SONBRA: uma base de dados para a classificação automática de gêneros brasileiros

João Marcos de Oliveira Santana[1], **Anderson Vinícius Ribeiro Amorim**[1], **Marcos Abreu de Oliveira**[1], **Cláudio Gomes** [1, 2]

[1]Grupo de pesquisa de COmputação mUsical e Tecnologias Emergentes (COUT-E)
Ciências da Computação - Universidade Federal do Macapá (UNIFAP)

[2]Pós-graduação em Ciência da Computação - Universidade Federal de Pernambuco (UFPE)

joaosantana1998,marcosabr97}@gmail.com, {andyreborn@hotmail.com.br, claudiorogerio@unifap.br

Abstract. *This work presents SONBRA, a model for automatic classification of the musical genres Bossa Nova, Forró, Forró Piseiro, Pagode, Samba, Samba-Enredo, and Sertanejo. To achieve this, a music dataset was created containing 10,000 fragments of 30-second audio clips, with 785 parameters extracted from the following features: Chroma CENS, Chroma CQT, Chroma STFT, Fourier Tempograma, Mel-espectrograma, MFCC, RMS, Spectral Bandwith, Spectral Centroid, Spectral Roll-Off, Tempograma, Tonnetz, and ZCR. From each track, 30-second segments were taken from the beginning, middle, end, middle-to-beginning, and middle-to-end. Experiments were conducted using the computational models Decision Tree, KNN, MLP, Random Forest, SVM, XGBoost and an ensemble model, specifically a Voting system. The Voting system achieved the best performance, with an accuracy of* ***83.2%****, while the worst performance was observed with the Decision Tree model, at* ***62.7%****. The most effective results were obtained from combinations of the features MFCC, Mel, Tempogram, and Chroma CENS.*

Keywords — machine learning, music information retrieval, music genre classification

Resumo. *Este trabalho apresenta o SONBRA, um modelo de classificação automática dos gêneros musicais Bossa Nova, Forró, Forró Piseiro, Pagode, Samba, Samba-Enredo e Sertanejo. Para isso, criou-se uma base de dados musical contendo 10000 fragmentos de faixas musicais de 30s com 785 parâmetros das seguintes características extraídas: Chroma CENS, Chroma CQT, Chroma STFT, Fourier Tempograma, Mel-espectrograma, MFCC, RMS, Spectral Bandwith, Spectral Centroid, Spectral Roll-Off, Tempograma, Tonnetz e ZCR. A partir de cada faixa musical retirou-se fragmentos de 30s do início, meio, fim, do meio ao início e do meio ao final. Foram realizados testes com os modelos computacionais Decision Tree, KNN, MLP, Random Forest, SVM, XGBoost e um ensemble, mais especificamente um sistema de Voting. O Voting obteve o melhor desempenho com 83,2% de acurácia, enquanto o pior foi do Decision Tree com 62,7%. E as características que melhor geraram resultados foram combinações entre MFCC, Mel, Tempograma e Chroma CENS.*

Palavras-chave — aprendizado de máquina, recuperação de informação musical, classificação de gêneros musical

1. Introdução

Além do seu uso para o entretenimento, a música constitui uma forma de expressão, comunicação e mobilização, ligadas diretamente a contextos socioculturais e ideológicos de uma determinada época. Sua estrutura é composta por harmonia, melodia, ritmo, timbre e textura que expressam sua diversidade sonora. A categorização por gêneros pode ser definida por esses elementos, mas também por aspectos culturais e históricos como origem geográfica, tradições e influências artísticas. Em um contexto nacional, a riqueza musical vem através da interculturalidade de influências, como a indígena, africana e europeia, que foram a base para o desenvolvimento de diversos estilos musicais marcantes para a identidade do país [1].

Aplicações e serviços utilizam dados extraídos de músicas para classificar, organizar e fazer sugestões musicais aos usuários. Nesse contexto, as ferramentas de classificação de gêneros musicais se tornam essenciais para encontrar similaridades de áudio. Essa área de pesquisa é amplamente estudada, com pesquisas resultando em bases de dados, sistemas e modelos de aprendizado de máquina [2]. No entanto, a maioria dessas é focada em música internacional e em gêneros mais populares como pop e rock. A classificação de gêneros musicais brasileiros sofre uma carência de bases de dados balanceadas e com quantidade de amostras relevante de gêneros regionais, pois as que contêm gêneros nacionais possuem poucas amostras ou são mal balanceadas, não sendo efetivas para a criação de modelos de aprendizado de máquina eficientes [3].

