Barbara Poliana Jaber Martins Ferreira

Aprendizado de Máquina aplicado à classificação de gêneros musicais a partir de letras de música

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Bacharela em Engenharia de Computação.

Centro Federal de Educação Tecnológica de Minas Gerais – CEFET-MG

Departamento de Computação

Curso de Engenharia da Computação

Orientador: Daniel Hasan Dalip

Coorientador: Ismael Santana Silva

Belo Horizonte

2020

Centro Federal De Educação Tecnológica de Minas Gerais
Curso de Engenharia de Computação
Avaliação do Trabalho de Conclusão De Curso

Aluna: Barbara Poliana Jaber Martins Ferreira

Título do trabalho: Aprendizado de Máquina aplicado à classificação de gêneros musicais a partir de letras de música
Data da defesa: 24 de novembro de 2020
Horário: 10:30

Link da Defesa: https://meet.google.com/yiv-awge-hgg

O presente Trabalho de Conclusão de Curso foi avaliado pela seguinte banca:

Professor Dr. Daniel Hasan Dalip – Orientador
Departamento de Computação
Centro Federal de Educação Tecnológica de Minas Gerais

Professor Me. Ismael Santana Silva – Co-orientador
Departamento de Computação
Centro Federal de Educação Tecnológica de Minas Gerais

Professor Me. Guilherme Lopes de Oliveira - Membro da banca de avaliação
Departamento de Computação
Centro Federal de Educação Tecnológica de Minas Gerais

Professor Dr. Thiago de Souza Rodrigues – Membro da banca de avaliação
Departamento de Computação
Centro Federal de Educação Tecnológica de Minas Gerais

Dedico este trabalho a todos que acreditaram que ele tomaria forma.

Agradecimentos

Ao DECOM, do CEFET-MG, pelo ambiente amigável e livre. E a todos os professores do departamento, por contribuírem com conhecimento para o meu desenvolvimento.

Ao meu orientador Prof. Dr. Daniel Hasan Dalip, pelo empenho à elaboração deste trabalho, pelo conhecimento ensinado de maneira bem-humorada, como também pelas palavras de acolhimento e coragem para que eu concluísse essa jornada.

Ao meu co-orientador Prof. Me. Ismael Santana, pelo olhar analítico e perspicaz que permitiu que este trabalho fosse construído de forma criativa, por ter acreditado neste trabalho desde o princípio e pelos conselhos nos momentos difíceis.

Aos professores membros da minha banca, Prof. Me. Guilherme Oliveira e Prof. Dr. Thiago Souza, por aceitarem o convite de participar da banca avaliativa e pelas correções significativas que pontuaram para a melhoria deste trabalho.

Aos meus pais, Nara e Ricardo, por serem meus primeiros professores, por todo apoio aos meus estudos.

A minha madrasta, Gláucia, pela sensibilidade que sempre me trouxe força e coragem.

À toda minha família, Jaber e Martins Ferreira, pelo incentivo e apoio incondicional ao meu caminho.

Ao Sinval, por toda parceria e paciência que tornam esse caminho comum uma fortaleza.

Aos meus amigos, por tornarem esse caminho mais leve e por compartilharem as alegrias e dificuldades ao longo do curso, até quando não fazia sentido o que eu estava falando.

E a todos que direta ou indiretamente acreditaram neste trabalho e participaram da minha formação.

Meus sinceros agradecimentos.

"Music is the universal language of mankind and poetry their universal pastime and delight."

(Henry Wadsworth Longfellow)

Resumo

A classificação de gêneros de música tem sido aplicada como uma forma de caracterizar gêneros específicos e similares entre si, em função da larga de gêneros existentes e da massa de dados de música crescente disponível na Internet. No entanto, a classificação de gêneros ainda é um desafio, uma vez que a quantidade de informações musicais disponíveis é muito grande, e diversos fatores culturais e musicais definem um gênero musical. Diferente de outros trabalhos que utilizam de informações como áudio, para classificação de músicas, esse trabalho utiliza apenas as letras de música. O intuito foi de aplicar o Processamento Natural de Linguagem e criar atributos relacionados às letras de música a fim de comparar formas de classificar os dez gêneros da base de dados utilizada. O *dataset* utilizado foi treinado com os métodos de SVM (Support Vector Machine), Random Forest e Naive Bayes, utilizando a Classificação Tradicional e Classificação Hierárquica, que por sua vez exigiu um agrupamento dos gêneros musicais. Esse agrupamento foi embasado na literatura de similaridade de gêneros musicais e de impressões da autora. O principal objetivo foi de comparar as formas de classificação, dados os métodos utilizados e as representações das letras como os atributos criados e o Bag of Words (BOW). Dentre as representações avaliadas, a Classificação Hierárquica apresentou um desempenho melhor com os atributos criados, em relação ao Bag of Words. Como contribuição deste trabalho têm-se a representação dos atributos criados, e análise de da relevância dos mesmos, assim como das classificações feitas.

Palavras-chave: Classificação Hierárquica, classificação supervisionada, gêneros de música, SVM, Random Forest, Naive Bayes

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

Atributos	*Representação de atributos das músicas*
BOW	*Bag of Words*
KNN	*K-Nearest Neighbors*
NB	*Naive Bayes*
MIR	Music Information Retrieval
PLN	*Processamento de Linguagem Natural*
RF	*Random Forest*
RIM	*Recuperação de Informação de Música*
SVM	*Support Vector Machine*
TF-IDF	*Term Frequency - Inverse Document Frequency*
TPE	*Tree-Structured Parzen Estimator*

Sumário

1 Introdução

Há poucas décadas atrás, para escutar uma música era preciso ou ligar numa estação de rádio, ou ter uma fita com suas músicas selecionadas e um aparelho estéreo a sua disposição. Enquanto uma fita cassete cabia apenas uma hora de música, hoje em dia temos a nossa disposição inúmeras plataformas de *streaming*, com músicas a um clique de serem escutadas. É inegável que a evolução dos aparelhos de áudio trouxe não só novas formas de escutar mas também de como conhecer músicas.

Ao mesmo tempo que a disponibilidade das músicas aumentam, elas passam a fazer parte de uma massa de informação crescente. Em função disso, o estudo desse conjunto crescente de dados é relevante tanto para a área de Recuperação de Informação de Música (RIM) (ou *Music Information Retrieval, MIR*, em inglês) quanto para estudos de aspectos culturais de gêneros musicais, como no trabalho de Tsaptsinos (2017). A RIM pode ser baseada em diferentes tipos de informação da música como áudio, letras, gênero, artistas, entre outras características. Entre essas, as letras e os gêneros foram o foco de estudo deste trabalho.

A tarefa de avaliar o gênero de uma música a partir de sua letra pode ser trivial para humanos, mas não é para uma máquina. Essa é uma tarefa que envolve diferentes campos de conhecimento além de, muitas vezes, uma percepção pessoal e cultural (SCHEGLOFF, 1997). O que reforça assim a relevância do estudo dos gêneros musicais baseado na particularidades de cada um, mas também de suas similaridades, visto que existem gêneros extremamente específicos, mas também gêneros mais conhecidos e plurais.

Diante desse cenário, neste trabalho é proposta uma metodologia de classificação dos gêneros musicais, utilizando uma Classificação Hierárquica, que é uma classificação em dois níveis de hierarquia. Enquanto que a Classificação Tradicional é a classificação usual para um conjunto de atributos e uma classe, para o qual se retorna um resultado. E a fim de validar que a Classificação é relevante para classificação dos gêneros, os gêneros musicais da base de dados foram agrupados com o objetivo de criar uma hierarquia para sua classificação. Essa hierarquia é a utilizada para predição dos gêneros reais das músicas.

Ambas as metodologias, a hierárquica e a tradicional, aplicam e avaliam os métodos de Aprendizado de Máquina SVM, Random Forest e Naive Bayes para dois tipos de representação dos dados. Uma delas foi criada pela autora e contêm 17 atributos distintos, feitos a partir de aspectos diferentes das letras de música, que são explicados no Capítulo 4. A outra é a representação por *Bag of Words* que é explicada no Capítulo 2.

Com o objetivo final de estabelecer um comparativo entre metodologias e das representações na classificação dos gêneros musicais, este trabalho teve como contribuição

os atributos das músicas elaborados nele e comparativo final da Classificação Hierárquica e Classificação Tradicional. Além disso, aprimorou os conhecimentos da autora a cerca dos tópicos de seu escopo de Aprendizado de Máquina e de Música.

