partir disso, são definidos os vetores de suporte, que representam as margens entre o hiperplano e o ponto mais próximo de cada classe. Em casos em que os dados não são linearmente separáveis, as SVMs utilizam uma função *kernel* para mapear os dados em um espaço de maior dimensionalidade, o espaço de características. Nesse novo espaço, as covariáveis se tornam linearmente separáveis, permitindo que o hiperplano ideal seja calculado (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

O MLP é uma rede neural composta por uma camada de entrada, uma ou mais camadas ocultas com neurônios artificiais interconectados, e uma camada de saída que fornece as probabilidades de cada classe. Nas camadas ocultas, cada neurônio aplica uma função de ativação aos dados de entrada, processando e transmitindo essas informações para os neurônios seguintes. Na camada de saída, uma função de custo é utilizada para medir o quão próximo o modelo está da resposta correta. Com base

nesse cálculo, o algoritmo de retropropagação (*backpropagation*) ajusta os pesos de todos os neurônios das camadas anteriores, direcionando-os para valores que minimizem o erro e se aproximem da saída desejada. Esse processo é realizado inúmeras vezes, até que o modelo alcance um nível de precisão desejável (GÉRON, 2019; KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006).

Após escolher o algoritmo para a construção do modelo, o passo seguinte é selecionar os hiperparâmetros ideais para obter os melhores resultados. Os métodos para essa seleção serão detalhados na subseção 2.2.4.

2.2.3 Métricas de avaliação do modelo

No estudo de Sturm (2012), são apresentadas as métricas mais utilizadas nos trabalhos analisados. Dentre as métricas, 82% dos estudos utilizaram a acurácia média, 25% empregaram a revocação, 10% adotaram a precisão e 4% fizeram uso do f1-score. Ainda existem outras formas de avaliar modelos, como a matriz de confusão. A seguir, uma descrição de cada uma dessas métricas:

- Matriz de Confusão: é uma matriz quadrada que mostra a distribuição das previsões do modelo entre as classes avaliadas. Dessa maneira, é possível observar visualmente quais classes tem maior e menor dificuldade de serem classificadas. Quando uma classificação é realizada em múltiplas classes, é necessário analisar cada classe individualmente, cada uma contra as outras. Cada linha, que representa uma classe, são exibidas para qual classes as amostras foram classificadas. Disso derivam dois conceitos: **verdadeiros positivos**, que indicam as previsões corretas e; **falsos negativos**, que são previsões erradas, isto é, as amostras da classe analisadas foram classificadas como outras classes. Outras duas definições são de **falso positivos** e **verdadeiros negativos**. A primeira diz respeito as amostras que foram erroneamente classificadas como a classe analisada, a quantidade dessas previsões pode ser visualizada na coluna correspondente. A segunda, são o restante das previsões, aquelas que não se relacionam a classe analisada e não foram classificadas como tal. Assim é possível fazer uma análise geral das previsões dos modelos acerca dos gêneros e desvendar quais são os que mais são problemáticos para o modelo (SHIROL; KATHIRESAN, 2023).

- Acurácia: é uma forma de medir percentualmente os acertos de um modelo (SHIROL; KATHIRESAN, 2023). Por sua vez, a acurácia média é obtida calculando-se a média das diversas tentativas de teste de modelos, e geralmente, associa-se o desvio padrão a métrica (STURM, 2012). A acurácia média pode ser usada para medir a eficiência de treinamentos usando validação cruzada.

- Precisão: uma forma de verificar dentre as previsões realizadas para uma classe, qual o percentual de acerto. Definida por: $P = VP/(VP + FP)$, em que VP são os verdadeiros positivos e FN os falsos positivos. Dessa maneira é possível determinar se um modelo tem muitos falsos positivos (SHIROL; KATHIRESAN, 2023).

- Revocação: é uma medida que indica qual o percentual de acerto de um modelo para uma classe. Definida como: $R = VP/(VP + FN)$, no qual FN são os falsos negativos (STURM, 2012).

- F1-score: essa medida é muito útil quando há desbalanceamento entre as classes, levando em consideração tanto a precisão e a revocação (SHIROL; KATHIRESAN, 2023). Dessa forma, o f1-score é elevado quando uma classe apresenta tanto alta precisão quanto alta revocação, indicando um bom equilíbrio do modelo na sua classificação. Por outro lado, se a precisão for alta e a revocação for baixa, ou vice-versa, o f1-score será reduzido, evidenciando um desequilíbrio significativo. Essa métrica reflete esse comportamento por ser uma média harmônica entre precisão e revocação, conforme fórmula abaixo:

$$F1 = \frac{2 * P * R}{P + R} \tag{2}$$

2.2.4 Hiperparâmetros e Otimizações

A construção de modelos de aprendizado de máquina para classificação de gêneros musicais envolve a definição de diversos parâmetros ajustáveis, conhecidos como hiperparâmetros. A otimização de hiperparâmetros consiste em buscar valores em um conjunto que minimizem os erros do modelo, ou seja, que aumentem a acurácia (YU; ZHU, 2020). Quando feita de forma eficaz, a otimização pode impactar diretamente na precisão e eficiência do treinamento dos modelos de aprendizado de máquina, assim como a configuração inadequada desses hiperparâmetros pode comprometer a capacidade do modelo de generalizar corretamente na validação para novos dados (HOSSAIN; TIMMER, 2021).

A busca pelos melhores hiperparâmetros pode ser realizada por diversas abordagens, cada uma com suas vantagens e desafios. A seguir, são descritas algumas dessas técnicas: busca manual, busca aleatória e a busca em grade.

A abordagem de busca manual de hiperparâmetros é baseada na experiência do usuário. Para a tarefa, este faz uma escolha de possíveis valores e manualmente testa-os até chegar a um resultado satisfatório (HOSSAIN; TIMMER, 2021).

A busca em grade é uma abordagem em que são definidos um conjunto de valores para os hiperparâmetros e, a partir disso, são realizados diversos treinos e

avaliações dos modelos, buscando o melhor desempenho. Ao final do processo, são listados os valores que fizeram o modelo ter o melhor desempenho. Um problema desse método é que, dependendo do número de valores e de hiperparâmetros, pode tornar o número de testes muito grande, por isto não é indicado para grandes conjuntos. Outra forma é a busca aleatória, no qual os hiperparâmetros são otimizados de acordo com certas distribuições independentes de hiperparâmetros, e a escolha dos valores no espaço de busca é realizada de forma aleatória (YU; ZHU, 2020).

A escolha da técnica de otimização depende de fatores como o tamanho do conjunto de dados, a complexidade do modelo e os recursos computacionais disponíveis. Ressalta-se que uma combinação das técnicas pode ser usada. No contexto da classificação de gêneros musicais, a aplicação de métodos eficientes de ajuste de hiperparâmetros permite melhorar a precisão do modelo na categorização automática de faixas musicais.

2.3 Trabalhos relacionados

No trabalho de Farajzadeh, Sadeghzadeh e Hashemzadeh (2023), foi introduzido o PMG-Net como uma rede neural convolucional projetada para classificar gêneros musicais persas. Primeiramente, foi criada a base de dados PMG-Data, contendo 500 músicas divididas em cinco gêneros: monody, pop, rap, rock e tradicional. Dois experimentos foram conduzidos com o PMG-Net. O primeiro experimento foi realizado sem aumento de dados, utilizando segmentos de 32 segundos de cada música e validação cruzada em 5-*folds*, alcançando uma acurácia média de 76% e uma acurácia máxima de 79%. O segundo experimento foi realizado com aumento de dados, em que cada faixa foi dividida em seis segmentos, totalizando 3 mil amostras. A etapa de pré-processamento utilizou MFCC para extração de características. A rede foi treinada com diferentes números de camadas e parâmetros, obtendo uma acurácia máxima de 86%. Por fim, os autores compararam o PMG-Net com algoritmos tradicionais de aprendizado de máquina, como SVM, KNN, *Random Forest* e XGBoost. O MFCC e o STFT foram utilizados como entradas para os algoritmos, com o SVM usando STFT apresentando o melhor resultado, alcançando 65% de acurácia.

O trabalho de Silva e Gomes (2022) apresentou um classificador automático para músicas populares da região amazônica. Para isso, os autores criaram a base de dados AMPOP com mil músicas, contendo 125 faixas de cada gênero: brega, andino, carimbó, cúmbia, merengue, pasillo, salsa e vaqueirada. De cada faixa, 788 parâmetros foram extraídos em três pontos temporais (início, meio e fim das músicas). Os modelos utilizados para teste foram KNN, MLP, Classificador de Vetores de Suporte (em inglês: *Support Vector Classifier* - SVC) e XGBoost, alcançando acurácias de 57,58%, 56,79%, 61,33% e 61,17%, respectivamente. Os autores concluíram que o modelo SVC apresentou

os melhores resultados nos cenários avaliados.

O *Latin Music Database* (SILLA JR.; KOERICH; KAESTNER, 2018) foi proposto para introduzir um conjunto de dados abrangente, projetado para preencher lacunas de pesquisa em MIR para músicas latino-americanas. A base, também conhecida pelas siglas LMD, inclui mais de 3.227 faixas distribuídas em dez gêneros, como salsa, tango e samba, anotadas com metadados incluindo artista, álbum, ano e gênero. O processo de rotulagem foi realizado por dois especialistas humanos, ao invés de assumir que as classificações fornecidas por sites são a verdade absoluta. Este conjunto de dados suporta diversas tarefas de MIR, incluindo classificação de gêneros e análise de ritmos.

No trabalho de Tzanetakis e Cook (2002), os autores criaram o banco de dados GTZAN para classificação de gêneros musicais, consistindo em dez gêneros: clássico, country, disco, hip-hop, jazz, rock, blues, reggae, pop e metal. Para cada gênero, foram coletadas 100 amostras de 30 segundos. Utilizando os algoritmos *Gaussian Classifier*, *Gaussian Mixture Model* (GMM) e KNN, características como timbre, ritmo e tonalidade foram analisadas. Os resultados de classificação foram calculados usando validação cruzada em 10-*folds* com 100 iterações. O algoritmo que alcançou a maior acurácia foi o GMM, com jazz apresentando a maior taxa de acurácia por gênero, 75%, enquanto rock teve o menor desempenho, com 40%.

O conjunto de dados de áudio *Free Music Archive* (FMA), desenvolvido por Deferrard *et al.* (2017), contém 16 gêneros principais e 161 subgêneros. Para melhor acessibilidade, os autores dividiram o conjunto em várias categorias: o conjunto completo, com 106.574 músicas; o conjunto grande, com clipes de 30 segundos de cada faixa; o conjunto médio, com músicas pertencentes apenas a gêneros principais, resultando em 25 mil amostras de 30 segundos; e o conjunto pequeno, com mil músicas para cada um dos oito gêneros mais ouvidos. Os autores destacam que o conjunto pequeno é semelhante ao GTZAN, mas inclui metadados. Além disso, apontam que o conjunto é tendencioso em relação a gêneros experimentais, rock e eletrônicos.