Diante dessa lacuna, este trabalho propõe a criação de uma base de dados estruturada e balanceada contendo amostras dos gêneros populares brasileiros: Bossa Nova, Forró, Forró Piseiro, Funk, Pagode, Samba Samba-Enredo e Sertanejo. A base será composta por características musicais extraídas de faixas musicais obtidas através de plataformas de *streaming*, sendo disponibilizada publicamente para a comunidade acadêmica. Para validar a eficácia e aplicabilidade em tarefas de classificação automática, a base de dados experimental foi usada para treino de modelos de aprendizado de máquina e avaliação em 6 algoritmos comuns na categorização musical, buscando fortalecer a representação de gêneros regionais em estudos de classificação musical e incentivar novas pesquisas e aplicações nessa área.

2. Trabalhos relacionados

O trabalho de [4] apresentou o PMG-NET, uma rede neural convolucional projetada para classificar gêneros musicais persas. Para isso, desenvolveu-se uma base de dados (PMG-Data) contendo 500 músicas divididas em cinco gêneros: monody, pop, rap, rock e internacional. Os experimentos foram divididos em dois, onde o primeiro utilizou amostras de 32 segundos de cada faixa e 5-folds de validação cruzada, com acurácia média de 76% e máxima de 79%; o segundo dividiu cada faixa em seis segmentos, totalizando 3 mil amostras. As características foram extraídas com MFCC. O treinamento da rede obteve acurácia máxima de 86%. Os autores também compararam o PMG-NET com algoritmos tradicionais de aprendizado de máquina, com o SVM usando STFT apresentando o melhor resultado com 65% de acurácia.

O AMPOP, de [5], é uma base de dados desenvolvida para a criação de um classificador automático de músicas populares da região amazônica. A base contém mil músicas, com 125 faixas para cada gênero: brega, andino, carimbó, cúmbia, merengue, pasillo, salsa e vaqueirada. Foram extraídos 785 parâmetros do início, meio e fim das faixas. Dos modelos usados para testar o classificador, o SVC obteve a maior acurácia e o MLP a menor, com 61,33% e 56,79%, respectivamente.

A *Latin Music Database* (LMD), de [6], contém 3.227 faixas, distribuídas em dez gêneros musicais latino-americanos, como salsa, tango e samba, com metadados de artista, álbum e gênero, rotulados por dois especialistas. A LMD suporta diversas tarefas de recuperação de informação musical, incluindo classificação de gêneros e análise de ritmos, porém não se encontra disponível para utilização.

O banco de dados GTZAN, dos autores [7], possui 100 amostras de 30 segundos para cada um dos dez gêneros musicais: clássico, country, disco, hip-hop, jazz, rock, blues, reggae, pop e metal. Os autores analisaram características musicais com os algoritmos Gaussian Classifier, Gaussian Mixture Model (GMM) e KNN, calculando os resultados com validação cruzada em 10-folds com 100 interações. O GMM obteve maior acurácia com jazz, enquanto rock obteve o menor, com 75% e 40% respectivamente.

O *Free Music Archive* (FMA) é um conjunto de dados desenvolvido por [8], contendo 16 gêneros principais e 161 subgêneros. Os autores dividiram o conjunto em três categorias: completo, com 106.574 faixas de gênero e subgênero; médio, com 25 mil amostras de 30 dos gêneros principais; e pequeno, com mil músicas para os oito gêneros mais ouvidos. Os autores também apontam que o conjunto é tendencioso em relação aos gêneros rock e eletrônico.

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

Tabela 1: SONBRA e trabalhos relacionados

A BMNet-5, é uma rede neural profunda projetada por [9] para classificar música bengalis. 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 [10]. A rede possui 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*. O modelo alcançou o desempenho ideal em 187 *epochs* com tamanho de lote de 5, com 90,32% de acurácia.

3. Base de Dados SONBRA

Para a criação da base de dados, primeiramente selecionou-se os gêneros musicais brasileiros, depois houve a coleta de faixas e a extração de características musicais. Estas etapas estão descritas a seguir.

