

Universidade Federal de Minas Gerais

Mineração de Dados - Trabalho Prático 2 - Agrupamentos

Sistema de Recomendação de Artistas baseado em Clusters de
Acordes e Letras de Músicas

Professor: Wagner Meira

Aluno: Pedro Lóes

Data: 19-12-2021

Conteúdo

| | |
|--|-----------|
| 1 - Entendimento do Negócio | 2 |
| 1.1 Objetivos do Negócio | 3 |
| 1.2 Avaliação da Situação | 3 |
| 1.3 Objetivos da Mineração de Dados | 3 |
| 1.4 Descrição do Projeto | 4 |
| 2 - Entendimento dos Dados | 4 |
| 2.1 Coleta dos Dados | 4 |
| 2.2 Descrição dos Dados | 5 |
| 2.3 Verificação da Qualidade e Integridade dos Dados | 7 |
| 2.4 Exploração dos Dados | 7 |
| 3 Pré Processamento dos Dados | 9 |
| 3.1 Matriz Esparsa da Modelagem | 10 |
| 4 - Modelagem | 10 |
| 4.1 Matriz de Distâncias | 10 |
| 4.2 Modelo | 10 |
| 5 Avaliação do Modelo | 11 |
| 5.1 Dendrograma de Resultados | 11 |
| 6 Implantação | 12 |
| 6.1 Aplicativo Shiny Clusters Musicais Asom | 12 |
| 7 Fontes e Referencias | 13 |
| 7.1 Links | 13 |

1 - Entendimento do Negócio

No ano de 1999 surgiu o primeiro aplicativo de compartilhamento de músicas gratuito na internet no formato mp3 chamado Napster. Apesar de sua popularidade entre estudantes e consumidores de música com padrão aquisitivo baixo, o programa passou por diversos problemas relacionados a direitos autorais.

Em 2001 o serviço foi encerrado porém ficou claro para a indústria fonográfica que o compartilhamento seria um caminho a ser explorado devido a grande adesão de usuários.

Em 2003 a Apple lançou a iTunes Store, uma biblioteca de músicas online que trazia um modelo de negócio viável e mais adequado aos interesses da indústria fonográfica. Cada música custava \$0.99 e a Apple repassava parte dos lucros às gravadoras e consequentemente aos artistas.

Apesar de plataformas de streaming gratuito continuarem a operar ilegalmente, a quantidade de vírus que poderiam ser transmitidos por essas plataformas forçou grande parte dos usuários a migrar para a loja iTunes.

No ano de 2005 o aplicativo Pandora mesclou os serviços streaming gratuito com anúncios e planos de inscrições mensais para remover as propagandas. Além disso, a plataforma utilizou sistemas de recomendação de músicas baseado nos dados dos usuários e seus padrões de consumo.

Em 2008 esse modelo de negócio foi solidificado por plataformas como Spotify, Tidal e YouTube que aprimoraram a compactação dos arquivos, disponibilizaram uma vasta coleção de artistas, possibilitaram a criação de listas e desenvolveram uma precificação com inscrição mensal adequada para viabilidade do negócio.

Na segunda década do século 20 a indústria de streaming teve como seu foco principal a melhoria da experiência do usuário oferecendo novas funcionalidades e interações inteligentes com os usuários.

1.1 Objetivos do Negócio

A indústria de streaming utiliza dados dos usuários para compreender padrões de comportamento com intuito melhorar a performance dos sistemas de inteligência artificial.

Os modelos de sistemas de recomendações utilizados pelo setor se baseiam no comportamento dos gostos dos usuários. Um dos grandes desafios do negócio no presente é minerar os arquivos de áudio.

Mesmo com sofisticadas técnicas de compactação, os arquivos possuem uma quantidade muito grande de informação. Tal fato gera um custo computacional inviável para agrupar artistas por padrões musicais no formato de áudio.

O objetivo deste projeto foi utilizar acordes e letras de músicas em formato de texto com intuito de detectar padrões musicais e líricos e poder determinar grupos ou clusters de artistas.

1.2 Avaliação da Situação

A elaboração deste projeto envolveu a extração de dados não estruturados das letras e cifras de músicas que sistetizam as informações do áudio em formato de texto. Esse processo consistiu na extração das palavras das letras e acordes das músicas no formato de texto.

O sítio [CifraClub](#) foi utilizado para recuperar as informações de milhares de músicas de diversos gêneros.

Para a execução do projeto foi utilizada uma máquina com processador Intel I7-9700 de 3GhZ e 16 GB de memória do tipo RAM com sistema Linux Ubuntu Lts 20.04.

A plataforma de desenvolvimento integrado Vscode, bem como a linguagem Python serão utilizadas para extrair os dados. A linguagem R e o IDE RStudio serão utilizados para a análise descritiva e desenvolvimento de diversos modelos de aprendizado de máquinas não supervisionado do tipo agrupamentos.