A pesquisa de Hasib *et al.* (2022) apresentou o BMNet-5, uma rede neural profunda projetada para classificação multiclasse de gêneros musicais bengalis, alcançando uma acurácia de 90,32%. O estudo focou em seis gêneros: *nazrulgeeti*, *rabindra sangeet*, *polligeeti*, música de banda, hip-hop e *adhunik gaan*, utilizando um conjunto de dados de 1.742 músicas, com aproximadamente 250 a 350 músicas para cada gênero, criado por Mamun (MAMUN *et al.*, 2019). O BMNet-5 consiste em cinco camadas com 512, 256, 128, 64 e sete nós, respectivamente. Os autores extraíram nove características principais: ZCR, *Spectral Centroid*, MFCC, *Spectral Roll Off*, Frequência Cromática, *Spectral Bandwidth*, Fluxo Espectral, *Pitch* e *Tempo*. Além disso, médias e variâncias dessas características foram computadas, resultando em 29 atributos numéricos no vetor de entrada. Após 250-300 *epochs*, o modelo alcançou desempenho ideal em 187 *epochs* com tamanho de

lote de 50.

Um estudo relevante para a classificação de gêneros musicais é apresentado por Sousa, Pereira e Veloso (2016), que aborda dois problemas principais: a seleção de características musicais para alcançar altas taxas de acurácia na classificação de gêneros e a falta de um conjunto de dados representativo para músicas regionais brasileiras. Os autores desenvolveram o Brazilian Music Dataset (BMD), contendo 30 músicas para cada um dos sete gêneros: forró, rock, repente, MPB, brega, sertanejo e disco. Para a extração de características, os autores utilizaram seis categorias de descritores: espectrais, domínio do tempo, ritmo, efeitos sonoros, tonalidade, histograma de batidas, entre outras. Para o desenvolvimento do modelo, os autores utilizaram o algorítimo SVM, aplicando a técnica de validação cruzada com 5-*folds*, repetindo o processo 30 vezes com diferentes divisões aleatórias dos dados para garantir a consistência dos resultados. Essa abordagem resultou em uma acurácia média de 86,11%, com o gênero MPB apresentando a maior taxa de acurácia individual, alcançando 93,54%, e os gêneros com menor desempenho foram disco, com 62,06%, e rock, com 67,64%.

Tabela 1 – Proposta comparada com os trabalhos relacionados

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	Sim	3227	Não
Hasib	6	Sim	1742	Não
BMD	7	Sim	210	Sim
Proposta	**8**	**Sim**	**10000**	**Sim**

3 Proposta

A proposta foi composta por duas etapas distintas. A primeira consistiu na criação da Base de Dados, enquanto a segunda envolveu a avaliação da aplicação dessa Base para o treinamento de modelos computacionais. Esses modelos, ao receberem trechos musicais, têm a capacidade de identificar os gêneros musicais utilizados neste estudo. A seguir, na seção 3.1 são descritos os gêneros musicais escolhidos para este trabalho, e em seguida são apresentadas as duas etapas desta proposta: a criação da base na seção 3.2 e a avaliação da Base de Dados com Aprendizado de Máquina na seção 3.3.

3.1 Gêneros Musicais

Para este trabalho foram escolhidos os gêneros Samba, Bossa Nova, Samba-Enredo, Pagode, Funk, Forró, Forró Piseiro e Sertanejo. Segue-se uma breve descrição do histórico e algumas de suas características musicais.

O **Samba** é um gênero musical de origem afro-brasileira, que ganhou popularidade no Rio de Janeiro no início do século 20, embora tenha suas raízes no estado da Bahia. Sua origem é discutível, pois poucos registros históricos são encontrados, sendo mais aceita a tese de que surgiu da mistura dos gêneros populares tango e maxixe.

Duas vertentes do Samba ficaram conhecidas, a primeira que foi chamada de Estilo Antigo, e era tocado em rodas usando instrumentos como piano, cavaquinho, flauta, clarinetes, instrumentos de corda e metais, isto é, instrumentos de origem europeia. A segunda vertente é o Novo Estilo, que também é conhecido como Estácio, pois é associado a autores que frequentavam o bairro Estácio de Sá no Rio de Janeiro. Neste estilo foi introduzido o refrão improvisado e o uso de instrumentos percussivos de origem africana ou criados no Brasil, como os pandeiros, as cuícas, os surdos e os tamborins, e além desses, eram usados violões e também os cavaquinhos.

Por ser um gênero que em sua origem era tocado nos morros e associado à cultura da malandragem, foi-lhe atrelado um baixo valor cultural e assim foi repudiado pelas classes mais abastadas. Porém, com a popularização do carnaval de rua e dos botequins, o samba foi sendo mais aceito por todas as classes sociais e o Novo Estilo (Estácio) foi tido como o samba-raiz. Com o advento da indústria fonográfica no Brasil, o samba gravado tornou-se diferente, apresentando uma estrutura de refrão e segundas partes, ou seja, perdeu-se a característica de improvisação (QUADROS JÚNIOR, 2019).

Segundo Quadros Júnior (2019), o **Samba-Enredo** é um tipo de Samba, que

surgiu como uma música característica das Escolas de Samba. Tem como objetivo dramatizar uma temática durante os desfiles de carnaval. Com letras elaboradas, o Samba-Enredo narra um enredo específico e é acompanhado de uma melodia marcante. Essa variante do Samba ganhou notoriedade e tornou-se um elemento essencial nas festas de Carnaval do Brasil.

O **Pagode**, considerado uma versão do Samba, surgiu no Rio de Janeiro no final da década de 70, com o grupo os Originais do Samba com seus cantos uníssonos e presença marcante de elementos percussivos. Tornaram-se referência e seu estilo foi usado por outros grupos, como o Fundo de Quintal, que lançou seu primeiro disco em 1980 e então foram incorporados instrumentos que se tornaram marca do Pagode: o tantã, o repique de anel e o banjo com braço de cavaquinho. Nos anos finais da década de 80, surgiu o chamado Pagode Romântico, uma versão mais comercial, em que grupos começaram a incorporar elementos da música pop nacional e internacional. Quanto à sonoridade, reduziram as polirritmias, formando uma música mais "limpa", adicionaram teclados eletrônicos e os artistas performavam danças coreografadas. A temática geralmente era sobre finais felizes de histórias de amor, embora mais pra frente tenham sido criadas letras mais erotizadas e com duplo sentido (QUADROS JÚNIOR, 2019; TROTTA, 2011; TROTTA, 2016).

Por sua vez, a **Bossa Nova** originou-se no fim da década de 50 no Rio de Janeiro, combinando influências do Samba, *Jazz* e do *Blues*. Uma tentativa de modernização da música popular brasileira, que seria aceita pelos eruditos e também seria popular, podendo ser exportada como um produto brasileiro. Dessa forma, é um gênero que possui uma melodia sofisticada e um ritmo encorpado, esta é uma característica do Samba e outros ritmos populares do Brasil. O canto, na Bossa Nova, pode ser considerado mais introspectivo, em que a voz do cantor não é mais a maior presença, alinhando-se com os outros instrumentos. Portanto, o cantor tem que se adequar ao arranjo. O Violão foi um instrumento que ganhou destaque, trazendo uma batida que realça a síncope e a dissonância e, junto com a voz, formam a parte mais importante da interpretação. As letras possuem uma temática leve e alegre, contrastando com a melancolia do Samba-Canção (uma tentativa anterior de modernização da música popular brasileira) e passar essa mensagem torna-se atributo,também, da melodia (QUADROS JÚNIOR, 2019; TEIXEIRA, 2011).

A história do **Funk Brasileiro** remonta ao gênero homônimo, Funk Americano, que tinha elementos do *Groove*, Rock e *Rhythm & Blues*. Esse gênero utilizava instrumentos elétricos, arranjos sofisticados e elementos musicais refinados. Sua transformação no Funk Nacional teve início quando o estilo tornou-se popular no Brasil, na década de 1970. Divulgado pelo rádio, chegou inicialmente às festas da Zona Sul carioca e, posteriormente, ao subúrbio, sendo tocado em festas comunitárias organizadas por Big Boy

e Ademir Lemos. Os *Disc Jockeys* (DJs), responsáveis por comandar as músicas dessas festas, utilizavam discos importados ou adquiridos no comércio informal nos chamados melódromos. Outra forte influência foi o *Miami Bass*, uma evolução do hip-hop popular na Flórida, caracterizada por sons com graves intensos e letras consideradas impróprias por seu conteúdo sexual explícito. Assim como o Funk Americano, essa vertente foi bem recebida nos bailes, e logo os frequentadores começaram a criar versões das letras em português, conhecidas como melôs, marcando o início da influência brasileira sobre o gênero.

Já na década de 1980, o DJ Marlboro retornou de uma viagem a Londres com o objetivo de desenvolver um Funk genuinamente brasileiro. Iniciou sua carreira como produtor e passou a produzir os já mencionados melôs, com letras em português e utilizando as batidas características do *Miami Bass*. Outro *Disc Jockey* influente foi Dennis DJ, responsável por expandir o gênero e afastá-lo da forte influência americana. O Funk Brasileiro consolidou-se nos anos 1990 com produções integralmente nacionais. Diferente do gênero que o originou, o Funk Brasileiro incorporou um ritmo eletrônico com batidas intensas, mixagem marcante e letras jocosas. Assim, tornou-se um importante veículo de expressão cultural da periferia brasileira (BEZERRA, 2017; VIANA, 2010).

O **Sertanejo** surgiu como uma forma de urbanização da música caipira, presente no interior paulista e em outros estados do centro-sul. A música caipira utilizava instrumentos como viola-triângulo, adufe, rabeca, reco-reco de chifre, surdo, caixa, tarol e pandeiro. Posteriormente, com a popularização do Sertanejo, houve mudanças nesses instrumentos: elementos percussivos como a caixa, o surdo, o tarol e o adufe deixaram de ser utilizados, sendo substituídos por sanfona, prato de metal, bateria, violão e guitarras. O gênero consolidou uma forte tradição de ser interpretado por duplas ou trios e herdou da música caipira o cantar anasalado.

Quanto à sua origem, o grande responsável foi Cornélio Pires, que criou a Turma Caipira Cornélio Pires e apresentou-se no interior paulista. Com o grande sucesso, Cornélio viajou a São Paulo para gravar alguns discos, o que impulsionou outros artistas a realizarem gravações, popularizando o gênero. Outro fator importante para sua divulgação foi o rádio, que trouxe uma necessidade de maior mercantilização das músicas. Assim, foram introduzidas mudanças temáticas, priorizando aspectos urbanos em detrimento dos temas rurais e do campo. Já na década de 1980, a participação na televisão levou alguns artistas ao estrelato, ampliando ainda mais a presença do sertanejo no Brasil. O gênero, ao longo do tempo, recebeu influências de diversos ritmos estrangeiros e nacionais, como a Guarânia paraguaia, a Rancheira mexicana e a Jovem Guarda. Isso resultou na incorporação de novos instrumentos como harpas paraguaias, trompetes, guitarras, teclados, baterias, teclados eletrônicos e contrabaixos. No entanto, nem todos esses instrumentos se tornaram fixos, sendo utilizados apenas por algumas

duplas que, atualmente, costumam contar com bandas de apoio. Para este trabalho, não foram selecionadas músicas relacionadas ao subgênero conhecido como Sertanejo Universitário (ALONSO, 2011; CALDAS, 1987; QUADROS JÚNIOR, 2019).