A estrutura deste trabalho é dividida nos Capítulos: Capítulo 2 onde é demonstrada a Fundamentação Teórica, Capítulo 3 com os Trabalhos Relacionados, Capítulo 4 com a Metodologia, Capítulo 4, Capítulo 5 com os experimentos e resultados e por último no Capítulo 6 as conclusões deste trabalho.

2 Fundamentação Teórica

Neste capítulo são apresentados os métodos de aprendizado de máquina que são utilizados no trabalho na Seção 2.1, e sobre a representação dos dados, na Seção 2.2, de Processamento de Linguagem Natural (PLN).

2.1 Métodos de Aprendizado de Máquina

Segundo Mitchell (1997), diz-se que um programa de computador aprende uma experiência E, com respeito a algum tipo de tarefa T e performance P, se sua performance P nas tarefas em T, na forma medida por P, melhoram com a experiência E. Esse processo é denominado aprendizado de máquina (ou *Machine Learning*, em inglês) que é um campo de estudo de reconhecimento de padrões por parte de um programa (ou máquina). (RUSSELL; NORVIG, 2002)

Dada a definição acima, entende-se uma tarefa T como um problema a ser resolvido por meio de um (ou mais) método(s) de aprendizado de máquina. Para isso, é necessária a experiência E que são os dados de treinamento dos métodos. Assim, dada uma resolução de uma tarefa com um ou mais métodos, avalia-se a performance P de cada método por meio de métricas de avaliação (como a acurácia, por exemplo), que medem o desempenho de aprendizado.

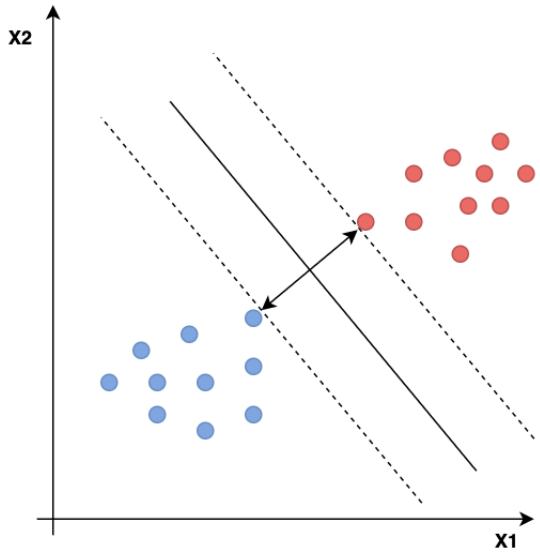
Os métodos de aprendizado de máquina utilizados neste trabalho são explicados a seguir. Já as métricas de avaliação são explicadas no Capítulo 4.

2.1.1 Support Vector Machine (SVM)

O método Máquina de Vetores de Suporte, (do inglês, *Support Vector Machine (SVM)*) é um método de aprendizado de máquina supervisionado proposto por Cortes e Vapnik (1995). O método já foi muito utilizado na literatura tradicional de aprendizado de máquina para classificação e análise de regressão com diferentes tipos de dados, dentre eles textos e hiper-textos. (MAYER; RAUBER, 2011)

O SVM recebe como entrada um conjunto de dados de treinamento pertencentes às classes X1 e X2 e constrói um modelo que classifica cada entrada como X1 ou X2. Essa classificação feita pelo modelo do SVM é como a representação de pontos no espaço, demonstrada na Figura 1, em que os pontos de cada classe dividem um mesmo espaço amplo.

Figura 1 – SVM



Fonte: A autora.

O modelo SVM mapeia esses pontos no espaço e prediz a qual classe cada um pertence, baseado na posição em que os pontos são colocados. Para isso, ele encontra uma linha de separação entre os pontos, chamada de *hiperplano*, que tem como objetivo maximizar a margem, que é a distância entre os pontos mais próximos das duas classes X1 e X2, minimizando o erro. O hiperparâmetro responsável por variar o cálculo dessa margem é o Custo C, que quanto mais alto, a otimização do SVM escolhe uma margem menor, o que pode facilitar a classificação de uma quantidade maior de pontos. E quanto mais baixo C, a otimização escolhe uma margem maior o que pode deixar de classificar muitos pontos (RUSSELL; NORVIG, 2002).

A classificação exposta na Figura 1 é binária, dado que só existem duas classes possíveis. Para realizar classificação de multi-classes, como proposto por Crammer e Singer (2001), o problema é reduzido a múltiplos problemas de classificação binária. Ou seja, é feita uma combinação da implementação do modelo SVM de classificação binária para que várias classificações binárias sejam feitas entre n classes do problema original.

Formas comuns para redução de problemas multi-classes são: diferenciar uma classe em relação a todas as outras, abordagem chamada de um-contra-todos, (do inglês, *one-versus-all*) ou diferenciar entre cada par de classes, abordagem chamada de um-contra-um, (do inglês, *one-versus-one*).

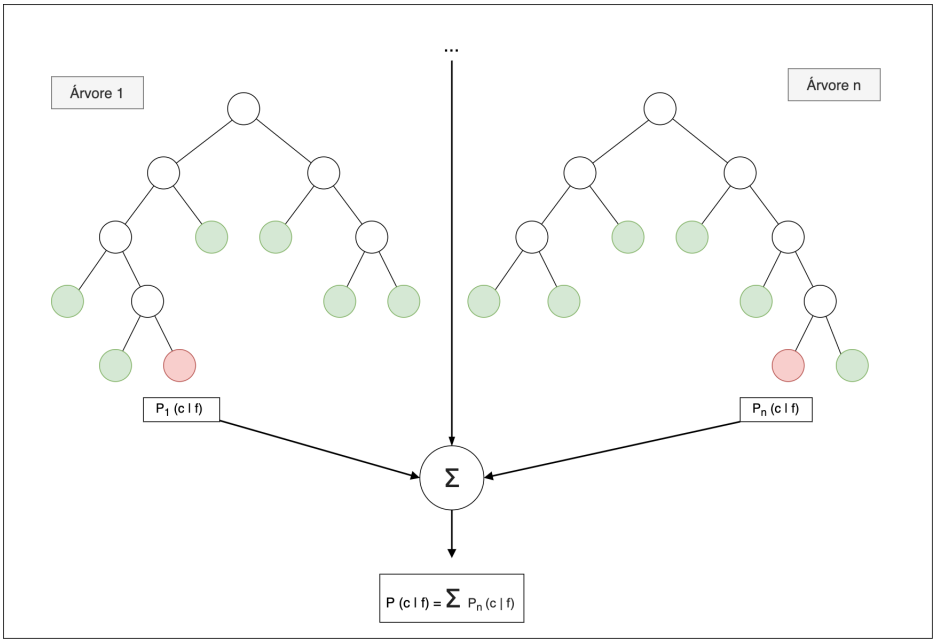
2.1.2 Random Forest

O método de Floresta Aleatória (ou *Random Forest*, em inglês) é um método de aprendizado de máquina supervisionado que consiste na implementação de múltiplas

árvores de decisão criadas aleatoriamente e cujos resultados de decisão são combinados (em inglês, *ensemble*) para obtenção de um resultado cuja predição seja a melhor. (RUSSELL; NORVIG, 2002)

Na Figura 2 é demonstrada uma Random Forest de n árvores que devem classificar um conjunto de dados. As árvores são montadas de tal forma que cada nodo, que não é folha, representa um atributo diferente. E os nós folhas são as decisões da árvore. Dessa forma, cada nodo folha é uma decisão, representados na Figura 2 pelas cores verde e vermelha. Dado o resultado previsto de alguma instância, os resultados das árvores são combinados, por meio de maioria de voto das decisões de cada árvore, por exemplo. Dessa forma, o resultado é então somado e combinado para encontrar o melhor resultado em performance da floresta. Essa soma e combinação de resultados é ilustrada na Figura 2 pelas folhas de cor vermelha que representam uma mesma decisão (ou predição) feita em cada árvore.

Figura 2 – Random Forest - duas árvores de decisão



Fonte: A autora.

É possível variar as formas como a Random Forest é criada, utilizando-se da variação de hiperparâmetros. Esses são responsáveis pelo número de árvores geradas, o número máximo de atributos por árvore ou o número mínimo de amostras para dividir um nodo interno, por exemplo (PEDREGOSA et al., 2011).

2.1.3 Naive Bayes

O método Naive Bayes é um algoritmo utilizado para classificação que se baseia no Teorema de Bayes com classificadores probabilísticos, como proposto por Zhang (2004).