Finalmente para a produção ou implantação será produzido uma aplicação com a biblioteca shiny da linguagem R para a visualização da análise exploratória dos dados e para utilização de modelos interativos que permitam a experimentação de diferentes funções de distância entre os clusters para definir o parâmetro link para definir o algoritmo de agrupamentos dos artistas.

1.3 Objetivos da Mineração de Dados

As técnicas de extração, transformação e limpeza dos dados foram utilizadas para recuperar as informações da web estruturadas em html, limpar o ruído e formatar as informações de cifras e letras, transformar os dados para o formato de matriz esparsa e finalmente para o formato matriz de distâncias requisitado pelo algoritmo de agrupamentos.

O algoritmo de cluster hierárquico foi a técnica mineração de dados utilizada para construir agrupamentos do tipo aprendizado de máquinas não supervisionado. Esse algoritmo foi implementado com objetivo de

descobrir grupos de artistas semelhantes utilizando as informações de palavras das letras e acordes das músicas sintetizadas no formatos de músicas cifradas.

O banco de dados não possui a variável resposta e portanto não ocorreu o particionamento do banco. Para verificar a qualidade do modelo foi desenvolvida uma ferramenta gráfica do tipo dendograma do modelo iterativo para possibilitar a inspeção visual dos agrupamentos gerados em função da medida de distância escolhida para formar os clusters.

1.4 Descrição do Projeto

O projeto consistiu na etapa de entendimento do negócio onde foi realizada uma pesquisa histórica. As etapas de extração, entendimento e preparação dos dados, onde as informações foram recuperadas, exploradas e verificadas. A etapa de pré-processamento, onde foi utilizada a técnica one hot encoding. As etapas de modelagem e avaliação do modelo onde foram aplicadas as técnicas de clusterização. E finalmente, a etapa de implantação do projeto, onde foi desenvolvido um aplicativo web.

2 - Entendimento dos Dados

Esta etapa cobriu a extração dos dados, a descrição de suas meta informações, e verificações de sua qualidade para identificar problemas como observações faltantes ou ausência de integridade. Finalmente a análise exploratória foi realizada para investigar tendências e ou padrões aparentes.

2.1 Coleta dos Dados

A coleta de dados de letras e acordes das músicas do sítio CifraClub foi realizada com o pacote BeautifulSoup da linguagem Python contemplando 10.000 músicas de 100 artistas. Nesta etapa foram utilizadas as bibliotecas requests, BeautifulSoup, Pandas e Numpy para extração e armazenamento das informações.

Script da Função *extraí_cifra* na linguagem Python

```
# Define função extraí_cifra com parametro artista
def extraí_cifra(artista):
    url = "https://www.cifraclub.com.br/" + artista + "/"
    reqs = requests.get(url)
    soup = BeautifulSoup(reqs.text, 'lxml')
    aux = [a['href'] for a in soup.find_all('a', {'class':"art_music-link"}, href=True)]
    musicas = [x for x in aux if "letra" not in x]
    enderecos = ['https://www.cifraclub.com.br' + musica for musica in musicas]
    repertorio = {}
    for endereco in enderecos:
        reqs = requests.get(endereco)
        soup = BeautifulSoup(reqs.text, 'lxml')
        chave = endereco.replace(url,"").replace("/", "")
        repertorio[chave] = [element.text for s in soup.find_all('pre')
                             for element in s.findChildren('b')]
    df = pd.DataFrame(dict([ (k,pd.Series(v))
                             for k,v in repertorio.items() ])).melt().dropna()
    df.columns = ["Músicas", "Acordes"]
    df.to_csv(artista + "_acordes" + ".csv")
    return()
```

Script da Função *extraí_letra* na linguagem Python

```
# Define função extraí_letra com parametro artista
def extraí_letra(artista):
    url = "https://www.cifraclub.com.br/" + artista + "/"
    reqs = requests.get(url)
    soup = BeautifulSoup(reqs.text, 'lxml')
    titulo = [a['href']]
    for a in soup.find_all('a', {'class': "art_music-link"}, href=True)]
    musicas = [x for x in titulo if "letra" not in x]
    enderecos = ['https://www.cifraclub.com.br' + musica + "letra"
                 for musica in musicas]
    letras = {}
    for endereco in enderecos:
        reqs = requests.get(endereco)
        soup = BeautifulSoup(reqs.text, 'lxml')
        aux = [str(element.nextSibling)
                for s in soup.find_all("div", {"class": "letra"})
                for element in s.findChildren('br')]
        aux2 = [frase.split() for frase in aux]
        chave = endereco.replace(url, "").replace("/", "").replace("letra", "")
        letras[chave] = [item for sublista in aux2 for item in sublista]
        df = pd.DataFrame(dict([ (k, pd.Series(v))
                                for k, v in letras.items() ])).melt().dropna()
        df.columns = ["Musicas", "Palavras"]
        df.to_csv(artista + "_letras" + ".csv")
    return()
```

Script do Algoritmo de Extração

```
# Extraí cifras e letras dos artistas
artistas = ["adoniran-barbosa", ..., "zeze-di-camargo-e-luciano"]
for artista in artistas:
    extraí_cifra(artista)
    extraí_letra(artista)
```

- Foram armazenados 2 arquivos planos separados por vírgulas para cada artista, indexados por nome das músicas para identificar os acordes e as letras de cada música para cada artista.

