



TOCANTINS
GOVERNO DO ESTADO



Alunos: Iago Batista Antunes Leobas

Disciplina: Mineração de Dados

Data: 20/11/2022

Relatório Sistemas Neuro-Fuzzy

Dado um conjunto de dados adequado, analise-se:

- **Estrutura:** atributos, perspectiva e contribuição
- **Classificadores:** Previsão e padrões
- **Opções de testes:** Individualmente e conjunto (treinamento e teste)
- **Métricas de validação:** Instâncias corretas e incorretas, média de erro e precisão

Tema Dataset: Top 50 músicas do Spotify em 2021

Link do kaggle:

<https://www.kaggle.com/datasets/equinxx/spotify-top-50-songs-in-2021?resource=download>

Resolução:

Inicialmente, é realizada a importação do dataset e apresentação dos dados:

O dataset escolhido tem como foco, as top 50 músicas mais ouvidas do spotify em 2021.

As colunas encontradas no dataset são:

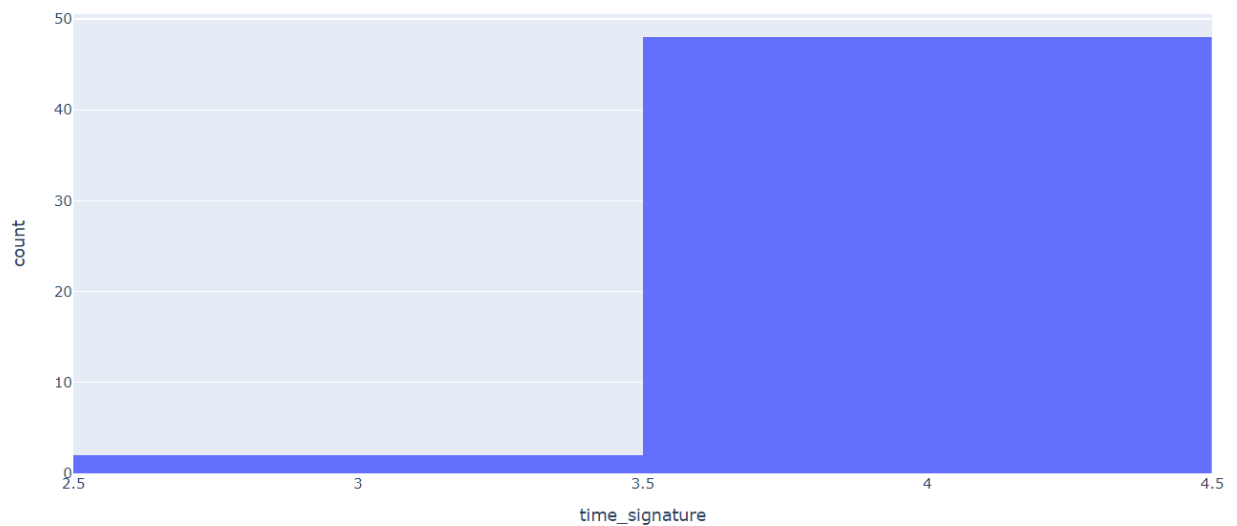
- Artist_name - Nome do artista ou grupo
- Track_name - Nome da música
- Track_id - Id dado a música
- Popularity - Popularidade, quanto maior o valor, mais popular é a música
- Danceability - Quanto maior o valor, mais fácil é dançar essa música

