



Como criar um Sistema de Recomendação de Músicas? **Simplificado.**

Leonardo Mauro P. Moraes, MSc.
Senior Artificial Intelligence Engineer



THE
DEVELOPER'S
CONFERENCE





Agenda



- **Contexto**
O que quero fazer? Por que?
- **Teoria**
Como fazer?
- **Estudo de Caso**
Resultados do experimento.



1. Contexto

Música

- *This Is Your Brain On Music | How Music Benefits The Brain | YouCurious? (YouTube)*
- Libera dopamina, causando satisfação.
- Altera o humor; *e.g.*,
canções felizes e tristes.
- Ajuda a ignorar os sinais cerebrais de fadiga usando 7% menos oxigênio ao se exercitar.



Música

- Muitos tocadores de música
 - *Music players*
- Competem via playlists, recomendação...
 - música certa para você!



Proposta - O que quero fazer?



Dadas as milhares de músicas, como podemos recomendar músicas **similares** as músicas ouvidas anteriormente?

Ideia de [Guilherme Muzzi](#)



Metodologia - Como fazer?

- Encontrar um conjunto de dados
 - Spotify Song Attributes
- Como encontrar músicas similares?
 - Similaridade
 - Algoritmo
- Experimento e resultados





Metodologia - Dataset

- Spotify Song Attributes, [Kaggle](#), do GeorgeMcIntire;
 - 2017 músicas com atributos extraídos Spotify (API).
- Algumas características interessantes:
 - *acousticness* – uso de instrumentos (oposto de eletrônica),
 - *danceability* – quanto fácil de dançar,
 - *energy* – quanto a música é enérgica,
 - *valence* – quanto positiva é a música,
 - *instrumentalness, liveness, speechiness*.





2. Teoria





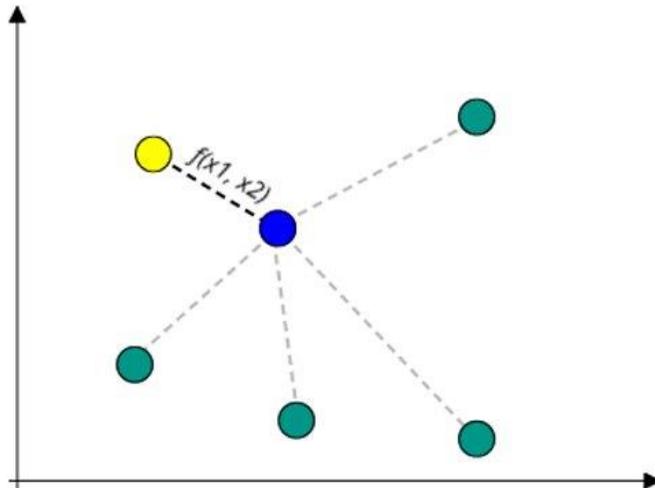
Teoria - Similaridade

Similaridade é um métrica que calcula o quanto dois objetos possuem formas, valores similares, ou estão próximos.

Considere que as músicas são pontos num espaço, desta forma, podemos computar a similaridade de duas músicas medindo a distâncias entre elas, utilizando uma função de distância $f(x_1, x_2)$.



Teoria - Similaridade



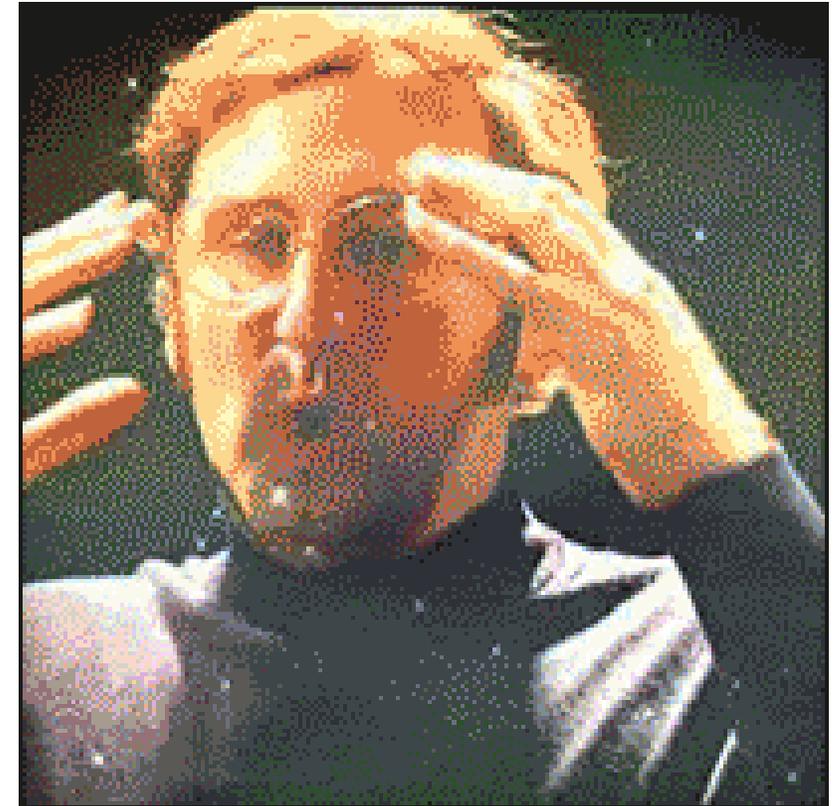
Considere que músicas são pontos num espaço; ou seja, vetores de características.