A palavra **Forró** tem duas possíveis origens. A mais aceita é a de que se originou de uma abreviação do termo africano *forrobodó*, que significa festa ou bagunça. O termo, portanto, foi usado para referir-se às danças e ritmos tocados nestes lugares, como: baião, coco, xaxado, xote, marcha junina, entre outros. Com a migração nordestina para as regiões centro-sul do Brasil, essas festas foram disseminadas pelos imigrantes nordestinos como uma forma de lembrar da sua terra natal. Esse processo teve grande contribuição de Luiz Gonzaga, que rodou o Brasil tocando baião, xote, xaxado, arrasta-pé, e outros. Consequentemente, Forró passou a ser a denominação de todos esses ritmos musicais. Com a contribuição de Jackson do Pandeiro, o Forró passou também a ser a denominação de um ritmo próprio, dando uma dualidade ao termo. Os instrumentos que o caracterizam são: acordeão, triângulo e zabumba, porém não se resume somente a esses. Com a popularização e movimento de urbanização do gênero, outras variações surgiram, como o chamado Forró MPB que adicionou violão, bateria e instrumentos elétricos como a guitarra e baixos. Mais recentemente, houve acréscimo de teclados e sintetizadores no chamado Forró Eletrônico (DIAS; DUPAN, 2022; SILVA, 2022).

O **Forró Piseiro** é uma variante do Forró Eletrônico. Este tem como característica o uso de instrumentos elétricos como guitarras e teclados, baterias, vários vocalistas, dançarinos, andamento mais acelerado, ritmo e musicalidade diferentes do que o Forró Tradicional. O Forró Piseiro tem estrutura semelhante, porém com bases rítmicas e efeitos eletrônicos reproduzidos por teclados e sintetizadores, e uso de guitarras e sanfona (DIAS; DUPAN, 2022).

3.2 Criação da Base de Dados

A base de dados SONBRA (Sonoridades Brasileiras) foi organizada de maneira a garantir o balanceamento entre os gêneros musicais, de forma que cada gênero tenha o mesmo número de amostras (1.250 amostras por gênero). Esse balanceamento é essencial para evitar vieses durante a etapa de treinamento de futuros modelos de aprendizado de máquina (JAPKOWICZ; STEPHEN, 2002). Além disso, a base de dados foi organizada em um formato acessível e estruturado, contendo todas as amostras e suas respectivas características musicais extraídas, possibilitando o uso em estudos futuros e em sistemas de classificação automática de gêneros musicais.

A base de dados final, devido a questões de direitos autorais, não contém os arquivos de áudio em si, mas disponibiliza as características extraídas de cada amostra, permitindo sua utilização em aplicações acadêmicas e tecnológicas. O processo da

criação da base está demonstrado visualmente na figura 1. Em sequência, estão descritas as etapas da criação da base.

3.2.1 Coleta das músicas

Nesta etapa, foram obtidas amostras musicais por meio da plataforma de *streaming* YouTube, selecionando músicas que representassem oito gêneros populares brasileiros: Bossa nova, Forró, Funk, Pagode, Forró Piseiro, Samba, Samba-Enredo e Sertanejo. Para cada um desses gêneros, foram coletadas 250 músicas, totalizando 2 mil faixas musicais. A seleção foi baseada em critérios que priorizam gravações em ambientes controlados, de modo a garantir a melhor qualidade de som possível para o processamento dos dados. Essas músicas foram extraídas com o auxílio da ferramenta Pytube[1], sendo inicialmente obtidas em formato MP4 e, posteriormente, convertidas para o formato WAV, utilizando o software FFmpeg[2], pois oferece maior fidelidade ao sinal de áudio, facilitando a extração posterior das características musicais.

3.2.2 Fragmentação

Um sinal de áudio contém muita variabilidade ao longo do tempo, dessa forma, vários fragmentos permitem capturar diversas nuances de uma faixa de música (SILLA JR.; KAESTNER; KOERICH, 2007). A fragmentação também permite economizar recursos computacionais, visto que extrair características de áudio de uma faixa completa demandaria muito processamento (COSTA; VALLE; KOERICH, 2004). Portanto, após a coleta, iniciou-se a segunda etapa, em que cada faixa de música foi fragmentada em 5 amostras de 30 segundos:

- *begin*: os 30 primeiros segundos;
- *middle*: 15 segundos antes e depois do meio da música;
- *middle*_1: os 30 segundos antes do meio da música;
- *middle*_2: os 30 segundos após o meio da música, e;
- *end*: os últimos 30 segundos;

Em vez de analisar as músicas em sua totalidade, a divisão em segmentos menores permite aumentar o número de amostras disponíveis para o treinamento dos modelos, contribuindo para a eficiência do aprendizado, melhorando a detecção de padrões musicais específicos de cada gênero. Sendo assim, cada gênero musical ficou

[1] https://pytube.io
[2] https://www.ffmpeg.org

com 1.250 amostras (250 x 5 tipos de fragmentos), totalizando 10 mil amostras (1.250 x 8 gêneros) no conjunto de dados final.

3.2.3 Extração das características musicais

A extração das características musicais consistiu no processamento das amostras utilizando a biblioteca librosa, conforme mostrado na figura 1. Os seguintes extratores foram aplicados às amostras de 30 segundos: *Tempograma, Mel-spectrograma, Fourier Tempograma, Chroma STFT, Chroma CENS, Chroma CQT, MFCC, Tonnetz, RMS, ZCR, Spectral Roll Off, Spectral Centroid* e *Spectral Bandwidth*.

Cada faixa de áudio passou pela extração de características. As características *Tempograma*, *Fourier Tempograma*, *Chroma STFT*, *Chroma CENS Chroma CQT*, *Tonnetz*, *RMS* e *Spectral Centroid* foram extraídas com os argumentos padrões da biblioteca librosa. Para o *Mel-spectograma*, o argumento **n_mels** foi definido como 128, definindo o número de bandas geradas. O mesmo foi feito para o *MFCC*; nesse caso, o argumento chama-se **n_mfcc** e foi definido como 20. Para o *ZCR* foi realizada a extração igualmente aos outros, fez-se o cálculo das estatísticas (média e mediana), mas também foi realizada a soma das características e foi dado o nome de *ZCR sum*. Para o *Spectral Roll Off* foram feitas três extrações, cada uma com um valor diferente do argumento **porcent**, sendo os valores: 0,01; 0,85 e 0,99. Procedimento semelhante foi aplicado ao *Spectral Bandwith*. Nesse caso, variou-se o argumento **p** de 2 a 12, implicando em 11 extrações.

A extração das características resultou em descrições numéricas, que podem ser escalares, vetoriais ou matriciais. Para tratar essas variáveis, de forma que se usasse menos recursos computacionais, foram realizadas análises estatísticas, conforme proposto na base GTZAN, em que se fez a extração da média e da variância das características extraídas. Para este trabalho, considerou-se a média e a mediana dos valores de uma determinada característica, resultando em duas bases de dados, uma para cada análise, como ilustra a figura 1.

A tabela 2 mostra quantos parâmetros pertencem a cada característica. Esses parâmetros foram adicionados na base de dados, em que suas linhas são as faixas musicais e as colunas são os parâmetros.

Para identificação das faixas, as três primeiras colunas da base de dados são: a de classe, que identifica a qual gênero musical a faixa pertence; arquivo, que é o nome da faixa de áudio e divisão, que aponta de qual parte da música original aqueles parâmetros foram extraídos, podendo ser *begin*, *end*, *middle*, *middle_1* e *middle_2*. Um exemplo da estrutura é mostrado na tabela 3.

Figura 1 – Processo de criação da base

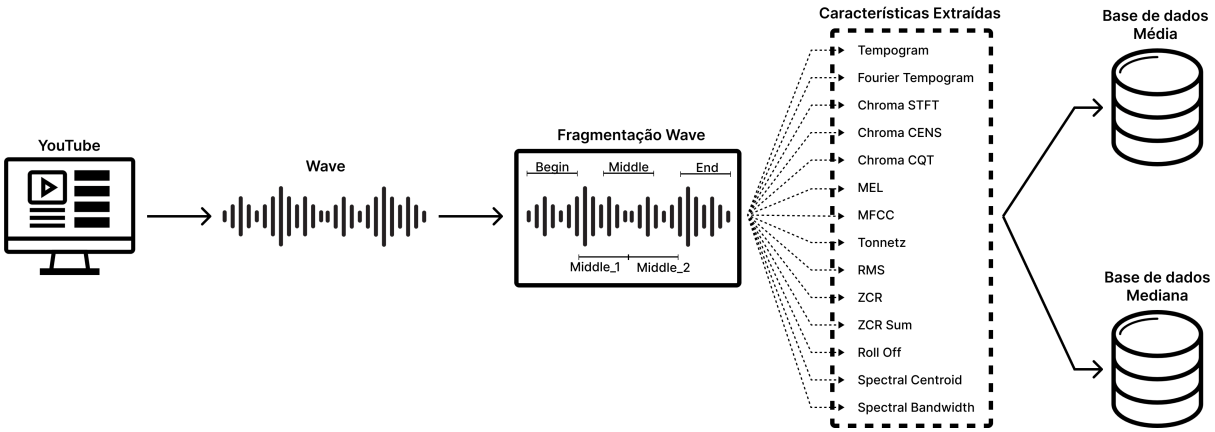


Tabela 2 – Características Extraídas e número de amostras

Característica Extraída	**Número de Colunas**
Fourier Tempograma	193
Tempograma	384
Chroma STFT	12
Chroma CENS	12
Chroma CQT	12
Melspectograma	128
MFCC	20
Tonnetz	6
RMS	1
ZCR	1
ZCR Sum	1
Spectral Roll Off	3
Spectral Centroid	1
Spectral Bandwith	11
Total	**785**

Tabela 3 – Estrutura da base de dados

classe	**arquivo**	**divisao**	**fourier_tempogram-median0**	**...**	**spectral_bandwidth-median12**
sertanejo	faixa1	middle	210.0185+0j	...	7936.223682
pagode	faixa2	middle	123.6754+0j	..	8055.215747

3.3 Avaliação com Modelos de Aprendizado de Máquina

Para avaliar se é possível usar a Base de Dados experimental para treinar modelos que reconheçam gêneros musicais, foi realizado o treino de diversos modelos computacionais em três etapas e, ao fim, escolheu-se um modelo vencedor. Usou-se a acurácia como métrica de avaliação e somente a base de dados com as médias das características foi usada para essa avaliação.

A base de dados, composta por 10 mil amostras, foi dividida em dois subconjuntos principais: **Base de Desenvolvimento**, com 7 mil amostras (70% do total), e **Base de Teste**, com 3 mil amostras (30% do total).

A Base de Desenvolvimento, por sua vez, foi subdividida em duas partes:

- **Base de Treinamento**: composta por 5.600 amostras (80% da Base de Desenvolvimento).
- **Base de Validação**: composta por 1.400 amostras (20% da Base de Desenvolvimento).

