



# ***FLG5143 - Bioclimatologia***

## ***Machine Learning***

Fernando Xavier

06/06/2019

# ***Machine Learning***

- Objetivos
  - Introduzir o tema Aprendizado de Máquina
  - Apresentar os principais conceitos relacionados
  - Demonstrar algumas aplicações no campo da Bioclimatologia

# ***Machine Learning***

- Programação
  - Introdução
  - Conceitos Básicos
  - Exemplos de aplicações
  - Referências
  - Links Úteis
  - Dúvidas
  - Material extra para estudo



# ***Machine Learning***

## Introdução

# Introdução

Vamos aquecer!

X	Y
1	2
2	4
3	6
4	8
5	?

# Introdução

X	Y
1	2
2	4
3	6
4	8
5	10

**Qual seria o modelo para achar o valor de Y para um novo valor de X?**

# Introdução

X	Y
1	2
2	4
3	6
4	8
5	10

O modelo para esse conjunto de dados é  $Y = 2 * X$

# ***Introdução***

- Como vocês chegaram ao valor correto do  $Y$  para  $X=5$ ?



# Introdução

- Como vocês chegaram ao valor correto do  $Y$  para  $X=5$ ?
  - Olharam os valores anteriores
  - Descobriram um padrão
  - Aplicaram o padrão aprendido para uma nova entrada

# Introdução

## E agora?

```
[{'a': 651, 'x': 473, 'y': 474, 'z': 622}, {'a': 1090, 'x': 634, 'y': 198, 'z': 69}, {'a': -1064, 'x': 303, 'y': 149, 'z': 907}, {'a': 1732, 'x': 967, 'y': 521, 'z': 399}, {'a': 3454, 'x': 814, 'y': 954, 'z': 111}, {'a': 2482, 'x': 614, 'y': 856, 'z': 350}, {'a': -1060, 'x': 8, 'y': 96, 'z': 678}, {'a': 380, 'x': 926, 'y': 64, 'z': 369}, {'a': 1031, 'x': 460, 'y': 517, 'z': 490}, {'a': -863, 'x': 588, 'y': 169, 'z': 979}, {'a': -1644, 'x': 122, 'y': 36, 'z': 937}, {'a': 773, 'x': 385, 'y': 516, 'z': 580}, {'a': 866, 'x': 584, 'y': 506, 'z': 618}, {'a': 1115, 'x': 748, 'y': 169, 'z': 70}, {'a': 200, 'x': 406, 'y': 346, 'z': 622}, {'a': 767, 'x': 678, 'y': 607, 'z': 866}, {'a': -604, 'x': 447, 'y': 271, 'z': 932}, {'a': 2490, 'x': 266, 'y': 838, 'z': 145}, {'a': 2077, 'x': 709, 'y': 948, 'z': 738}, {'a': 1299, 'x': 409, 'y': 392, 'z': 143}, {'a': 2729, 'x': 416, 'y': 877, 'z': 159}, {'a': 1768, 'x': 719, 'y': 453, 'z': 155}, {'a': 1615, 'x': 64, 'y': 839, 'z': 483}, {'a': -534, 'x': 55, 'y': 211, 'z': 611}, {'a': 577, 'x': 700, 'y': 357, 'z': 597}, {'a': 2221, 'x': 288, 'y': 655, 'z': 16}, {'a': 2086, 'x': 760, 'y': 874, 'z': 648}, {'a': -393, 'x': 616, 'y': 253, 'z': 884}, {'a': 1846, 'x': 37
```

# Introdução

## E agora?

```
[{'a': 651, 'x': 473, 'y': 474, 'z': 622}, {'a': 1090, 'x': 634, 'y': 198, 'z': 69}, {'a': -1064, 'x': 303, 'y': 149, 'z': 907}, {'a': 1732, 'x': 967, 'y': 521, 'z': 399}, {'a': 3454, 'x': 814, 'y': 954, 'z': 111}, {'a': 2482, 'x': 614, 'y': 856, 'z': 350}, {'a': -1060, 'x': 8, 'y': 96, 'z': 678}, {'a': 380, 'x': 926, 'y': 64, 'z': 369}, {'a': 1031, 'x': 460, 'y': 517, 'z': 490}, {'a': -863, 'x': 588, 'y': 169, 'z': 979}, {'a': -1644, 'x': 122, 'y': 36, 'z': 937}, {'a': 773, 'x': 385, 'y': 516, 'z': 580}, {'a': 866, 'x': 584, 'y': 506, 'z': 618}, {'a': 1115, 'x': 748, 'y': 169, 'z': 70}, {'a': 200, 'x': 406, 'y': 346, 'z': 622}, {'a': 767, 'x': 678, 'y': 607, 'z': 866}, {'a': -604, 'x': 447, 'y': 271, 'z': 932}, {'a': 2490, 'x': 266, 'y': 838, 'z': 145}, {'a': 2077, 'x': 709, 'y': 948, 'z': 738}, {'a': 1299, 'x': 409, 'y': 392, 'z': 143}, {'a': 2729, 'x': 416, 'y': 877, 'z': 159}, {'a': 1768, 'x': 719, 'y': 453, 'z': 155}, {'a': 1615, 'x': 64, 'y': 839, 'z': 483}, {'a': -534, 'x': 55, 'y': 211, 'z': 611}, {'a': 577, 'x': 700, 'y': 357, 'z': 597}, {'a': 2221, 'x': 288, 'y': 655, 'z': 16}, {'a': 2086, 'x': 760, 'y': 874, 'z': 648}, {'a': -393, 'x': 616, 'y': 253, 'z': 884}, {'a': 1846, 'x': 37
```

**Agora não é tão simples, correto?**

# Introdução

**E se um computador executasse os mesmos passos que fizemos no primeiro exemplo?**

```
[{'a': 651, 'x': 473, 'y': 474, 'z': 622}, {'a': 1090, 'x': 634, 'y': 198,
'z': 69}, {'a': -1064, 'x': 303, 'y': 149, 'z': 907}, {'a': 1732, 'x': 967,
'y': 521, 'z': 399}, {'a': 3454, 'x': 814, 'y': 954, 'z': 111}, {'a': 2482,
'x': 614, 'y': 856, 'z': 350}, {'a': -1060, 'x': 8, 'y': 96, 'z': 678}, {'a':
380, 'x': 926, 'y': 64, 'z': 369}, {'a': 1031, 'x': 460, 'y': 517, 'z': 490},
{'a': -863, 'x': 588, 'y': 169, 'z': 979}, {'a': -1644, 'x': 122, 'y': 36,
'z': 937}, {'a': 773, 'x': 385, 'y': 516, 'z': 580}, {'a': 866, 'x': 584,
'y': 506, 'z': 618}, {'a': 1115, 'x': 748, 'y': 169, 'z': 70}, {'a': 200,
'x': 406, 'y': 346, 'z': 622}, {'a': 767, 'x': 678, 'y': 607, 'z': 866},
{'a': -604, 'x': 447, 'y': 271, 'z': 932}, {'a': 2490, 'x': 266, 'y': 838,
'z': 145}, {'a': 2077, 'x': 709, 'y': 948, 'z': 738}, {'a': 1299, 'x': 409,
'y': 392, 'z': 143}, {'a': 2729, 'x': 416, 'y': 877, 'z': 159}, {'a': 1768,
'x': 719, 'y': 453, 'z': 155}, {'a': 1615, 'x': 64, 'y': 839, 'z': 483},
{'a': -534, 'x': 55, 'y': 211, 'z': 611}, {'a': 577, 'x': 700, 'y': 357, 'z':
597}, {'a': 2221, 'x': 288, 'y': 655, 'z': 16}, {'a': 2086, 'x': 760, 'y': 87
4, 'z': 648}, {'a': -393, 'x': 616, 'y': 253, 'z': 884}, {'a': 1846, 'x': 37
```

# Introdução

**E se um computador executasse os mesmos passos que fizemos no primeiro exemplo?**

```
[{'a': 651, 'x': 473, 'y': 474, 'z': 622}, {'a': 1090, 'x': 634, 'y': 198, 'z': 69}, {'a': -1064, 'x': 303, 'y': 149, 'z': 907}, {'a': 1732, 'x': 967, 'y': 521, 'z': 399}, {'a': 3454, 'x': 814, 'y': 954, 'z': 111}, {'a': 2482, 'x': 614, 'y': 856, 'z': 350}, {'a': -1060, 'x': 8, 'y': 96, 'z': 678}, {'a': 380, 'x': 926, 'y': 64, 'z': 369}, {'a': 1031, 'x': 460, 'y': 517, 'z': 490}, {'a': -863, 'x': 588, 'y': 169, 'z': 979}, {'a': -1644, 'x': 122, 'y': 36, 'z': 937}, {'a': 773, 'x': 385, 'y': 516, 'z': 580}, {'a': 866, 'x': 584, 'y': 506, 'z': 618}, {'a': 1115, 'x': 748, 'y': 169, 'z': 70}, {'a': 200, 'x': 406, 'y': 346, 'z': 622}, {'a': 767, 'x': 678, 'y': 607, 'z': 866}, {'a': -604, 'x': 447, 'y': 271, 'z': 932}, {'a': 2490, 'x': 266, 'y': 838, 'z': 145}, {'a': 2077, 'x': 709, 'y': 948, 'z': 738}, {'a': 1299, 'x': 409, 'y': 392, 'z': 143}, {'a': 2729, 'x': 416, 'y': 877, 'z': 159}, {'a': 1768, 'x': 719, 'y': 453, 'z': 155}, {'a': 1615, 'x': 64, 'y': 839, 'z': 483}, {'a': -534, 'x': 55, 'y': 211, 'z': 611}, {'a': 577, 'x': 700, 'y': 357, 'z': 597}, {'a': 2221, 'x': 288, 'y': 655, 'z': 16}, {'a': 2086, 'x': 760, 'y': 874, 'z': 648}, {'a': -393, 'x': 616, 'y': 253, 'z': 884}, {'a': 1846, 'x': 37
```

