

# Modelo de aprendizado de máquina para classificação de gêneros musicais populares da região amazônica legal internacional

Douglas Silva<sup>1</sup>, Cláudio Gomes<sup>1</sup>

<sup>1</sup>Grupo de Computação Musical e Tecnologias Emergentes (COUT-e)  
Ciência da Computação – Departamento de Ciências Exatas e Tecnologias  
Universidade Federal do Amapá (UNIFAP)  
Caixa Postal 261 – 68.906-970 – Macapá – AP – Brazil

douglasdasilvasousa8@gmail.com, claudiorogério@unifap.br

**Abstract.** *The musical preference of today's society presents a continuous appreciation for international musical genres at the preconception of national or local ones. Music is one of the most used means of communication for the facilitation and ergonomics in the organization of the social structure, influencing the lifestyle, tastes and interpersonal coexistence. Several applications with musical concept within several tools use music classifiers. This current work presents a machine learning model of automatic classification of popular Amazonian music genres. Introducing the new version, the database contains the popular musical genres: andean, brega, carimbó, cumbia, merengue, pasillo, salsa and vaqueirada, from the region of the Legal Amazon International. The created database has 125 tracks for each rhythm with 788 parameters. These parameters were extracted in three temporal versions: beginning, middle and end of the song. For tests, the accuracy of the KNN, SVC, MLP and XGB models was analyzed, obtaining 57.58%, 56.79%, 61.33%, and 61.17%, respectively. Concluding that the support vector machine (SVC) model presented the best robustness for the scenarios used.*

**Resumo.** *A preferência musical da sociedade possui continua estima por gêneros musicais internacionais em detrimento aos nacionais ou locais. A música é um dos meios de comunicação mais utilizados para a facilitação e ergonomia na organização da estrutura social influenciando o estilo de vida, gostos e convivência interpessoal. Diversos aplicativos com conceito musical dentro de várias ferramentas utilizam os classificadores musicais. Por diversos motivos, os classificadores musicais possuem pouca ou nenhuma informação sobre gêneros musicais locais. Este trabalho atual apresenta um modelo de aprendizado de máquina de classificação automático de gêneros musicais populares amazônicos. Apresentando a nova versão, a base de dados contém os gêneros musicais populares: andino, brega, carimbó, cúmbia, merengue, pasillo, salsa e vaqueirada, oriundos da região da Amazônia Legal Internacional. A base de dados confeccionada apresenta 125 faixas para cada ritmo com 788 parâmetros. Esses parâmetros foram extraídos em três versões temporais: início, meio e final da música. Para testes, analisou-se a acurácia a partir de validação cruzada dos modelos de KNN, MLP, SVC e XGB obtendo 57.58%, 56.79%, 61.33% e 61.17%, respectivamente. Concluindo que o modelo em máquina de vetor suporte (SVC) apresentou a melhor robustez para os cenários empregados.*

## 1. Introdução

A globalização afeta a preferência musical da sociedade que tem continuo estima por ritmos ou gêneros musicais internacionais em detrimento à nacionais ou locais. A música é composta, teoricamente, por ritmo, melodia e harmonia. De várias teorias, esses elementos substanciais auxiliam na construção da composição ou intenção musical, que são as bases para gêneros musicais [Fabbri and Pinho 2017]. Assim, o ritmo é o elemento contínuo que manifesta a marcação temporal regular de uma música, como por exemplo bater palmas ou estalar os dedos [Schreibman et al. 2015]. Além do ritmo, há ainda a melodia que é uma sequência de notas tocadas por instrumentos. Por fim, o último elemento básico da música é a harmonia, responsável em acompanhar a melodia a partir de notas tocadas simultâneas e sequenciais (acordes). A harmonia dá sentimento para a música, é responsável em gerar músicas com aspectos alegres, melancólicos, tristes etc [Geleilate and Marcelino 2010, Bigliassi et al. 2013]. Portanto, música apresenta grande complexidade de uso subjetivo pois é uma forma de expressão invisível, não palpável e ainda maleável ao longo do tempo tornando-se difícil a tarefa de descrevê-la, rotulá-la ou classificá-la.

O Brasil consome em maior quantidade, música popular brasileira que a música internacional, indicando a possibilidade de “reinvenção brasileira” a cada momento [Menezes Bastos 2007]. Menezes [Menezes Bastos 2007] e Júnior [Junior 2021] afirmam a relevância da contribuição da música popular brasileira para a compreensão sobre as músicas populares do mundo e possíveis rearranjos dessa constituição. Dessa forma, o estudo sobre músicas regionais populares da Amazônia tem grande valor para o cenário musical regional, nacional e internacional.