Como trata-se de um método de probabilidade condicional, consideremos um exemplo a seguir, onde uma instância a ser classificada é representada por um vetor E =(x_1,..., x_n) com n representando o número de atributos (variáveis independentes) e c sendo a classe, temos:

$$P(c \mid E) = \frac{P(E \mid c)P(c)}{P(E)} \tag{2.1}$$

onde $P(c \mid E)$ é a probabilidade condicional de c acontecer dado que E seja verdade, $P(E \mid c)$ é a probabilidade condicional de E acontecer dado que c seja verdade, e $P(E)$ e $P(c)$ são as probabilidades marginais dos eventos c e E, respectivamente (ANNIS, 2006).

Naive Bayes é a forma mais simples de uma rede Bayesiana, na qual todos os atributos são independentes dado um valor de uma classe variável. Isso é chamado de independência condicional (ZHANG, 2004).

2.2 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) tem como propósito tornar possível o processamento da linguagem humana para o entendimento de computadores (RUSSELL; NORVIG, 2002).

Para que haja um PLN é necessário que o texto seja traduzido para a máquina e existem diversos modelos possíveis para tal tarefa. Na seção abaixo, segue a explicação da representação de *Bag of Words*.

2.2.1 Bag of Words

O *Bag of Words* (BOW) consiste na representação do texto como um conjunto de palavras, sem considerar o significado de cada uma, ou sua ordem, mas considerando sua multiplicidade, ou seja, um vocabulário (BAEZA-YATES; RIBEIRO-NETO, 1999).

O vocabulário de k termos $V = \{t_1, t_2, t_3, ..., t_k\}$ em um documento d_j, esse documento d_j é representado por um vetor de pesos $(w_{1,j}, w_{2,j}, ...w_{k,j})$ em que $w_{i,j}$ representa o peso do termo i no documento j, onde $j = 1, ..., N$, onde N é o número de documentos sob análise.

O peso do termo pode ser calculado de várias formas como a frequência de um termo i na coleção de N documentos, (f_{ij}) e o *TF-IDF* (*term frequency - inverse document frequency*, em inglês) explicado a seguir.

A métrica *TF-IDF* (*term frequency - inverse document frequency*, em inglês) é o produto de seu TF (frequência do termo) e de seu IDF (o inverso da frequência na coleção). Assim, para calcularmos o *TF-IDF*, primeiramente calculamos a frequência de um termo no documento (TF) e, logo após, multiplicamos pelo inverso da frequência no documento (IDF) (BAEZA-YATES; RIBEIRO-NETO, 1999).

Para calcular o TF do termo i do documento j, TF_{ij}, faz-se:

$$TF_{ij} = log(f_{ij}) \tag{2.2}$$

em que f_{ij} é a frequência de um termo i no documento j.

Para calcular o IDF (*Inverse Document Frequency*, em inglês) do termo i, do documento j, usa-se o *log*, com o objetivo de suavizar valores muito altos. O IDF_i é demonstrado na Equação 2.3:

$$IDF_i = log(\frac{N}{n_i}) \tag{2.3}$$

em que N é o número de documentos da coleção e n_i é o número de documentos em que esse termo i ocorre. Espera-se que quanto mais discriminativo o termo, em menos documentos esse termo irá ocorrer e, consequentemente, o IDF deste termo será mais alto (BAEZA-YATES; RIBEIRO-NETO, 1999).

Por fim, o produto das duas métricas acima resulta então no *TF-IDF* do termo i no documento j, o qual é dado por:

$$TFIDF_{ij} = TF_{ij} \times IDF_i \tag{2.4}$$

3 Trabalhos Relacionados

A classificação de gêneros de música é um tópico desafiador de Recuperação de Informação de Músicas (*Music Information Retrieval* - MIR, em inglês), dado a multiplicidade de gêneros musicais existentes e as diversas características próprias de cada artista e música. A similaridade entre gêneros musicais foi demonstrada por Whitman e Lawrence (2002) sendo ela relacionada tanto à sonoridade de cada música, quanto ao conteúdo de suas letras. Essas últimas, por serem textos de caráter literário, estão intimamente relacionadas à percepção do eu lírico em questão, o que é diretamente ligado a uma construção cultural (SCHEGLOFF, 1997).

Tendo em vista a abordagem de classificação de gêneros musicais a partir apenas do estudo de letras de música como fonte de informação, como as músicas são textos, esse tipo de estudo é um tópico de estudo de Processamento Natural de Linguagem (*Natural Language Processing* - NLP, em inglês). A partir de análises do corpo do texto de músicas: como o vocabulário de cada uma, diferentes separações de segmentos de música, seja por palavra seja por segmentos de versos e estrofes, é possível demonstrar características diferentes dos gêneros musicais inclusive entre línguas de origem distintas, como foi o tema do trabalho de Mahedero et al. (2005).

No artigo Mayer, Neumayer e Rauber (2008), que é o trabalho com maior similaridade com este, também tem como objetivo a classificação de gêneros de música a partir de suas letras. Os métodos de aprendizado de máquina utilizado são Support Vector Machines (SVM), com kernel linear e polinomial, k-Nearest Neighbour (kNN), Árvores de Decisão (Decision Tree) e Naive Bayes (NB). São propostos diferentes atributos, como número de palavras por linha e número de palavras únicas por linha, dentre outras que serviram como base para a elaboração dos atributos coletadas neste presente trabalho, no Capítulo 4.

Existem outras abordagens na literatura que combinam outras fontes de dados de música, como áudio e cifras Kolchinsky et al. (2017), além das letras de música. Um trabalho mais recente de Mayer e Rauber (2011) demonstrou como ao acrescentar os áudio das músicas, além de suas letras, melhorou a acurácia do método Support Vector Machines (SVM) para classificação de gêneros musicais. O comparativo desses trabalhos é feito na Tabela 1.

Tabela 1 – Comparativo dos Trabalhos Relacionados

Artigo	**Representação**	**Método de Análise**
Mahedero et al. (2005)	Letras de músicas (bag-of-words)	Statistical Identification of language Dunning (1996)
Mayer, Neumayer e Rauber (2008)	Atributos (features) de letras de músicas	SVM, Random Forest, KNN, Naive Bayes
Kolchinsky et al. (2017)	Letras (bag-of-words) e cifras de músicas	labMT (biblioteca de análise de sentimentos)

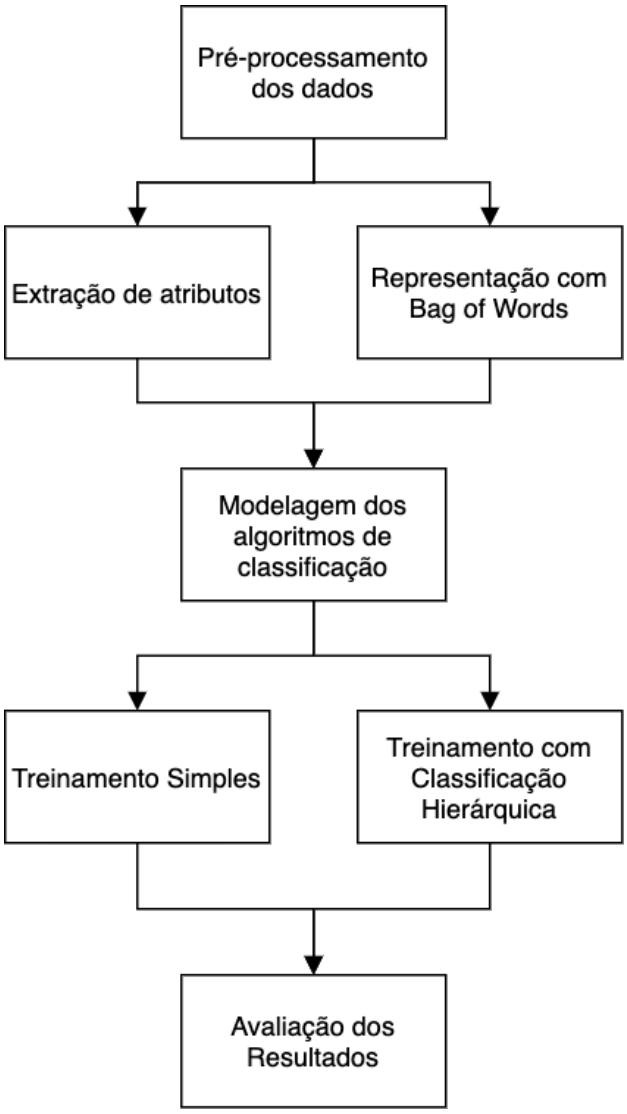
Ainda neste contexto de Recuperação de Informação de Músicas, existe uma contribuição crescente de análise de emoções em músicas. Dado que a expressão da música é um ato natural da expressão emocional (MEYER, 1956). O trabalho de Yang e Lee (2010) utilizou apenas de letras de música para uma classificação multi-classe de emoções, cujas categorias de emoções foram originárias do estudo de psicologia de Clark e Tellegen (1999), e propôs um processamento estatístico do texto das músicas que foram transformadas num vetor de 182 características psicológicas. O resultado final do trabalho de Clark e Tellegen (1999) foi o banco de dados gerado com todas essas caractaterísticas.