**Nesse caso, a máquina estaria executando uma tarefa de aprendizado!**



# ***Machine Learning***

## Conceitos Básicos

# Conceitos Básicos

- O que é
  - *O aprendizado de máquina é uma aplicação de inteligência artificial (IA) que fornece aos sistemas a capacidade de aprender e melhorar automaticamente a partir da experiência sem serem explicitamente programados. (Fonte: Expert System)*

# Conceitos Básicos

- O que é Machine Learning
  - Termo definido em 1959 por Arthur Samuel
  - Portanto, não é um assunto novo
  - Por que o Gartner afirma: "*machine learning, one of the hottest concepts in technology*" ?



# Conceitos Básicos

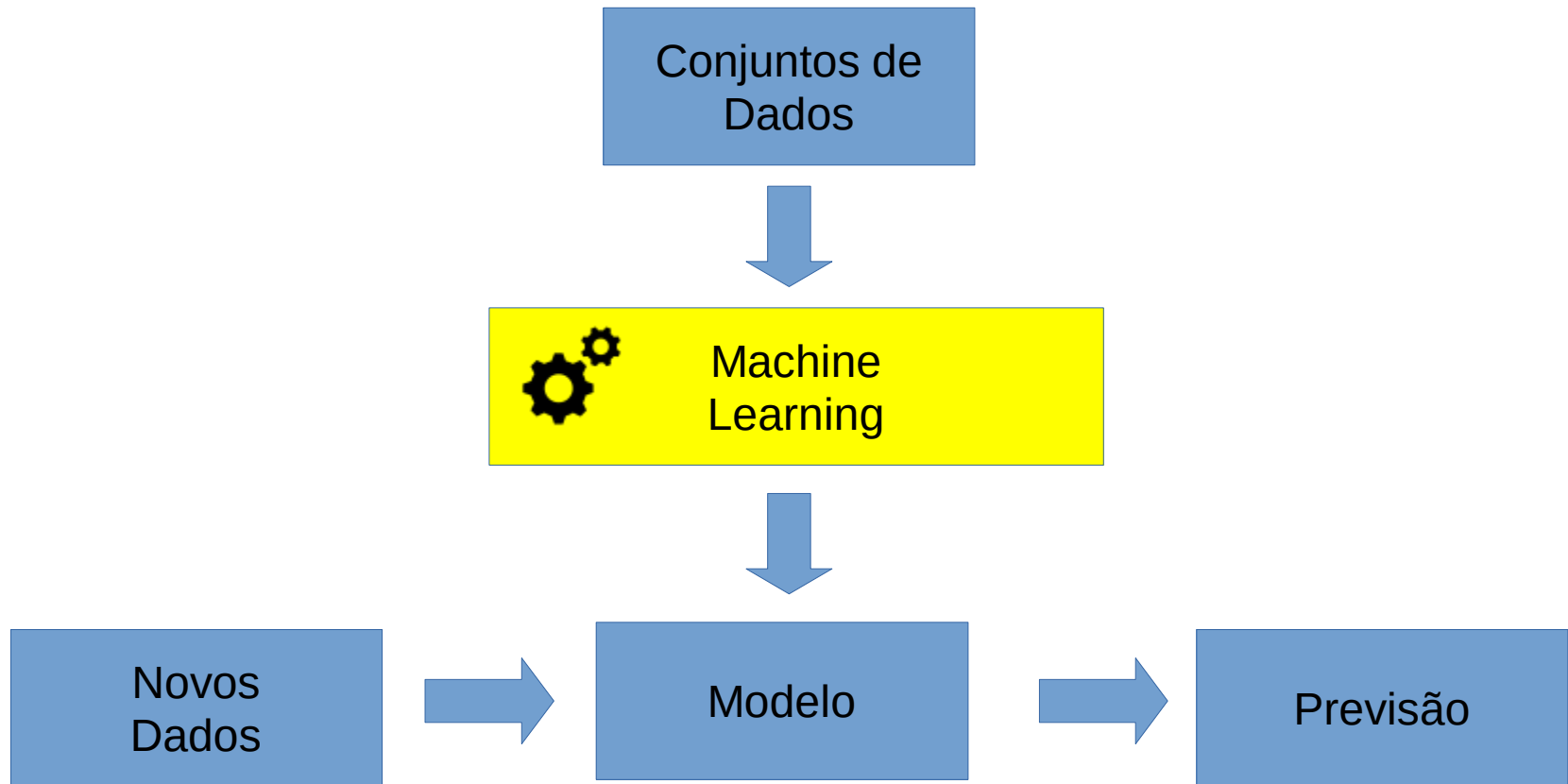
- Algumas razões
  - Desenvolvimento dos carros autônomos que, naturalmente, impulsionam o desenvolvimento dos algoritmos de aprendizado de máquina
  - Cenário de Big Data, com imenso e diverso volume de dados sendo gerado em velocidades cada vez maiores
  - Uso de dispositivos interconectados (IoT, Smart Cities, etc)

# Conceitos Básicos

- No que consiste uma aplicação de aprendizado de máquina?
  - Objetivo definido
  - Fluxo de tarefas
  - Conjuntos de dados
  - Algoritmos