Figura 2 – Divisão base de dados para avaliação

3.3.1 Construção do modelo de aprendizado de máquina

Os modelos de aprendizado de máquina usados nesta proposta foram: *Decision Tree*, KNN, MLP, *Random Forest*, SVM e XGBoost. Estes foram usados nas etapas 1 e 2 descritas na subseção 2.2.4. Para a etapa 3, foi criado um *ensemble* que combina os seis classificadores citados anteriormente em um sistema de *Voting*, onde a faixa musical é definida de acordo com uma votação entre os resultados de cada modelo.

3.3.2 Hiperparâmetros e otimizações

Neste estágio, os modelos foram submetidos a dois métodos de treino, em que se almejou encontrar os melhores hiperparâmetros e características de entrada para o modelo computacional. Usou-se aqui a técnica de busca em grade, com a implementação encontrada no Scikit-Learn (PEDREGOSA *et al.*, 2011). A tabela 4 expõe os modelos, os hiperparâmetros e os valores testados de cada um. Em cada etapa, subseções 3.3.2.1 e 3.3.2.2, o seguinte processo foi executado: inicialmente, os testes foram conduzidos individualmente para cada característica musical extraída (*Tempograma, Fourier Tempograma, MFCC, Mel, ZCR, ZCR Sum, Chroma STFT, Chroma CENS, Chroma CQT, Tonnetz, RMS, Spectral Centroid, Spectral Bandwith* e *Spectral Roll Off*) e para cada modelo foram anotadas as cinco melhores: *Mel, MFCC, Tempograma, Chroma CENS* e *Chroma CQT*. Em seguida, foram realizados experimentos combinando pares (combinações 2 a 2) e trios (combinações 3 a 3) de características, buscando identificar possíveis interações que pudessem contribuir para uma representação mais rica e eficiente dos dados, por exemplo: *Tempograma*; *Tempograma* com *MFCC* e; *Tempograma*, *MFCC* e *Mel*. Ao mesmo tempo, nessa etapa também houve a otimização de hiperparâmetros, removendo ou ajustando os valores que não estavam alcançando resultados minimamente satisfatórios. Ressalta-se que só foram usados 45 dos 384 parâmetros do *Tempograma*. A seguir estão descritos os dois métodos de treinamento e critérios de seleção de modelos em cada etapa:

3.3.2.1 Etapa 1 — Validação cruzada

A validação cruzada em *k-folds* é uma técnica em que se divide a base em *k-folds*, isto é, em k partes, e treina-se o modelo com uma parte (k-1) e a outra parte é usada para validá-lo (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Para selecionar os melhores modelos, foi realizada uma **validação cruzada estratificada com 10-*folds*** e foi usada, exclusivamente, a divisão de Treino da Base de Desenvolvimento. A estratificação possibilita que as classes (gêneros musicais) estejam representadas igualmente nos *folds* de teste e de validação, impedindo um desequilíbrio que impactaria no resultado final (SAMMUT; WEBB, 2011). A acurácia média foi calculada para comparar o desempenho entre os modelos, e a de maior valor foi selecionada para representar cada abordagem de treinamento.

3.3.2.2 Etapa 2 — Validação

Treino realizado com o conjunto de dados Treino e para a validação usou-se o conjunto de Validação. Para selecionar um modelo de cada algoritmo de aprendizagem, adotou-se o seguinte critério: o modelo que teve acurácia mais próxima da acurácia média obtida do modelo correspondente da etapa anterior.

3.3.2.3 Sistema de votação

Adicionalmente aos modelos computacionais já citados, um sistema de votação (GÉRON, 2019) foi proposto. Para essa tarefa, foi selecionado o modelo de cada algoritmo que obteve melhor acurácia na validação cruzada (subseção 3.3.2.1). Dessa forma, combinaram-se os modelos: KNN, MLP, SVM, *Random Forest*, XGBoost, *Decision Tree* em um sistema de votação (*voting*), que foi submetido aos métodos de decisão abaixo:

- **Votação Hard**: a decisão final é baseada na maioria dos votos, ou seja, cada modelo palpita um gênero e vence aquele que tiver mais votos.
- **Votação Soft**: a decisão final é ponderada pelas probabilidades preditas pelos modelos.

Os testes seguiram as etapas da subseção 3.3.2, exceto pela combinação de características. Nesta abordagem, foram utilizadas as combinações em trio que obtiveram a maior acurácia em cada modelo individual. Mediante validação cruzada (subseção 3.3.2.1) e validação (subseção 3.3.2.2), dois modelos foram selecionados (um por etapa).

3.3.3 Etapa Final — Avaliação Conjunta

Ao todo, foram treinados 14 modelos, sendo dois de cada tipo de modelo computacional: um referente à etapa de validação cruzada e outro à validação. Esses modelos foram treinados com a Base de Desenvolvimento e avaliados utilizando a Base de Teste. As acurácias finais foram então obtidas, e o modelo com o melhor desempenho foi considerado o mais adequado para este estudo. Por fim, foram selecionados sete modelos computacionais com base na acurácia. Foi gerada uma matriz de confusão, um gráfico que apresenta as acurácias por gênero e um *class predict error* para cada modelo. A partir da matriz de confusão, foram calculadas a acurácia, precisão média, revocação média e f1-score médio para os sete modelos computacionais selecionados e para cada gênero individualmente, somente para o modelo que teve maior acurácia.

Tabela 4 – Hiperparâmetros e espaço amostral para busca em grade

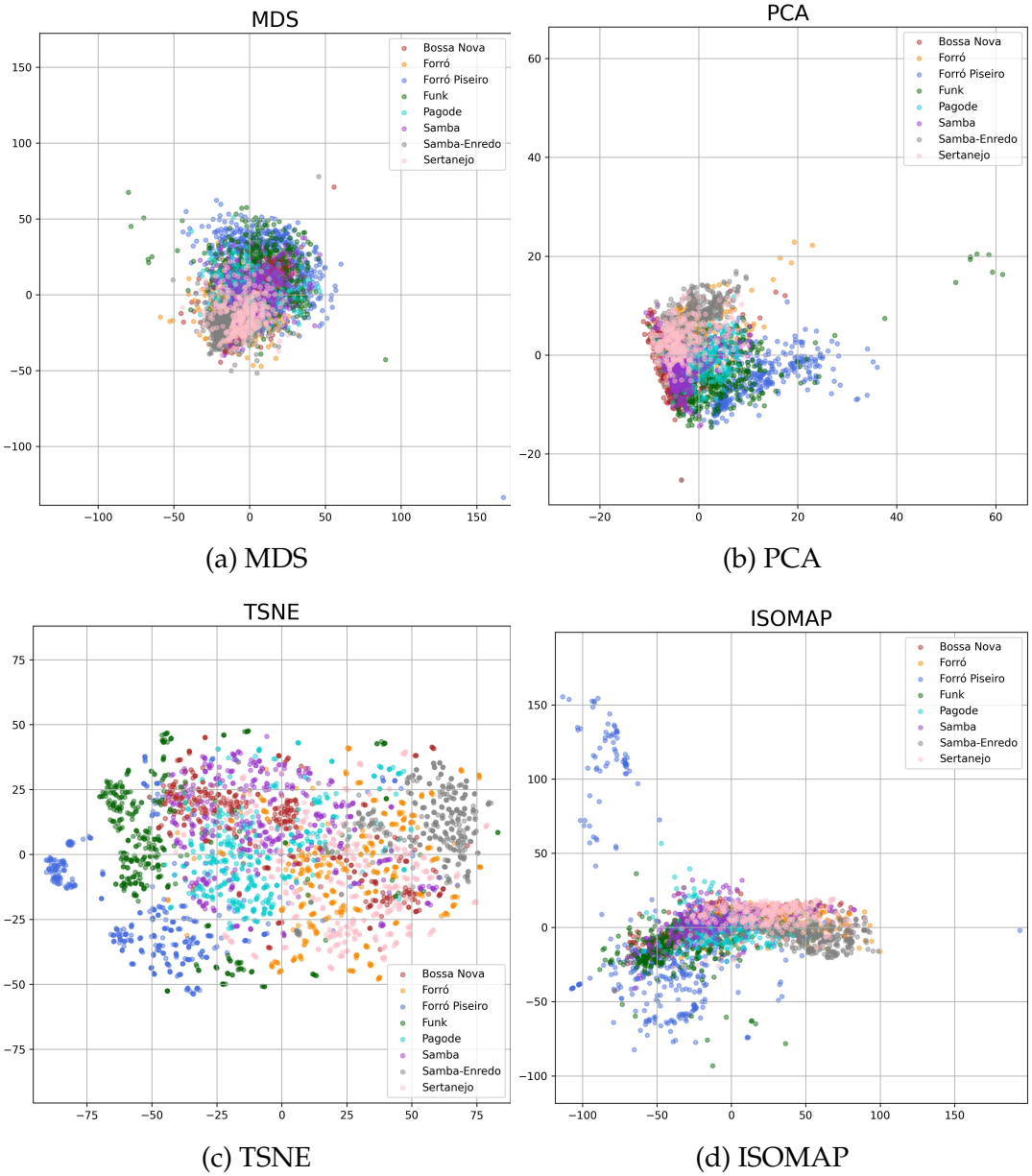
Modelo	**Hiperparâmetros**	**Espaço Amostral**
KNN	k	5..20
	metric	canberra, manhattan, minkowski, braycurtis
	weight	distance, uniform
Decision Tree	criterion	gini, entropy
	max_depth	1..20
	max_features	auto, sqrt, log2
MLP	activation	relu, tanh
	solver	adam, lbfgs
	hidden_layer_sizes	(180, 150, 50), (200, 150, 100), (130, 100, 60)
Random Forest	criterion	gini, entropy, log_loss
	max_features	sqrt, log2
	n_estimators	25, 100, 150
	max_depth	1, 9, 10, 16, 10, 101
SVM	kernel	rbf, poly
	C1	1.0, 1.5, 2.0, 2.5, 9
	gamma	0.9
	decision_function_shape	ovo, ovr
XGBoost	eval_metric	mlogloss, merror, mae
	objective	multi:softmax, multi:softprob
	max_depth	3, 6, 9
Voting	strategy	soft, hard
	estimators	KNN, Decision Tree, MLP, Random Forest, SVM, XGBoost

4 Resultados

4.1 Análise Preliminar

Para visualizar quão sobrepostos estão os gêneros, e a dificuldade de classificá-los, foram usadas algumas técnicas de redução de dimensionalidade e de visualização. A figura 3 exibe os resultados em duas dimensões para *Multidimensional Scaling* (MDS), *Principal Component Analysis* (PCA), *T-Distributed Stochastic Neighbor Embedding* (TSNE) e *Isometric Mapping* (ISOMAP).

Figura 3 – Redução de dimensionalidade



(a) MDS

(b) PCA

(c) TSNE

(d) ISOMAP

No MDS (figura 3a) e no PCA (figura 3b) há uma maior concentração central dos gêneros, com muita sobreposição entre os gêneros. É possível observar alguns agrupamentos em algumas regiões para alguns gêneros como Sertanejo, Samba-Enredo e Funk, sendo o Funk o mais bem definido dentre os gêneros no PCA.

