

TÓPICOS EM CIÊNCIA DE DADOS PARA O ESPORTE

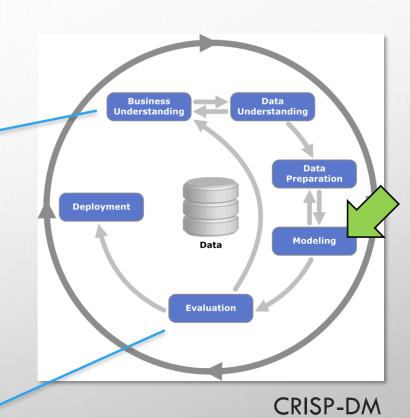
MACHINE LEARNING: AGRUPAMENTO

DIEGO RODRIGUES DSC

INFNET

CRONOGRAMA

DIA	NÚMERO	ÁREA	AULA	TRABALHOS
30/1/2024	1	Intro	Introdução a Disciplina e Organização do Ambiente	
1/2/2024	2	Dados	Coleta de Dados e Sensoriamento	
6/2/2024	3	Estatística	Variáveis Aleatórias	Grupos
8/2/2024	4		Análise Exploratória	
15/2/2024	5		Estatísticas para Ranqueamento	
20/02/2024	6		Ranqueamento Estatístico : ELO	
22/02/2024	7		Ranqueamento Estatístico : Glicko	
27/2/2024	8		Ranqueamento Estatístico : TrueSkill	
29/2/2024	9		Ranqueamento Estatístico : XELO	Base de Dados
5/3/2024	10	ML	Modelos de Aprendizado de Máquina	
7/3/2024	11		Machine Learning: Classificação	
12/3/2024	12		Machine Learning: Regressão	
14/3/2024	13		Machine Learning: Agrupamento	Pesquisa
19/3/2024	14		Machine Learning: Visão Computacional	
21/3/2024	15	Esportes	Aplicações & Artigos: Esportes Independentes	Modelo
26/3/2024	16		Aplicações & Artigos: Esportes de Objeto	
28/3/2024	17		Aplicações & Artigos: Esportes de Combate	
2/4/2024	18		Aplicações & Artigos : Betting	
4/4/2024	19	Workshop	Workshop	
9/4/2024	20		Apresentações de Trabalhos I	Apresentação
11/4/2024	21		Apresentações de Trabalhos II	



AGENDA

- PARTE 1 : TEORIA
 - CONCEITOS
 - MÉTRICAS DE DISTÂNCIA
 - ALGORITMOS
- PARTE 2 : AGRUPAMENTO DE JOGADORES DO FIFA22

AMBIENTE PYTHON

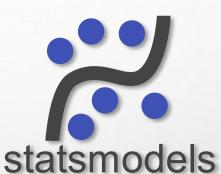


4. Variáveis Aleatórias



5. Visualização

6. Estimação e Inferência



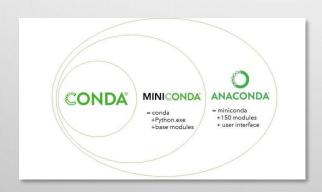


7. Machine Learning





1. Editor de Código



2. Gestor de Ambiente



3. Ambiente Python do Projeto

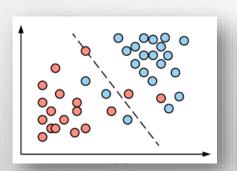


3. Notebook Dinâmico

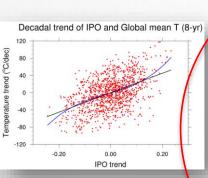


CONCEITOS

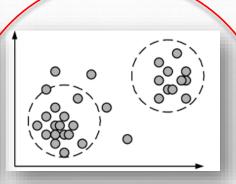
PARADIGMAS DE MODELAGEM ESTATÍSTICA



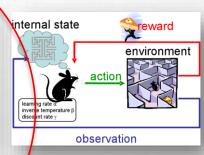
SUPERVISIONADO – CLASSIFICAÇÃO



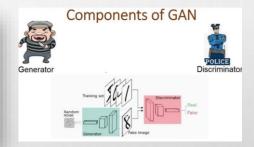
SUPERVISIONADO – REGRESSÃO



NÃO SUPERVISIONADO



APRENDIZADO POR REFORÇO



GENERATIVO

Agrupamento (Clustering)

Um bebê consegue **agrupar objetos por cor, tamanho, formato** e muitos outros atributos que ele pode observar nos objetos.