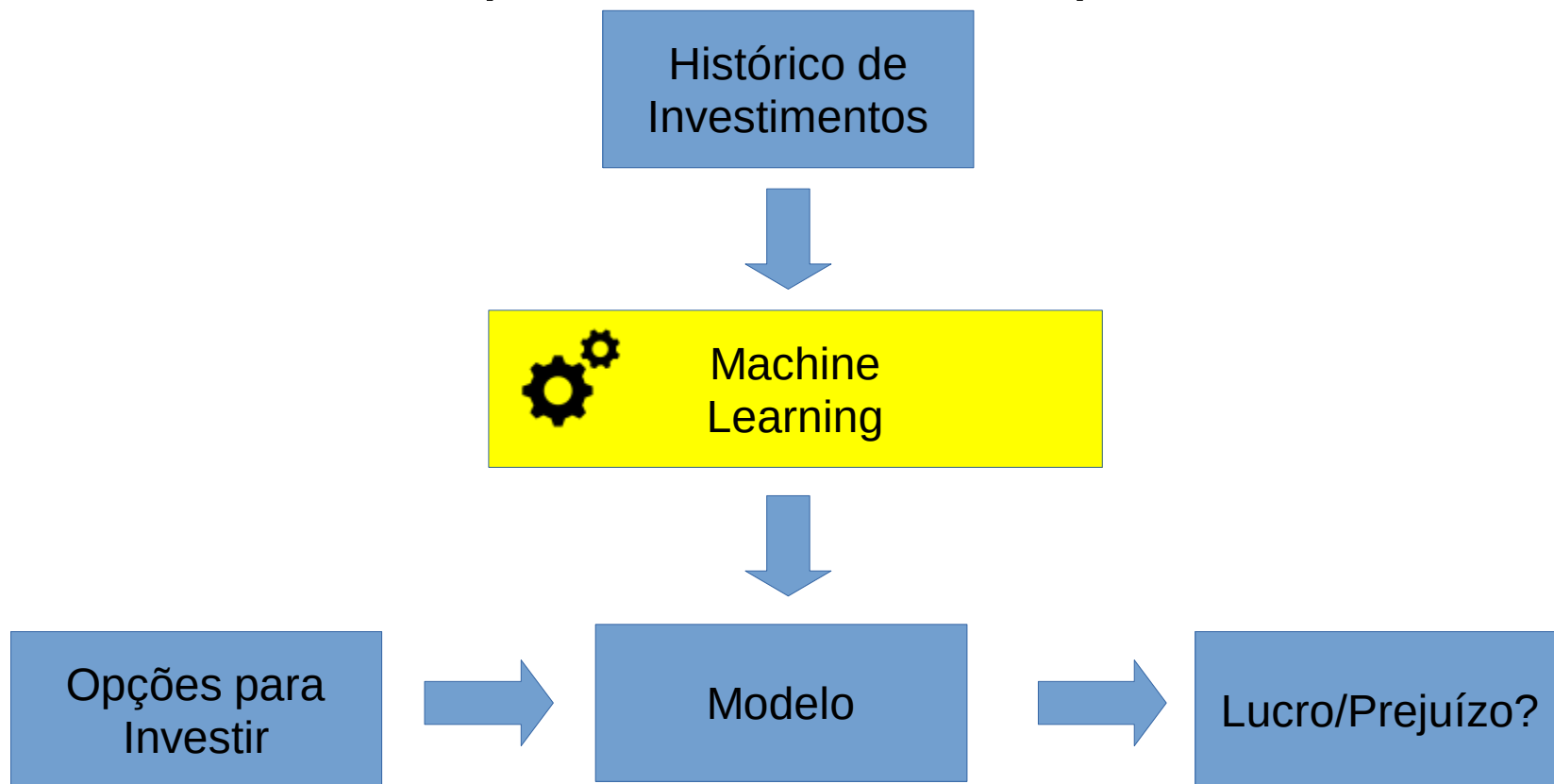
# Conceitos Básicos

- Um fluxo de aprendizado de máquina



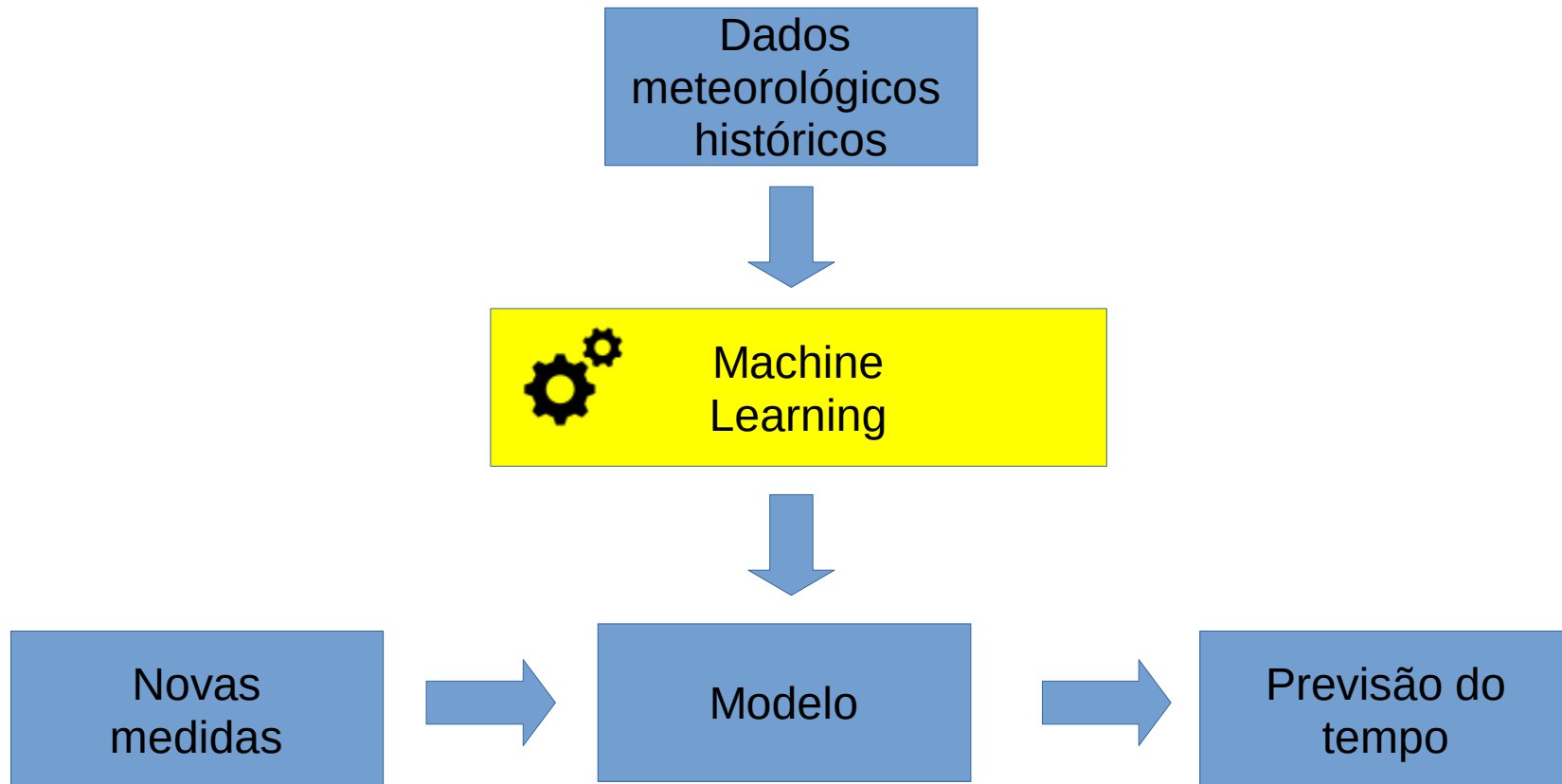
# Conceitos Básicos

- Um fluxo de aprendizado de máquina



# Conceitos Básicos

- Um fluxo de aprendizado de máquina





# ***Conceitos Básicos***

- Para que se usa
  - Reconhecimento de padrões
  - Análise exploratória
  - Geração de modelos preditivos

# Conceitos Básicos

- Reconhecimento de Padrões
  - Qual é o perfil dos clientes que mais compram o produto X?
  - O que está causando o surto de tal doença?
  - Quem são os eleitores do candidato Y?

# Conceitos Básicos

- Análise Exploratória
  - Existe a predominância de um grupo de pessoas entre meus compradores?
  - Quais as características do meu conjunto de dados sobre medições meteorológicas?
  - Não existe obrigatoriedade de perguntas iniciais. Muitas vezes, elas surgem ao explorar os dados



# Conceitos Básicos

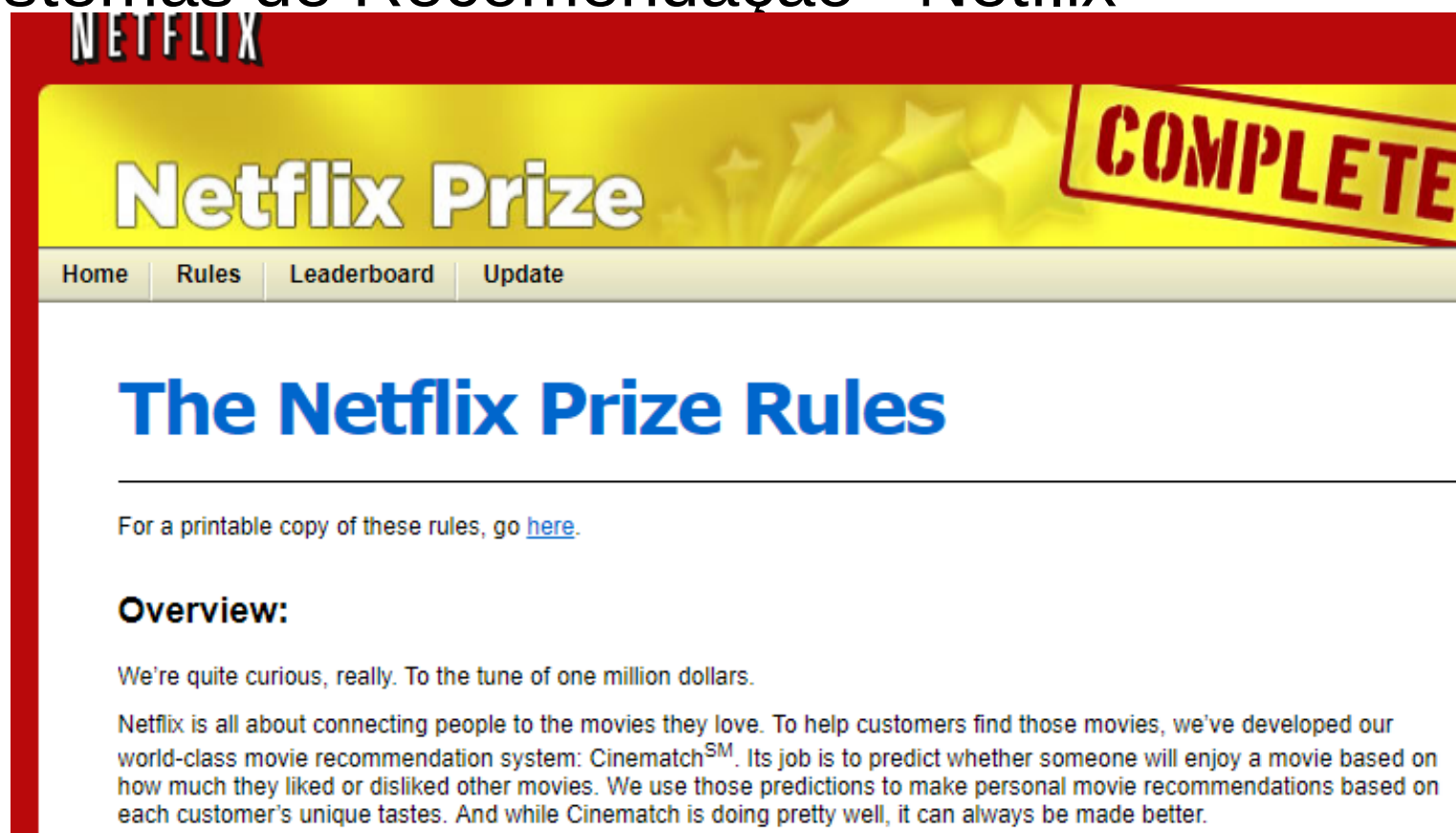
- Geração de modelos preditivos
  - Dado o investimento  $X$  nas ações A, B e C, qual o retorno esperado?
  - Com essa taxa de crescimento da população e o atual padrão de consumo, em quanto tempo faltará alimentos no planeta?
  - Como será a produção do alimento  $X$  caso a temperatura aumente em  $3^{\circ}\text{C}$ ?