2.2 Descrição dos Dados

A descrição dos dados consistiu no levantamento de meta informações sobre o banco de dados tais como, formato dos arquivos, número de observações, número de atributos (Acordes e Palavras) de cada música e de cada artista, tipo de dados dos atributos de cada música.

Também foi realizado o levantamento do dicionário de palavras e acordes que compõem as grandes matrizes binárias de informações que serão construídas com artistas nas linhas como observações e palavras e acordes nas colunas como atributos.

Tabela da Amostra dos Arquivos de 10 Artistas

| Id | Artistas | Arquivos de Letras | Arquivos de Acordes |
|----|-------------------|------------------------------|-------------------------------|
| 6 | anitta | anitta_letras.csv | anitta_acordes.csv |
| 19 | cartola | cartola_letras.csv | cartola_acordes.csv |
| 38 | emicida | emicida_letras.csv | emicida_acordes.csv |
| 54 | luan-santana | luan-santana_letras.csv | luan-santana_acordes.csv |
| 69 | nelson-cavaquinho | nelson-cavaquinho_letras.csv | nelson-cavaquinho_acordes.csv |
| 88 | trio-nordestino | trio-nordestino_letras.csv | trio-nordestino_acordes.csv |

- A lista de arquivos de letras apresentou 93 arquivos, portanto, dos 100 artistas indicados para extração somente 93% das letras destes artistas foram corretamente recuperadas.
- A lista de arquivos de acordes apresentou 96 arquivos, portanto, dos 100 artistas indicados para extração somente 96% dos acordes destes artistas foram corretamente recuperados.

Importação dos dados e Junção.

Os arquivos foram importados e unificados em 2 data frames no formato longo. Ou seja, as informações de artistas e músicas foram repetidas para cada acorde.

O primeiro data frame possui na primeira e a segunda colunas as informações ou atributos que indicam a que música e artista cada acorde pertence.

| Index | Artistas | Musicas | Acordes |
|--------|---------------------------|---------------|---------|
| 1 | adoniran-barbosa | trem-das-onze | D7 |
| 2 | adoniran-barbosa | trem-das-onze | Dm |
| 3 | adoniran-barbosa | trem-das-onze | Gm |
| 617504 | zeze-di-camargo-e-luciano | yesterday | Bb |
| 617505 | zeze-di-camargo-e-luciano | yesterday | C |
| 617506 | zeze-di-camargo-e-luciano | yesterday | F |

- O data frame de acordes apresentou 617.506 observações de acordes.

O segundo data frame possui na primeira e a segunda colunas as informações ou atributos que indicam a que música e artista cada palavra pertence.

| Index | Artistas | Musicas | Palavras |
|---------|---------------------------|---------------|----------|
| 1 | adoniran-barbosa | trem-das-onze | Sinto |
| 2 | adoniran-barbosa | trem-das-onze | muito |
| 3 | adoniran-barbosa | trem-das-onze | amor, |
| 1262431 | zeze-di-camargo-e-luciano | yesterday | eu |
| 1262432 | zeze-di-camargo-e-luciano | yesterday | None |
| 1262433 | NA | NA | NA |

- O data frame de letras apresentou 1.262.433 palavras de letras das músicas. A palavra amor seguida de “,” indicou a possibilidade de limpeza das palavras para remover caracteres indesejados. A última observação apresentou valores ausentes em todos os campos.

2.3 Verificação da Qualidade e Integridade dos Dados

A verificação da qualidade dos dados consistiu na identificação de dados faltantes e inspeção da estrutura dos dados.