Observa-se no TSNE (figura 3c) que o Funk, Samba-Enredo e Forró Piseiro estão concentrados nas bordas, sem presença relevante de outros gêneros, revelando uma facilidade de separação dos gêneros. Por outro lado, o Samba e Bossa Nova se sobrepõem, juntamente com o Pagode, este em menor grau. O Forró e o Sertanejo compartilham a mesma região. Por fim, é notável uma maior dispersão dos cinco gêneros: Bossa Nova, Forró, Sertanejo, Samba e Pagode, revelando que estes possuem características parecidas e de maior dificuldade de separação.

No ISOMAP (figura 3d) é possível observar alguns agrupamentos de gêneros: Sertanejo e Samba-Enredo, com maior concentração na faixa central, sugerindo que esses gêneros têm características mais coesas. O destaque fica para o Forró Piseiro que possui uma dispersão maior, porém sem se sobrepor muito a outros gêneros.

4.2 Base de Dados

Conforme mencionado no capítulo 3, primeiramente foi criada a base de dados. Por questões de direitos autorais, a base de áudio não será publicada, apenas a base de dados[1].

4.3 Avaliação Modelos

Os modelos selecionados por cada etapa estão descritos na tabela 5. A coluna "selecionado por" contém dois valores possíveis **V** e **VC**, que se referem aos modelos selecionados pelos critérios de Validação Cruzada (subseção 3.3.2.1) e Validação (subseção 3.3.2.2), respectivamente. Os sistemas de votação (subseção 3.3.2.3) também estão descritos na tabela. A coluna "diferença" contém o valor absoluto da diferença entre a acurácia obtida na validação e a acurácia média da validação cruzada para o modelo com os mesmos hiperparâmetros mostrados na respectiva linha. Essa informação é apresentada somente para os modelos selecionados na etapa 2 (subseção 3.3.2.2).

[1] A Base de dados deste link contém apenas 4000 amostras, pois a base completa será publicada posteriormente: <https://github.com/andyreborn1/sonbra_music_database>

Tabela 5 – Melhores acurácias de Validação Cruzada (CV) e de Validação (V)

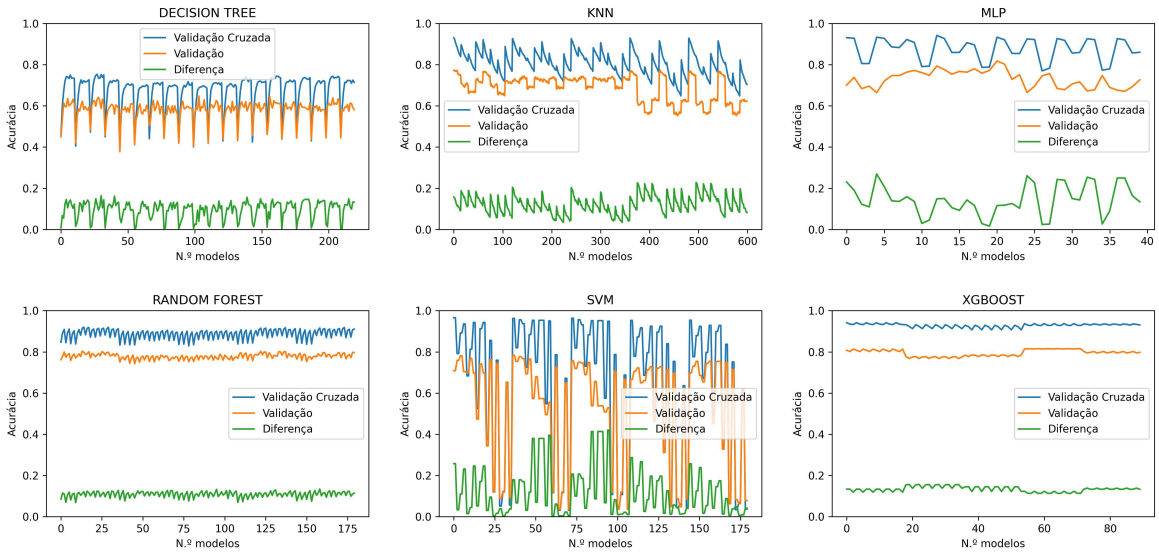
Id	Modelo	ACC VC	STD VC	ACC V	Diferença	Características	Hiperparâmetros	Selecionado por
1	Decision Tree	0,438	0,041	0,440	0,002	Chroma CQT Mel Tempograma	criterion: entropy max_depth: 3 max_features: log2	V
2	**Decision Tree**	**0,753**	**0,012**	**0,631**		**Mel** **MFCC** **Tempograma**	**criterion: entropy** **max_depth: 18** **max_features: sqrt**	**CV**
3	KNN	0,717	0,025	0,684	0,033	Chroma CENS MFCC Tempograma	k: 19 metric: mikowski; weight: uniform	V
4	**KNN**	**0,930**	**0,010**	**0,773**		**Mel** **MFCC** **Tempograma**	**k: 5;** **metric: canberra** **weight: distance**	**CV**
5	MLP	0,788	0,022	0,772	0,016	Chroma CQT MFCC Tempograma	activation: relu solver: lbfgs hidden_layers_sizes: (150, 100, 50)	V
6	**MLP**	**0,942**	**0,007**	**0,794**		**Chroma CENS** **MFCC** **Tempograma**	**activation: tanh** **solver: adam** **hidden_layers_size: (180, 150, 75)**	**CV**
7	Random Forest	0,838	0,013	0,770	0,068	Mel MFCC Tempograma	criterion: gini max_features: sqrt n_estimators: 125 max_depth: 9	V
8	**Random Forest**	**0,920**	**0,014**	**0,799**		**Mel** **MFCC** **Tempograma**	**criterion: log_loss** **max_features: sqrt** **n_estimators: 100** **max_depth:15**	**CV**
9	SVM	0,617	0,026	0,609	0,008	Chroma CENS Mel Tempogram	kernel: sigmoid C: 1; gamma: auto decision_function_shape: ovr	V
10	**SVM**	**0,966**	**0,005**	**0,709**		**Mel** **MFCC** **Tempograma**	**kernel: rbf** **C:9;** **gamma: 0.9;** **decision_function_shape: ovo**	**CV**
11	XGB	0,929	0,008	0,743	0,186	Chroma CENS Mel Tempogram	eval_metric: mae objective: multi:softprob max_depth: 3	V
12	**XGB**	**0,944**	**0,005**	**0,743**		**Mel** **MFCC** **Tempogram**	**eval_metric: mlogloss** **objective: multi:softprob** **max_depth: 3**	**CV**
13	**Voting**	**0,957**	**0,009**	**0,806**	**0,151**	**Chroma CENS** **Mel** **Tempogram**	**strategy: soft**	**V**
14	Voting	0,966	0,007	0,771		Mel MFCC Tempogram	strategy: soft	CV

ACC - Acurácia; VC - Validação Cruzada; STD - Desvio Padrão; V - Validação

A estabilidade dos modelos é definida como a variação mínima entre os resultados de acurácia entre validação cruzada e de validação, como mostrado na figura 4. Foi observado que o *Random Forest* e o SVM exibiram, respectivamente, a maior e a menor estabilidade.

Durante a etapa de validação cruzada, a combinação de características de *Mel*, *MFCC* e *Tempograma* apresentou os melhores resultados em cinco dos seis modelos analisados, enquanto o sexto modelo obteve seu melhor desempenho com a combinação de *Chroma CENS*, *Mel* e *Tempograma*. Por esse motivo, essas duas configurações de características foram utilizadas como parâmetros de entrada para avaliar o modelo de *voting*.

Figura 4 – Comparativo de acurácia entre Validação Cruzada e Validação



Os resultados da Etapa Final de Avaliação (subseção 3.3.3) são mostrados na tabela 6, que contém sete modelos, um para cada modelo computacional: KNN, *Decision Tree*, MLP, *Random Forest*, SVM, *XGBoost* e *Voting*. A coluna **Id** refere-se à coluna homônima da tabela 5, que contém as demais informações dos modelos. Estes foram os que melhor performaram, de acordo com acurácia: *Voting* com 83,1%; XGBoost com 82,5% e; *Random Forest* com 81,3%. Diante disso, o *Voting* foi selecionado como o melhor modelo. Por fim, foram calculadas as matrizes de confusão de cada um dos sete modelos, que podem ser visualizadas na figura 5.

Tabela 6 – Acurácias avaliação cojunta

Id	**Modelo**	**Acurácia**
2	Decision Tree	0,627
4	KNN	0,771
6	MLP	0,749
8	Random Forest	0,813
10	SVM	0,799
12	XGBoost	0,825
13	**Voting**	**0,832**

Já na figura 6 é possível identificar visualmente quais modelos apresentaram maior taxa de erro nas previsões. No lado esquerdo do gráfico, estão representados os gêneros musicais reais das amostras, enquanto no lado direito estão os gêneros previstos pelo modelo. As linhas que conectam esses dois lados ilustram as correspondentes

previsões. Além disso, o gráfico informa o percentual de acerto dos modelos para cada gênero individualmente, facilitando a avaliação do desempenho por categoria.

A partir disso, é possível identificar que os gêneros com maior dificuldade de serem classificados foram: Bossa Nova, Forró, Pagode, Samba e Sertanejo que têm seus desempenhos por modelo resumidos na tabela 7. Em negrito estão destacados os menores desempenhos por modelo e apresenta-se a média, mediana e desvio padrão de cada gênero. O ritmo Sertanejo teve o pior desempenho em 3 modelos - MLP, SVM e XGBoost - e obteve a menor média e mediana. O Forró obteve o maior desvio padrão, implicando que teve o desempenho mais instável na ao ser classificado. Dessa forma, o gênero mais difícil de ser classificado no geral foi o Sertanejo, seguido por Samba, Bossa Nova e Forró.