Diferentes maneiras de organizar os objetos são diferentes **estruturas de agrupamentos** existentes em uma amostra de dados.



De quantas maneiras estes blocos podem ser organizados em grupos?

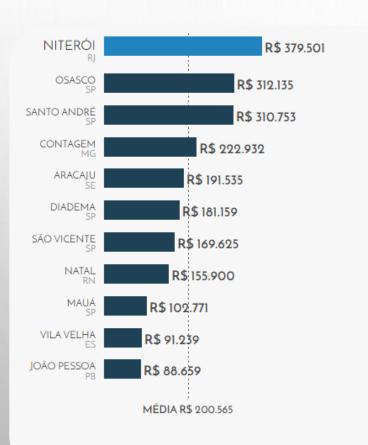
Um modelo de agrupamento é usado para identificar grupos, ou estruturas de agrupamentos, nos dados.

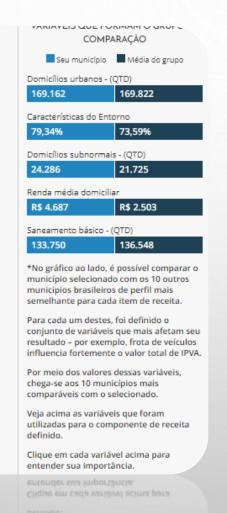
Aprendizado Não-Supervisionado

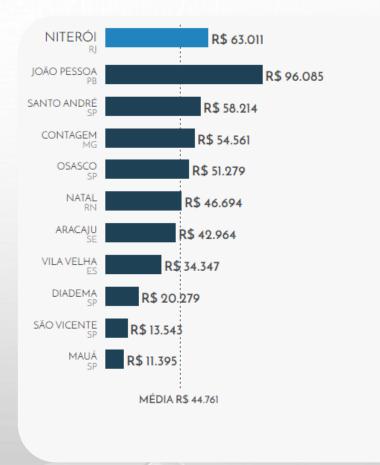
Não existe um **conhecimento "a priori" dos grupos** contidos nos dados. Algoritmos de agrupamento dependem fortemente de uma definição de "**distância**" ou "**similaridade**" entre as observações.

REPRESENTAÇÃO: COMO ENCONTRAR OS 10 MUNICÍPIOS MAIS SIMILARES A NITERÓI?











VARIAVEIS QUE FORMAM O GRUPO

município selecionado com os 10 outros municípios brasileiros de perfil mais semelhante para cada item de receita.

Para cada um destes, foi definido o conjunto de variáveis que mais afetam seu resultado – por exemplo, frota de veículos influencia fortemente o valor total de IPVA.

Por meio dos valores dessas variáveis, chega-se aos 10 municípios mais comparáveis com o selecionado.

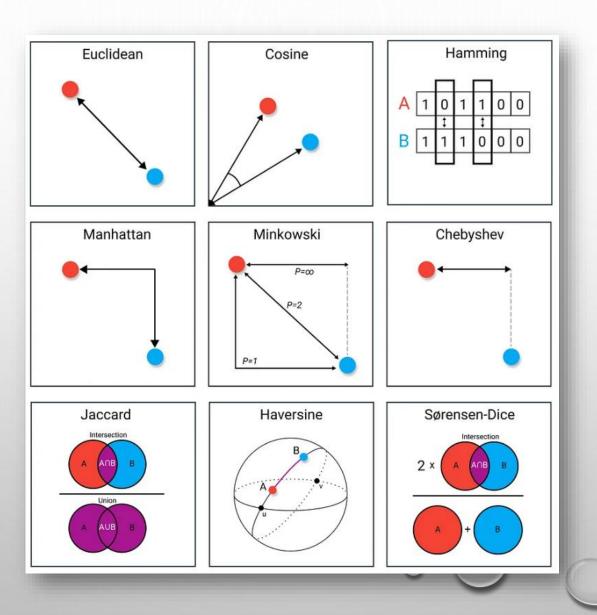
Veja acima as variáveis que foram utilizadas para o componente de receita definido.