Dados Faltantes e Limpeza

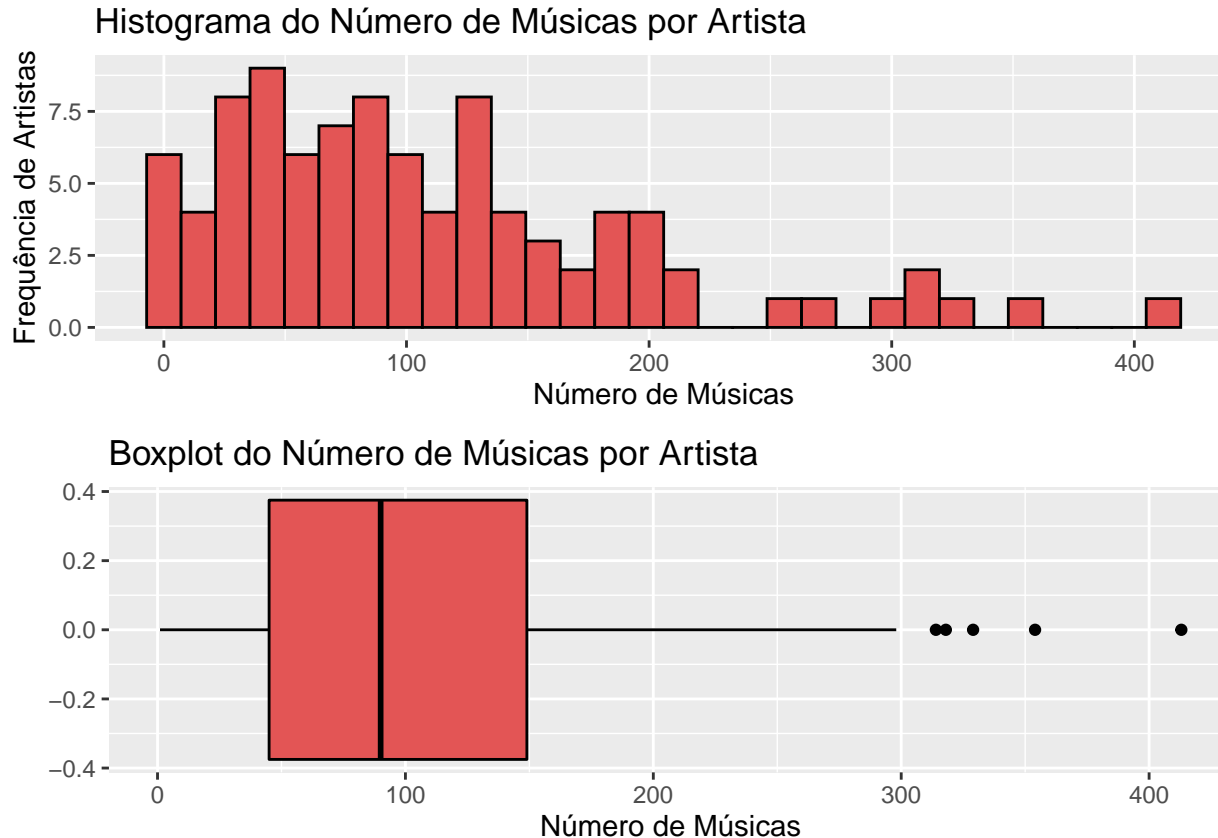
O data frame de acordes não apresentou nenhuma observação com informação faltante e portanto nenhuma observação foi removida. O data frame de letras apresentou apenas uma observação ou palavra com informação faltante. A mesma foi analisada e constatou-se que a mesma também não possuía informação sobre o artista e a letra e portanto essa observação foi removida.

Pontuação e acentuação foram removidas. A padronização das palavras para letras minúsculas foi utilizada para eliminar caracteres que não fossem alfanuméricos. Desta forma buscou-se salientar os estilos e padrões líricos dos artistas que poderiam contribuir para o agrupamento dos artistas.

2.4 Exploração dos Dados

A exploração dos dados consistiu na produção e análise de gráficos e tabelas para identificar padrões aparentes que auxiliassem na etapa de modelagem e contribuíssem para comparar padrões musicais e líricos de artistas. Outra utilidade da análise descritiva foi a inspeção do equilíbrio da quantidade de músicas por artista para evitar viés no modelo.

