



<UDFTECHDAY//>

Introdução a Machine Learning

Prof. Me. Fernando Xavier

02/10/2018

Introdução a Machine Learning

- Objetivos
 - Conhecer os principais conceitos relacionados ao Aprendizado de Máquina
 - Conhecer algumas das tecnologias existentes
 - Exercitar a aplicação de aprendizado de máquina em casos reais

Introdução a Machine Learning

- Programação
 - Conceitos Básicos
 - Estudos de Casos
 - Exercícios

Introdução a Machine Learning

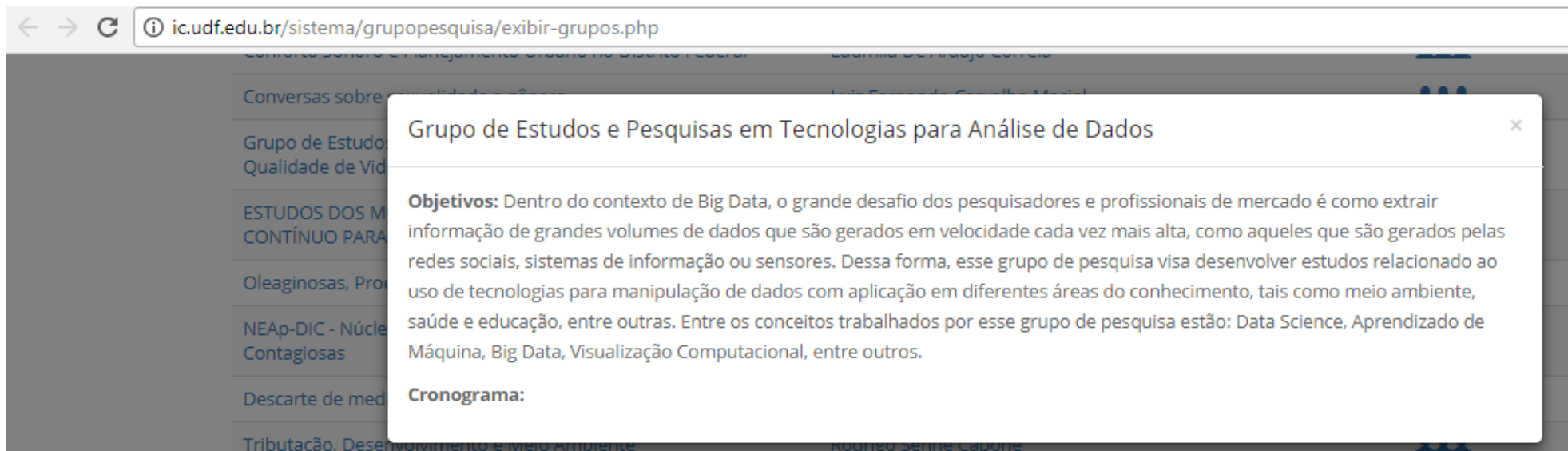
- O que usaremos
 - Materiais de referência
 - Python/Jupyter/Weka
 - Alguns datasets

Introdução a Machine Learning

- Sobre mim
 - Professor e pesquisador
 - Mestre pela UNIRIO (2016) com aplicação de *Data Science* na geração de modelos na área de Recursos Hídricos
 - Doutorando da Escola Politécnica - USP
 - Um eterno estudante

Introdução a Machine Learning

- Nosso grupo de pesquisa



The screenshot shows a web browser window with the URL ic.udf.edu.br/sistema/grupospesquisa/exibir-grupos.php. A modal window is open over the page content, displaying the following information:

Grupo de Estudos e Pesquisas em Tecnologias para Análise de Dados

Objetivos: Dentro do contexto de Big Data, o grande desafio dos pesquisadores e profissionais de mercado é como extrair informação de grandes volumes de dados que são gerados em velocidade cada vez mais alta, como aqueles que são gerados pelas redes sociais, sistemas de informação ou sensores. Dessa forma, esse grupo de pesquisa visa desenvolver estudos relacionado ao uso de tecnologias para manipulação de dados com aplicação em diferentes áreas do conhecimento, tais como meio ambiente, saúde e educação, entre outras. Entre os conceitos trabalhados por esse grupo de pesquisa estão: Data Science, Aprendizado de Máquina, Big Data, Visualização Computacional, entre outros.

Cronograma:

Introdução a Machine Learning

Vamos aquecer!

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

Introdução a Machine Learning

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

Qual seria o modelo usado para qualquer entrada nova de X?

Introdução a Machine Learning

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

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

Introdução a Machine Learning

Qual seria o modelo usado para qualquer entrada nova de X?

{x:159,y:515}{x:112,y:829}{x:647,y:527}{x:626,y:514}{x:871,y:803}{x:650,y:638}{x:52}{x:298,y:937}{x:198,y:800}{x:845,y:138}{x:896,y:833}{x:378,y:987}{x:225,y:319}{y:505}{x:902,y:889}{x:208,y:122}{x:516,y:774}{x:582,y:878}{x:811,y:444}{x:945,y:55,y:478}{x:914,y:156}{x:878,y:642}{x:494,y:946}{x:142,y:260}{x:562,y:522}{x:681}{x:336,y:920}{x:492,y:174}{x:714,y:548}{x:140,y:272}{x:861,y:760}{x:704,y:918}{x:87}{x:460,y:913}{x:509,y:409}{x:928,y:490}{x:573,y:268}{x:921,y:990}{x:556,y:732}{y:570}{x:589,y:879}{x:360,y:204}{x:346,y:626}{x:535,y:192}{x:101,y:963}{x:412,y:93,y:232}{x:522,y:465}

Introdução a Machine Learning

- Como vocês chegaram ao valor correto do Y para $X=2$?