- Energy - A energia de uma música, quanto maior o valor, mais enérgica é a música
- Key - O tom em que a faixa está. Os inteiros são mapeados para as notas usando a notação padrão de Classe de Pitch. Por exemplo, 0 = C, 1 = C#/D ♭, 2 = D e assim por diante. Se nenhuma chave foi detectada, o valor é -1 (intervalo: -1; 11)
- Loudness (dB) - Quanto maior o valor, mais alta a música
- Mode - Indica a modalidade (maior ou menor) de uma faixa, o tipo de escala da qual deriva seu conteúdo melódico. Maior é representado por 1 e menor é 0
- Speechiness - Quanto maior o valor, mais palavras faladas a música contém
- Acousticness - Quanto maior o valor, mais acústica é a música
- Instrumentalness - o número de vocais em uma música. Quanto mais próximo o valor de 1,0, mais instrumental é a música
- Liveness - Quanto maior o valor, maior a probabilidade de a música ser uma gravação ao vivo
- Valence - Quanto maior o valor, mais clima positivo para a música
- Tempo - O andamento geral estimado de uma faixa em batidas por minuto (BPM)
- Duration - Duração da música em ms
- Time signature - Uma assinatura de compasso estimada. A fórmula de compasso (medidor) é uma convenção de notação para especificar quantas batidas há em cada compasso. A fórmula de compasso varia de 3 a 7, indicando fórmulas de compasso de "3/4" a "7/4"

Para diminuir as colunas encontradas, a única retirada a princípio do dataset é a coluna com o ID da música (Track_id), esse que não interessa inicialmente à observação.

Também é feita a verificação de valores nulos, esses que não foram encontrados dentro do dataset.

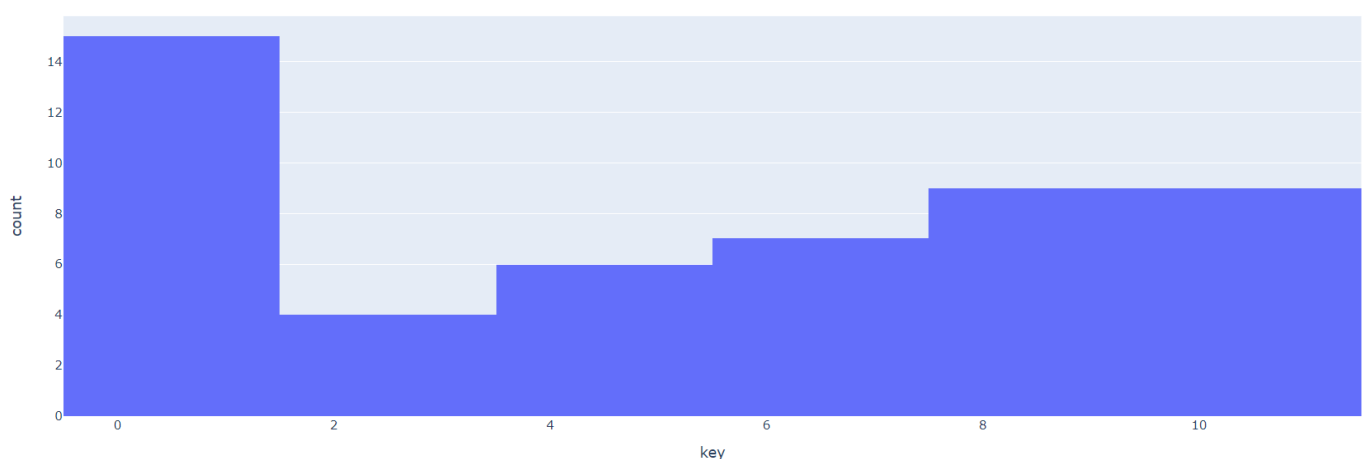
A primeira observação feita foi realizada com o uso de histograma para verificar como o "Time signature" se repete nas músicas top 50 de 2021.



A primeira observação a se fazer é o padrão de músicas com compassos de 4/4, no caso 48 das 50 músicas, e apenas 2 músicas com 3/4.