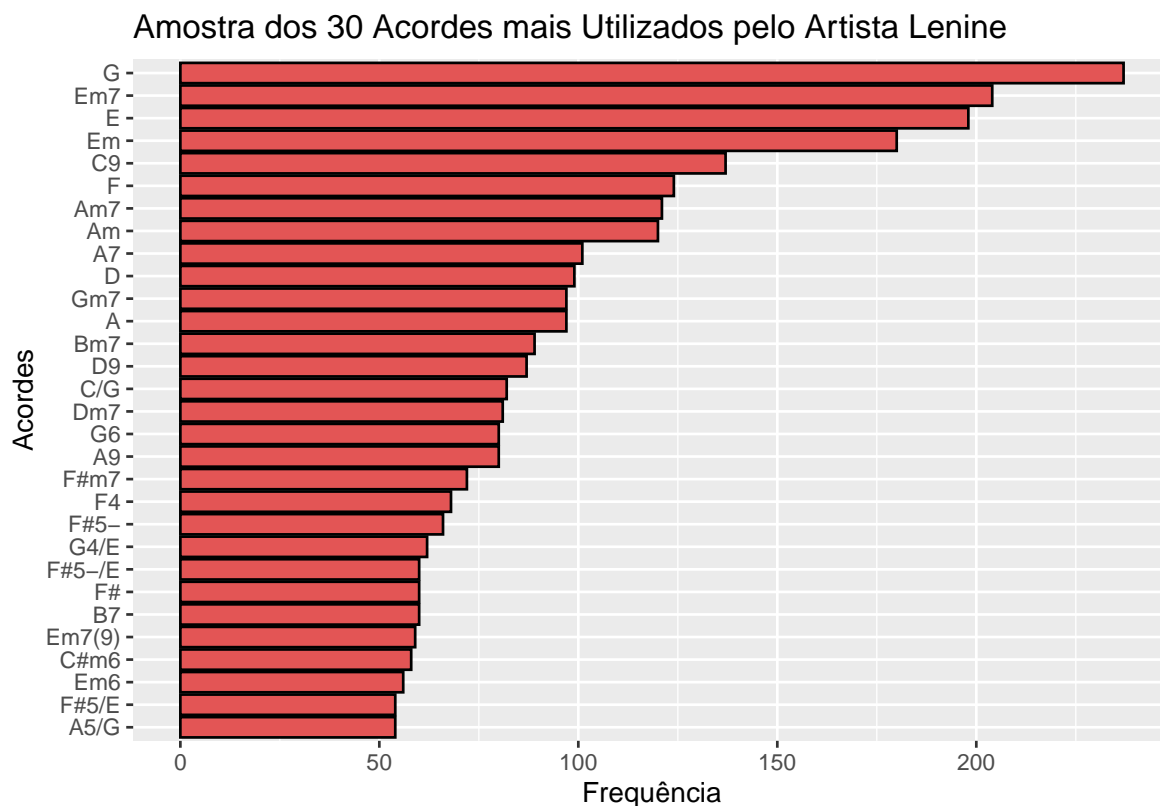
Distribuição da Contagem de Músicas por artista



- A distribuição do número de músicas por artista observada no histograma indicou cauda assimétrica à direita o que levou ao diagnóstico de assimetria positiva. A análise da distribuição também permitiu a conclusão de que a maioria dos artista apresentou até 150 músicas e alguns poucos artistas apresentaram mais do que 300 músicas.
- Os artistas GOG e RZO apresentaram somente 1 música cada e o artista Caetano Veloso, com 413 músicas, foi o artista que apresentou o maior número de músicas. O primeiro quartil de 45, e o terceiro de 149 indicaram a caixa do gráfico boxplot que representa a massa central de 50% da amostra. A mediana de 90 e a média de 110 forneceram uma perspectiva da influência dos artistas com muitas músicas sobre as medidas de centralidade dos dados.
- Finalmente o gráfico do tipo box plot indicou a presença de 5 artistas com número de músicas discrepante no sentido de possuírem muito mais músicas que o restante dos artistas da amostra. Para balancear a amostra foram escolhidos apenas os artistas com mais do que 50 músicas.