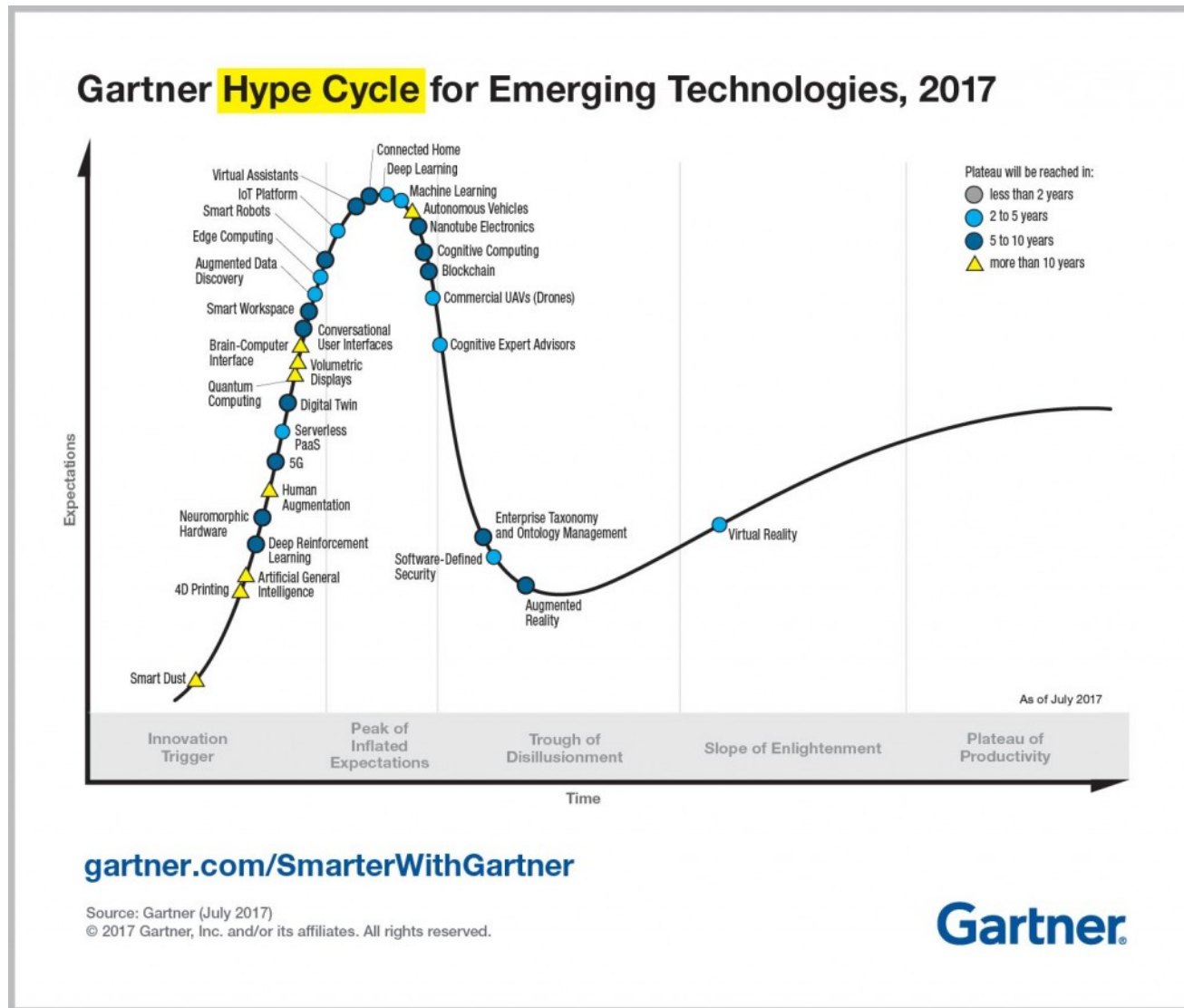
Introdução a Machine Learning

- Como vocês chegaram ao valor correto do Y para $X=2$?
 - Olharam os valores anteriores
 - Descobriram um padrão
 - Aplicaram esse padrão para uma nova entrada
 - Resumidamente, essa é a base de aprendizado de máquina!