A partir do conhecimento desses trabalhos relacionados e da diversidade de abordagens possíveis para execução deste trabalho, foi escolhido seguir a linha do artigo Mayer, Neumayer e Rauber (2008). A diferença entre a abordagem de Mayer, Rauber e Neumayer e a deste trabalho é que neste faz-se um comparativo da classificação hierárquica com a não hierárquica dos mesmos métodos de aprendizado, com exceção do KNN.

4 Metodologia

Neste capítulo são apresentadas as principais características da Metodologia adotada para o desenvolvimento do trabalho. A Seção 4.1 foca na representação dos dados e a Seção 4.2 na Classificação Hierárquica. As etapas dessa metodologia são demonstradas na Figura 3.

Figura 3 – Metodologia



Fonte: A autora.

4.1 Representação dos dados

A base de dados escolhida foi retirada do Kaggle (2018), fruto da coleta de 380.000 letras de músicas do MetroLyrics (2020) em 2018, com 10 gêneros musicais: Country, Folk, Hip-Hop, Indie, R&B, Jazz, Pop, Eletrônico, Rock e Metal. O motivo da escolha dessa base de dados foi o volume da mesma. Porém, neste trabalho, a quantidade de músicas consideradas para o avaliação dos algoritmos foi reduzida para 18.000. Em função do balanceamento por agrupamento de gêneros feito para Classificação Hierárquica, explicada na Seção 4.2.

O primeiro passo foi um pré-processamento das letras para limpar o texto, para isso foram removidos todos os caracteres que não fossem letras ou números e foram tratadas todas as letras para serem minúsculas. O segundo passo foi a criação da representação de atributos, explicada a seguir.

Considere-se um conjunto de músicas X = $\{x_1, x_2, ..., x_n\}$. Cada música é representada por um conjunto de m Atributos A = $\{a_1,..., a_m\}$, de tal forma que, $x_i = \{a_{i1}, a_{i2}, ..., a_{im}\}$ é o vetor que representa x_i, em que cada a_{ij} é o valor de um atributo A_j na música x_i. Neste trabalho, as métricas definidas pela autora medem estatísticas diferentes das letras música, tendo sido elaboradas com base no trabalho de Mayer e Rauber (2011) e de ideias sobre o valor da informação contida nas letras de música dada sua estrutura de texto. Para a compreensão dos atributos criados, considera-se o conceito de *stopwords* como palavras que possuem pouco significado do ponto de vista semântico, tais como preposições, artigos, conjunções e outros.

Nesta proposta, assume-se que o acesso aos dados de treinamento é na forma $\{(x_1,g_1), (x_2,g_2), ..., (x_n,g_n)\}$, em que cada par (x_i,g_i) representa a letra da música e seu gênero correspondente. As descrições dos atributos extraídos das músicas seguem abaixo:

1. **Quantidade de palavras únicas, sem *stopwords***: Número de termos únicos da música, sem contar as *stopwords*.

2. **Quantidade de palavras únicas, com *stopwords***: Número de termos únicos da música, contando as *stopwords*.

3. **Ruído de informação (do inglês, *Information to Noise* 1 e 2)**: Divisão entre as relações encontradas para Atributo 1 e Atributo 2.

4. **Densidade de palavras únicas sem *stopwords* na música**: Divisão entre a quantidade de termos únicos sem *stopwords* pelo número total de termos da música com *stopwords*.

5. **Quantidade de termos únicos total normalizada**: Total de termos únicos da música com *stopwords* e normalizado em relação ao número de termos únicos. A

Equação 4.1 referencia o cálculo feito para a normalização:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4.1}$$

em que X_{min} e X_{max} são os valores mínimo e máximo em X (número de termos únicos).

6. **Quantidade de versos únicos total normalizada**: Total de versos únicos da música e normalizado conforme a Equação 4.1.

7. **Média de termos por estrofe na música**: Média de termos únicos por estrofe na música, considerando *stopwords*, e que cada estrofe seja um segmento de 4 versos.

8. **Média de termos únicos por versos, com *stopwords***: Média de termos únicos por versos, considerando *stopwords*.

9. **Desvio padrão de termos únicos por versos, com *stopwords***: Desvio-padrão de termos únicos por versos, considerando *stopwords*.

10. **Quantidade máxima de termos únicos por versos normalizado, com *stopwords***: Quantidade máxima de termos únicos por versos normalizado, considerando *stopwords*.

11. **Quantidade mínima de termos únicos por versos normalizado, com *stopwords***: Quantidade mínima de termos únicos por versos normalizado, considerando *stopwords*.

12. **Média de termos únicos por versos, sem *stopwords***: Média de termos únicos por veros, sem considerar *stopwords*.

13. **Desvio padrão de termos únicos por versos, sem *stopwords***: Desvio-padrão de termos únicos por veros, sem considerar *stopwords*.

14. **Quantidade máxima de termos únicos por versos normalizado, sem *stopwords***: Quantidade máxima de termos únicos por versos normalizado, sem considerar *stopwords*.

15. **Quantidade mínima de termos únicos por versos normalizado, sem *stopwords***: Quantidade mínima de termos únicos por versos normalizado, sem considerar *stopwords*.

16. **Índice de positividade da música**: Divisão entre a frequência de termos considerados palavras positivas, em relação ao número total de termos sentimentais da música. Esse vocabulário foi elaborado pela autora e é composto pela soma de todos os termos sentimentais positivos e negativos, que são demonstrados na Tabela 2.

17. **Índice de negatividade da música**: Divisão entre a frequência de termos considerados palavras negativas, em relação ao número total de termos sentimentais da música.

Tabela 2 – Vocabulário para índices de positividade e negatividade

Índice de	Descrição
positividade	$baby, confidence, confident,$ $cute, dear, faith, good,$ $honey, hope, hot, life,$ $like, love, peace, right,$ $safe, solitude, truth$
negatividade	$bad, broken, cold, cry,$ $dead, death, fear, guilty,$ $hate, helpless, hopeless, hurt,$ $lie, lies, isolation, lonely,$ $pain, sad, sadness,$ $tears, war, wrong$

Diante do exposto, a representação de cada um desses atributos é referenciada neste trabalho como Atributo $1, 2, 3, .., 17$ um conjunto de todos eles juntos é referenciado como Atributos. Já a representação com Bag of Words é referenciada apenas como Bag of Words.

4.2 Classificação Hierárquica

A Classificação Tradicional de Aprendizado de Máquina não possui uma estrutura que trabalhe originalmente com hierarquias ou classes agrupadas. No entanto, é possível elaborar estratégias desse tipo para certos tipos de problema, conforme foi demonstrado por Nakano et al. (2017). Assim, neste trabalho, foi elaborada a proposta de agrupar alguns gêneros musicais a fim de gerar uma Classificação Hierárquica dos mesmos, com base na similaridade entre os gêneros musicais demonstrada no trabalho de Pampalk, Flexer e Widmer (2005).

4.2.1 Agrupamento dos gêneros

Para cada agrupamento $g \in G$, em que G é o conjunto de agrupamentos musicais feitos, g é um grupo de um ou dois gênero agrupados. Dessa forma, existem apenas dois níveis na hierarquia de classificação. No primeiro nível dela, classifica-se a música como parte de um agrupamento g, e no segundo e último nível, classifica-se a música como parte de um único gênero musical. Esses agrupamentos são explicitados na Tabela 3 assim como os gêneros que permaneceram sem agrupamento.