# ***Conceitos Básicos***

- Algumas aplicações conhecidas
  - Prevenção de fraudes
  - Processamento de imagens
  - Sistemas de recomendação
  - Carros autônomos
  - Cenários de mudanças climáticas
  - Identificação de spams

# Conceitos Básicos

- Sistemas de Recomendação - Netflix

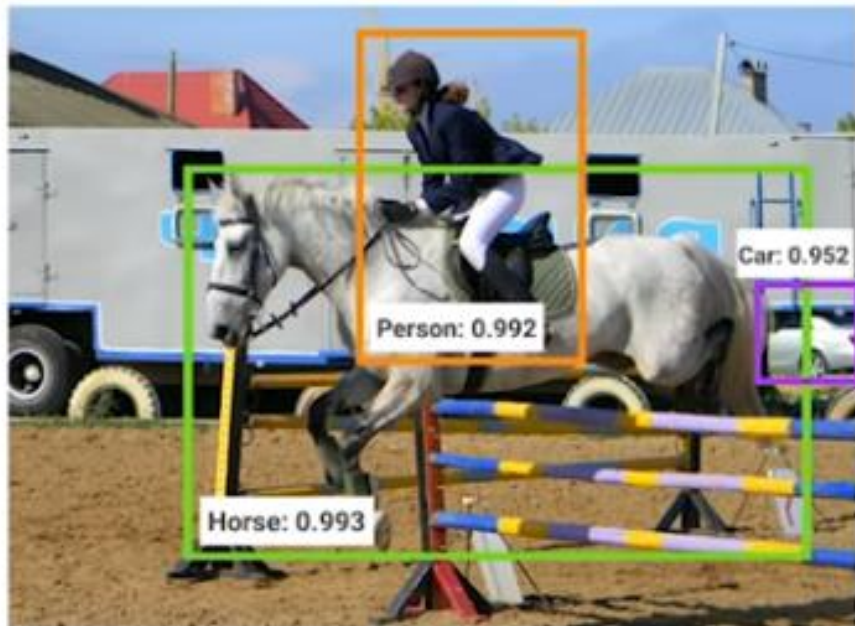


Fonte: <https://www.netflixprize.com/rules.html>

**Machine Learning - Fernando Xavier**

# Conceitos Básicos

- Processamento de imagens



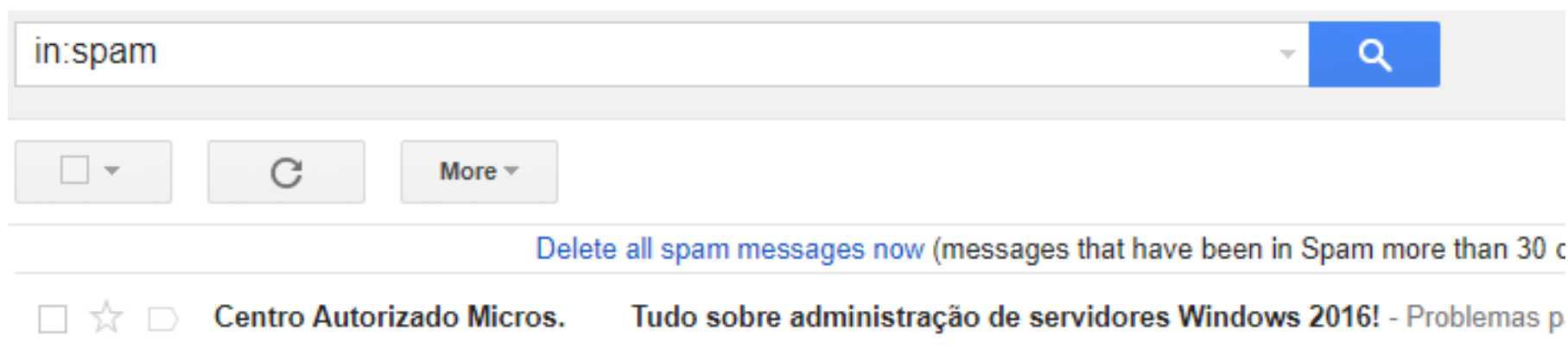
Fonte: <https://developers.google.com/machine-learning/practica/image-classification/>

# ***Conceitos Básicos***

- Carros Autônomos
  - NVIDIA AI Car Demonstration
  - <https://www.youtube.com/watch?v=-96BEoXJMs0>

# Conceitos Básicos

- Classificação de E-mail como Spam
  - Dada uma série de características, um programa de email identifica que aquele email é um spam



# Conceitos Básicos

- Classificação de E-mail como Spam
  - Um algoritmo de ML poderia aprender novas características
  - Ou seja, o algoritmo de classificação é treinado constantemente para melhorar a sua classificação de spam
  - Fazemos a mesma coisa quando queremos treinar um algoritmo para determinada tarefa

# Conceitos Básicos

- Métodos de Aprendizado
  - Supervisionado: Passa-se um conjunto de valores, que são usados para que o algoritmo aprenda
  - Não-Supervisionado: O algoritmo deve buscar os padrões sem o uso de gabaritos
  - Aprendizado por Reforço: descoberta por tentativa/erro dos melhores caminhos para a melhor recompensa
- [https://www.youtube.com/watch?v=L4KBBAwF\\_bE](https://www.youtube.com/watch?v=L4KBBAwF_bE)

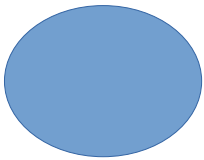


# Conceitos Básicos

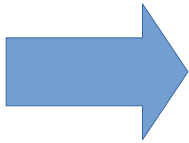
- Método Supervisionado



Quadrado



Elipse



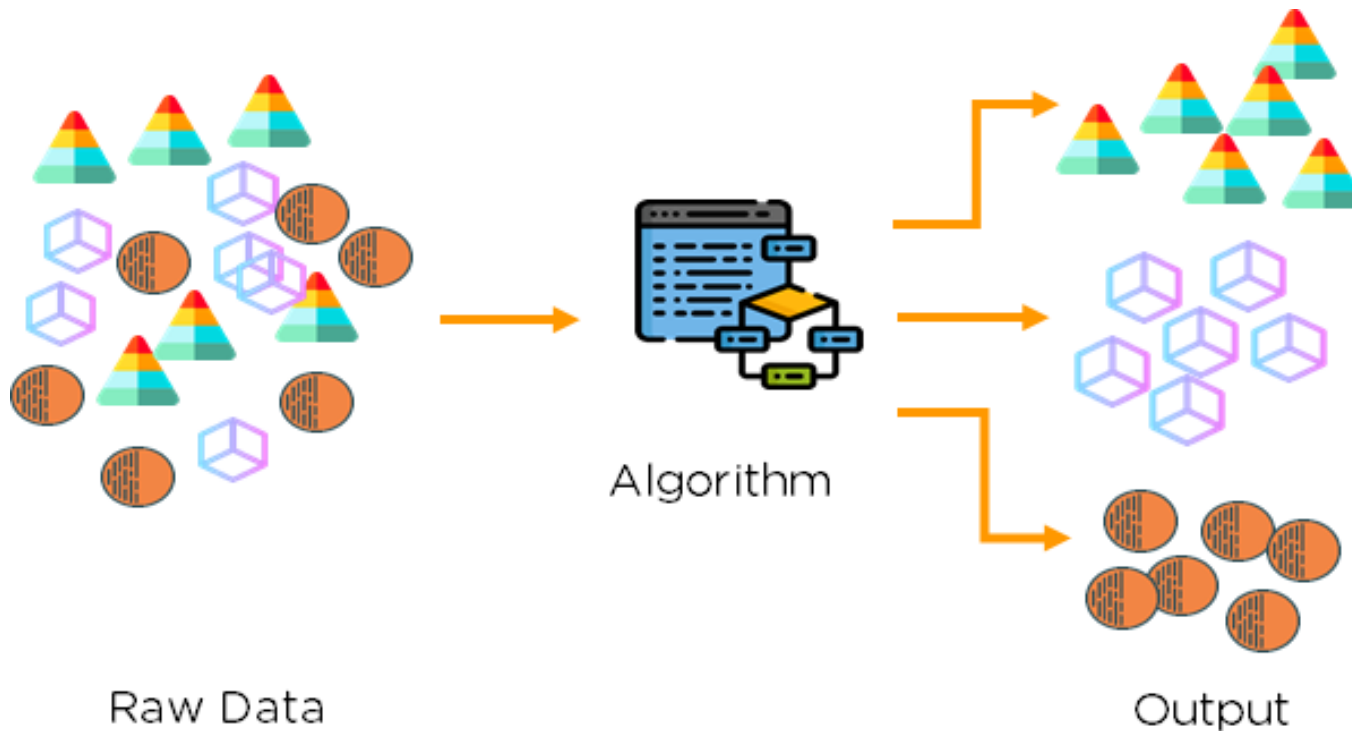
Seta



?