3.1. Seleção dos gêneros musicais

A música brasileira contém diversos ritmos populares e para este trabalho foram escolhidos: Samba, Samba-Enredo, Pagode, Bossa Nova, Funk, Forró, Forró Piseiro, e Sertanejo. O **Samba** é de origem afro-brasileira, que tem forte uso de instrumentos percussivos e com um síncope característica . Dele surgiram outros ritmos como o **Samba-Enredo**, que é uma variação usada nos populares desfiles de escola de samba e o **Pagode** que nos anos 90 tornou-se mais popular com a comercialização musical mais moderna [1]. A **Bossa Nova** surgiu como uma tentativa de modernizar a música brasileira, para isso usou-se de elementos característicos dos ritmos brasileiros, mas também influencias dos ritmos estrangeiros como *Jazz* e *Blues*. Outro gênero marcado pela influência estrangeira foi o **Funk**, que se tornou popular nas periferias do Rio de Janeiro através de festas em que o público criava suas próprias letras em cima das músicas do *Funky* americano e, posteriormente, do *Miami Bass* [11, 12]. Por sua vez, o **Forró**, surgiu como nome de uma festa em que se tocavam vários ritmos. Dessa maneira, tornou-se o nome de um conjunto de ritmos, e mais a frente, o nome de um gênero, sendo característico o uso da Zabumba, Triângulo e Sanfona. O **Forró Piseiro** é uma variante do Forró que utiliza instrumentos eletrônicos como teclados e sintetizadores, guitarras além da sanfona e outros instrumentos [13]. Por último, o **Sertanejo** é uma modernização da musica caipira, originária do interior de São Paulo e outros estados do Centro-Sul, sendo característico uso de duplas e trios para sua execução e viola (violão) [1, 14, 15].

3.1.1. Criação da base de dados

A partir da plataforma *Youtube* foram selecionadas 2000 faixas (250 por ritmo), priorizando gravações em estúdio para garantir melhor qualidade sonora e consistência. Posteriormente, as músicas foram obtidas por meio da ferramenta *Pytube*[1] no formato MP4. Para diminuir eventuais interferências e ruídos, as faixas foram convertidas para o formato WAV usando a ferramenta *ffmpeg*[2]. Para a análise, foram selecionadas amostras de 30 segundos de diferentes partes de cada faixa, incluindo o início (*begin*), meio (*middle*), do meio para o início (*middle_1*), do meio para o final (*middle_2*) e final (*end*). A figura 1 ilustra as etapas seguidas para a criação da base de dados. Dessa forma, as 250 faixas de cada gêneros foram transformadas em 1250 fragmentos (250 x 5 tipos de fragmentos) de 30s. Assim, a base de dados musical possui ao todo 10000 (1250 X 8 gêneros) fragmentos musicais de 30s, que serão usadas apenas para fins acadêmicos. A base pode ser acessada em: `https://drive.google.com/drive/folders/1sDHdirK2w9ZmH8_F45KQ36Z6n_RKhA8C`.

3.1.2. Extratores musicais

De acordo com a figura 1, utilizando a biblioteca *Librosa*[3] para a construção da base de dados, aplicou-se nas amostras de 30 segundos os extratores musicais de *Tempograma, Mel-espectrograma, Fourier Tempograma, Chroma STFT, Chroma CENS, Chroma CQT, MFCC, Tonnetz, RMS, ZCR, Spectral Roll Off, Spectral Centroid* e *Spectral Bandwidth.*

A extração das características gera uma representação numérica, que pode assumir a forma de escalares, vetores ou matrizes. Com o intuito de diminuir recursos computacionais, foram realizadas análises estatísticas considerando a média e a mediana como valores de uma determinada característica. Isto resultou em

[1] https://pytube.io/en/latest/
[2] https://www.ffmpeg.org/
[3] https://librosa.org/doc/latest/index.html

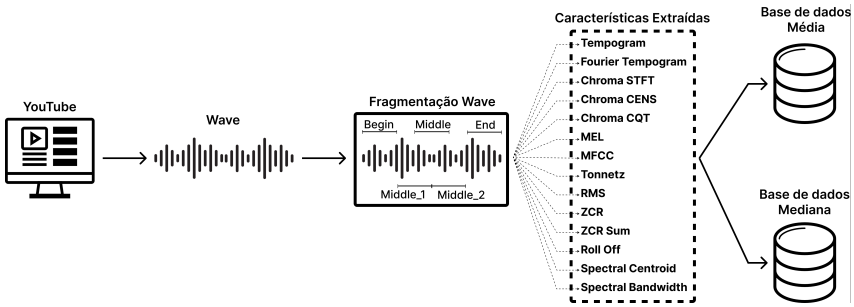

Figura 1: Visão geral das etapas para a construção da base de dados

duas bases de dados, como ilustra a figura 1. Este processo impossibilita que os dados sejam retornados a originalidade, ou seja, é impossível reconverter-los para áudio.