Tabela 3 – Agrupamento de gêneros

Gênero original	**Gênero agrupado**	**Número do grupo g**
Country	Country	1
Folk	Country	
Hip-Hop	Hip-Hop	2
Indie	Indie	3
R&B	Jazz	4
Jazz	Jazz	
Pop	Pop	5
Eletrônico	Pop	
Rock	Rock	6
Metal	Rock	

Os gêneros Hip-Hop e Indie são os gêneros que não foram agrupados pois, dados os 10 gêneros musicais da base de dados, não haviam gêneros similares a eles, em relação às características mostradas por Pampalk, Flexer e Widmer (2005) e às impressões de estilo da autora. Ainda assim, esses gêneros passam pelo processo de verificação de qual grupo pertencem, durante a Classificação Hierárquica, assim como os agrupados. Esse processo é explicado na Subseção 4.2.2.

A Tabela 4 apresenta a quantidade de músicas para cada gênero após o balanceamento da base de dados feito para que os grupos tivessem a mesma quantidade de músicas entre si. Como na base de dados original o *Indie* possuía apenas 3140 músicas, reduziu-se o tamanho dos grupos para 3000 músicas para cada um.

Tabela 4 – Quantidade de músicas por gênero e grupo

Gênero original	**N**	**N por grupo**	**Grupo**
Country	2622	3000	1
Folk	378		
Hip-Hop	3000	3000	2
Indie	3000	3000	3
R&B	882	3000	4
Jazz	2118		
Pop	2530	3000	5
Eletrônico	480		
Rock	2461	3000	6
Metal	539		

Quantidade N de músicas por gênero e grupo

4.2.2 O Algoritmo de Classificação Hierárquica

A forma que foi desenvolvida a Classificação Hierárquica neste trabalho está ilustrada no Algoritmo 1 que mostra como é feita a filtragem das bases de treino e de teste para a predição do gênero real da música. Para compreender o Algoritmo 1, na Tabela 5 são descritos os significados de cada variável.

Antes do ciclo de repetição iniciar, é feita a classificação no primeiro nível. Inicialmente, antes do primeiro laço de repetição, cria-se um modelo M_1 para a predição com base no grupo de gênero. Nesse momento, $\hat{y}^{tr}$ possui as predições do primeiro nível.

Para iniciar as predições no segundo nível, considera-se g um agrupamento do conjunto de agrupamentos G, no laço de maior escopo. No primeiro laço dentro desse, é preparada a base de treino para o segundo nível, a partir da filtragem dos gêneros preditos na classificação do primeiro nível, ou seja, os grupos de gênero.

A seguir, para preparar a base de teste, para cada instância de treino $x_i \in X$, se o vetor de predições na instância do grupo i pertencer ao grupo g, então ele é filtrado para base de testes.

Finalizada a preparação do treino e teste, é criado o modelo M_2 que é treinado com a matriz de treino completa do grupo g, $X^{tr[g]}$ e o vetor de classe alvo de treino do grupo g, $y^{[g]}$. Depois é feita a predição da base de testes do grupo g, $X^{[g]}$, com o vetor de predições de teste $\hat{y}^t$, e o resultado é inserido no vetor de predições de teste do grupo g, $\hat{y}^{t[g]}$. Por fim, o resultado das predições é dado por $\hat{y}^f$.

Tabela 5 – Descrição das variáveis do Algoritmo

Termo	Descrição
y^{tr}	Vetor de classe alvo do treino
$y^{tr[i]}$	Vetor de classe alvo, considerando o i-ésimo grupo de gênero
$y^{\hat{tr}}$	Vetor das predições do teste, do primeiro nível
$\hat{y}_i$	Vetor das predições do teste, considerando o i-ésimo grupo de gênero.
X^{tr}	Matriz do treino completa
$X^{tr[i]}$	Matriz do treino completa, considerando apenas intâncias do i-ésimo grupo de gêneros
X	Atributos do teste completa
$X^{[i]}$	Atributos do teste completa, considerando apenas intâncias do i-ésimo grupo de gêneros
M_1	Modelo primeiro nível
M_2	Modelo segundo nível
G	Conjunto de grupos
$\hat{y}^f$	Vetor de predições final

Algoritmo 1: Classificação Hierárquica

início
- $M_1 \leftarrow criamodelo(\mathrm{X}^{tr}, y^{tr})$
- $\hat{y}^{tr} \leftarrow prediz(\mathrm{X})$
- início das predições do segundo nível por grupo
- **para cada** g *em* G **faça**
  - prepara o treino, considerando apenas instâncias em que o y_i^{tr} pertence ao grupo g
  - **para cada** $x_i \in X^{tr}$ **faça**
    - **se** $y_i^{tr} = g$ **então**
      - $X^{tr[g]} \leftarrow \mathrm{X}^{tr[g]} \cup \{x_i\}$
      - $y^{tr[g]} \leftarrow \mathrm{y}^{tr[g]} \cup \{y_i^{tr[g]}\}$
    - **fim**
  - **fim**
  - prepara o teste, considerando apenas instâncias em que o $y^{tr[i]}$ pertence ao grupo g
  - **para cada** $x_i \in X$ **faça**
    - **se** $\hat{y}_i = g$ **então**
      - $X^{[g]} \leftarrow \mathrm{X}^{[g]} \cup \{x_i\}$
    - **fim**
  - **fim**
  - $M_2 \leftarrow criamodelo(\mathrm{X}^{tr[g]}, y^{tr[g]})$
  - $\hat{y}^{[g]} \leftarrow prediz(\mathrm{X}^g)$
  - $\hat{y}^f \leftarrow \hat{y}^f \cup \hat{y}^{[g]}$
- **fim**

fim

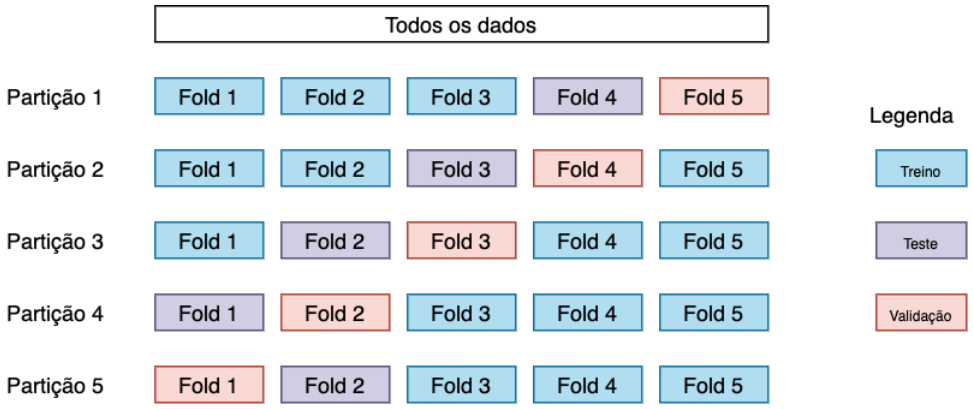
4.3 Metodologia de Avaliação

Para avaliar o desempenho do modelo de classificação feito, foi utilizada a Validação Cruzada *K-fold* (*K-fold Cross Validation*, em inglês). Esse método consiste na divisão de um conjunto de dados em K-folds (ou subconjuntos) que são conjuntos de: treino, teste e validação (BAEZA-YATES; RIBEIRO-NETO, 1999).

A Figura 4 apresenta um exemplo de uma validação cruzada com 5 *folds*, de mesmo tamanho, divididos em: 3 *folds* de treino, 1 *fold* de teste e 1 *fold* de validação, para cada partição. O *fold* de validação é utilizado para estimação de parâmetros dos *folds* de treino e que são aplicados no *fold* de teste. Dessa forma, o treino é utilizado para treinar o modelo e a validação é utilizada para avaliar o modelo para cada hiperparâmetro variado.

Os valores de K variam de acordo com o *trade-off* entre viés e variância. Sendo que 5 e 10, são os valores que são geralmente utilizados, por terem se mostrado, os melhores *trade-offs*, segundo Baeza-Yates e Ribeiro-neto (1999). Neste trabalho, utilizou-se o valor de $K = 5$, gerando partições assim como no exemplo ilustrado na Figura 4.

Figura 4 – Validação Cruzada



Fonte: A autora.

4.4 Métricas de Avaliação

No estudo de projetos de classificação, espera-se melhorar o desempenho dos resultados ao longo do tempo e para estudar os aspectos como a sua generalização, as características de convergência dos dados e o desempenho computacional, são necessárias métricas de avaliação (RUSSELL; NORVIG, 2002).