Introdução a Machine Learning

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

Introdução a Machine Learning



Introdução a Machine Learning

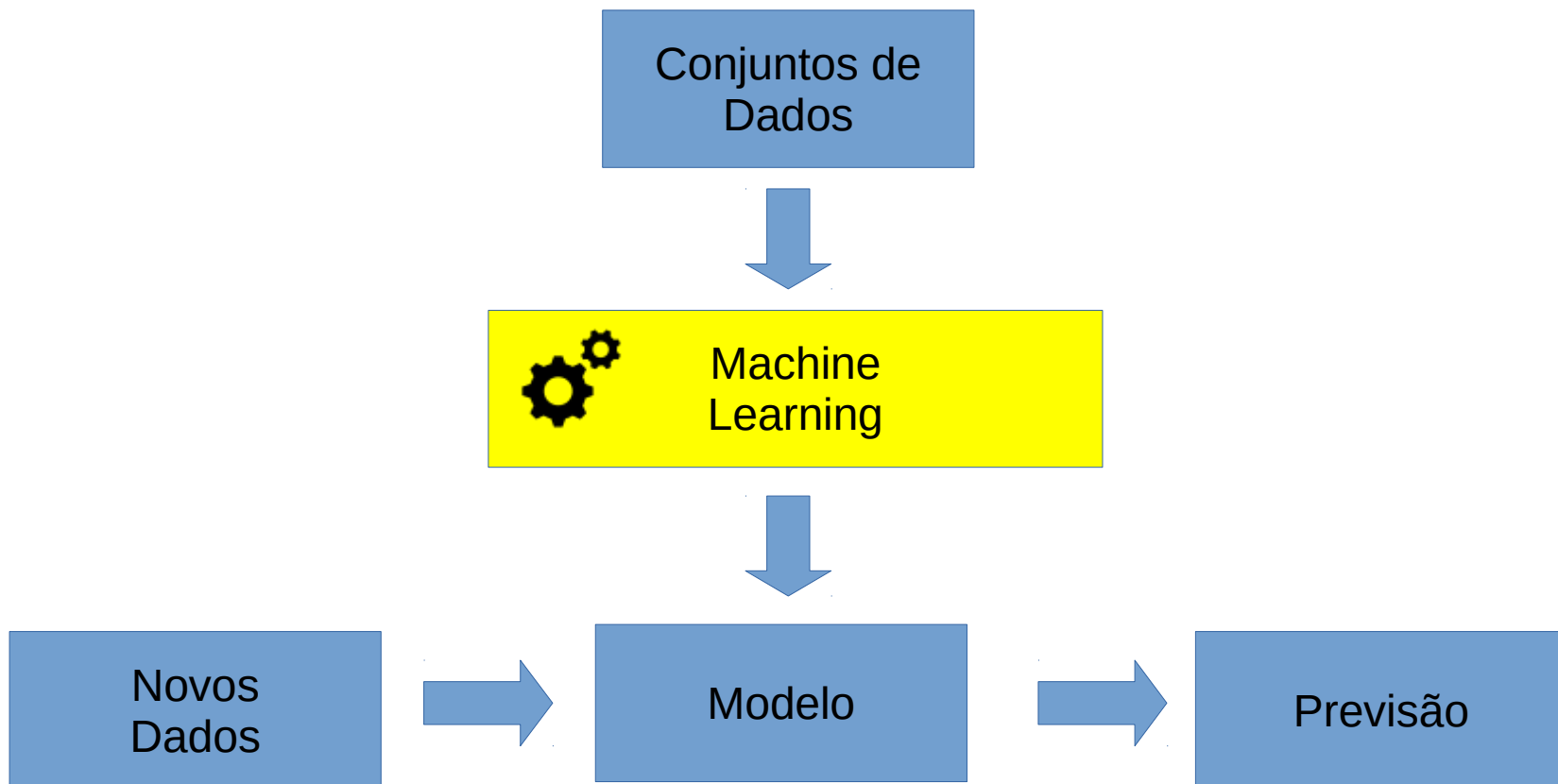
- O que é Machine Learning
 - Termo definido em 1959 por Arthur Samuel
 - Não é, portanto, um assunto novo
 - Por que o Gartner afirma: "*machine learning, one of the hottest concepts in technology*" ?

Introdução a Machine Learning

- Algumas possíveis 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 e inteligentes (IoT, Smart Cities, etc)