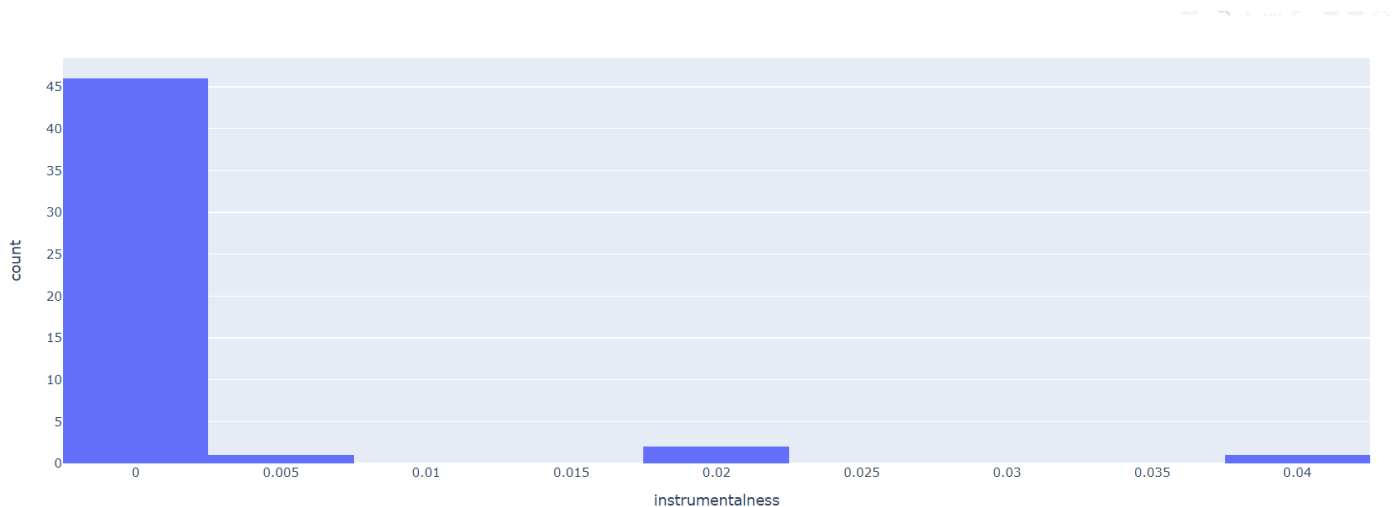
O que se pode analisar é a preferência dos ouvintes em padrões já familiares, no caso, compassos que não fogem do usual ouvido.

Outro dado observado é a prevalência dos dois tons de “C” (Dó) e “C#” (Dó sustenido) como mais utilizados nas top 50 músicas do spotify.



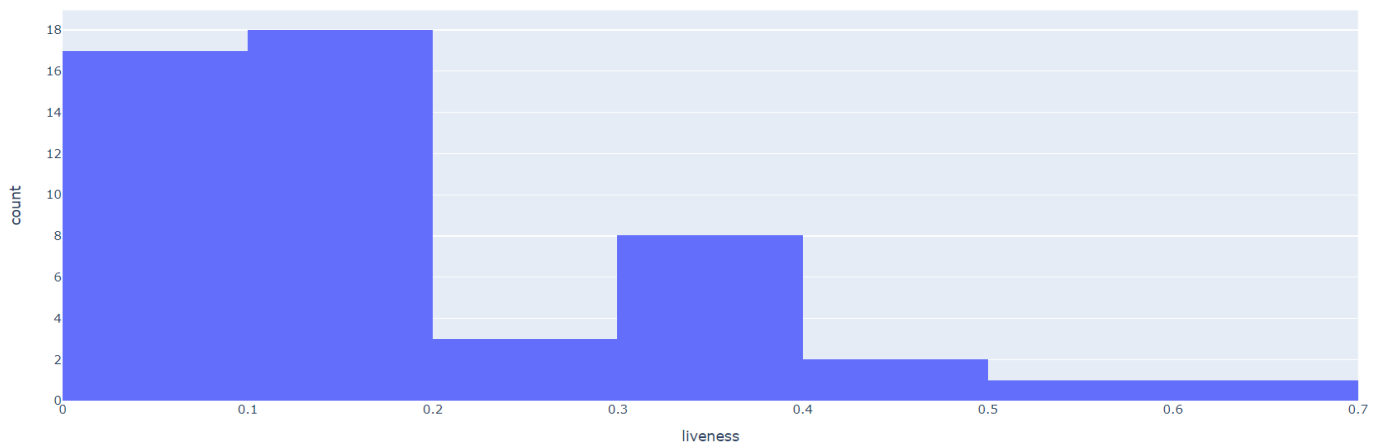
Foi realizada observações de diversos dados, e um que teve bastante destaque foi o “instrumentalness”, mostrando que existe uma tendência a não instrumentalização

das músicas. Por isso é difícil ver músicas de vertente clássicas ou apenas instrumentais entrando nas top músicas mais ouvidas.