O *tempograma* é uma representação que evidencia os tempos mais relevantes em batidas por minuto (BPM) ao longo da música e pode ser extraído por diferentes métodos, a partir da detecção de mudanças significativas no sinal de áudio por meio de uma função chamada *novelty*, que identifica os momentos em que notas são tocadas. Entre as técnicas utilizadas estão o Fourier Tempograma, que enfatiza o que se chama de harmônicos de tempo, e o Tempograma Autocorrelacionado, que enfatiza os subharmônicos [16].

O *mel-espectrograma (Mel)* utiliza-se da escala mel para representar os sons conforme a percepção humana, enquanto o Mel-*Frequency Cepstrum Coeficient* (MFCC) utiliza o mesmo conceito, porém usa coeficientes cepstrais, aplicando uma Transformada Discreta do Cosseno ao espectro mel [17].

O *Chroma*, baseado nas doze classes de altura da escala temperada, representa notas que compartilham identidade sonora mesmo em diferentes oitavas. É extraído por técnicas como *Chroma STFT*, *Chroma CQT* e *Chroma CENS*, esta última aplicando normalização e suavização para reduzir ruídos e enfatizar estruturas relevantes [18, 16].

As características espectrais complementam a análise dos sinais musicais. O *Spectral Centroid* indica a concentração média das frequências [19], enquanto o *Spectral Bandwidth* e o *Spectral Roll-Off* descrevem a dispersão e a distribuição de energia no espectro [17].

Extratores de domínio temporal fornecem informações sobre as propriedades temporais dos sinais de áudio ao longo do tempo. O *Zero Crossing Rate* (ZCR) mede as transições de sinal em um *frame*, associadas a ruído, enquanto o *Root Mean Square (RMS)* calcula a energia geral das amplitudes, auxiliando na detecção da presença e intensidade do som [17].

O Tonnetz captura relações harmônicas entre notas e acordes em um contexto tonal, principalmente das quintas perfeitas e terças menores e maiores, através de um vetor Tonal Centroid de seis dimensões [20].

4. Avaliação

A base de dados (10.000 amostras) foi dividida para ser usada na avaliação. Primeiramente em duas bases: desenvolvimento (70%) e teste (30%). A primeira, base de desenvolvimento, foi novamente dividida em duas: base de treino (80%) e base de validação (20%). A divisão da base de dados pode ser vista na figura 2. Após isso, iniciou-se a avaliação, em busca do modelo que tivesse melhor acurácia.

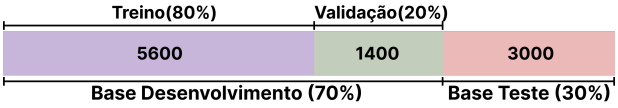

Figura 2: Divisão Base

4.1. Modelos computacionais

Dentre as diversas opções de modelos computacionais foram usados nesta abordagem aqueles que tiveram um desempenho razoável, sendo estes: *Decision Tree*, *K-Nearest Neighbors* - KNN, *Multilayer Perceptron* - MLP, *Random Forests*, *Support Vector Machines* - SVM e *eXtreme Gradient boosting* (XGBoost).

As *Decision Trees* são modelos hierárquicos que particionam os dados com base em critérios pré-definidos, organizando as decisões em uma estrutura binária cujas folhas representam as classes finais [21, 22]. O KNN é um modelo de classificação e regressão que classifica uma instância com base na maioria das classes de seus vizinhos mais próximos, pressupondo que eles tendem a pertencer à mesma classe [23]. Por sua vez, o MLP representa um modelo de rede neural com múltiplas camadas interligadas, onde os dados são processados por funções de ativação e ajustados por retropropagação para minimizar os erros de classificação. Esse processo iterativo permite à rede aprender representações progressivamente mais precisas ao longo do tempo [24, 22].

Já As *Random Forests* melhoram esse processo utilizando a técnica *ensemble*, empregando um conjunto de árvores construídas sobre subconjuntos de dados, combinando o resultado das árvores [21]. Ao passo que as SVMs realizam a classificação ao identificar o hiperplano que melhor separa as classes, utilizando vetores de suporte para definir as margens dessa separação. Quando os dados não são linearmente separáveis, uma função *kernel* é aplicada para mapeá-los em um espaço de maior dimensionalidade, onde a separação se torna possível [22]. Por fim, o XGBoost é uma extensão otimizada do *gradient boosting*, que utiliza paralelismo e armazenamento em cache para acelerar o treinamento, sendo especialmente eficaz em bases de dados volumosas [25].