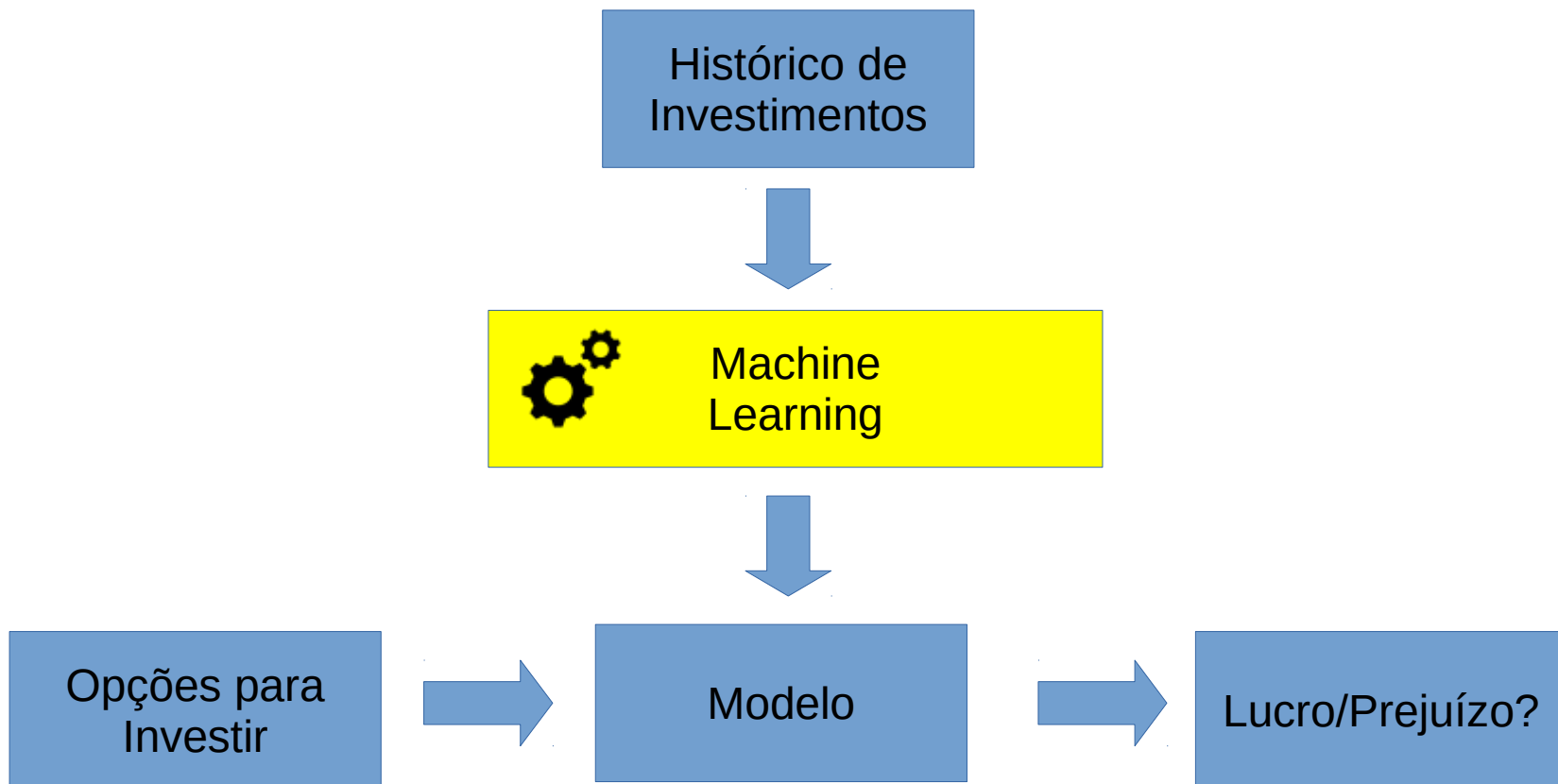
Introdução a Machine Learning

- Um fluxo de aprendizado de máquina



Introdução a Machine Learning

- Um fluxo de aprendizado de máquina



Introdução a Machine Learning

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

Introdução a Machine Learning

- 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?

Introdução a Machine Learning

- 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

Introdução a Machine Learning

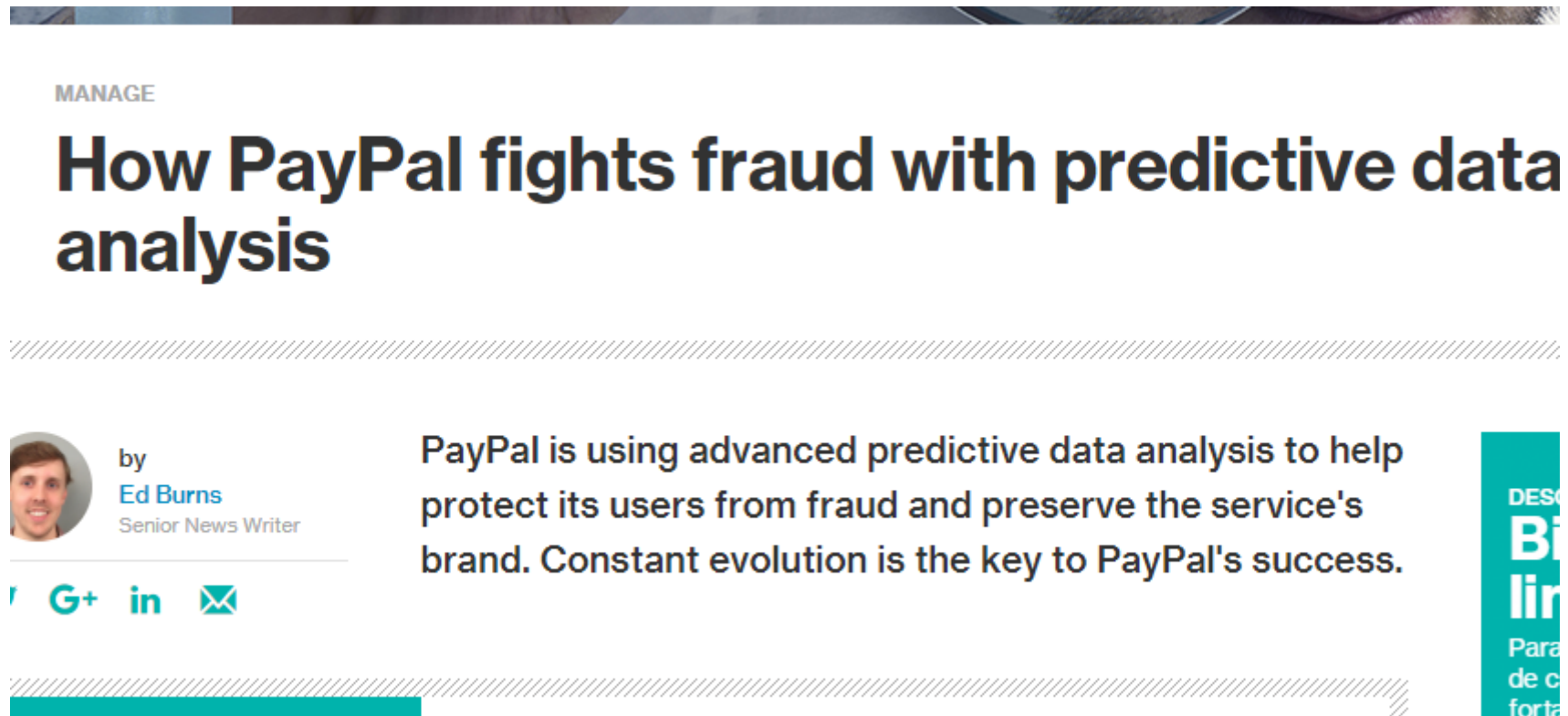
- Geração de modelos preditivos
 - Dado o investimento X nas ações A, B e C, qual o retorno esperado?
 - Com essa taxa de crescimento da população e o atual padrão de consumo, em quanto tempo faltará alimentos no planeta?