Com base em intenções, objetivos ou valores históricos e culturais, um gênero musical é caracterizado a partir de propriedades comuns ou similares relacionadas à instrumentação, estrutura rítmica e conteúdo harmônico extraídos a partir de várias músicas [Schreibman et al. 2015]. Portanto, cada manifestação rítmica possui ações distintas para o ouvinte, como por exemplo, para dançar, festejar, descansar, ajudar em momentos de solidão, na tristeza etc [Bigliassi et al. 2013].

Atualmente, diversos aplicativos utilizam-se de diversas ferramentas inteligíveis como os classificadores para indicar, prever ou sugerir novas músicas para seus ouvintes. Basicamente, a cada nova interação do usuário ao aplicativo, os classificadores adquirem novos conhecimentos e aprendem novas opções musicais para os ouvintes. Assim, tanto o usuário quanto os classificadores são duplamente consumidores de um sistema com propósitos similares e complementares [Schreibman et al. 2015].

Para classificar gêneros musicais é necessário, inicialmente, confeccionar uma base de dados, que em grande parte é constituído de ação manual, para prosseguir através de diversos modelos computacionais de processamento e classificação [Aucouturier and Pachet 2003, Sturm 2014]. Na literatura há várias arquiteturas que realizam o processo automático de classificação de gêneros musicais, porém a maioria não possui informações de gêneros musicais regionais ou, por diversos motivos, estão apenas engajados na classificação de gêneros musicais internacionais [Dim et al. 2019, Mondelli et al. 2018]. Assim, ouvintes que buscam ritmos regionais ou locais podem ter grande desconforto de sugestões por limitações de classificadores [Holzapfel et al. 2018].

Este trabalho propõem uma nova versão de um modelo de classificação de gêneros musicais populares amazônicos. Os ritmos musicais da amazônicos envolvem os países do Bolívia, Brasil, Colômbia, Equador, Guiana Francesa, Peru e Venezuela. Inicialmente, criou-se uma base de dados a partir da plataforma de *streaming YouTube* com total de 1000 faixas musicais, sendo 125 para cada gênero a partir de trechos musicais do início, meio e final da música. Para esta versão, os principais gêneros analisados foram: andino, brega, carimbó, cúmbia, merengue, pasillo, salsa e vaqueirada.

O artigo está organizado na seguinte forma. A seção 2 apresenta discussões de trabalhos relacionados. Na seção 3, apresenta-se o modelo de classificação automático de gêneros musicais amazônicos e os resultados obtidos e, por fim, a seção 4, comenta-se as conclusões e trabalhos futuros.

## 2. Trabalhos Relacionados

O trabalho de Guggari et al. [Guggari et al. 2020] apresenta uma abordagem de classificação de ritmos a partir de árvore de decisão. Realizou comparativos entre os algoritmos de floresta aleatória, C4.5, bagging e adaboost em 15 base de dados. A quantidade de classes e parâmetros utilizados para a extração de características musicais foi dependente de cada base de dados diferente. Comparando-se as técnicas de classificação com as bases de dados, os autores obtiveram diferentes resultados, em que no pior ao melhor caso alcançou-se 50% a 100% de acurácia. Porém, não pode-se afirmar qual a técnica mais indicada para a criação de um classificador de ritmos, pois cada base de dados possuía quantidades de amostras, parâmetros e classes de ritmos diferentes.

O trabalho de Karunakaran e Arya [Karunakaran and Arya 2018] apresenta um classificador baseado nas técnicas de K-vizinhos mais próximos, máquina de vetor suporte, *naive Bayes*, redes neurais, análise do discriminante quadrático e lógica *fuzzy*. Investigou os gêneros de pop, rock e músicas eletrônicas nas bases de dados GTZAN [Sturm 2013] e *Free Music Archive* [Defferrard et al. 2016] com acurácia de 90% e 70%, respectivamente.

O trabalho de Silla Jr. et al. [Silla Jr et al. 2007] realizou um classificador automático de músicas latina. Para isso, confeccionou a base de dados intitulada *Latin Music Database* (LMD) [Silla Jr et al. 2008] que possui 3160 músicas de 10 gêneros musicais: tango, bolero, batchata, salsa, merengue, axé, forró, sertaneja, gaúcha e pagode. Esta base possui três opções: utilização de 30 segundos iniciais, 30 segundos intermediário e 30 segundos finais de cada música. O classificador apresentou melhor resultado com características utilizando o segmento intermediário de cada da faixa musical. Realizou vários comparativos entre os modelos de redes neurais, máquina de vetor suporte e *naive bayes* na qual dependendo do cenário obteve em seu melhor caso a acurácia de 65,73%.