Número de Acordes Mais Frequentes

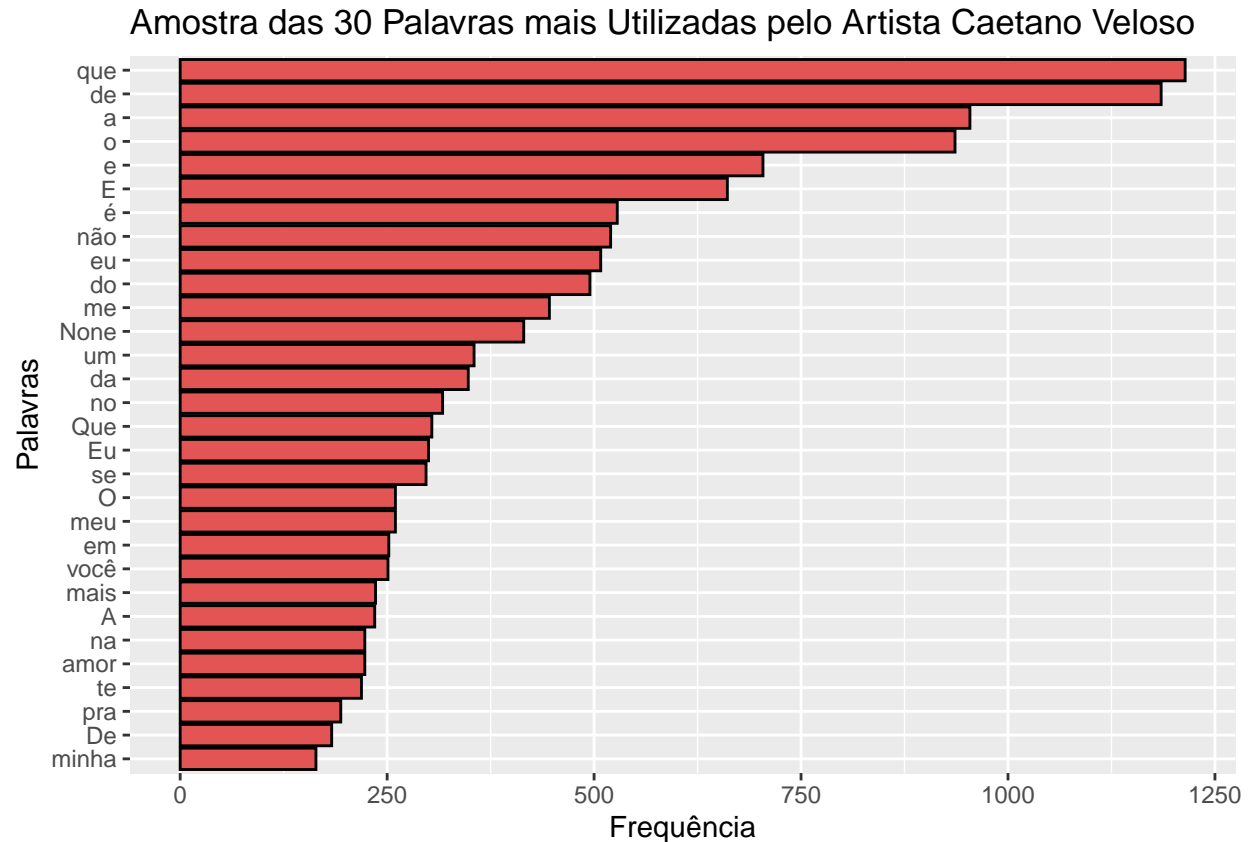
A análise do número de acordes mais utilizados por um determinado artista permitiu explorar quais os acordes foram mais utilizados por um determinado artista.



- A análise do artista Lenine indicou preferência pelos acordes Sol maior, Mi menor com sétima, Mi e Mi menor com mais do que 150 observações. Tal fato ressaltou a preferência pelo acorde Mí e suas variantes.
- A fase de implantação gráfica permitiu a interação para a escolha do artista e para a escolha do número de acordes que determinaria o tamanho do n-grama, ou seja, a contagem do número de vezes que se pode observar uma sequência de acordes. Essa ferramenta permitiu a comparação dos padrões musicais dos artistas.

Número de Palavras Mais Frequentes

A análise do número de palavras mais utilizadas por um determinado artista permitiu explorar quais as palavras foram mais utilizadas por um determinado artista.



- A proposição de subordinação “de”, os artigos “o” e “a”, o pronome “que” usado para se referir ao que foi dito anteriormente e a conjunção de ligação “e”, foram as palavras mais utilizadas pelo artista.
- A fase de implantação gráfica permitiu a interação para a escolha do artista e para a escolha do número de palavras que determinaria o tamanho do n-grama, ou seja, a contagem do número de vezes que se pode observar uma sequência de palavras. Essa ferramenta possibilitou comparar os padrões literários dos artistas.

3 Pré Processamento dos Dados

A etapa de pré- processamento dos dados consistiu na transformação dos dois data frames utilizando a técnica de geração de matrizes esparsas one hot encoding. Esta técnica consiste na alteração do formato dos data frames de longos para largos.

O novo data frame de acordes passou a ter a contagem do número de vezes em que um acorde aparece nas colunas ou atributos. E as palavras também foram transformadas em atributos ou colunas. Portanto, após a transformação, a matriz dispôs as músicas como observações nas linhas e os acordes e cifras nas colunas como atributos.

3.1 Matriz Esparsa da Modelagem

| | abajur | abala | abalada | abalado | Gsus4 | Gsus4/B | Gsus4/D | Gsus4/F# |
|---------------------------|--------|-------|---------|---------|-------|---------|---------|----------|
| lenine | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| los-hermanos | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| luan-santana | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| toquinho | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| trio-nordestino | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| victor-leo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| vinicius-de-moraes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ze-ramalho | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| zeca-pagodinho | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| zeze-di-camargo-e-luciano | 2 | 0 | 0 | 0 | 8 | 0 | 0 | 0 |

4 - Modelagem

4.1 Matriz de Distâncias

O algoritmo de clusterização considera a matriz de distâncias entre as músicas para formar os agrupamentos. A etapa de implantação permitiu o experimento com diversas métricas para o cálculo da matriz de distâncias.

A métrica escolhida para ilustrar o modelo foi a distância binária. Essa métrica é similar à distância de Jaccard onde os artistas(vetores) são considerados como bits binários, portanto, elementos diferentes de zero estão ligados e iguais a zero desligados.