# Conceitos Básicos

- Método Não-Supervisionado



Fonte: <https://www.quora.com/What-is-supervised-and-unsupervised-in-machine-learning>



# ***Conceitos Básicos***

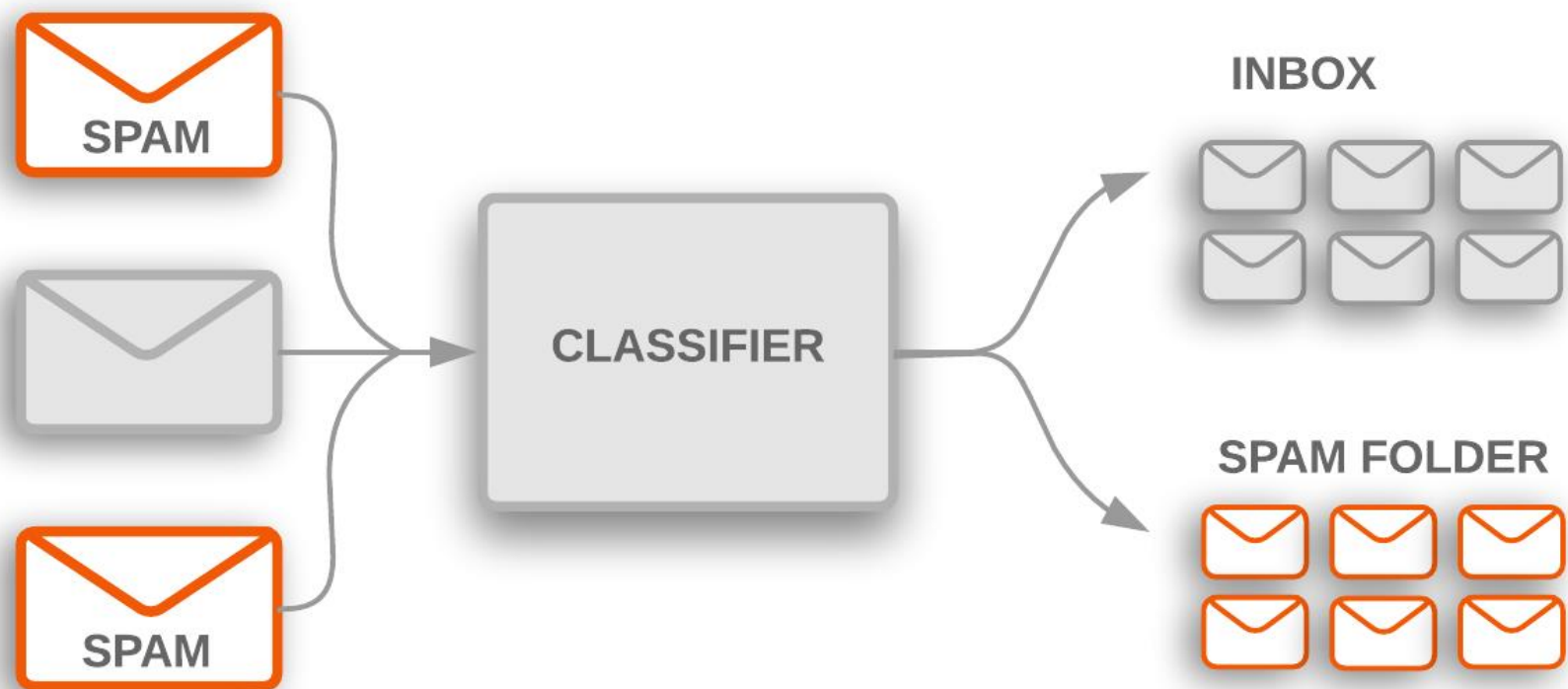
- Alguns Tipos de Algoritmos
  - Classificação
  - Clusterização
  - Regressão

# Conceitos Básicos

- Tipos de Algoritmos - Classificação
  - Uso de rótulos (classes) para os dados
    - Spam/Não-spam
    - Fraude/Não-Fraude
    - Carro/moto/Caminhão
    - Pessoa/objeto

# Conceitos Básicos

- Tipos de Algoritmos - Classificação



Fonte: <https://developers.google.com/machine-learning/guides/text-classification/>

# Conceitos Básicos

- Tipos de Algoritmos - Clusterização
  - Dados são agrupados de acordo com similaridades

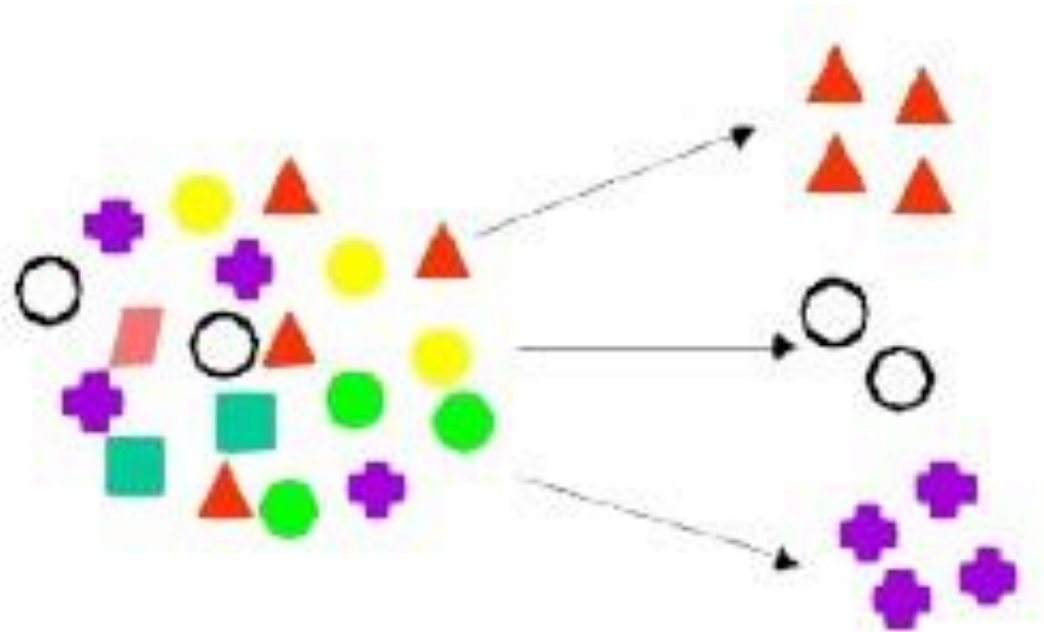


Figura: <https://www.quora.com/What-is-clustering>

# Conceitos Básicos

- Tipos de Algoritmos – Regressão
  - Dados são rotulados com valores
  - Usados para achar um modelo que descreva os valores do conjunto de dados
    - $Y = 2 * X$
    - Do exemplo do início da apresentação:  
 $a = x + 3 * y - 2 * z$

# Conceitos Básicos

- Datasets, instâncias e características
  - Datasets: São conjuntos de dados nos quais os dados podem estar estruturados (em atributos) ou não

```
Estacao;Data;Hora;DirecaoVento;VelocidadeVentoMedia;VelocidadeVentoMaximaMedia;EvaporacaoPiche;EvapoBHPotencial;EvapoBHReal;  
InsolacaoTotal;NebulosidadeMedia;NumDiasPrecipitacao;PrecipitacaoTotal;PressaoNivelMarMedia;PressaoMedia;TempMaximaMedia;  
TempCompensadaMedia;TempMinimaMedia;UmidadeRelativaMedia;VisibilidadeMedia;  
82024;31/01/1996;0000;5;2.569565;5;178.7;;;6.48913;6;57.5;;;32.604348;27.734783;23.608696;84.790323;;  
82024;29/02/1996;0000;5;2.283908;8;149.7;;;146.5;6.91954;8;38.7;;;32.793103;27.943448;23.737931;82.422414;;
```



# Conceitos Básicos

- Datasets, instâncias e características
  - Instância: É cada registro do dataset

A	B	C	D	E	F
rodas	porta	janela	classe	passageiros	instância
4	4	sim	carro	4	1
4	1	sim	carro	1	2
4	1	sim	carro	2	3
4	4	sim	carro	3	4
4	4	sim	carro	4	5
6	3	sim	carro	3	6
4	4	sim	carro	2	7

# Conceitos Básicos

- Datasets, instâncias e características
  - Característica: É cada atributo do registro (rodas, porta, janela, passageiros)