Tabela 7 – Gêneros com menor desempenho geral

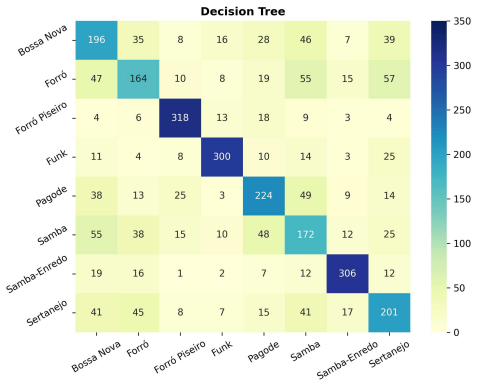
Modelo	**Bossa Nova**	**Forró**	**Pagode**	**Samba**	**Sertanejo**
Decision Tree	52,3%	**43,7%**	59,7%	45,9%	53,6%
KNN	**57,3%**	76,3%	78,7%	72,0%	68,8%
MLP	74,1%	50,4%	91,2%	54,7%	**48,5%**
Random Forest	72,3%	73,6%	84,8%	**66,9%**	74,1%
SVM	66,4%	87,2%	65,1%	84,8%	**57,1%**
XGBoost	75,2%	84,8%	79,2%	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%

Tabela 8 – Gêneros com melhor desempenho geral

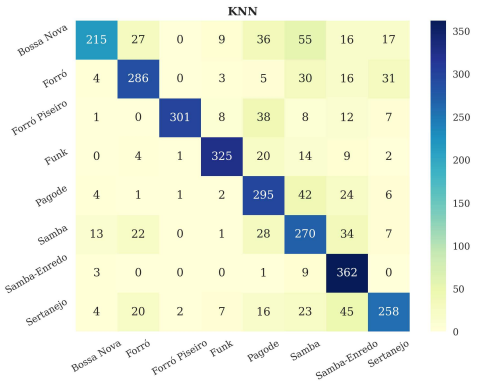
Modelo	**Forró Piseiro**	**Funk**	**Samba-Enredo**
Decision Tree	**84,8%**	80,0%	81,6%
KNN	80,3%	86,7%	**96,5%**
MLP	93,9%	**96,00%**	90,1%
Random Forest	90,9%	**94,9%**	93,1%
SVM	89,9%	**95,7%**	93,3%
XGBoost	94,9%	**95,7%**	93,6%
Voting	**94,4%**	92,8%	92,8%
Média	89,87%	**91,69%**	91,57%
Mediana	90,9%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

Em contrapartida, os gêneros com melhor desempenho foram: Forró Piseiro, Funk e Samba-Enredo. As informações resumidas do desempenho de cada gênero estão descritas na tabela 8. Fica claro que o Funk foi o gênero que melhor performou no geral, porém comparando-o com os outros, nota-se que há um maior equilíbrio entre estes gêneros, sendo o Samba-Enredo o que teve menor desvio padrão.