## 3. Classificador de ritmos populares da amazônia

Inicialmente, para classificar gêneros musicais é necessário confeccionar uma base de dados, que em grande parte é constituído de ação manual, e em seguida prosseguir para a etapa de classificação a partir de diversos modelos computacionais.

Na literatura há vários modelos que realizam o processo automático de classificação de gêneros musicais, porém a maioria não possui informações de gêneros

musicais populares da amazônia ou, por diversos motivos, estão apenas engajados na classificação de gêneros musicais internacionais. Assim, ouvintes que buscam ritmos regionais ou locais podem ter grande desconforto de sugestões por limitações de classificadores.

A subseção 3.1 descreve as características dos gêneros musicais selecionados. A subseção 3.2 apresenta as características utilizadas para a construção da base de dados dos gêneros musicais populares da amazônia. E, por fim, a subseção 3.3 apresenta as abordagens utilizadas para o modelo de classificador de gêneros musicais da amazônia.

### **3.1. Seleção dos gêneros musicais populares da Amazônia**

A partir da plataforma de *streaming YouTube* realizou a seleção de, 125 faixas musicais para cada ritmo de: andino, brega, carimbo, cúmbia, merengue, pasillo, salsa e vaqueirada totalizando 1000 faixas. Para cada gênero musical, utilizou-se mais de 7 grupos ou bandas musicais diferentes. Além disso, utilizou-se dois critérios de seleção das faixas musicais: qualidade sonora (gravações em ambientes fechados ou estúdios) e taxa de visualização acima de 10.000, para indicar que as faixas musicais sejam amplamente conhecidas.

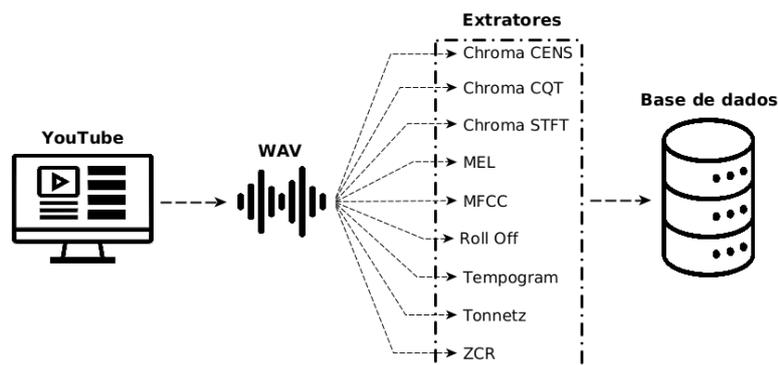
As músicas andinas são gêneros populares na Cordilheira dos Andes que inclui os países da Bolívia, Equador e Peru. Trazido de zona rural da Colômbia, a cúmbia possui várias vertentes regionais em todo o país com características de folclore e celebrações festivas. Já nos países da Colômbia, Guianas, República Dominicana, Panamá e região norte do Brasil o merengue apresenta conotação de celebração festiva bastante popular. O pasillo é outro ritmo de origem colombiana com grandes variações encontradas, principalmente, nos países da Colômbia, Panamá e Equador. A salsa é um dos ritmos mais bailados do mundo com várias influências rítmicas com rica sonoridade nos países da Colômbia, Porto Rico, México, Venezuela e Cuba.

O gênero do brega é um ritmo característico da região norte do Brasil, especificamente do estado do Pará, com influências de vários gêneros musicais com características dançantes oriundos da América Central, reconhecido como patrimônio cultural do Brasil. Outro ritmo da região norte do Brasil é o carimbó que apresenta características indígenas e influências das cúmbias de folclore e reconhecido como patrimônio cultural do Brasil. A vaqueirada, ritmo popular do estado de Amazonas no norte do Brasil, apresenta características de ritmos de ciranda, toada e com encenação popular com diversas influências indígenas, africanas também reconhecido como patrimônio cultural do Brasil.