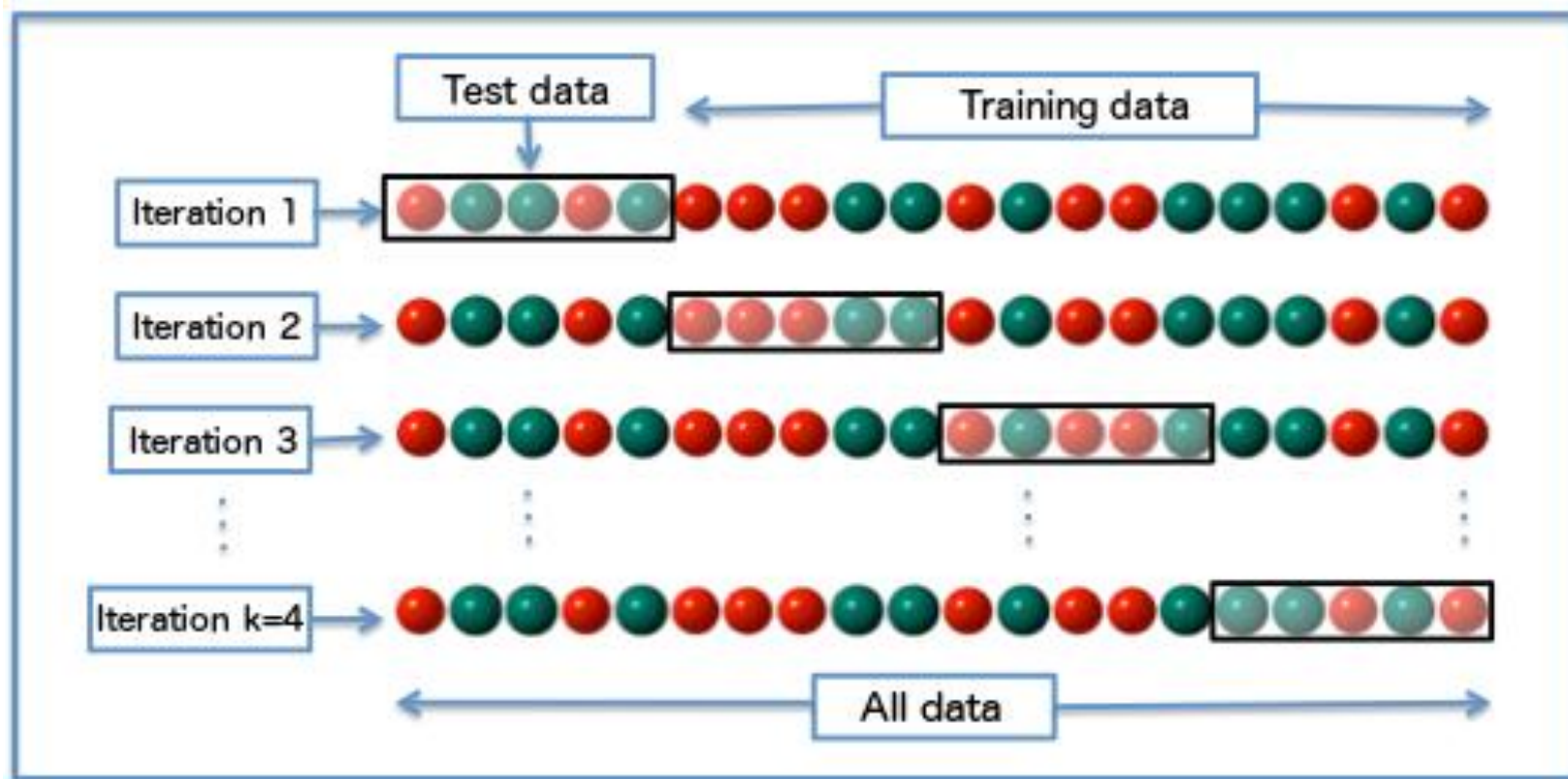
A	B	C	D	E	F
rodas	porta	janela	classe	passageiros	instância
4	4	sim	carro	4	1
4	1	sim	carro	1	2
4	1	sim	carro	2	3
4	4	sim	carro	3	4
4	4	sim	carro	4	5
6	3	sim	carro	3	6
4	4	sim	carro	2	7

# Conceitos Básicos

- Conjuntos de Treino e Teste
  - São conjuntos de dados (datasets) que são divididos em subconjuntos para treinar o algoritmo e outro para testar o algoritmo
  - Existem diversas estratégias para particionamento do dataset em treino/teste
  - Uma das mais famosas é a validação cruzada

# Conceitos Básicos

- Validação Cruzada



Fonte: Genome.tugraz.at. (2006).

# Conceitos Básicos

- Ciclo de Vida de ML (Yufeng G, 2017):
  - Coletar dados
  - Preparar os dados
  - Escolher um algoritmo
  - Treinar
  - Avaliar
  - Ajustar
  - Predizer
- Esses passos variam de autor para autor, mas geralmente tratam da mesma coisa



# ***Machine Learning***

## Exemplos de aplicações

# ***Exemplos de aplicações***

- Exemplos de aplicações na Bioclimatologia
  - Estimativa da evapotranspiração
  - Equação para temperatura do solo
  - Identificação de tipos climáticos

# ***Exemplos de aplicações***

- Exemplo 1: Estimativa da Evapotranspiração
  - Existem diversos métodos para estimativa, como a equação de Penman-Monteith, método de referência da FAO
  - No entanto, esse método requer variáveis que nem sempre podem ser obtidas [Majidi et al 2015]
  - Outros requerem menos variáveis mas podem ter sua eficácia limitada sob certas condições, como o de Thornthwaite, que subestima a evapotranspiração em locais secos [Camargo et al 1999]



# ***Exemplos de aplicações***

- Exemplo 1: Estimativa da Evapotranspiração
  - Outros métodos para estimativa da evapotranspiração são baseados em dados de sensoriamento remoto
  - Esses métodos têm como vantagem a possibilidade de estimativa para grandes áreas
  - No entanto, podem ter baixa precisão dependendo das condições locais [Liou and Kar 2014]

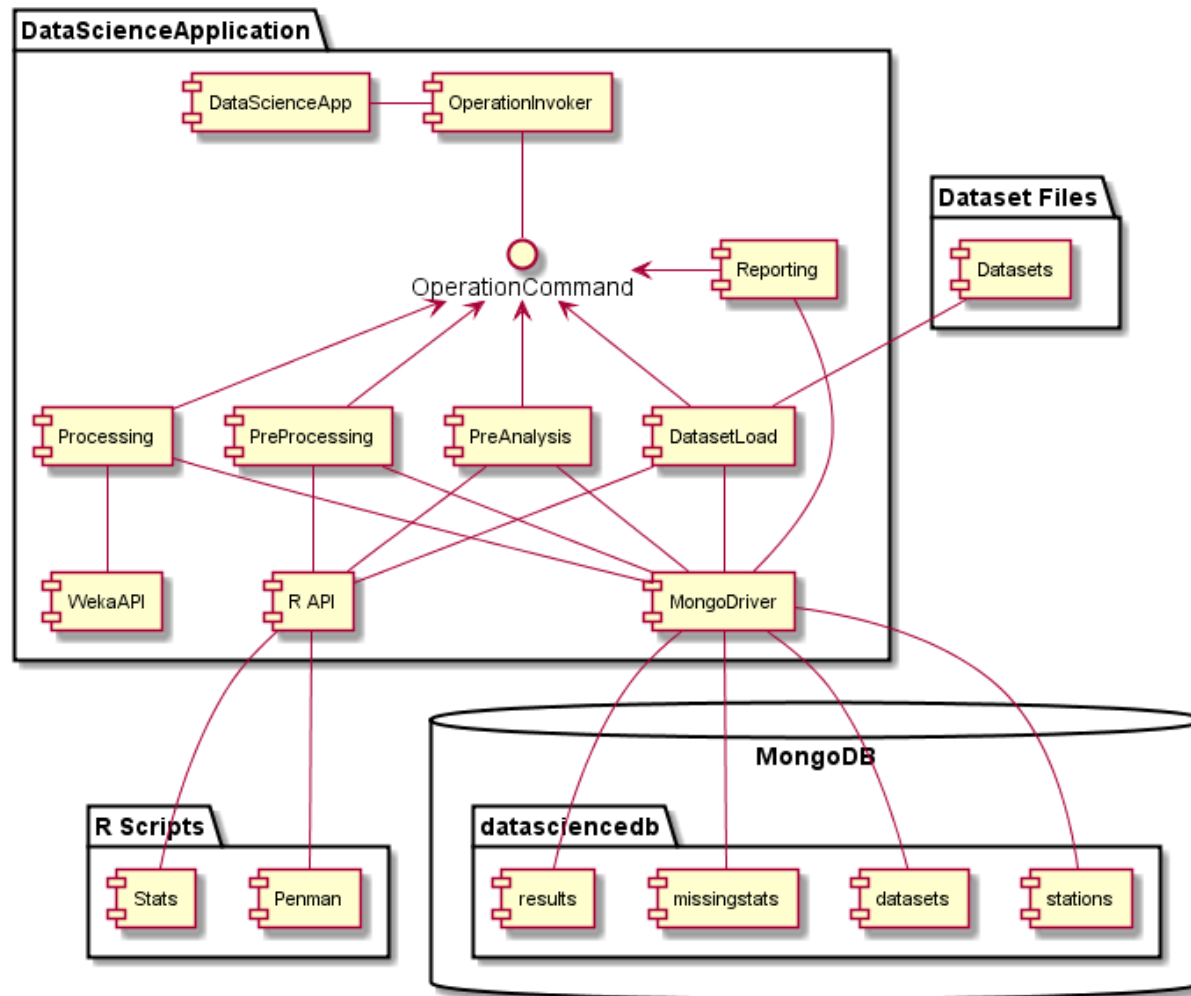
# Exemplos de aplicações

- Exemplo 1: Estimativa da Evapotranspiração
  - A escolha de um método para estimar a evapotranspiração pode estar relacionada a:
    - Disponibilidade dos valores das variáveis
    - Condições locais
    - Área onde será feita a estimativa
    - Precisão
  - Essa escolha representa, muitas vezes, um *trade-off* entre a simplificação e a precisão na estimativa