Considere-se um conjunto C de elementos da classe c e P um conjunto de elementos preditos como a classe c. As métricas de avaliação considerando essa formulação de

conjuntos, são demonstradas a seguir:

$$\text{precisão(c)} = \frac{\mid C_c \cap P_c \mid}{\mid P_c \mid} \quad (4.2)$$

$$\text{recovação(c)} = \frac{\mid C_c \cap P_c \mid}{\mid C_c \mid} \quad (4.3)$$

$$F_1 = 2 \cdot \frac{\text{precisão} \cdot \text{revocação}}{\text{precisão} + \text{revocação}} \quad (4.4)$$

$$MacroF_1 = \frac{1}{\text{total de classes}} \cdot \sum_{n=i} F_1 \quad (4.5)$$

A matriz de confusão é uma representação possível para demonstração da acurácia de um modelo de classificação. Ela consiste em uma tabela que relata o número da predições feitas tendo as classes reais, no eixo y, e as classes preditas, no eixo x (RUSSELL; NORVIG, 2002). Neste trabalho, ela é utilizada para a análise por classe no Capítulo 5.

5 Avaliação Experimental

Neste capítulo é explicado o desenvolvimento do trabalho, incluindo os experimentos realizados e os resultados obtidos, considerando os conceitos e a metodologia apresentados nos capítulos anteriores.

5.1 Experimentos

Os experimentos foram feitos com os métodos de Aprendizado de Máquina escolhidos: o SVM, o Random Forest e Naive Bayes, cujas implementações são da biblioteca de Python, Scikit-learn de Pedregosa et al. (2011).

5.1.1 Metodologia Experimental

Os experimentos para a representação dos Atributos foram feitos com Classificação Hierárquica e Classificação Tradicional, para cada método de Aprendizado de Máquina com a representação de Atributos. Para variar os hiperparâmetros dos métodos foi utilizado o *Tree-Structured Parzen Estimator (TPE)* é uma abordagem que escolhe hiperparâmetros de forma aleatória porém enviesado. Ele é enviesado porque existe uma probabilidade maior de que o TPE escolha valores para hiperpârametros em regiões do plano que ainda não foram exploradas ou regiões do plano que já foram exploradas mas possuem um bom resultado (BAEZA-YATES; RIBEIRO-NETO, 1999).

Para cada experimento, considera-se a configuração de Validação Cruzada de 5 *folds*. Sendo que para cada *fold*, variou-se 100 vezes os hiperparâmetros de cada método pelo TPE. Cada vez que se varia um hiperpâmetro, trata-se de um ensaio cuja implementação foi feita com a biblioteca Optuna proposta por Akiba et al. (2019). Ou seja, para cada *fold* de um dado experimento, são feitos 100 ensaios, com valores de hiperpâmetros variados pelo TPE. Os valores utilizados pelo TPE para variação de cada método para essa configuração de experimento estão demonstrados na Tabela 6.

Tabela 6 – Variação de hiperparâmetros com representação de Atributos

Método	**Variação de hiperparâmetro**
SVM	*Custo C: 2^-5 a 2^15*
Random Forest	*Número mínimo de instâncias para poda: 10% a 50%,* *Número máximo de atributos: 10% a 90%* *Número de árvores: 5 a 100*
Naive Bayes	Nenhum

Os hiperparâmetros variados tem significados diferentes para cada método. O Custo C é variado no SVM, conforme a Tabela 6, aumentando o expoente de 2 em 2, como proposto por Hsu, Chang e Lin (2003). Esse Custo C, como explicado no Capítulo 2 determina o comportamento da margem do algoritmo, por isso sua variação determina a classificação do mesmo.

Para o Random Forest é variado o *Número mínimo de instâncias para poda* que, supondo um valor de 10%, significa que se o tivermos menos de 10% de instâncias para um nodo, ele vira uma folha, ou seja, acontece uma poda, o que reduz o *overfiting* (RUSSELL; NORVIG, 2002). O *Número máximo de atributos*, supondo um valor de 10%, significa que as árvores do Random Forest serão criadas com apenas 10% dos atributos das músicas. E o *Número de árvores* de árvores criadas é o que o nome do hiperparâmetro diz por si só.

Por fim, para o Naive Bayes não foi feita variação de parâmetros, e foi utilizado o modelo gaussiano do método sem variação de parâmetros, o que é adequado para o tipo de classificação proposta, segundo Russell e Norvig (2002).

Para os experimentos da representação de Bag of Words, nem todos os hiperparâmetros dos métodos de Aprendizado de Máquina foram variados. O SVM seguiu o mesmo padrão de variação da Tabela 6, porém para o Random Forest o número de árvores não variou e foi mantido como 10.

5.2 Avaliação dos resultados

Para reportar os resultados dos experimentos, aplicou-se nos *folds* de teste a melhor combinação de parâmetros para os n ensaios em cada partição da Validação Cruzada, e mediu-se o resultado dessa performance. Assim, o resultado final reportado de cada experimento é dado então pela média dos resultados das 5 partições da sua Validação Cruzada.

Para a representação de Atributos foram executados 100 ensaios e para representação de Bag of Words foi executado apenas um, com exceção somente do SVM, para o qual

foram feitos 2 ensaios. O fato do número de ensaios para Bag of Words ser muito menor do que para os Atributos, pode ter relação com a diferença entre os resultados. Porém, o Bag of Words naturalmente exige um tempo de execução maior assim como demanda mais memória. Por isso, foi decidido não executar mais ensaios para ele.

5.2.1 Análise de performance dos métodos

A métrica escolhida para avaliação dos métodos foi a Macro-F1, visto que ela é mais adequada para contextos de base de dados desbalanceadas, segundo Baeza-Yates e Ribeiro-neto (1999). Como os gêneros quando não agrupados estão desbalanceados, a Macro-F1 é adequada para este contexto. Na Tabela 7 encontram-se os resultados dos experimentos realizados neste trabalho, com as representações por Atributos e por Bag of Words, para cada uma das classificações: Hierárquica e Tradicional.

	Atributos		**Bag of Words**	
Método	*Hierárquico*	*Tradicional*	*Hierárquico*	*Tradicional*
SVM	43,73%	42,11%	32,18%	35,78%
RF	**58,67%**	56,53%	22,42%	14,53%
NB	46,06%	38,33%	22,85%	23,51%

Tabela 7 – Macro-F1 dos Experimentos

Conforme demonstrado na Tabela 7, para a representação por Bag of Words, com Classificação Tradicional, o método com melhor resultado é o SVM com apenas $35,78\%$ de Macro-F1, contra $14,53\%$ e $23,51\%$ do Random Forest e Naive Bayes, respectivamente. Sendo que, a variação de hiperparâmetros do SVM foi mantida para dois ensaios, o que determina que essa variação otimiza o resultado final. Porém, o método Naive Bayes, não tem hiperparâmetros variados, para nenhum de seus experimentos e seu resultado com a representação de Atributos para Classificação Tradicionsl é de $38,33\%$, que é maior que seu resultado $23,51\%$, para Bag of Words.

Outra característica comum dos experimentos com Naive Bayes é que como para este nenhum parâmetro é variado, seus ensaios tem sempre os mesmos parâmetros, o que qualitativamente é encarado como um ensaio apenas, mesmo quando são feitos 100 ensaios para o mesmo experimento. Logo, pode-se concluir que a representação dos Atributos é, em média, melhor que a representação por Bag of Words.

Ao comparar as metodologias de Classificação, levando em consideração apenas a representação de Atributos, percebe-se que os resultados da Macro-F1 para o modelo hierárquico, são todos melhores que os do tradicional. O que por fim, valida a ideia inicial

deste trabalho de que a Classificação Hierárquica tem um desempenho melhor para o problema de classificação em questão.

O resultado baixo do Bag of Words pode ser em função da menor variação dos hiperparâmetros. Mesmo assim, considerando o tempo de execução de ambas as representações, a de Atributos é mais vantajosa do que o Bag of Words. Só que não se descarta a possibilidade de, em trabalhos futuros, executar os experimentos de Bag of Words com variações maiores para obtenção de melhores resultados.

Com o objetivo de analisar a classificação por classe predita, o experimento de melhor resultado para representação de Atributos, é analisado na subseção a seguir. E na Seção 5.3 é demonstrada a relevância dos Atributos propostos neste trabalho.

5.2.2 Análise por classe do melhor método

Esta seção apresenta o resultado por classes do melhor metodo tanto na abordagem hirarquica e na tradicional.

O método Random Forest, foi o método que teve melhor desempenho entre os experimentos da representação de Atributos, com $58,67\%$ de Macro-F1 para Classificação Hierárquica e $56,53\%$ para Tradicional. Por isso, ele foi o método escolhido para análise por classe de predição, que é feita a seguir pela explicação de suas matrizes de confusão demonstradas na Figura 5. Em que o eixo x representa a classe (gênero) predita pelo modelo e o eixo y a classe (gênero) real das músicas.