Introdução a Machine Learning

- Algumas aplicações conhecidas
 - Prevenção de fraudes
 - Sistemas de recomendação
 - Cenários de mudanças climáticas
 - Identificação de spams

Introdução a Machine Learning

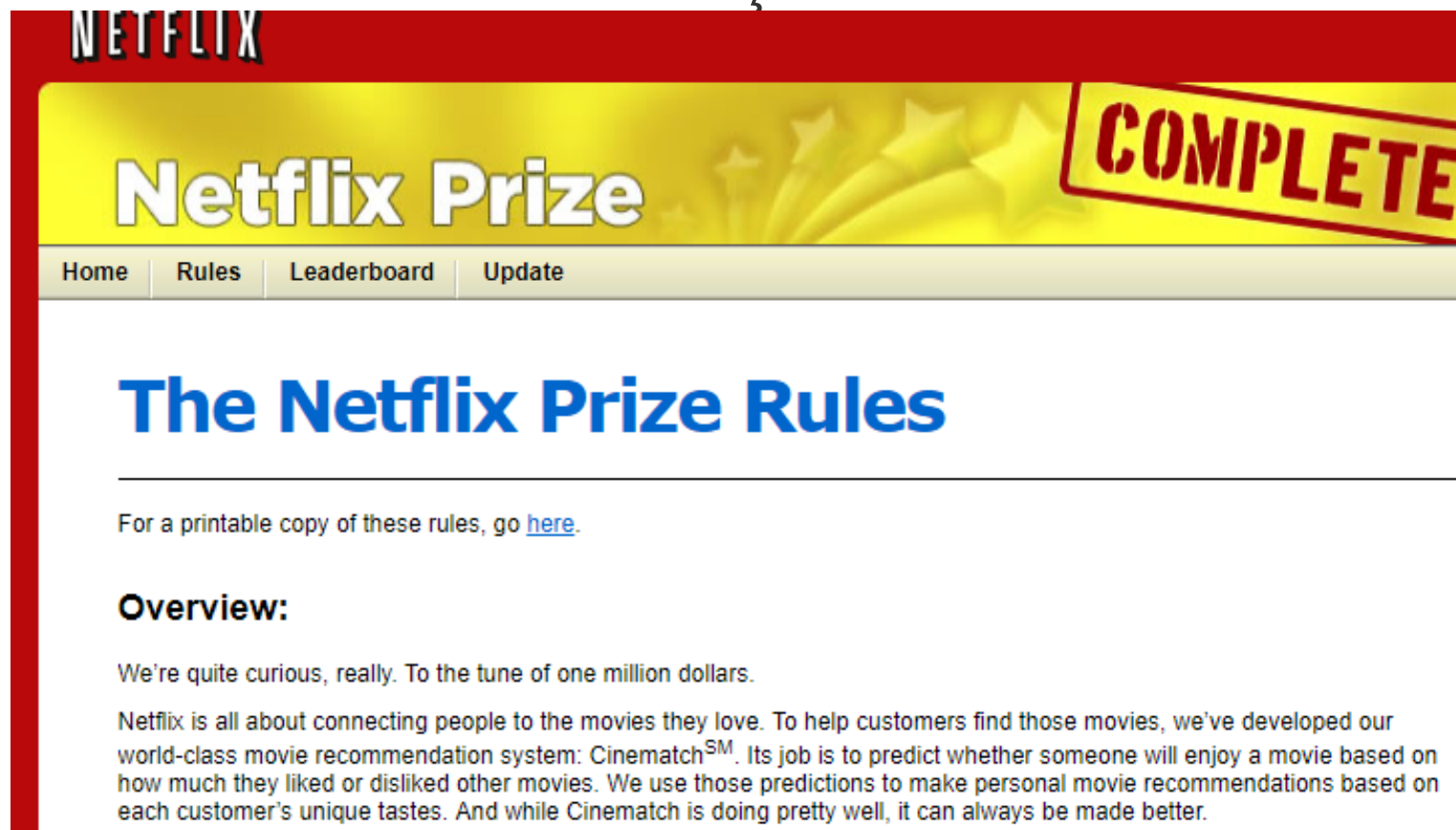
- Prevenção de Fraudes



Fonte: <https://searchbusinessanalytics.techtarget.com/feature/How-PayPal-fights-fraud-with-predictive-data-analysis>