A distância é a proporção de atributos que estão ligados em relação aos outros bits onde pelo menos um está ligado. Para ilustrar a matriz de distância foram exibidas as distâncias entre os primeiros 5 artistas.

| | adoniran-barbosa | alceu-valenca | amado-batista | ana-carolina | angra |
|------------------|------------------|---------------|---------------|--------------|----------|
| adoniran-barbosa | 0.0000 | 1830.883 | 4711.090 | 579.8552 | 2031.759 |
| alceu-valenca | 1830.8828 | 0.000 | 3299.130 | 1634.9196 | 1551.299 |
| amado-batista | 4711.0899 | 3299.130 | 0.000 | 4413.3244 | 3804.075 |
| ana-carolina | 579.8552 | 1634.920 | 4413.324 | 0.0000 | 1875.650 |
| angra | 2031.7586 | 1551.299 | 3804.075 | 1875.6500 | 0.000 |

4.2 Modelo

O modelo de clusters hierárquicos escolhido considerou o método Ward.D. O critério de variância mínima de Ward minimiza a variância total dentro de cada cluster.

As distâncias iniciais do cluster no método de variância mínima de Ward são, portanto, definidas como sendo a distância euclidiana quadrada entre os pontos.

- $d_{i,j} = d(\{X_i\}, \{X_j\}) = |X_i - X_j|^2$

O algoritmo executa uma análise de cluster hierárquica usando o conjunto de dissimilaridades para os objetos que estão sendo agrupados, neste caso os artistas. Inicialmente, cada objeto é atribuído ao seu próprio cluster e, em seguida, o algoritmo prossegue de forma iterativa. Em cada estágio dois clusters mais semelhantes são unificados, o processo continua até que haja apenas um único cluster.

5 Avaliação do Modelo

Os algoritmos de clusters fazem parte de um conjunto de técnicas de aprendizado de máquinas onde os dados não possuem a variável resposta que permite avaliar o desempenho dos modelos.

Possíveis soluções são realizar a classificação manual subjetiva de todas as músicas em gêneros ou considerar uma amostra heterogênea de artistas que claramente pertencem a estilos diferentes, porém com uma quantidade também razoável de artistas dentro de um mesmo gênero para verificar os grupos sugeridos.

A segunda solução foi utilizada com uma amostra de 60 artistas e a inspeção dos agrupamentos resultantes foi realizada por meio da técnica gráfica de cluster denominada dendrograma.

5.1 Dendrograma de Resultados

A avaliação dos agrupamentos foi realizada com a construção de um gráfico do tipo dendrograma no formato circular e poda em 20 clusters para determinar os 20 grupos ou estilos de artistas mais semelhantes.

Dendrograma dos Clusters de Artistas



- De acordo com o dendrograma e considerando os gêneros musicais samba, MPB, sertanejo, axé, baião, rock, pop rock, samba rock, metal, pode-se observar que os agrupamentos ficaram relativamente bem definidos com alguns poucos artistas apresentando classificação pouco intuitiva.
- Contudo é válido destacar que gêneros musicais são rótulos subjetivos e o consumo de música contemporâneo apresenta perfis mais ecléticos o que favorece a sugestão de músicas de artistas de gêneros diferentes que talvez possuam padrões musicais e líricos que sejam semelhantes e agradem aos usuários das plataformas de streaming.

- O cluster com Clara Nunes, Gonzaguinha, Dorival Caymmi, Arlindo Cruz, João Nogueira e Paulinho da Viola representou os artistas do gênero samba de raiz.
- O cluster com Luiz Gonzaga, Dominginhos e Trio Nordestino representou o gênero baião.
- O cluster Fernando e Sorocaba, Gustavo Lima, Jorge e Matheus, Henrique e Juliano e Marília Mendonça representou o gênero sertanejo.
- O cluster Jorge Ben Jor, Ana Carolina e Seu Jorge representaram o gênero Samba Rock.
- O cluster com Sepultura e Angra representou o agrupamento de bandas de metal ou com letras em inglês.
- O cluster com Raimundos Capital Inicial e Legião Urbana representou o rock nacional.
- Alguns cluster apresentaram agrupamento de artista pouco intuitivo como por exemplo o cluster de Axé com banda Eva e Araketu, misturados com o gênero sertanejo representado por luan santana, bruno e marrone e Zezé di camargo e Luciano e Victor E Leo e ainda pop rock com Nando Reis e Barão Vermelho.