4.2. Hiperparâmetros e otimizações

Cada modelo computacional foi submetido ao processo de busca da melhor combinação de hiperparâmetros e extratores nas etapas descritas abaixo. Primeiramente todos os extratores foram testados individualmente, e foram selecionados os cinco que tiveram acurácia: *Mel, MFCC, Tempograma, Chroma CENS* e *Chroma CQT*. Posteriormente foram combinados em pares, e por fim em trios, como nos exemplos:*Mel* com *MFCC* e; *Mel, MFCC, Tempograma*.

4.2.1. Etapa 1 - Validação Cruzada

Nesta etapa, os modelos foram treinados (base de treino) com validação cruzada estratificada de 10 *folds*, e para cada tipo, aqueles que tiveram maior acurácia média foram selecionados.

4.2.2. Etapa 2 - Validação

Os modelos foram treinados (base de treino) e validados (base de validação), e o representante de cada modelo computacional foi escolhido com base na acurácia mais próxima da acurácia média na etapa 1.

4.2.3. Sistema *Voting*

Foi proposto um sistema de votação (*Voting*) com os 6 modelos selecionados anteriormente na etapa 1, testando as abordagens ***hard*** (voto por maioria) e ***soft*** (média das probabilidades). Os extratores escolhidos — MFCC, Tempograma, Mel e Chroma CENS — foram os que apresentaram melhor desempenho na etapa 1. Ambos os sistemas passaram pelas mesmas etapas de avaliação anteriores.

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

Tabela 2: Melhores acurácias etapas 1 (V) e 2 (CV)

4.3. Avaliação Final

Para avaliação final, os 14 modelos foram treinados com a base de desenvolvimento e testados com a base de teste. No fim, foram selecionados 7 modelos, um de cada modelo computacional e o modelo que obteve a melhor acurácia foi considerado o melhor modelo.

5. Resultados

Os modelos selecionados como os melhores das etapas 1 (validação cruzada) e 2 (validação) estão descritos na tabela 2, que além das acurácias, apresentam os hiperparâmetros e os extratores de cada modelo.

Os resultados da avaliação final estão demonstrados na tabela 3, que além da acurácia, indicam a precisão média, revocação média e f1-score médio, A coluna Id refere-se ao modelo correspondente na tabela 2, em que foram destacados para melhor identificação. Quanto a avaliação final, o modelo que teve o melhor resultado foi o *Voting* com acurácia 0,832. Dentre os modelos restantes, o XGBoost obteve o melhor desempenho com acurácia de 0,825. Por sua vez, Árvore de Decisão teve a pior performance com 0,627 de acurácia.

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

Tabela 3: Avaliação final dos modelos

A figura 3 mostra a matriz de confusão *Voting*. Observa-se que os gêneros que o modelo melhor previram foram: forró piseiro, funk e samba-enredo. Já o pior foi o Samba, seguido de perto de bossa nova e forró. Nota-se que o Samba teve maior quantidade de falsos negativos, enquanto o sertanejo teve um número maior de previsões e concentrou o maior número de falsos positivos.

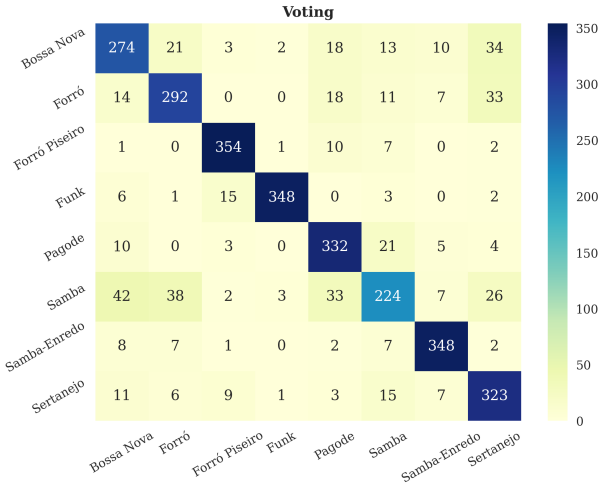

Figura 3: Matriz de Confusão Voting

6. Conclusões

A base de dados foi testada com diferentes modelos de aprendizado de máquina, e os resultados demonstraram sua viabilidade. O modelo com melhor desempenho foi o sistema de votação *Voting*, que atingiu uma acurácia de 83,2%, seguido pelo XGBoost, com 82,5%. Em contrapartida, o *Decision Tree* obteve o pior resultado, com 62,7% de acurácia. Gêneros como Forró Piseiro, Funk e Samba-Enredo foram os mais facilmente classificados, enquanto Samba e Sertanejo apresentaram maiores dificuldades.