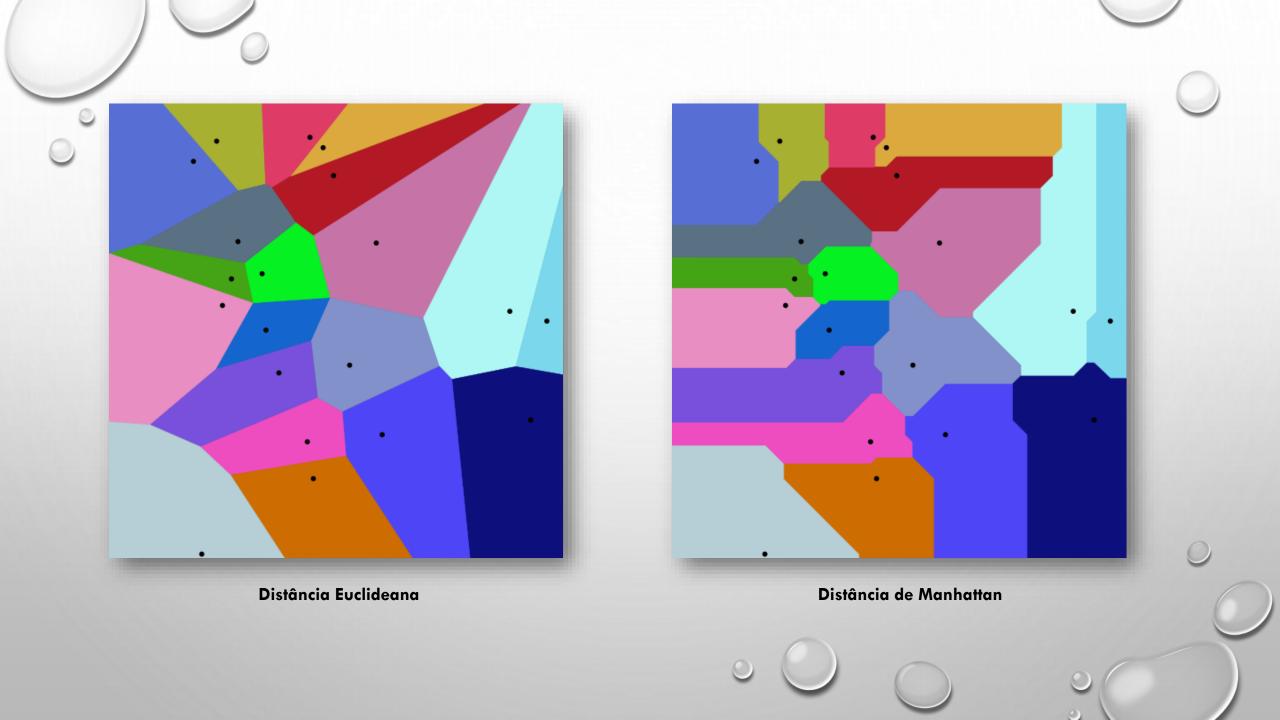
Clique em cada variável acima para entender sua importância.

lique em cada variavel acima para ntender sua importância.

veja acima as variaveis que foram utilizad.