- Cada ponto é uma música;
- $f(x_1, x_2)$ é uma função de distância;
- Qual é a música mais similar a azul?
É a música amarela.
- Resolvido! Só precisamos de uma função.



Teoria - Distâncias

- Momento “**Mind-Blowing**” ...
- Cada tipo de dados tem um conjunto de funções de distâncias
 - Imagem, texto, áudio
 - Series temporais, fluxos
 - Vetores, ou pontos

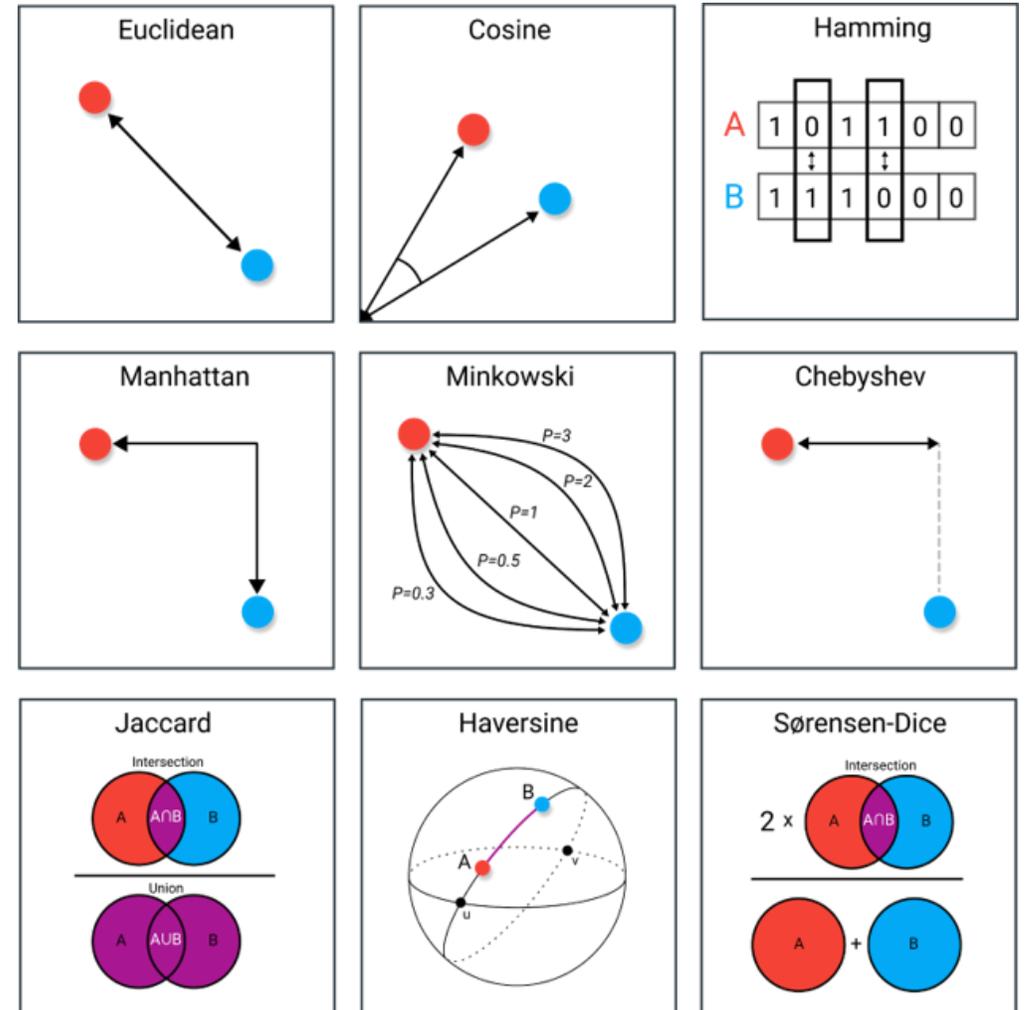


[Mind Blow Wow GIF](#)

Teoria - Distâncias



- 9 Distance Measures in Data Science
- Métricas de similaridade
- Calculando similaridade entre strings
- *Em Python, R, Julia, etc...*





Teoria - Distâncias

- Vetores - pontos

$$song_i = (attr_1, attr_2, \dots, attr_m)$$

- Família Minkowski
 - Distância Manhattan
 - Distância Euclidiana
 - ...
 - Distância Chebyshev