Introdução a Machine Learning

- Sistemas de Recomendação - Netflix

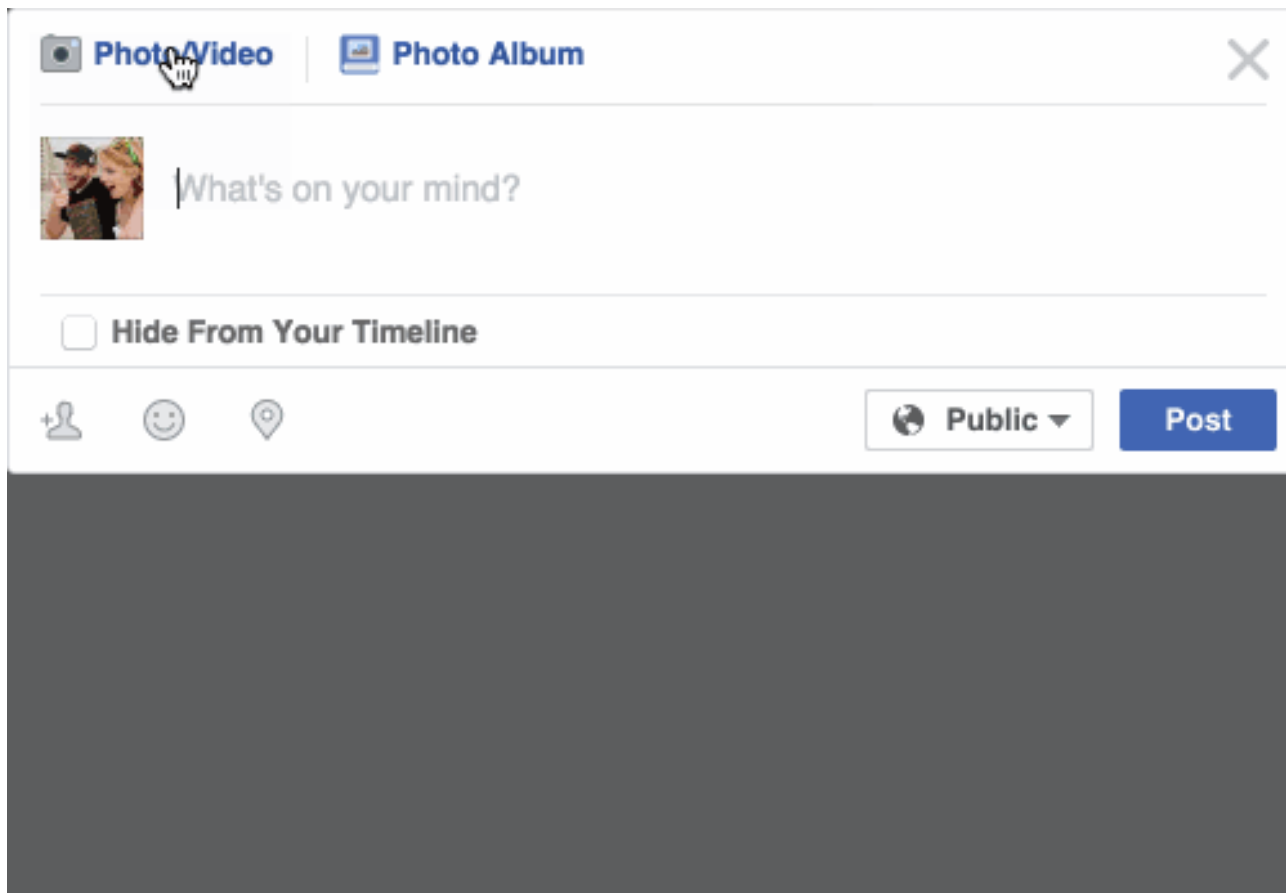


The screenshot shows the Netflix Prize website. At the top, there is a red banner with the Netflix logo and a yellow banner with the text "Netfix Prize" and a "COMPLETED" stamp. Below the banner is a navigation menu with links for "Home", "Rules", "Leaderboard", and "Update". The main heading is "The Netflix Prize Rules". Below the heading, there is a link to a printable copy of the rules. The "Overview:" section begins with the text: "We're quite curious, really. To the tune of one million dollars. Netflix is all about connecting people to the movies they love. To help customers find those movies, we've developed our world-class movie recommendation system: CinematchSM. Its job is to predict whether someone will enjoy a movie based on how much they liked or disliked other movies. We use those predictions to make personal movie recommendations based on each customer's unique tastes. And while Cinematch is doing pretty well, it can always be made better."