É notório que dentre as 50 músicas mais ouvidas de 2021, nenhuma foi considerada 0,5 de Instrumentalidade, ou seja, nenhuma música que compõe a lista das mais populares tem 50% de sua composição de melodia apenas com instrumentos(s).

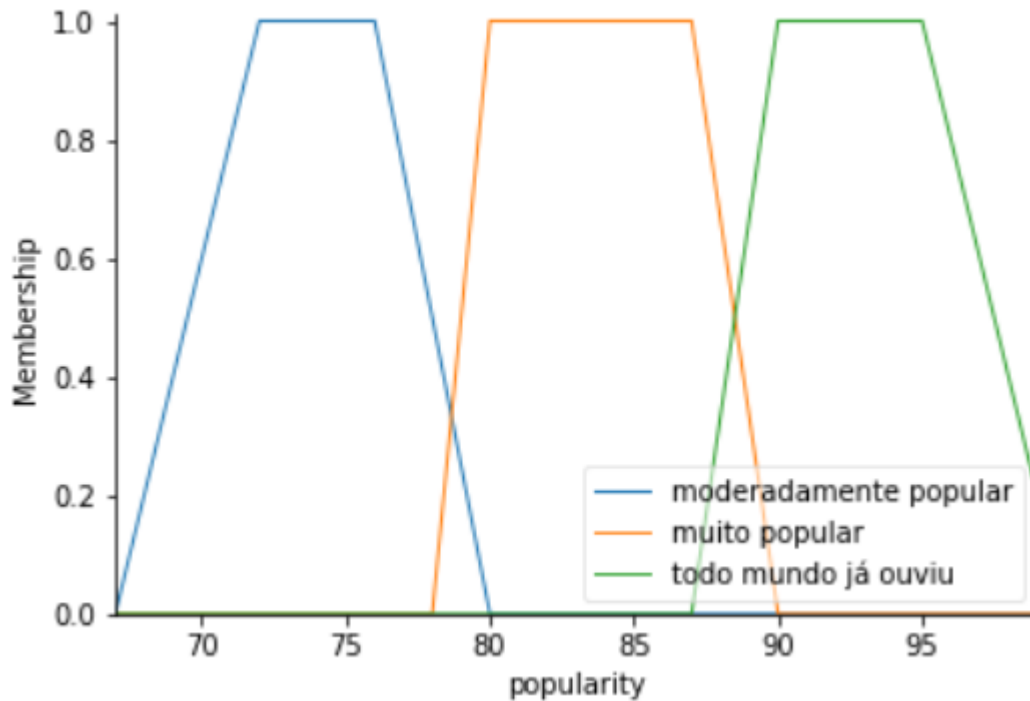
Outro dado interessante é a grande quantidade de músicas gravadas em estúdio profissional e menos em shows. O “liveness” mostra a tendência de músicas não serem mais espontâneas e cada vez mais roteirizadas.



Trabalhando agora com a parte de testes, foi realizada a etapa de testes com o conjunto inteiro, no caso o “size” igual a 1 pois um dos problemas encontrados ao trabalhar com o dataset pequeno é a dificuldade de extrair conhecimento de uma amostra pequena, por isso foi realizado com ele completo.

Em seguida, foi realizada a instalação do pacote “skfuzzy” para realizar o processo de métricas e validação.

Após dar o comando “describe” na coluna “popularity” e tirar informações importantes sobre essa coluna, foi feito então a divisão das métricas e categorização, em 'moderadamente popular' - 'muito popular' - 'todo mundo já ouviu'. Visto que todas já são músicas bem populares, por isso esse foi um dos problemas encontrados ao trabalhar com um dataset pequeno.



As métricas passadas para essa relação pode ser encontrada no colab na íntegra:

Link:

https://colab.research.google.com/drive/1lok1yk7_Inlvho0DN7xje1AXh7DJYWf_?usp=sharing