### **3.2. Criação da base de dados**

Como uma faixa musical pode apresentar variações de velocidade, tonalidade, vozes, e instrumentos ao percorrer da música, analisou-se trechos do início, do meio e do final de cada faixa. Assim, há três versões de bases de dados: uma base com 1000 amostras com apenas instantes iniciais, intermediários e finais. Optou-se por uso de 10 segundos para a extração de características musicais de tempograma, *chroma short-time Fourier transform* (*chroma\_stft*), *chroma Constant Q spectral Transform* (*chroma\_cqt*), RMS e *zero crossing rate* (ZCR) ao total de 788 parâmetros. características musicais. A figura 1 ilustra as etapas utilizadas para a construção da base de dados.

Alguns desses parâmetros citados possuem diversos dados de forma vetorial ou matricial. Dessa forma, apresentar a base com todos os dados seria bastante extensa e invi-



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

abiliza por grande complexidade para tarefas inteligíveis. Para essas variáveis, realizou-se a estatística de média, moda e mediana. Em testes, a mediana apresentou melhor acurácia comparada, exclusivamente, com a média e moda. Assim, para as variáveis matriciais ou vetoriais há o valor mediano de uma determinada característica.

O tempograma indica uma matriz de prevalência de certo tempo para cada momento temporal. Os testes mostraram que o uso do tempograma tiveram melhor acurácia caso fossem utilizados isoladamente. Portanto, o tempograma auxilia na identificação da velocidade dos ritmos. O chroma\_stft apresenta características relacionadas às oitavas musicais, assim, auxilia na caracterização de tonalidades musicais. O chroma\_cqt realiza de modo semelhante ao chroma\_stft porém em escala logarítmica com características estatísticas para cada amostra no tempo, considerando questões de articulação, arpegio e acordes musicais. Apesar de apresentarem características conceituais próximas, os resultados alcançados em conjunto foram melhores caso utiliza-se separadamente. Utilizou-se o *root-mean-square* (rms) devido a grande maioria das músicas não estarem equalizadas e por fim o *zero crossing rate* realiza avaliação de mudanças rápidas do sinal pelo cruzamento ao ponto de amplitude zero.

### 3.3. Classificador de gêneros musicais amazônicos

Conforme as considerações descritas na subseção 3.2, gerou-se a base de gêneros musicais populares amazônicos <sup>1</sup>. Em seguida, realizou-se um estudo dos parâmetros relevantes para o classificador, conforme a subseção 3.2. Em sequência, realizou-se as análises de modelos computacionais entre *K vizinhos mais próximos (KNN)*, *Gradiente booting (XGB)*, *multilayer perceptron (MLP)* e máquina de vetor suporte (SVC). Para todos os modelos, utilizou-se validação cruzada com 10 pastas. A partir de testes, os extratores de tempograma, chroma stft, chroma ctq, rms e zcr foram que obtiveram melhor acurácia, porém não utilizando todos os seus valores matriciais. A tabela 1 apresenta os índices utilizados para cada parâmetro.

No modelo *KNN* a partir de testes iniciais, obteve-se melhores resultados com os hiperparâmetros de 12 *k-vizinhos* e com análise de parâmetro considerando *distância*. O

<sup>1</sup><https://bit.ly/3keCFuK>

| Parâmetro   | Índices das matrizes |
|-------------|----------------------|
| Tempograma  | [1 - 45]             |
| Chroma stft | [1 - 5]              |
| Chroma cqt  | [1 - 12]             |
| RMS         | [1]                  |
| ZCR         | [1]                  |

**Tabela 1. Índices utilizados da matriz de cada parâmetro para os modelos computacionais**

Gradiente *boosting* (XGB) é um modelo computacional que realiza pequenas previsões para um conjunto de dados corrigidos temporalmente o predictor final. Assim, o XGB analisa a contribuição de cada predictor sequencial através de seu crescimento baseado no erro residual do predictor anterior. Com relação modelo de máquina vetor suporte (SVM), tenta encontrar a distância de linhas para a separação entre os dados das classes. O SVC é indicado para um volume de dados não muito grande, que não estejam sobrepostos, ou ainda com pouca quantidade de ruídos. Para o SVM, utilizou-se a função de decisão *ovr*, *kernel rbf* com taxa de *0.00205* e com penalidade *L2*.