# ***Exemplos de aplicações***

- Exemplo 1: Estimativa da Evapotranspiração
  - Objetivo: definir um modelo simplificado que fosse específico de cada local
  - Método: uso de aprendizado de máquina (regressão) com a partir de dados meteorológicos históricos
  - Dados: medidas históricas disponibilizadas pelo INMET
  - Ferramentas: software desenvolvido em Java para automatizar o processo para todas as estações do INMET

# Exemplos de aplicações



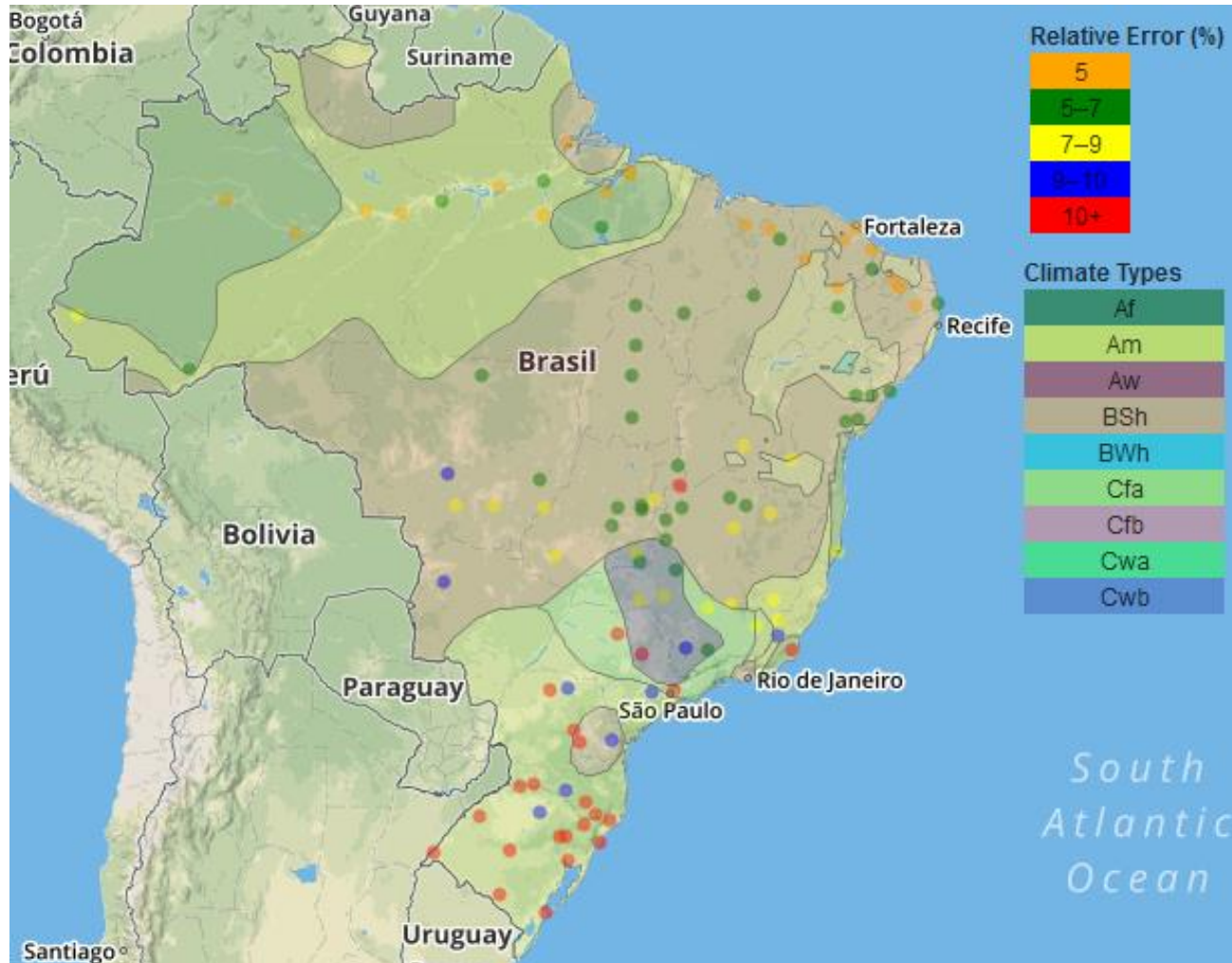
Aplicação desenvolvida (Xavier, 2016)

# Exemplos de aplicações

Station Name	Correla- tion	MAE	RMSE	RMAE	RRMSE	Models
ba_feiradesantana	0.952307	8.629320	10.05246	0.050863	0.059252	LM num: 1 => evp = 19.4011 * WA - 0.0814 * TI - 5.7426 * NM + 0.1108 * TP + 13.4312 * MT - 5.6783 * MiT - 149.643
ba_caravelas	0.926407	9.005550	11.29342	0.062249	0.078063	LM num: 1 => evp = 28.6745 * WA + 5.4723 * NM + 14.6622 * MT - 371.5959
ba_cipo	0.926505	8.361942	11.69584	0.046394	0.064892	LM num: 1 => evp = 26.6686 * WA + 5.5898 * NM + 0.1255 * TP + 17.0235 * MT - 8.1806 * MiT - 314.9227
ba_feiradesantana	0.952307	8.629320	10.05246	0.050863	0.059252	LM num: 1 => evp = 19.4011 * WA - 0.0814 * TI - 5.7426 * NM + 0.1108 * TP + 13.4312 * MT - 5.6783 * MiT - 149.643
ba_ituacu	0.884323	9.862507	12.34818	0.061732	0.077290	LM num: 1 => evp = 23.9634 * WA + 0.1823 * TI + 10.1044 * NM + 0.1548 * TP + 10.3698 * MT - 312.2645
ce_barbalha	0.864496	7.581262	9.486599	0.045526	0.056968	LM num: 1 => evp = 10.0855 * WA + 0.1411 * TI + 4.7999 * NM + 0.0705 * TP + 9.8497 * MT - 234.9181
ce_crateus	0.849791	8.562495	10.46908	0.044794	0.054768	LM num: 1 => evp = 0.0967 * TP + 12.78 * MT - 248.6993
ce_fortaleza	0.778497	5.991667	7.649599	0.035820	0.045732	LM num: 1 => evp = 2.6397 * WA + 0.1754 * TI + 4.9852 * NM + 0.0447 * TP + 12.1733 * MT - 302.9193 LM num: 2 => evp = 3.2996 * WA + 0.035 * TI + 0.0228 * TP + 12.3165 * MT - 237.8026

Alguns resultados

# Exemplos de aplicações



Alguns resultados

# ***Exemplos de aplicações***

- Exemplo 2: Equação para temperatura do solo
  - Objetivo: criar um modelo de cálculo da temperatura do solo a partir de dados meteorológicos
  - Método: uso de aprendizado de máquina (regressão) com os dados históricos
  - Dados: medidas coletadas pela estação do LCB
  - Ferramentas: Weka

# ***Exemplos de aplicações***

- Exemplo 2: Equação para temperatura do solo
  - Descrição: uso de regressão nos dados históricos para geração de um modelo para estimativa da temperatura do solo
  - Divisão dos datasets em:
    - Treino/teste do modelo
    - Validação do modelo



# Exemplos de aplicações

- Exemplo 2: Equação para temperatura do solo

Dataset	Diferença média previsto X real	Equação gerada
Janeiro/2018 (treino/teste)	1,66%	$T_{\text{solo}} = 0.6992 * T_{\text{ar}} + 0.0411 * \text{Umidade} - 0.0541 * \text{Vento} - 0.0229 * \text{Pressao} + 25.93$
Fevereiro/2018 (validação)	1,84%	$T_{\text{solo}} = 0.6992 * T_{\text{ar}} + 0.0411 * \text{Umidade} - 0.0541 * \text{Vento} - 0.0229 * \text{Pressao} + 25.93$
Julho/2018 (treino/teste)	7,46%	$T_{\text{solo}} = 0.626 * T_{\text{ar}} + 0.0406 * \text{Umidade} - 0.0417 * \text{Pressao} + 42.3829$
Agosto/2018 (validação)	6,89%	$T_{\text{solo}} = 0.626 * T_{\text{ar}} + 0.0406 * \text{Umidade} - 0.0417 * \text{Pressao} + 42.3829$

# ***Exemplos de aplicações***

- Exemplo 3: Identificação de tipos climáticos
  - Objetivo: identificar os tipos climáticos das localidades
  - Método: uso de aprendizado de máquina (clusterização) com os dados históricos
  - Dados: normais climatológicas (INMET)
  - Ferramentas: Python e QGis

# ***Exemplos de aplicações***

- Exemplo 3: Identificação de tipos climáticos
  - Descrição:
    - Uso de dados de chuva, umidade e temperatura de diversas localidades de modo a classificá-las em tipos climáticos
    - Cluster (grupos) definidos: 6, 9 e 12