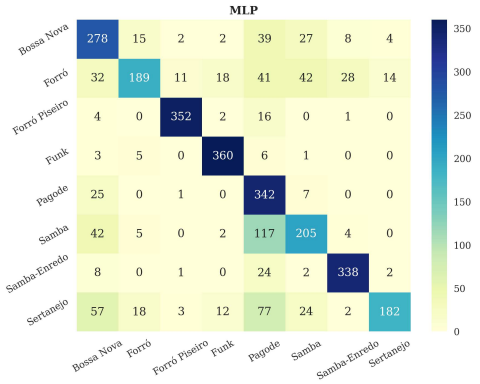
Figura 5 – Matriz de confusão



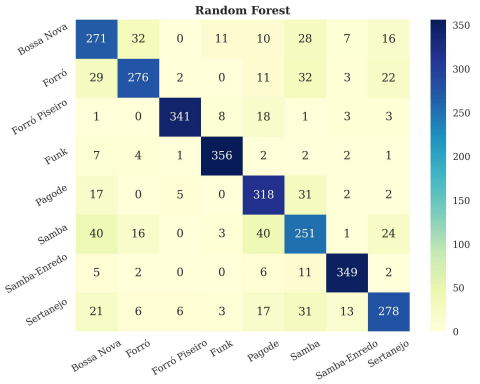
(a) *Decision Tree*



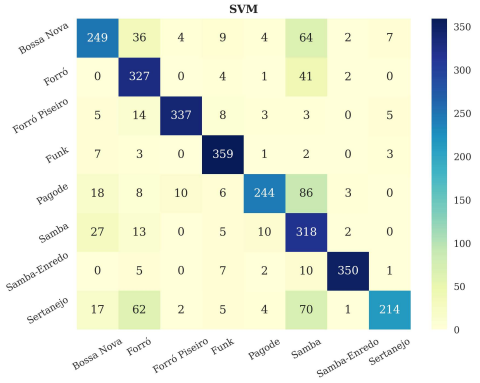
(b) KNN



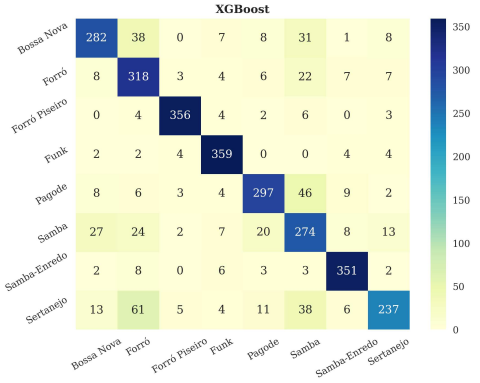
(c) MLP



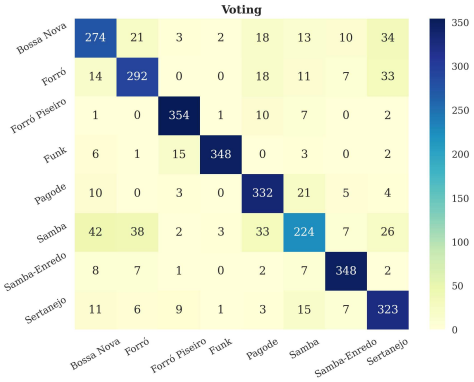
(d) *Random Forest*



(e) SVM

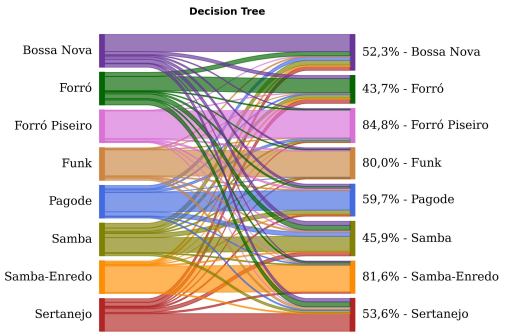


(f) XGBoost

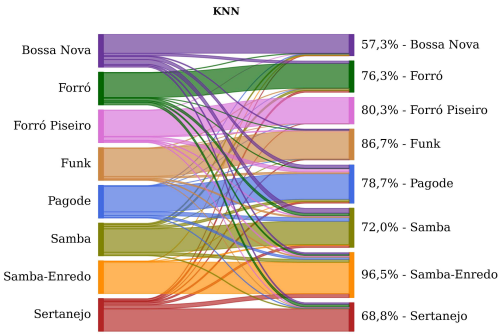


(g) *Voting*

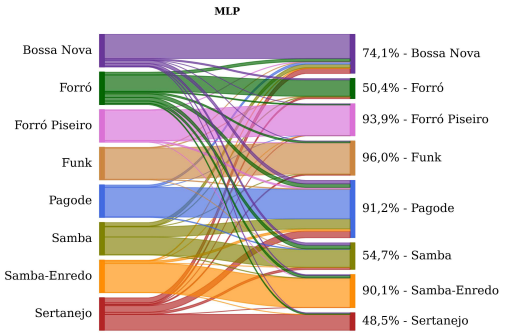
Figura 6 – Acurácia por classe



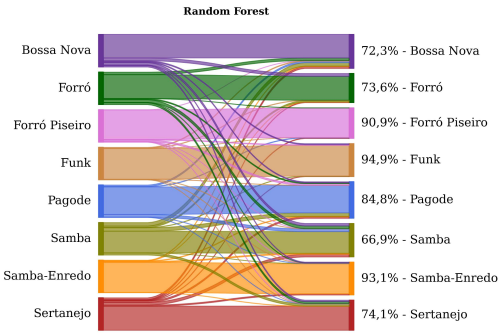
(a) *Decision Tree*



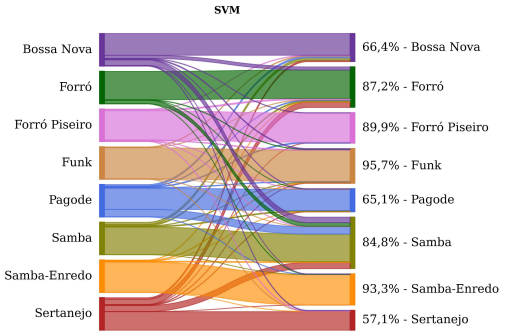
(b) KNN



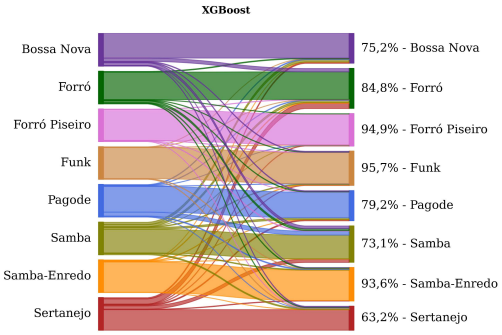
(c) MLP



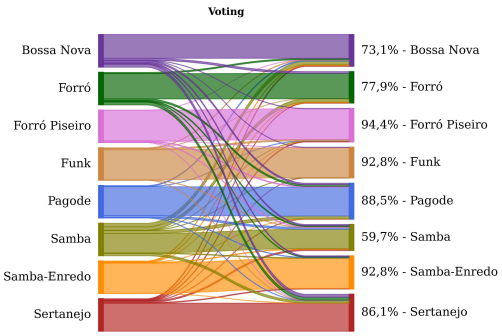
(d) *Random Forest*



(e) SVM

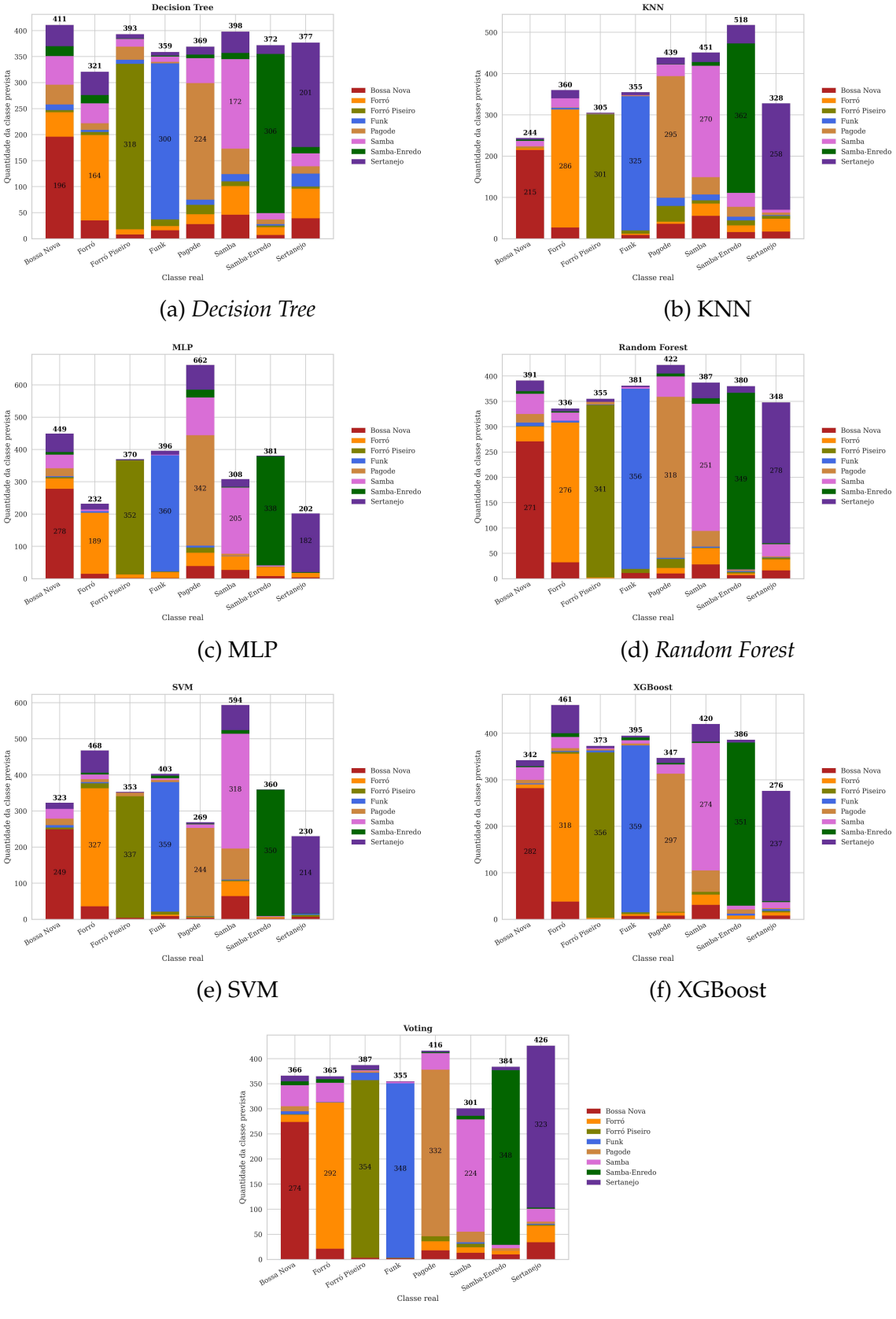


(f) XGBoost



(g) *Voting*

Figura 7 – *Class Predict Error*



(a) *Decision Tree*

(b) KNN

(c) MLP

(d) *Random Forest*

(e) SVM

(f) XGBoost

(g) *Voting*

A tabela 9 mostra a acurácia, precisão média, revocação média e f1-score médio de cada modelo. Os melhores, conforme citado anteriormente, foram: *Voting*, XGBoost e *Random Forest*, e o que menos performou foi o *Decision Tree*. Os modelos KNN, MLP e SVM têm precisão média maior que a revocação, o que indica que alguns gêneros podem ter obtido muitas previsões, a maior parte falsos positivos, enquanto outros receberam menos, porém com maior acurácia. Isso significa que esses modelos tem alta sensibilidade a um tipo de gênero. Para o KNN, a figura 7b expõe que Bossa Nova teve poucas previsões, porém corretas, enquanto Samba-Enredo recebeu maior parte das previsões, o que explica a menor revocação. A figura 7c revela que o ritmo Pagode recebeu o maior número de previsões, sendo muitas erradas, enquanto o Sertanejo o menor número, entretanto com bastantes acertos. Por fim, a figura 7e realça que no SVM, o Samba teve 594 previsões, e o Sertanejo, novamente teve poucas previsões, com uma alta revocação. Esses desequilíbrios, derrubaram a performance geral desses modelos. Os modelos que tiveram maior equilíbrio, indicado pelo f1-score, também foram os que tiveram melhor acurácia.

Tabela 9 – Métricas de desempenho dos modelos

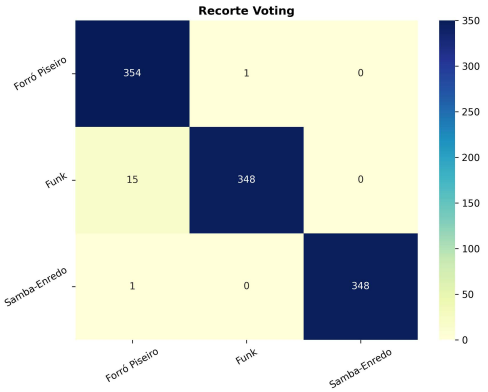
Modelo	**Acurácia**	**Precisão média**	**Revocação média**	**F1-score médio**
Decision Tree	0,627	0,628	0,627	0,627
KNN	0,771	0,792	0,771	0,771
MLP	0,749	0,783	0,749	0,745
Random Forest	0,813	0,816	0,813	0,814
SVM	0,799	0,833	0,799	0,801
XGBoost	0,825	0,832	0,825	0,824
Voting	0,832	0,831	0,832	0,83

Considerando a acurácia, o melhor modelo foi o *Voting*. A tabela 10 exibe as métricas do modelo para cada um dos gêneros musicais. Como neste caso o número de amostras é igual para todos os gêneros, a revocação pode ser considerada como acurácia. Por isso, observa-se que o Samba foi o gênero com menor acurácia, sendo, portanto, o pior desempenho, enquanto os gêneros com maior acurácia foram: Forró Piseiro, Funk e Samba-Enredo. Estes são os mesmos que na análise preliminar (seção 4.1 foram os gêneros que se mostraram mais fáceis de classificar de acordo com o TSNE (figura 3c) e que obtiveram melhor desempenho no geral entre todos os modelos.

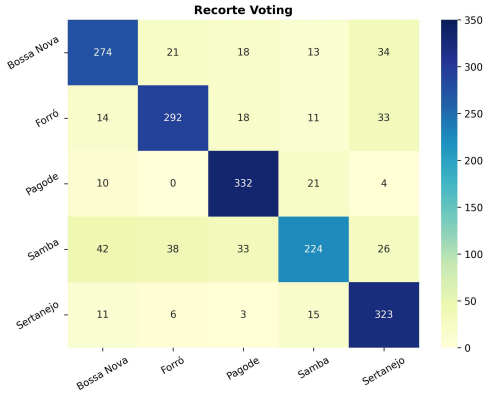
Para comparar os três gêneros musicais com melhor desempenho, foi realizado um recorte da matriz de confusão (Figura 5g), apresentando apenas os gêneros Forró Piseiro, Funk e Samba-Enredo, conforme ilustrado na Figura 8a. A partir desse recorte, foram calculadas as métricas de precisão, revocação e f1-score, além da acurácia geral. Os resultados estão descritos na tabela 11, que evidencia que o modelo foi capaz de

distinguir claramente os três gêneros: Forró Piseiro, Funk e Samba-Enredo. Destaque para o Samba-Enredo, que teve apenas um único falso negativo, evidenciado pelo f1-Score de 0,999. O Funk, por sua vez, foi o que menos desempenhou, com 15 amostras sendo previstas como Forró Piseiro. O número total de amostras classificadas entre as três classes é de 1067, próximos aos 1125 esperados, mostrando que a maioria das previsões esperadas para os três gêneros concentrou-se entre os três, e que poucas vezes foram confundidos com os outros cinco gêneros.

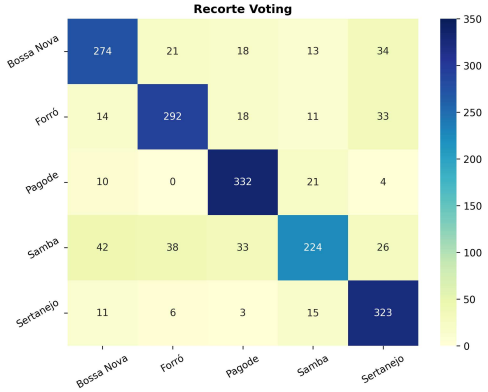
Figura 8 – Recortes matriz de confusão *voting*



(a) *Forró Piseiro, Funk e Samba-Enredo*



(b) *Bossa Nova, Forró, Pagode, Samba e Sertanejo*



(c) *Bossa Nova, Forró, Pagode e Sertanejo*

Seguindo o mesmo processo, foi realizado um novo recorte, desta vez abrangendo os gêneros Bossa Nova, Pagode, Samba, Samba-Enredo e Sertanejo, como apresentado na Figura 8b. As métricas calculadas para esse conjunto estão expostas na tabela 12, a qual evidencia que esses gêneros representaram uma maior dificuldade para o modelo, resultando na queda da acurácia de 98,4% para 79,6%. Dentre esses gêneros, o Samba apresentou o pior desempenho, com uma baixa revocação (0,607), uma diferença de 0,144 em relação à Bossa Nova (0,761), que obteve o segundo pior resultado.

Tabela 10 – 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

Tabela 11 – Métricas entre Forró Piseiro, Funk e Samba-Enredo para *Voting*

Gênero	Precisão	Revocação	F1-Score	Amostras
Forró Piseiro	0,957	0,997	0,977	355
Funk	0,997	0,959	0,978	363
Samba-Enredo	1	0,997	0,999	349
Acurácia				0,984

Tabela 12 – Métricas entre Bossa Nova, Forró, Samba, Pagode e Sertanejo para *Voting*

Gênero	Precisão	Revocação	F1-Score	Amostras
Bossa Nova	0,781	0,761	0,771	360
Forró	0,818	0,793	0,806	368
Pagode	0,822	0,905	0,861	367
Samba	0,789	0,617	0,692	363
Sertanejo	0,769	0,902	0,83	358
Acurácia				0,796

Um novo recorte foi realizado, entre os gêneros Bossa Nova, Forró, Pagode e Sertanejo, que pode ser visto na figura 8c. As métricas para este recorte são demonstradas na tabela 13. Constata-se que o modelo fez o total de 1393 previsões dos 1500 esperados para esse grupo. A acurácia aumentou de 79,6% para 87,7%. Isso demonstra a extrema dificuldade do modelo para classificar o gênero musical Samba.

Tabela 13 – Métricas entre Bossa Nova, Forró, Pagode e Sertanejo para *Voting*.

Gênero	**Precisão**	**Revocação**	**F1-Score**	**Amostras**
Bossa Nova	0,887	0,79	0,835	347
Forró	0,915	0,818	0,864	357
Pagode	0,895	0,96	0,926	346
Sertanejo	0,82	0,942	0,877	343
Acurácia				0,877

É importante destacar as principais características que se sobressaíram nos sete modelos avaliados na etapa final: *Mel*, *MFCC*, *Tempograma* e *Chroma CENS* — o último presente nos modelos MLP e *Voting*. Dessa maneira, o *Mel* e *MFCC* trazem informações espectrais, enquanto o *Tempograma* capta os ritmos e as estruturas temporais típicas de cada gênero, e por fim, o *Chroma CENS* abrange a progressão harmônica e características tonais dos ritmos. As informações sobre os extratores utilizados por cada modelo estão apresentadas na tabela 5, na qual os modelos com melhor acurácia na avaliação conjunta estão destacados em negrito.

5 Conclusão

Este trabalho teve como objetivo principal a criação de uma base de dados contendo gêneros musicais populares brasileiros, destinada ao estudo e à classificação automática desses gêneros. Para validar a eficácia da base de dados, foram aplicados diversos modelos de classificação, com o intuito de verificar se ela estava adequada para essa tarefa.

Os resultados indicam que a base de dados é, de fato, viável para a classificação de gêneros musicais. Dentre os modelos avaliados, o *Voting*, um sistema de votação, apresentou o melhor desempenho, alcançando 83,2% de acurácia, com os gêneros Forró Piseiro, Funk e Samba-Enredo sendo os mais fáceis de serem classificados, em contraste o modelo teve maior dificuldade de prever o gênero Samba. O modelo individual que melhor performou foi o XGBoost com 82,5% de acurácia. Por outro lado, o modelo *Decision Tree* obteve o pior resultado, com uma acurácia de 62,7%. Em relação aos gêneros musicais, no geral, três deles se destacaram por serem mais facilmente reconhecidos pelos modelos: Forró Piseiro, Funk e Samba-Enredo. Já o gênero Sertanejo apresentou maior dificuldade de classificação, seguido de Samba, Bossa Nova e Forró.