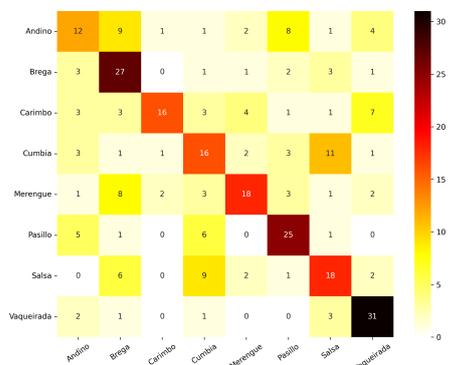
Com relação à redes neurais *perceptron* (MLP), é outro modelo computacional de decisão com objetivos de classificar, agrupar e ainda prever informações. A partir das entradas a rede de neurônios com suas interconexões realizam o aprendizado do dados. É um modelo bastante complexo e inerente de sua estrutura possuindo diversas variáveis controláveis para facilitar resultados esperados [Géron 2019]. Para o classificador, utilizou-se uma rede com 300, 1200 por 8 camadas escondidas utilizando *tangente hiperbólica*, *relu*, *tangente hiperbólica e softmax* com 100 *épocas* e *batch size* com tamanho 32.

A figura 2 apresenta os resultados de cada modelo computacional para o trecho do meio. O KNN, MLP, SVC e XGB obtiveram acurácia de 57.58.62%, 56.79%, 61.33%, 61.17%, respectivamente. Com relação à cumbia e merengue, percebe-se algumas falhas na classificação, pois esses gêneros influenciam os demais gêneros selecionados.

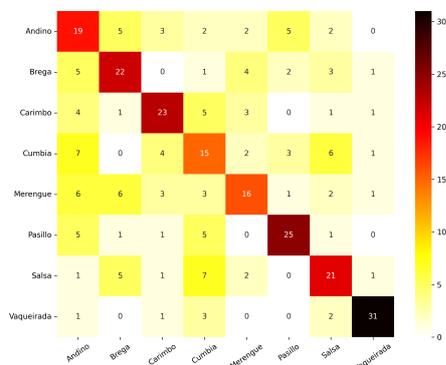
A tabela 2 apresenta o resumo de resultados de treino e teste por validação cruzada em 10 pastas considerando a acurácia média e o desvio padrão. Os modelos XBG e SVC que possuem por essência realizar separação de classes, considerando etapas temporais de aprendizado, apresentaram resultados semelhantes com relação ao resultado de acurácia quanto aos erros de classificação. Já os modelos KNN e MLP com características de aprendizado temporal, apresentaram evoluções de classificação dependente de seus respectivos hiperparâmetros. No geral, analisando a taxa de não acertos dos classificadores pode-se considerar aceitável pois, os ritmos que foram classificados possuem sonoridade muito próximas umas das outras como a brega, carimbó, cumbia, merengue e pasillo.

#### 4. Conclusões e trabalhos futuros

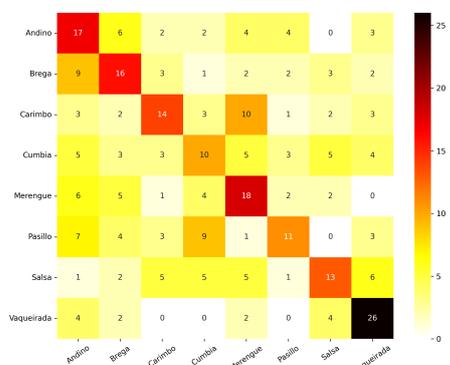
Conclui-se que o modelo de vetor suporte apresentou melhores resultados com relação aos demais modelos. Para trabalhos futuros, pretende-se analisar outros modelos de aprendizado de máquina com objetivo de alcançar melhorias na acurácia da classificação e ainda