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

Introdução a Machine Learning

- Reconhecimento de Face

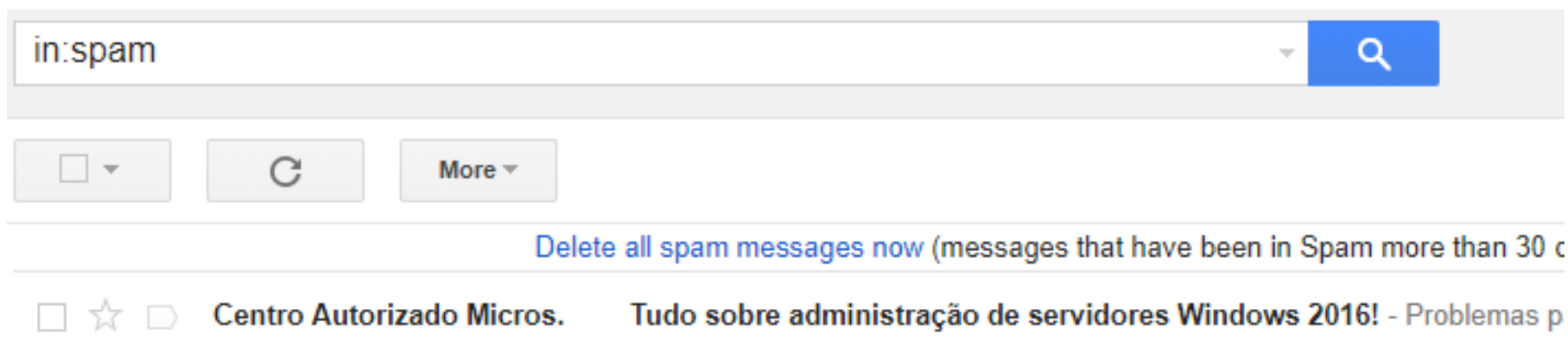


Introdução a Machine Learning

- Carros Autônomos
 - https://www.youtube.com/watch?v=mCj_C1NOVxw

Introdução a Machine Learning

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



Introdução a Machine Learning

- Classificação de E-mail como Spam
 - Algumas características pode ser usadas, como:
 - Servidor de envio
 - Falta de link para descadastrar
 - Formato do e-mail
 - Notificações de outros usuários
 - ...

Introdução a Machine Learning

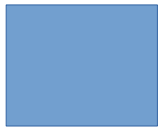
- Classificação de E-mail como Spam
 - Um algoritmo de ML poderia aprender novas características
 - Ou seja, o algoritmo de classificação é treinado constantemente para melhorar a sua classificação de spam
 - Fazemos a mesma coisa quando queremos treinar um algoritmo para determinada tarefa

Introdução a Machine Learning

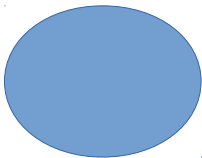
- Métodos de Aprendizado
 - Supervisionado: Passa-se um conjunto de valores, que são usados para que o algoritmo aprenda
 - Não-Supervisionado: O algoritmo deve buscar os padrões sem o uso de gabaritos
 - Aprendizado por Reforço: descoberta por tentativa/erro dos melhores caminhos para a melhor recompensa

Introdução a Machine Learning

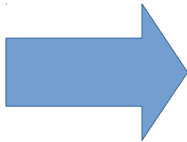
- Método Supervisionado



Quadrado



Elipse



Seta



?

Introdução a Machine Learning

- Método Não-Supervisionado



Introdução a Machine Learning

- Alguns Tipos de Algoritmos
 - Classificação
 - Clusterização
 - Regressão

Introdução a Machine Learning

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

Introdução a Machine Learning

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

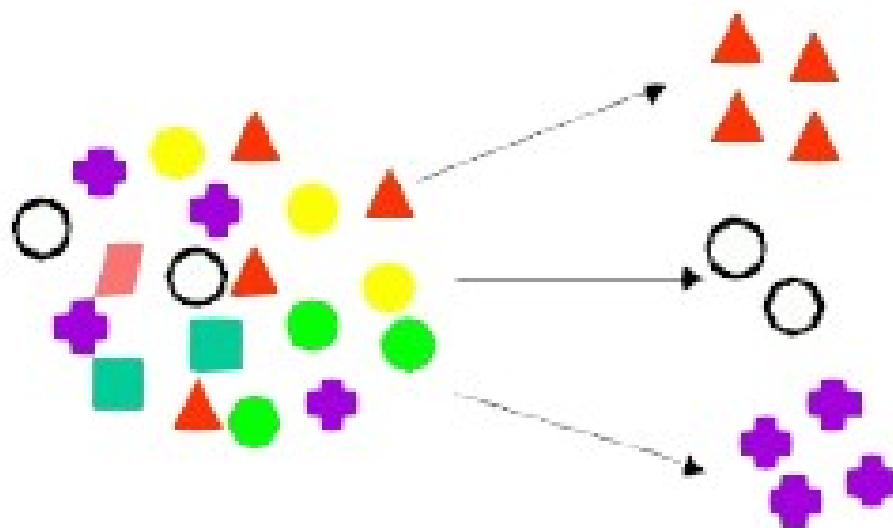


Figura: <https://www.quora.com/What-is-clustering>

Introdução a Machine Learning

- Tipos de Algoritmos – Regressão
 - Dados são rotulados com valores
 - Usados para achar um modelo que descreva os valores do conjunto de dados
 - $Y = 2 * X$

Introdução a Machine Learning

- Conceitos Básicos: Datasets
 - São conjuntos de dados nos quais os dados podem estar estruturados (em atributos) ou não estar estruturados

```
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;;
```