Este estudo contribui significativamente para o campo da classificação de gêneros musicais ao disponibilizar uma base de dados abrangendo os seguintes estilos populares brasileiros: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba, Samba-Enredo e Sertanejo. Cada faixa musical na base contém informações extraídas por meio de diversas técnicas, como: *Fourier Tempograma*, *Tempograma*, *Mel*, *MFCC*, *Chroma STFT*, *Chroma CQT*, *Chroma CENS*, *Tonnetz*, *ZCR*, *Spectral Centroid*, *Spectral Roll Off*, *Spectral Bandwidth* e *RMS*. Esses dados permitem uma análise detalhada das faixas musicais, abrangendo aspectos como timbre, harmonia e andamento.

Entretanto, algumas limitações devem ser consideradas. A categorização das músicas, feita manualmente, pode conter equívocos que impactam a precisão dos modelos. Além disso, não foi implementado um filtro para evitar que faixas do mesmo artista ou partes distintas de uma mesma faixa estivessem simultaneamente nos conjuntos de treino e teste, o que pode ter influenciado os resultados. Outro ponto é a alta repetição de artistas em determinados gêneros, o que pode ter facilitado a classificação nesses casos.

Para trabalhos futuros, sugere-se:

- A inclusão de outros gêneros musicais brasileiros para ampliar a abrangência da base de dados;

- A aplicação de redes neurais convolucionais para aprimorar a classificação dos gêneros;
- A remoção de características que não contribuam significativamente para a classificação;
- O uso de técnicas de aprendizado não supervisionado para identificar com maior precisão os limites entre os gêneros;
- A revisão da categorização manual das músicas, verificando se as faixas identificadas como Samba ou Pagode, por exemplo, realmente pertencem a essas categorias.

Essas propostas visam aprimorar a base de dados e fortalecer sua utilidade como ferramenta para o estudo da música brasileira e seus variados estilos.

Referências

ALONSO, G. **Cowboys do Asfalto**: Música sertaneja e modernização brasileira. 2011. Tese (Doutorado em História) — Programa de Pós-Graduação em História, Universidade Federal Fluminense, Niterói, 2011.

BENNET, R. **Uma breve História da Música**. Rio de Janeiro: Jorge Zahar Editor, 1986.

BEZERRA, J. **Funk: a batida dos bailes cariocas que contagiou o Brasil**. 1. ed. São Paulo: Panda Books, 2017.

BROWN, J. C. Calculation of a constant q spectral transform. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 89, n. 1, p. 425–434, 1991.

BÜHLMANN, P. Bagging, boosting and ensemble methods. **Handbook of Computational Statistics**, p. 39, 2012.

CALDAS, W. **O que é a música sertaneja?** São Paulo: Brasiliense, 1987.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. Nova York: Association for Computing Machinery, 2016. p. 785–794. Disponível em: <https://doi.org/10.1145/2939672.2939785>.

CONCEIÇÃO, J. *et al.* Applying supervised learning techniques to brazilian music genre classification. **2020 XLVI Latin American Computing Conference (CLEI)**, p. 102–107, 10 2020.

COSTA, C.; VALLE, J.; KOERICH, A. Automatic classification of audio data. In: **2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)**. [S.l.: s.n.], 2004. v. 1, p. 562–567 vol.1.

DEFERRARD, M. *et al.* **FMA: A DATASET FOR MUSIC ANALYSIS**. 2017. Disponível em: <https://arxiv.org/pdf/1612.01840.pdf>. Acesso em: 16 fev. 2023.

DIAS, I.; DUPAN, S. **O que é o Forró?**: um pequeno apanhado da história do forró. 3. ed. Campina Grande: Meroveu, 2022.

FARAJZADEH, N.; SADEGHZADEH, N.; HASHEMZADEH, M. Pmg-net: Persian music genre classification using deep neural networks. **Entertainment Computing**, v. 44, p. 100518, 2023. ISSN 1875-9521. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1875952122000428>.

GÉRON, A. **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow**: Concepts, tools, and techniques to build intelligent systems. 2. ed. Sebastopol: O'Reilly, 2019.

HARTE, C.; SANDLER, M.; GASSER, M. Detecting harmonic change in musical audio. In: **Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia**. New York: Association for Computing Machinery, 2006. p. 21–26. Disponível em: <https://doi.org/10.1145/1178723.1178727>.

HASIB, K. M. *et al.* Bmnet-5: A novel approach of neural network to classify the genre of bengali music based on audio features. **IEEE Access**, v. 10, p. 108545–108563, 2022.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2009. ISBN 978-0-387-84858-7.

HOSSAIN, M. R.; TIMMER, D. Machine learning model optimization with hyper parameter tuning approach. **Global Journal of Computer Science and Technology: D - Neural & Artificial Intelligence**, Global Journals, v. 21, n. 2, 2021. ISSN 0975-4172. Disponível em: <https://d1wqtxts1xzle7.cloudfront.net/98220794/2-Machine-Learning-Model-Optimization-libre.pdf>.

JANNOTI JÚNIOR, J. **Gêneros musicais em ambientações digitais.** Belo Horizonte: PPGCOM/UFMG, 2020.

JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem: A systematic study. **Intelligent Data Analysis**, IOS Press, v. 6, n. 5, p. 429–449, 2002.

KLAPURI, A.; DAVY, M. (Ed.). **Signal Processing Methods for Music Transcription**. New York: Springer, 2006.

KOTSIANTIS, S.; ZAHARAKIS, I.; PINTELAS, P. Machine learning: A review of classification and combining techniques. **Artificial Intelligence Review**, v. 26, p. 159–190, 11 2006.

MAMUN, M. A. A. *et al.* Bangla music genre classification using neural network. In: **2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)**. [S.l.: s.n.], 2019. p. 397–403.

MÜLLER, M. **Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications**. 1. ed. Cham: Springer, 2015.

MÜLLER, M.; BALKE, S. **Short-Time Fourier Transform and Chroma Features**. International Audio Laboratories Erlangen, 2018. Disponível em: <https://www.audiolabs-erlangen.de/content/05_fau/professor/00_mueller/02_teaching/2018s_apl/LabCourse_STFT.pdf>.

PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

QUADROS JÚNIOR, J. F. S. d. **Música Brasileira**. Belo Horizonte: PPGCOM/UFMG, 2019.

RAMRAJ, S. *et al.* Experimenting xgboost algorithm for prediction and classification of different datasets. **International Journal of Control Theory and Applications**, v. 9, n. 40, p. 651–662, 2016.

SAMMUT, C.; WEBB, G. I. (Ed.). **Encyclopedia of Machine Learning**. Boston: Springer, 2011.

SCARINGELLA, N.; ZOIA, G.; MLYNEK, D. Automatic genre classification of music content: a survey. **IEEE Signal Processing Magazine**, v. 23, n. 2, p. 133–141, 2006.

SCHEDL, M. **Automatically Extracting, Analyzing, and Visualizing Information on Music Artists from the World Wide Web**. 2008. Tese (Doutorado em Ciências Técnicas) — Johannes Kepler University Linz, 2008.

SCHÖRKHUBER, C.; KLAPURI, A. Constant-q transform toolbox for music processing. In: **7th Sound and Music Computing Conference (SMC)**. Barcelona: [s.n.], 2010.

SHARMA, G.; UMAPATHY, K.; KRISHNAN, S. Trends in audio signal feature extraction methods. **Applied Acoustics**, v. 161, p. 107201, 2020.

SHIROL, S.; KATHIRESAN, R. S. A comprehensive survey of music genre classification using audio files. **International Journal of Enhanced Research in Science, Technology & Engineering**, v. 12, p. 183–192, jun. 2023.

SILLA JR., C. N.; KAESTNER, C. A. A.; KOERICH, A. L. Automatic music genre classification using ensemble of classifiers. In: **2007 IEEE International Conference on Systems, Man and Cybernetics**. [S.l.: s.n.], 2007. p. 1687–1692.

SILLA JR., C. N.; KOERICH, A. L.; KAESTNER, C. A. A. The latin music database. In: **Proceedings of the 9th International Conference on Music Information Retrieval**. Philadelphia: ISMIR, 2018. p. 451–456.

SILVA, C. Cardoso da. Aspectos históricos das festas e festividades de forró no Brasil. **Em Tempo de Histórias**, v. 1, n. 40, set 2022. Disponível em: <https://periodicos.unb.br/index.php/emtempos/article/view/42122>.

SILVA, D.; GOMES, C. Modelo de aprendizado de máquina para classificação de gêneros musicais populares da região amazônica legal internacional. **Revista Eletrônica de Iniciação Científica em Computação**, v. 20, n. 4, dez. 2022. Disponível em: <https://journals-sol.sbc.org.br/index.php/reic/article/view/2772>.

SOUSA, J. Martins de; PEREIRA, E. T.; VELOSO, L. R. A robust music genre classification approach for global and regional music datasets evaluation. In: **2016 IEEE International Conference on Digital Signal Processing (DSP)**. [S.l.: s.n.], 2016. p. 109–113.

STURM, B. A survey of evaluation in music genre recognition. **Adaptive Multimedia Retrieval**, 2012.

SUTHAHARAN, S. **Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning**. Springer US, 2016. ISSN 2197-7968. ISBN 9781489976413. Disponível em: <http://dx.doi.org/10.1007/978-1-4899-7641-3>.

TEIXEIRA, P. B. **Do Samba à Bossa Nova: Uma Invenção de Brasil**. 2011. Dissertação (Mestrado em Estudos Literários) — Universidade Federal de Juiz de Fora, Juiz de Fora, 2011. Disponível em: <https://repositorio.ufjf.br/jspui/bitstream/ufjf/2128/1/pedrobustamanteteixeira.pdf>.

TROTTA, F. Música popular e qualidade estética: estratégias de valoração na prática do samba. **Intercom: Revista Brasileira de Ciências da Comunicação**, v. 34, n. 2, p. 117–138, 2011. Disponível em: <https://www.revistaintercom.org.br/index.php/revistaintercom/article/view/1726>.

TROTTA, F. Mussum, "Os Originais do Samba" e a sonoridade do pagode carioca. **Revista Famecos**, v. 23, n. 2, p. 1–15, 2016. Disponível em: <https://www.redalyc.org/articulo/4955>.

TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. **IEEE Transactions on Speech and Audio Processing**, v. 10, n. 5, p. 293–302, 2002.

VIANA, L. R. O funk no brasil: Música desintermediada na cibercultura. In: **Revista Sonora, Unicamp**. [S.l.: s.n.], 2010. v. 3, n. 5.

YU, T.; ZHU, H. Hyper-parameter optimization: A review of algorithms and applications. **ArXiv**, abs/2003.05689, 2020. Disponível em: <https://api.semanticscholar.org/CorpusID:212675087>.

ZAMPRONHA, M. d. L. S. **Da música, seus usos e recursos.** São Paulo: Editora UNESP, 2002.