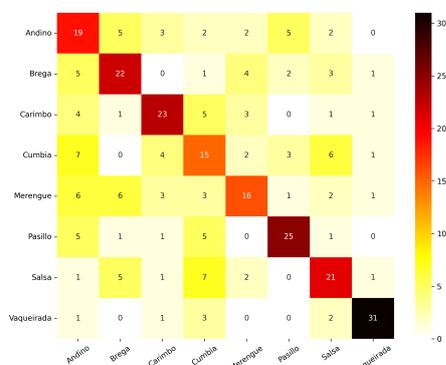
(a) KNN



(b) SVC



(c) MLP



(d) XGB

Figura 2. Matriz de confusão de testes entre modelos computacionais

|        |        | Acurácia  |       |        |       |
|--------|--------|-----------|-------|--------|-------|
| Áudio  | Modelo | Validação |       | Teste  |       |
|        |        | Média     | STD   | Média  | STD   |
| INÍCIO | KNN    | 44.96%    | 0.045 | 45.36% | 0.088 |
|        | MLP    | 41.27%    | 0.033 | 41.36% | 0.071 |
|        | SVC    | 46.27%    | 0.078 | 45.50% | 0.080 |
|        | XGB    | 49.99%    | 0.069 | 0.4801 | 0.088 |
| MEIO   | KNN    | 57.58%    | 0.062 | 53.59% | 0.064 |
|        | MLP    | 56.79%    | 0.066 | 20.90% | 0.001 |
|        | SVC    | 61.33%    | 0.064 | 56.56% | 0.085 |
|        | XGB    | 61.17%    | 0.039 | 57.87% | 0.053 |
| FIM    | KNN    | 45.54%    | 0.015 | 43.40% | 0.060 |
|        | MLP    | 42.56%    | 0.011 | 45.55% | 0.078 |
|        | SVC    | 48.69%    | 0.052 | 44.41% | 0.056 |
|        | XGB    | 48.28%    | 0.039 | 42.78% | 0.071 |

Tabela 2. Acurácia dos modelos computacionais em trechos musicais

desenvolver diferentes visões entre os ritmos populares da amazônia selecionados, como encontrar semelhanças computacionais entre ritmos semelhantes e diferentes. Além disso, adicionar faixas musicais para cada gênero musical afim de melhor compreender os ritmos por diversas características computacionais.

## Agradecimentos

Os autores agradecem o apoio financeiro pelo Programa Institucional de Bolsas de Iniciação Científica (PROBIC) da Universidade Federal do Amapá (UNIFAP).

## Referências

- Aucouturier, J.-J. and Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of new music research*, 32(1):83–93.
- Bigliassi, M., Altimari, L. R., and Ito, W. M. (2013). Ritmos e estilos musicais: um estudo descritivo das preferências e percepções no exercício físico. *Brazilian Journal of Biomotricity*, 7(4):165–173.
- Defferrard, M., Benzi, K., Vandergheynst, P., and Bresson, X. (2016). Fma: A dataset for music analysis. *arXiv preprint arXiv:1612.01840*.
- Dim, C., Alves, L., and Sousa, P. (2019). Predição de gêneros musicais utilizando técnicas de aprendizado de máquina. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 344–352. SBC.
- Fabbri, F. and Pinho, M. G. (2017). Uma teoria dos gêneros musicais: duas aplicações [tradução]. *Revista Vórtex*, 5(3).
- Geleilate, J. M. G. and Marcelino, D. M. N. (2010). A influência de estilos musicais no humor, tempo percebido e decisão de retorno em loja de consumo popular. *IV Encontro de Marketing da ANPAD*.
- Géron, A. (2019). *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & Tensor-Flow*. Alta Books.
- Guggari, S., Kadappa, V., Umadevi, V., and Abraham, A. (2020). Music rhythm tree based partitioning approach to decision tree classifier. *Journal of King Saud University-Computer and Information Sciences*.
- Holzappel, A., Sturm, B., and Coeckelbergh, M. (2018). Ethical dimensions of music information retrieval technology. *Transactions of the International Society for Music Information Retrieval*, 1(1):44–55.
- Junior, J. S. J. (2021). Nem original, nem cópia: versões musicais entre o mainstream e a periferia. *Números*.
- Karunakaran, N. and Arya, A. (2018). A scalable hybrid classifier for music genre classification using machine learning concepts and spark. In *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 128–135. IEEE.
- Menezes Bastos, R. J. d. (2007). *As contribuições da música popular brasileira às músicas populares do mundo: diálogos transatlânticos Brasil-Europa-África - Segunda parte*. Universidade Federal de Santa Catarina.

- Mondelli, M. L. B., Gadelha Jr, L. M., and Ziviani, A. (2018). O que os países escutam: Analisando a rede de gêneros musicais ao redor do mundo. In *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*. SBC.
- Schreibman, S., Siemens, R., and Unsworth, J. (2015). *A new companion to digital humanities*. John Wiley & Sons.
- Silla Jr, C. N., Kaestner, C. A., and Koerich, A. L. (2007). Automatic music genre classification using ensemble of classifiers. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 1687–1692. IEEE.
- Silla Jr, C. N., Koerich, A. L., and Kaestner, C. A. (2008). The latin music database. In *ISMIR*, pages 451–456.
- Sturm, B. L. (2013). The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use. *CoRR*, abs/1306.1461.
- Sturm, B. L. (2014). The state of the art ten years after a state of the art: Future research in music information retrieval. *Journal of new music research*, 43(2):147–172.