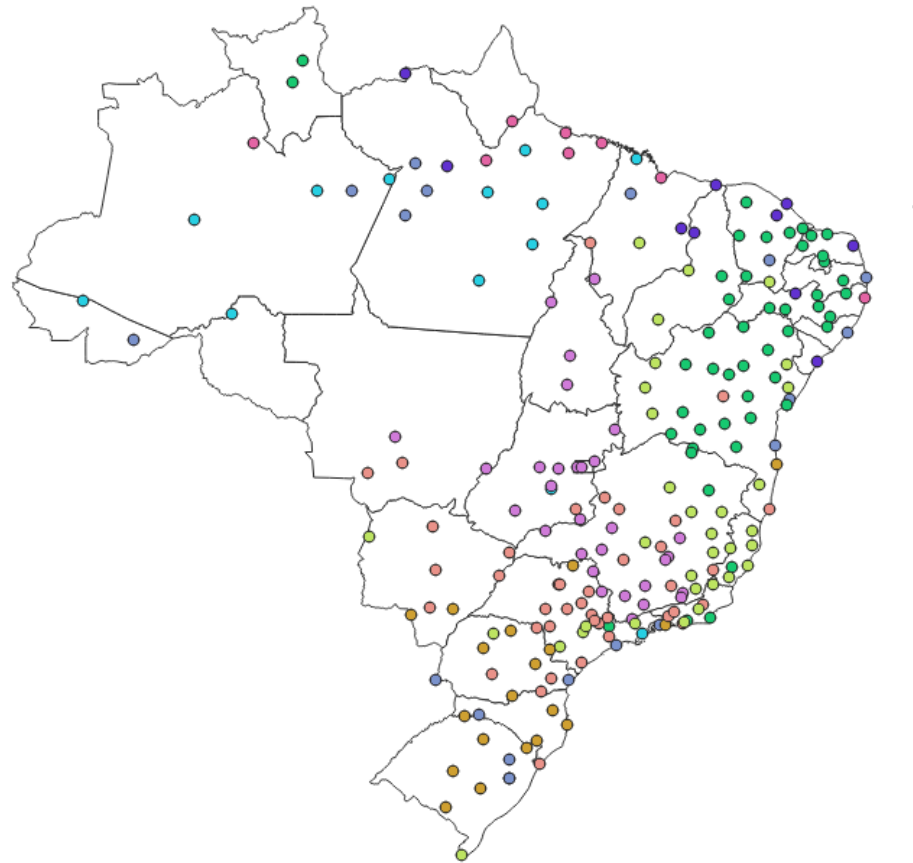
# Exemplos de aplicações

- Exemplo 3: Identificação de tipos climáticos com 6 clusters



# ***Exemplos de aplicações***

- Exemplo 3: Identificação de tipos climáticos com 9 clusters



# ***Exemplos de aplicações***

- Exemplo 3: Identificação de tipos climáticos com 12 clusters



# Links Úteis

- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.
- Alguns algoritmos e seus funcionamentos:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-choice>
- Diversos vídeos no YouTube sempre são úteis



# ***Muito obrigado!***

- Mantenha contato:
  - <https://www.fernandoxavier.com.br>
  - LinkedIn: <https://www.linkedin.com/in/fernando-xavier-3b4466b8/>
  - [https://www.researchgate.net/profile/Fernando\\_Xavier](https://www.researchgate.net/profile/Fernando_Xavier)



# Referências

- Camargo, A.P. de and Marin, F.R. and Sentelhas, P. C. and Picini, A.G., 1999, "Ajuste da equação de Thornthwaite para estimar a evapotranspiração potencial em climas áridos e superúmidos, com base na amplitude térmica diária", Revista Brasileira de Agrometeorologia, v. 7, Issue 2, pp. 251-257
- G, Y. (2017). *The 7 Steps of Machine Learning*. [online] Towards Data Science. Available at: <https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e> [Accessed 27 May 2019].
- Gartner.com. (2018). *5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018*. [online] Available at: <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/> [Accessed 27 May 2019].

# Referências

- Genome.tugraz.at. (2006). *ProClassify User's Guide - Cross-Validation Explained*. [online] Available at: <http://genome.tugraz.at/proclassify/help/pages/XV.html> [Accessed 27 May 2019].
- Liou, Yuei-An and Kar, Sanjib Kumar, 2014, “Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—A review”, *Energies*, v. 7, Issue 5, pp. 2821-2849, Multidisciplinary Digital Publishing Institute.
- Majidi, M. and Alizadeh, A. and Vazifedoust, M. and Farid, A. and Ahmadi, T., 2015, “Analysis of the Effect of Missing Weather Data on Estimating Daily Reference Evapotranspiration Under Different Climatic Conditions”, *Water Resources Management*, v. 29, Issue 7, pp. 2107-2124, Springer Netherlands.

# ***Referências***

- Xavier, F. (2016). Application of Data Science Techniques in Evapotranspiration Estimation. Dissertation (Master in Informatics). Federal University of the State of Rio de Janeiro, p. 95. 2016



# ***Machine Learning***

## Extras

# ***Machine Learning***

- Exercícios
  - Usaremos Python 3 + Jupyter Notebook (procure nos sites oficiais os procedimentos de instalação) e os pacotes scikit-learn e python-twitter
  - Para instalar nas suas máquinas (depois de Python e Jupyter instalados):
    - `pip install python-twitter`
    - `pip install -U scikit-learn`
  - Sugestão: analise as soluções disponibilizadas e crie suas próprias variações

# Machine Learning

- Exercício 1: Dados de Veículos - [GitHub](#)
  - Dataset que contém dados de veículos
  - Objetivo: criar um modelo para identificar se é carro ou moto

A	B	C	D	E
rodas	porta	janela	passageiros	classe
4	4	1	4	carro
4	1	1	1	carro
4	1	1	2	carro
4	4	1	3	carro
4	4	1	4	carro
6	3	1	3	carro
4	4	1	2	carro
4	3	1	3	carro
4	2	1	2	carro
3	4	1	4	carro

# Machine Learning

- Exercício 2: Análise de Sentimento (adaptado de [scikit-learn](#)) - [GitHub](#)
  - Dataset: contém frases e classificação
  - Objetivo: criar um modelo para avaliar posts

```
1 1 Corinthians otimo
2 1 Adorei o jogo
3 1 Estou muito feliziz com esse jogo de hoje
4 1 Estou feliz que ainda somos lideres
5 0 Jogou mal demais
6 0 Time eh muito ruim
7 0 Nao gostei desse jogo
```

# ***Machine Learning***

- Outros exercícios
  - Exercício 1: Regressão Linear: [GitHub](#)
  - Exercício 2: Regressão Linear: [GitHub](#)
  - Exercício 3: Análise de Sentimento: [GitHub](#)
  - Exercício 4: Avaliação de Vinho: [GitHub](#)



# Machine Learning

- Outros exercícios
  - Projetos prontos/adaptados de [http://scikit-learn.org/stable/auto\\_examples](http://scikit-learn.org/stable/auto_examples)
  - Ao carregar, execute para ver os resultados
  - Experimente alterar os valores nos arquivos de dados e veja os resultados
  - Considere que os exemplos podem ser usados para outras diversas aplicações, apenas trocando os conjuntos de dados utilizados

# Machine Learning

- Mais exercícios
  - Reconhecimento de imagens: [http://scikit-learn.org/stable/auto\\_examples/classification/plot\\_digits\\_classification.html#sphx-glr-auto-examples-classification-plot-digits-classification-py](http://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html#sphx-glr-auto-examples-classification-plot-digits-classification-py)
  - Predição usando validação cruzada: [http://scikit-learn.org/stable/auto\\_examples/plot\\_cv\\_predict.html#sphx-glr-auto-examples-plot-cv-predict-py](http://scikit-learn.org/stable/auto_examples/plot_cv_predict.html#sphx-glr-auto-examples-plot-cv-predict-py)
  - Escolha outros em: [http://scikit-learn.org/stable/auto\\_examples](http://scikit-learn.org/stable/auto_examples)

# ***Machine Learning***

- Exemplos para testar
  - Python: [http://scikit-learn.org/stable/auto\\_examples](http://scikit-learn.org/stable/auto_examples)
  - Python-Jupyter: <https://resources.oreilly.com/examples/9781783555130/>
  - Datasets: <https://www.kaggle.com/datasets>

# ***Machine Learning***

- Algumas ferramentas/bibliotecas para praticar
  - Weka: <https://www.cs.waikato.ac.nz/ml/weka/>
  - Jupyter: <https://try.jupyter.org/>
  - Scikit-learn: <http://scikit-learn.org/stable/>
  - Apache Mahout: <http://mahout.apache.org/>
  - Pandas: <https://pandas.pydata.org/>

# ***Machine Learning***

- Plataformas para Machine Learning
  - IBM Watson: <https://www.ibm.com/watson/br-pt/>
  - Google: <https://cloud.google.com/ml-engine/>
  - Amazon: <https://aws.amazon.com/pt/amazon-ai/>
  - Microsoft: <https://azure.microsoft.com/pt-br/services/machine-learning-studio/>

# ***Machine Learning***

- Conceitos Relacionados à Aprendizado de Máquina
  - Redes Neurais
  - Mineração de Dados
  - Inteligência Artificial
  - Deep Learning
  - Tensorflow
  - (pesquise e divirta-se!)