A aplicação das diversas técnicas de extração de características musicais, para capturar aspectos relevantes das amostras, permite análises mais aprofundadas e abre espaço para avanços na pesquisa em classificação musical de gêneros brasileiros.

No entanto, o estudo também apresenta limitações, como possíveis erros na categorização manual das faixas e a ausência de filtros para evitar a presença de músicas do mesmo artista nos conjuntos de treino e teste. Essas questões podem ter influenciado os resultados de classificação. Para trabalhos futuros, são sugeridas melhorias como a ampliação da base com mais gêneros, uso de redes neurais convolucionais, remoção de atributos

irrelevantes, técnicas não supervisionadas para identificar fronteiras entre gêneros e uma revisão mais criteriosa das classificações manuais.

Referências

[1] João Fortunato Soares de Quadros Júnior. *Música Brasileira*. PPGCOM/UFMG, Belo Horizonte, 2019.

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

[3] Júlia Conceição, Rosiane Freitas, Bruno Gadelha, Joao Kienen, Sergio Anders, and Brendo Cavalcante. Applying supervised learning techniques to brazilian music genre classification. *2020 XLVI Latin American Computing Conference (CLEI)*, pages 102–107, 10 2020.

[4] Nacer Farajzadeh, Nima Sadeghzadeh, and Mahdi Hashemzadeh. Pmg-net: Persian music genre classification using deep neural networks. *Entertainment Computing*, 44:100518, 2023.

[5] Douglas Silva and Cláudio Gomes. Modelo de aprendizado de máquina para classificação de gêneros musicais populares da região amazônica legal internacional. *Revista Eletrônica de Iniciação Científica em Computação*, 20(4), dez. 2022.

[6] Carlos Nascimento Silla, Jr., Alessandro L. Koerich, and Celso A. A. Kaestner. The latin music database. In *Proceedings of the 9th International Conference on Music Information Retrieval*, pages 451–456, Philadelphia, 2018. ISMIR.

[7] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[8] Michaël Deferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. Fma: A dataset for music analysis, 2017.

[9] Khan Md. Hasib, Anika Tanzim, Jungpil Shin, Kazi Omar Faruk, Jubayer Al Mahmud, and M. F. Mridha. Bmnet-5: A novel approach of neural network to classify the genre of bengali music based on audio features. *IEEE Access*, 10:108545–108563, 2022.

[10] Md. Afif Al Mamun, Imamul Kadir, AKM Shahariar Azad Rabby, and Abdullah Al Azmi. Bangla music genre classification using neural network. In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages 397–403, 2019.

[11] Júlia Bezerra. *Funk: a batida dos bailes cariocas que contagiou o Brasil.* Panda Books, São Paulo, 1 edition, 2017.

[12] Lucina Reitenbach Viana. O funk no brasil: Música desintermediada na cibercultura. In *Revista Sonora, Unicamp*, volume 3, 2010.

[13] Ivan Dias and Sandrinho Dupan. *O que é o Forró?* Meroveu, Campina Grande, 3 edition, 2022.

[14] Gustavo Alonso. *Cowboys do Asfalto*. Doutorado em história, Programa de Pós-Graduação em História, Universidade Federal Fluminense, Niterói, 2011.

[15] Waldenyr Caldas. *O que é a música sertaneja?* Brasiliense, São Paulo, 1987.

[16] Meinard Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, Cham, 1 edition, 2015.

[17] Anssi Klapuri and Manuel Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, New York, 2006.

[18] Meinard Müller and Stefan Balke. *Short-Time Fourier Transform and Chroma Features*. International Audio Laboratories Erlangen, 2018.

[19] Garima Sharma, Kartikeyan Umapathy, and Sridhar Krishnan. Trends in audio signal feature extraction methods. *Applied Acoustics*, 161:107201, 2020.

[20] Christopher Harte, Mark Sandler, and Martin Gasser. Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, page 21–26, New York, 2006. Association for Computing Machinery.

[21] Shan Suthaharan. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Springer US, 2016.

[22] Sotiris Kotsiantis, I. Zaharakis, and P. Pintelas. Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26:159–190, 11 2006.

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

[24] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly, Sebastopol, 2 edition, 2019.

[25] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 785–794, Nova York, 2016. Association for Computing Machinery.