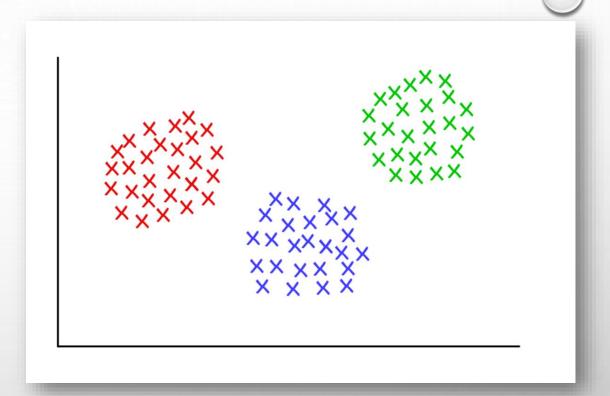
ALGUMAS MÉTRICAS DE DISTÂNCIA





AGRUPAMENTO: ALGORITMOS

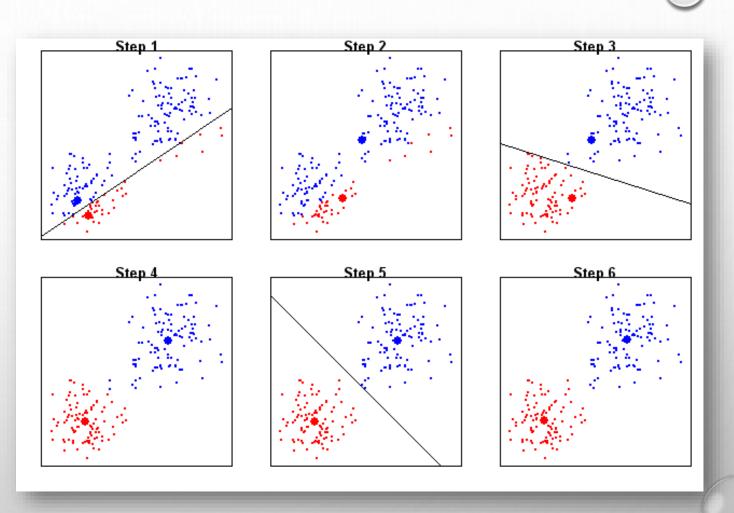
- 1)K-Means
- 2)Hieráquico
- 3)DBSCAN
- 4) Mapa Auto-Organizável



Além da escolha do algoritmo, os resultados do agrupamento dependem diretamente dos atributos e da **métrica escolhida para definir similaridade** entre os objetos.

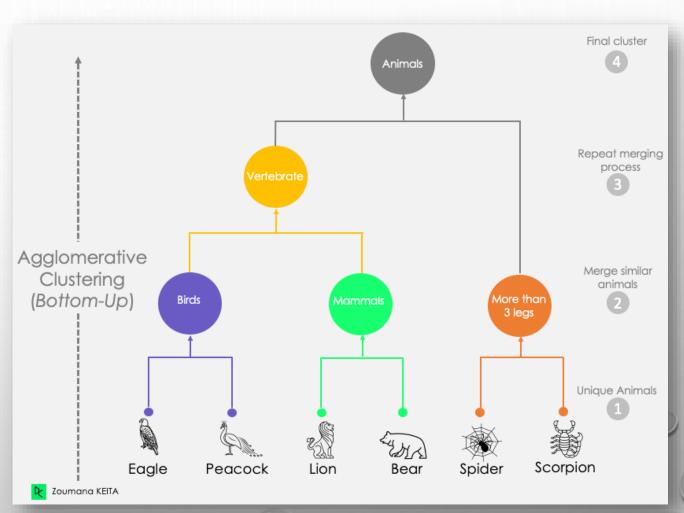
PARTICIONAMENTO: K-MEANS

K-means (ou K-médias) parte de K centroides (centros de agrupamento) e através de iteração, recalcula os centroides até que particione os dados em K grupos.



AGRUPAMENTO HIERÁRQUICO

Esse algoritmo depende de mais uma definição de "Linkage" para decidir como calcular similaridade entre um grupo e um indivíduo, e permite que diferentes arranjos de grupos sejam detectados, variando o limiar de distância.