6 Implantação

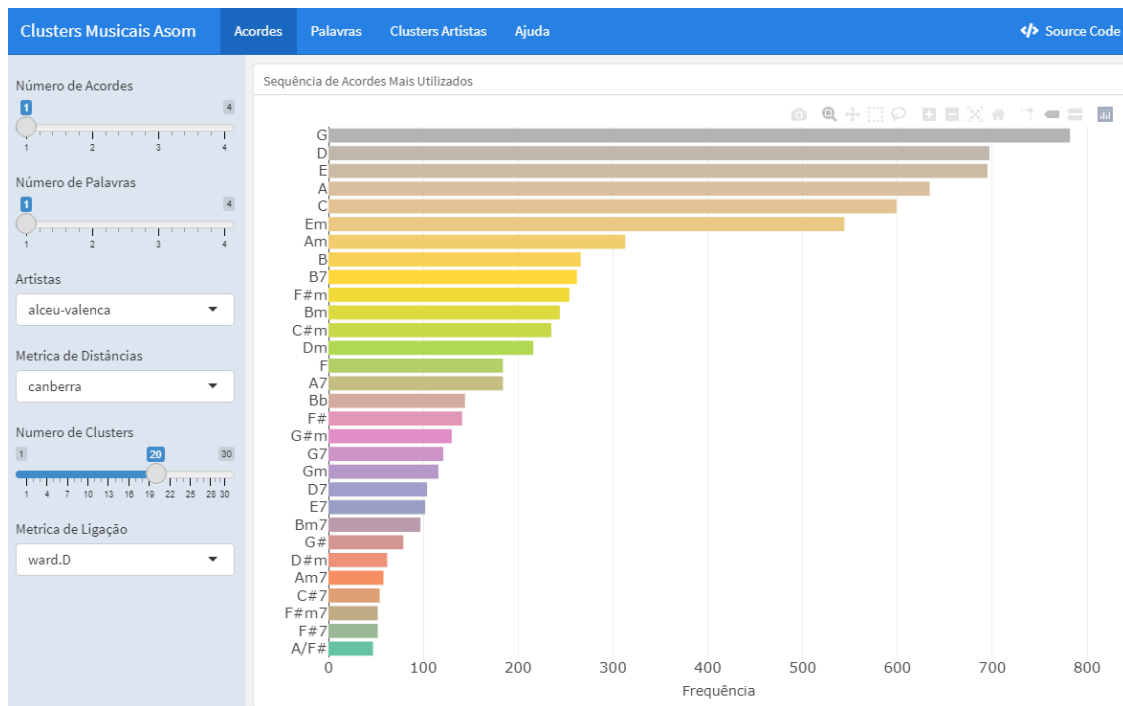
A implantação do melhor modelo para sugerir novas músicas aos usuários da plataforma de streaming poderá ser alocada nos servidores dos aplicativos mobile e desktop das plataformas.

Esta etapa deverá ser executada com a equipe de desenvolvimento para conectar o modelo ao banco de dados de cada usuário e automatizar o processo de sugestão iniciando desta forma o estágio de modelo em produção.

6.1 Aplicativo Shiny Clusters Musicais Asom

A etapa de implantação consistiu na construção de uma aplicação web denominada Asom. Esta aplicação foi desenvolvida com o pacote shiny da linguagem R, onde é possível interagir com a análise descritiva e experimentar modelos com diferentes métricas de distâncias e parâmetros de ligação dos agrupamentos. O aplicativo de clusters musicais Asom possui com os seguintes serviços:

- Barra lateral de Botões Interativos:
 - Botão Número de Acordes seleciona a quantidade de acordes (n-gramas de acordes).
 - Botão Número de Palavras seleciona a quantidade de palavras (n-gramas de palavras).
 - Botão Artista seleciona um artista para as análises descritivas individuais e comparativas.
 - Botão Métrica de Distâncias seleciona a métrica de distância para a matriz de distâncias.
 - Botão Número de Clusters seleciona o número de cluster que o algoritmo manter na poda.
 - Botão Métrica de Ligação seleciona qual vai ser o tipo de ligação do algoritmo de cluster.
- Menu:
 - Menu Acordes que oferecem a visualização de um gráfico de barras da contagem dos n-gramas dos acordes.
 - Menu Palavras que oferece a visualização de um gráfico de barras da contagem dos n-gramas das palavras.
 - Menu Cluster Musicais que oferece a visualização de um gráfico do tipo dendograma com os agrupamentos de artistas.
 - Menu Ajuda que oferece uma explicação sobre o funcionamento do aplicativo.



Link: [Cluster Músicais Asom](#)

7 Fontes e Referencias

As fontes das informações, dados e referências, foram disponibilizados abaixo no formato de links para os endereços url onde essas informações estão disponibilizadas em 20/12/2021.

7.1 Links

- [Zaki, M. J., Meira, W. Data Mining and Machine Learning: Fundamental Concepts and algorithms. Cambridge university press.](#)
- [CifraClub Acervo de Letras e Acordes de Músicas](#)
- [Dynaudio Historia do Streaming de Audio](#)
- [Linguagem Python](#)
- [Pacote Python Requests](#)
- [Pacote Python Beautiful Soup](#)
- [Linguagem R](#)
- [Pacote R ggplot2](#)
- [Pacote R stats](#)
- [Método de Ligação Ward](#)
- [GitHub](#)