As Figuras 5a e 5b representam as predições da Classificação Hierárquica no primeiro e segundo nível, respectivamente. Nota-se que na Figura 5a, no primeiro nível, praticamente todas os instâncias são preditas como de sua classe corretamente. Existe um erro, mas ele é pequeno.

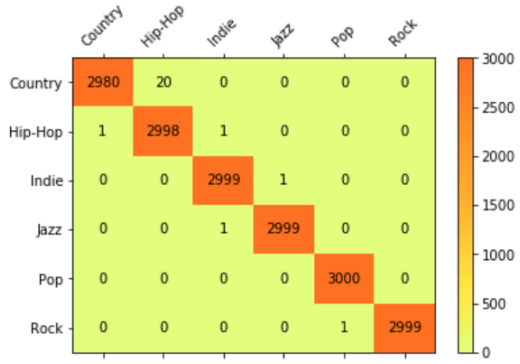
Já a Figura 5b, o segundo nível do mesmo experimento da Figura 5a, tem maior semelhança com a Figura 5c. O que acontece porque ambas correspondem aos resultados que predizem as classes de gêneros reais, não agrupados.

Ao analisar detalhadamente as Figuras 5b e 5c percebe-se que os gêneros não agrupados, Hip-Hop e Indie, são bem classificados nos dois casos e tem o mesmo erro. Mas é possível também observar a predição dos gêneros Pop e Eletrônico, somente como Pop, e ambos fazem parte do grupo Pop.

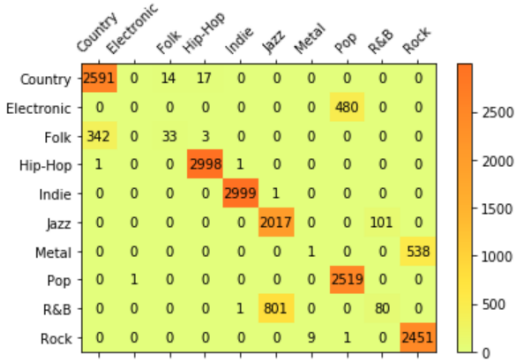
Outra situação que é visível na matriz da Figura 5b é a do gênero R&B em que pelo menos 80 instâncias que são R&B originalmente, foram classificadas como tal, enquanto que na Classificação Tradicional na Figura 5c, todas as instâncias R&B foram classificadas como Jazz, que é o gênero do agrupamento de R&B e Jazz. A mesma situação ocorre com o gênero o Folk, que teve 33 instâncias preditas corretamente no modelo hierárquico

enquanto que no tradicional, teve apenas 17.

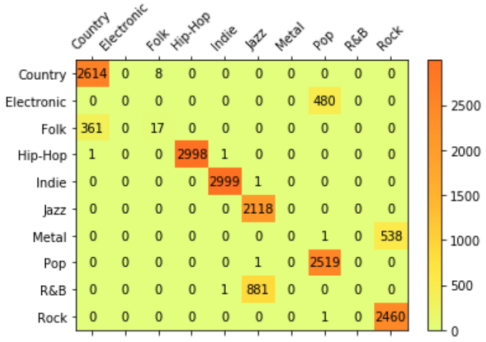
Diante do exposto, fica evidenciado que a Classificação Hierárquica possui um erro menor que a Classificação Tradicional nas predições. E por isso, conclui-se que obteve um melhor resultado. Mas esse resultado não se deve apenas à Classificação Hierárquica, pois como demonstrado na Tabela 7 a representação de Atributos foi a melhor representação, contribuindo para essa conclusões.



(a) Cl . Hierárquica - Primeiro nível



(b) Cl. Hierárquica - Segundo nível



(c) Cl. Tradicional

Figura 5 – Matrizes de Confusão

5.3 Relevância dos Atributos

A relevância dos Atributos propostos neste trabalho é discutida com base no *ranking* do Ganho de Informação (do inglês, *info gain*) dos Atributos, demonstrado na Tabela 8. Esse cálculo foi feito com o auxílio da ferramenta Weka, proposta por Hall et al. (2009). O Ganho de Informação é uma medida baseada em impureza, o qual usa entropia como medida de impureza (RUSSELL; NORVIG, 2002).

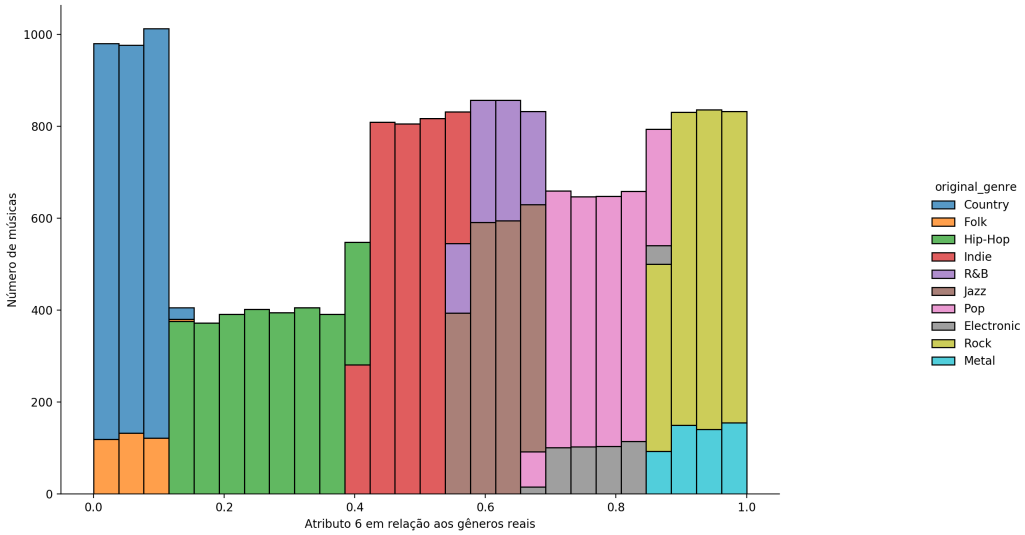
Conforme a Tabela 8 fica evidente que ganho de informação do Atributo 6, que é a relação da *Quantidade de versos únicos* de cada música, tem a maior relevância diante dos demais. Por isso, foi feito um gráfico que toma como referência a *Quantidade de versos únicos* e os gêneros reais das músicas. Esse gráfico está representado na Figura 6.

Atributo	**Ganho de Informação**
6	*2,585*
5	0,3187
2	0,3177
1	0,3163
4	0,2597
14	0,1192
10	0,1192
11	0,1192
7	0,1169
8	0,105
12	0,105
9	0,0877
13	0,0877
15	0,0824
16	0,0643
17	0,0606
3	0

Tabela 8 – Ranking de Ganho de Informação dos Atributos

Analisando o gráfico da Figura 6, percebe-se que existe a distribuição da *Quantidade de versos únicos* das músicas em relação a número de músicas, para cada um dos gêneros reais. Por exemplo, o gênero Country, representado pela cor azul escuro se encontra no início do gráfico, com uma quantidade versos relativamente menor que os demais gêneros, junto ao gênero Folk. O Hip-Hop, que engloba gêneros como Rap, Trap-Hop, dentre outros, possui quantidade de versos maior em relação ao gênero Country. No entanto, possui uma distribuição maior em relação aos demais gêneros, possivelmente por englobar gêneros que possuem menos versos, de um modo geral.

Figura 6 – Valor do Atributo 6 para os gêneros originais



Fonte: A autora.