https://www.datacamp.com/tutorial/introduction-hierarchical-clustering-python

DBSCAN: DENSIDADE

Density-Based Spatial Clustering of Applications with Noise

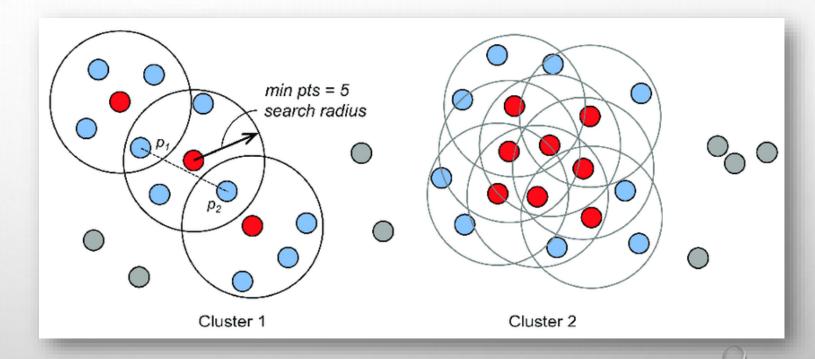
Baseado em uma similaridade

mínima e quantidade de vizinhos

para ser considerado um ponto

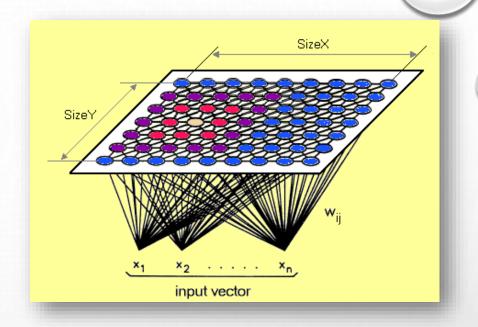
central, DBSCAN agrupa pontos que

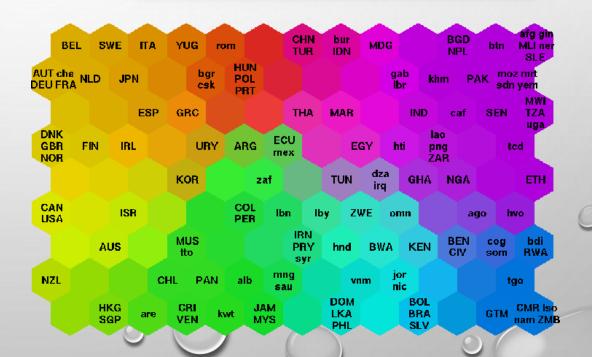
tenham vizinhos comuns.



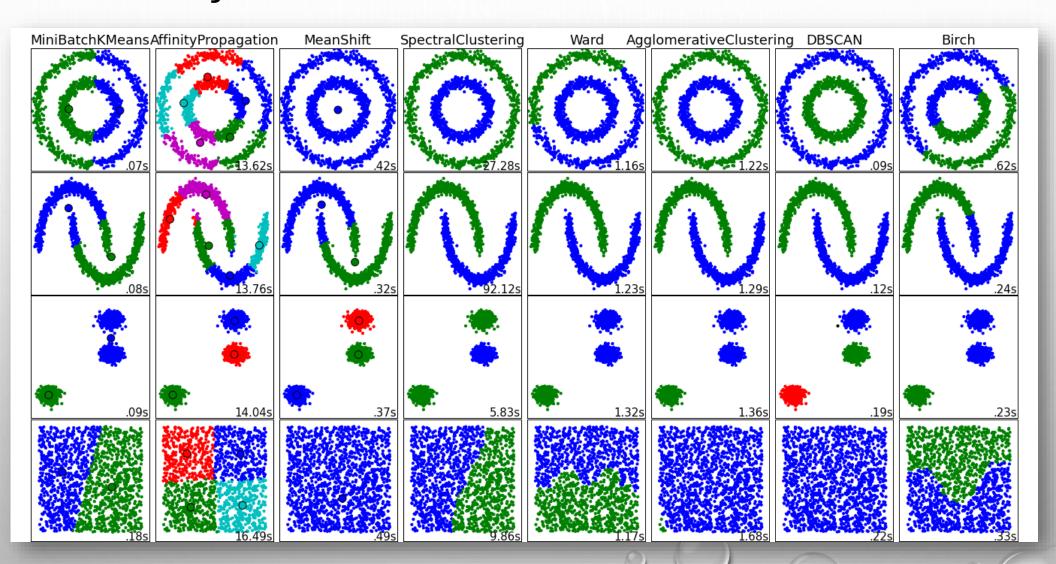
MAPA AUTO-ORGANIZÁVEL

Rede neural que traduz as distâncias no espaço original dos dados para um "espaço de proximidade" entre neurônios da rede, fazendo que dados N-dimensionais seja agrupados num plano bidimensional, de maneira não-linear.





COMPARAÇÃO DE ALGORITMOS: SKLEARN







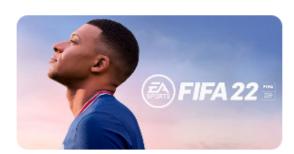
▲ 354

New Notebook



FIFA 22 complete player dataset

19k+ players, 100+ attributes extracted from the latest edition of FIFA



Data Card

Code (36)

Discussion (8)

About Dataset

Context

The datasets provided include the players data for the Career Mode from FIFA 15 to FIFA 22 ("players_22.csv"). The data allows multiple comparisons for the same players across the last 8 version of the videogame.

Usability ①

10.00

License

CC0: Public Domain



PRÓXIMA AULA LEITURA: VISÃO COMPUTACIONAL