Introdução a Machine Learning

- Conceitos Básicos: Instância
 - É cada registro do dataset

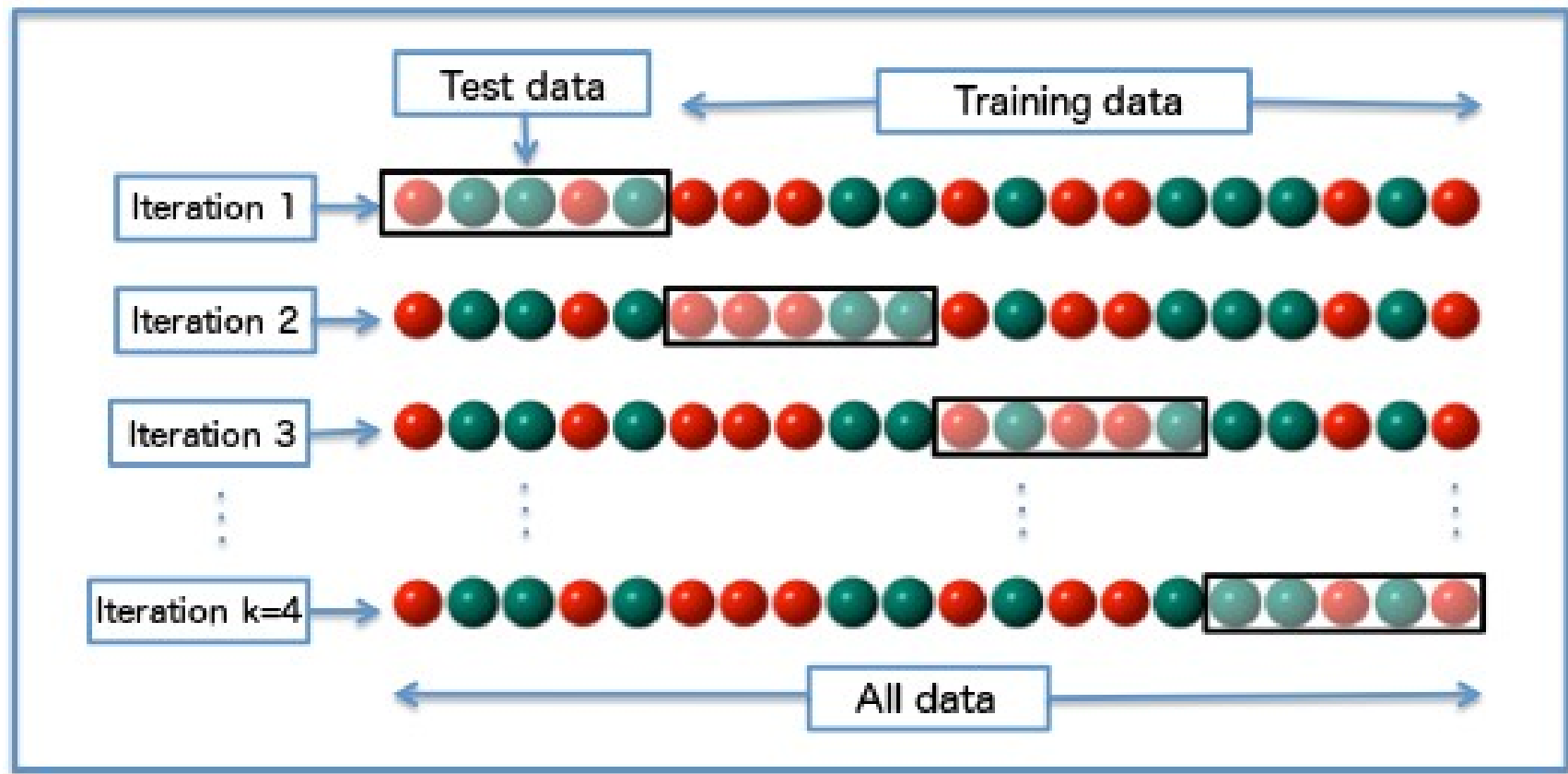
A	B	C	D	E	F
rodas	porta	janela	classe	passageiros	instância
4	4	sim	carro	4	1
4	1	sim	carro	1	2
4	1	sim	carro	2	3
4	4	sim	carro	3	4
4	4	sim	carro	4	5
6	3	sim	carro	3	6
4	4	sim	carro	2	7

Introdução a Machine Learning

- 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

Introdução a Machine Learning

- Conceitos Básicos: Validação Cruzada



Introdução a Machine Learning

- Etapas para aplicação
 - Sugeridas por Yufeng G
 - Coletar dados
 - Preparar os dados
 - Escolher um modelo
 - Treinar
 - Avaliar
 - Ajustar
 - Predizer
 - Obs: esses passos variam de autor para autor, mas geralmente tratam da mesma coisa

Introdução a Machine Learning

- Exercícios
 - Usaremos Python 3 + Jupyter Notebook (se quiser instalar no seu computador, procure nos sites oficiais)
 - Usaremos os pacotes scikit-learn e python-twitter (se o proxy deixar)
 - Para instalar nas suas máquinas (depois de Python e Jupyter instalados):
 - `pip install python-twitter`
 - `pip install -U scikit-learn`

Introdução a Machine Learning

- Exercício 1: Dados de Veículos
 - 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	

Introdução a Machine Learning

- Exercício 2: Análise de Sentimento (adaptado de [scikit-learn](#))
 - 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 feeliz 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
```

Introdução a Machine Learning

- Outros exercícios
 - Projetos prontos/adaptados de 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

Introdução a 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](#)

Introdução a 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
 - Predição usando validação cruzada:
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

Introdução a 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/>

Introdução a Machine Learning

- Exemplos para testar
 - Python:
http://scikit-learn.org/stable/auto_examples
 - Python-Jupyter:
<https://resources.oreilly.com/examples/9781783555130/>
 - Datasets: <https://www.kaggle.com/datasets>

Introdução a Machine Learning

- Plataformas 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/>

Introdução a Machine Learning

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

Introdução a Machine Learning

- Algumas Referências
 - 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

Introdução a Machine Learning

Muito obrigado!!

- Mantenha contato:
 - <https://www.fernandoxavier.com.br>
 - LinkedIn:
<https://www.linkedin.com/in/fernando-xavier-3b4466b8/>
 - Orkut: Error 404!