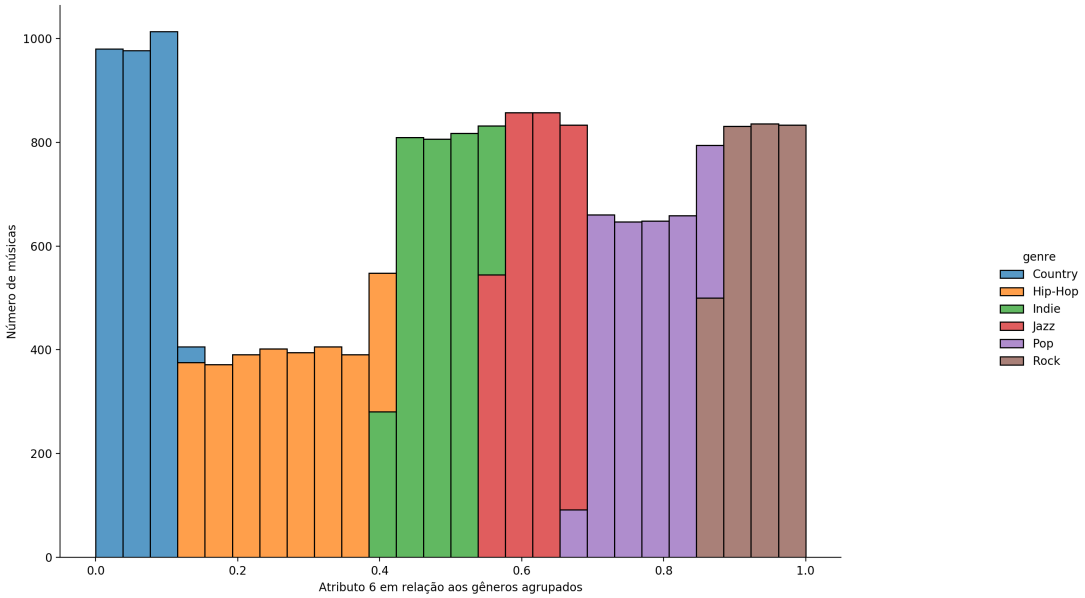
Dessa forma, nota-se um padrão ao longo do gráfico. Os gêneros que foram agrupados por terem semelhanças demonstradas no trabalho proposto por (PAMPALK; FLEXER; WIDMER, 2005), se encontram próximos na disposição do gráfico da Figura 6. O Metal, por sua vez, se encontra dividindo a posição de maior Ganho de Informação com o Rock, o que pode ser explicado em função de o Metal ter comumente um vocabulário maior, que outros gêneros musicais (MARTíN-GóMEZ; CáCERES, 2018).

Ainda sobre esse gráfico, é possível analisar o gênero Pop, que não tem um histórico de versos muito elaborados, no entanto, ele apresenta uma quantidade de versos representada pela faixa entre 0.6 e 0.83, aproximadamente. Isso pode ter acontecido pois o Atributo 6 contabiliza termos como *"yeah", "oh"* entre outros desses tipos, que pode tornar um verso único em relação ao outro simplesmente pela diferença do número de *"yeah"* presentes em cada um. Assim, pode-se ter um número maior de versos únicos, mesmo que eles sejam semânticamente idênticos. Essa mesma análise se aplica ao gênero Eletrônico.

Conforme a visualização da distribuição das quantidades de versos por gêneros reais, é possível observar que os gêneros que dividem as mesmas faixas de quantidade, como Rock e Metal, com os valores de 0.84 a 1.0, são gêneros que foram agrupados, confirmando assim, a validade dos agrupamentos por similaridade propostos neste trabalho.

A título de curiosidade, a Figura 7 ilustra a distribuição da *Quantidade de versos únicos* das músicas, ou Atributo 6, em relação aos gêneros agrupados.

Figura 7 – Valor do Atributo 6 para os gêneros agrupados



Fonte: A autora.

6 Conclusão

Por meio do desenvolvimento deste trabalho, foi possível identificar que o método de Classificação Hierárquica tem o desempenho melhor para a representação de atributos criados, em relação à Classificação Tradicional. A combinação dessa metodologia de classificação proposta com as representações relevantes como a da quantidade de versos, pode gerar desempenhos melhores, o que pode vir a ser um trabalho futuro.

Ao comparar as representações das músicas, a média dos resultados para Bag of Words foi menor do que a média para os Atributos propostos neste trabalho. No entanto, é preciso considerar o fato de que foram executados menos ensaios para Bag of Words e ainda que a lentidão de execução dessa representação já seja uma desvantagem dela, a variação de hiperparâmetros num número de ensaios maior é uma contribuição válida.

Como sugestão de trabalho futuros, pode-se realizar o balanceamento dos gêneros musicais para cada grupo e reforçar a criação de Atributos como o de quantidade de versos únicos por música. Também é possível utilizar outras combinações de métodos de Aprendizado de Máquina para predição dos gêneros como Redes Neurais, por exemplo. Outra possibilidade, seria combinar outros tipos de informação das músicas como o áudio, além de suas letras.

O impacto deste trabalho para Recuperação de Informação de Música é que a partir da análise dos Atributos e do entendimento de quais deles são mais relevantes, é possível facilitar e até melhorar a organização de problemas parecidos com o deste escopo. Como por exemplo, problemas de extração de informação, organização e classificação de música e até de recomendação de conteúdo musical, dentre outros.

Referências

AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2019.

ANNIS, D. Kendall's advanced theory of statistics, vol. 1: Distribution theory (6th ed.). alan stuart and j. keith ord; kendall's advanced theory of statistics, vol. 2a: Classical inference and the linear model (6th ed.). alan stuart and j. keith ord, and steven f. arnold. *Journal of the American Statistical Association*, v. 101, p. 1721–1721, 02 2006.

BAEZA-YATES, R.; RIBEIRO-NETO, B. Modern information retrieval. 07 1999.

CLARK, L.; TELLEGEN, A. On the dimensional and hierarchical structure of affect. *Lee Anna Clark*, v. 10, 07 1999.

CORTES, C.; VAPNIK, V. Support-vector networks. In: *Machine Learning*. [S.l.: s.n.], 1995. p. 273–297.

CRAMMER, K.; SINGER, Y. On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research*, JMLR. org, v. 2, p. 265–292, 2001. Disponível em: <http://jmlr.org/papers/volume2/crammer01a/crammer01a.pdf>.

DUNNING, T. Statistical identification of language. 01 1996.

HALL, M. et al. The WEKA data mining software: an update. *SIGKDD Explorations*, v. 11, n. 1, p. 10–18, 2009.

HSU, C.-w.; CHANG, C.-c.; LIN, C.-J. A practical guide to support vector classification chih-wei hsu, chih-chung chang, and chih-jen lin. 11 2003.

KAGGLE. 2018. <https://www.kaggle.com/gyani95/380000-lyrics-from-metrolyrics/data>. Acessado em: 2019-10-30.

KOLCHINSKY, A. et al. The minor fall, the major lift: Inferring emotional valence of musical chords through lyrics. *CoRR*, abs/1706.08609, 2017. Disponível em: <http://arxiv.org/abs/1706.08609>.

MAHEDERO, J. et al. Natural language processing of lyrics. In: . [S.l.: s.n.], 2005. p. 475–478.

MARTíN-GóMEZ, L.; CáCERES, M. N. Applying data mining for sentiment analysis in music. In: . [S.l.: s.n.], 2018. p. 198–205. ISBN 978-3-319-61577-6.

MAYER, R.; NEUMAYER, R.; RAUBER, A. Rhyme and style features for musical genre classification by song lyrics. In: . [S.l.: s.n.], 2008. p. 337–342.

MAYER, R.; RAUBER, A. Music genre classification by ensembles of audio and lyrics features. In: *ISMIR*. [S.l.: s.n.], 2011.

METROLYRICS. 2020. <https://www.metrolyrics.com/>. Acessado em: 2019-10-30.

MEYER, L. B. *Emotion and Meaning in Music*. [S.l.]: University of Chicago Press, 1956.

MITCHELL, T. M. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN 978-0-07-042807-2.

NAKANO, F. K. et al. Top-down strategies for hierarchical classification of transposable elements with neural networks. In: . [S.l.: s.n.], 2017. p. 2539–2546.

PAMPALK, E.; FLEXER, A.; WIDMER, G. Improvements of audio-based music similarity and genre classificaton. In: . [S.l.: s.n.], 2005. p. 628–633.

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

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach (2nd Edition)*. [S.l.]: Prentice Hall, 2002. Hardcover. ISBN 0137903952.

SCHEGLOFF, E. A. Whose text? whose context? *Discourse & Society*, v. 8, n. 2, p. 165–187, 1997. Disponível em: <https://doi.org/10.1177/0957926597008002002>.

TSAPTSINOS, A. Lyrics-based music genre classification using a hierarchical attention network. *CoRR*, abs/1707.04678, 2017. Disponível em: <http://arxiv.org/abs/1707.04678>.

WHITMAN, B.; LAWRENCE, S. Inferring descriptions and similarity for music from community metadata. *Proceedings of the 2002 International Computer Music Conference*, 01 2002.

YANG, D.; LEE, W.-S. Music emotion identification from lyrics. In: . [S.l.: s.n.], 2010. p. 624 – 629.

ZHANG, H. The optimality of naive bayes. In: BARR, V.; MARKOV, Z. (Ed.). *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*. [S.l.]: AAAI Press, 2004.