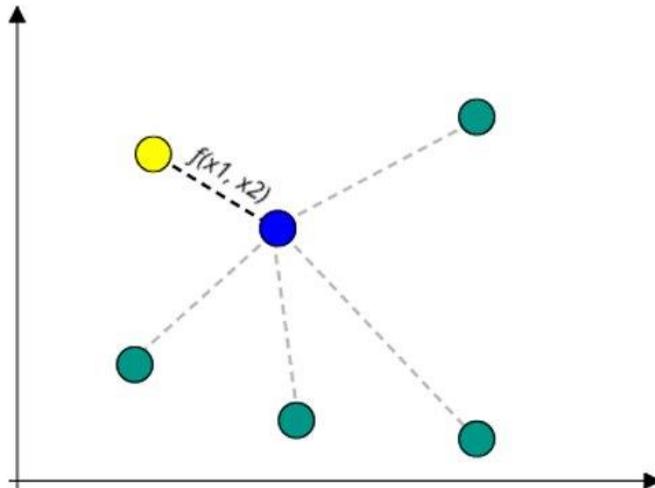
- Distância Euclidiana. Escolhida

- Uma linha reta;
- Teorema de Pitágoras;
- Todas as características possuem o mesmo peso.

$$f(x_1, x_2) = \sqrt{(x_1 - x_2)^2}$$



Teoria - Algoritmo



Considere que músicas são pontos num espaço; ou seja, vetores de características.

- k -nearest neighbors (k -NN)
 - Algoritmo de Inteligência Artificial
- Pseudocódigo
 - (1) Selecione um ponto de consulta
 - (2) Calcule as distâncias para as músicas
 - (3) Retorna as k mais similares



3. Experimentos

Experimento I - Músicas Similares



- [Avril Lavigne - Complicated](#) 
(como ponto de consulta)
- $k = 3$ (mais similares)

- (1) [DOLF - Fuck It All Up](#),
- (2) [Sam F - Limitless](#), e
- (3) [ASTR - Blue Hawaii](#)

Contemporâneas, eletrônicas,
e bem animadas!



Experimento I - Observações



De qualquer forma, a segunda e a terceira música não são tão similares ao pontos de consulta; isso pode acontecer pelo fato de:

(1) conjunto de dados pequeno; ou (2) não temos informações suficiente para comparar músicas, tais como ausência da melodia, ou das letra das músicas.





Experimento II - O que mais?

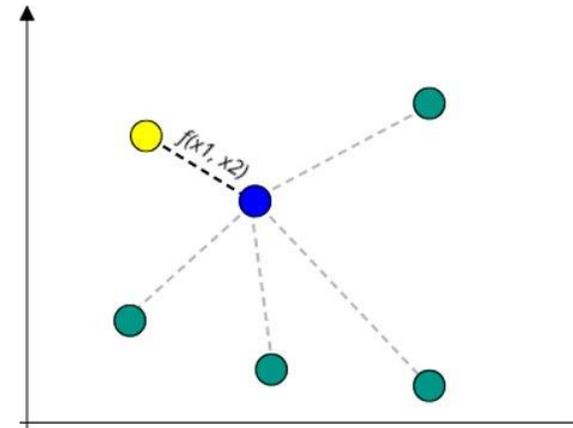
Isso é tudo que conseguimos fazer?

- Não! podemos inventar consultas...

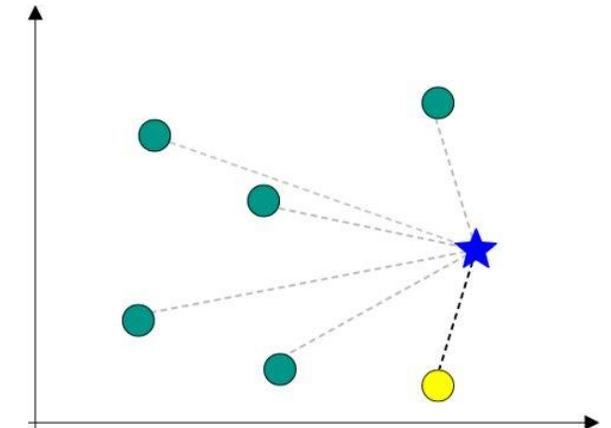
- (1) “Qual é a música mais animada?” ou
- (2) “Qual é a música menos animada?”

Como?

- Utilizando pontos sintéticos!



Ponto de consulta: música



Ponto de consulta: inventado



Experimento II - Música mais animada!



- [Gwen Stefani - Hollaback Girl](#)
(*música mais animada*)



“Qual é a música mais animada?”

- Atributos -
danceability, energy, e valence
- Ponto de consulta -
o maior valor possível para cada um dos atributos



4. Conclusão



Conclusão

- Distâncias de similaridade são abordagens simples para realizar consultas.
 - k-NN é utilizado em Banco de Dados
 - k-NN é utilizado em Inteligência Artificial
- Algoritmo de Recomendação Simples.
 - [Medium](#), [Kaggle](#)
- Outros experimentos...
 - (1) mudar a função de distância
 - (2) modificar o algoritmo
 - (3) criar novos pontos de consulta
 - (4) *(o que você tiver em mente)...*





Faça parte do time!

latam.sinch.com/careers/



@sinch.latam



linkedin.com/company/sinch



medium.com/wearesinch





Muito obrigado!

Leonardo Mauro P. Moraes

- Site - <http://leonardomauro.com>
- LinkedIn </in/leomaurodesenv/>

