



FIAP

Statistics for Machine Learning

Aula 15: Introdução Análise de Texto e Imagens

Prof. Jones Egydio

profjones.egydio@fiap.com.br



Objetivos

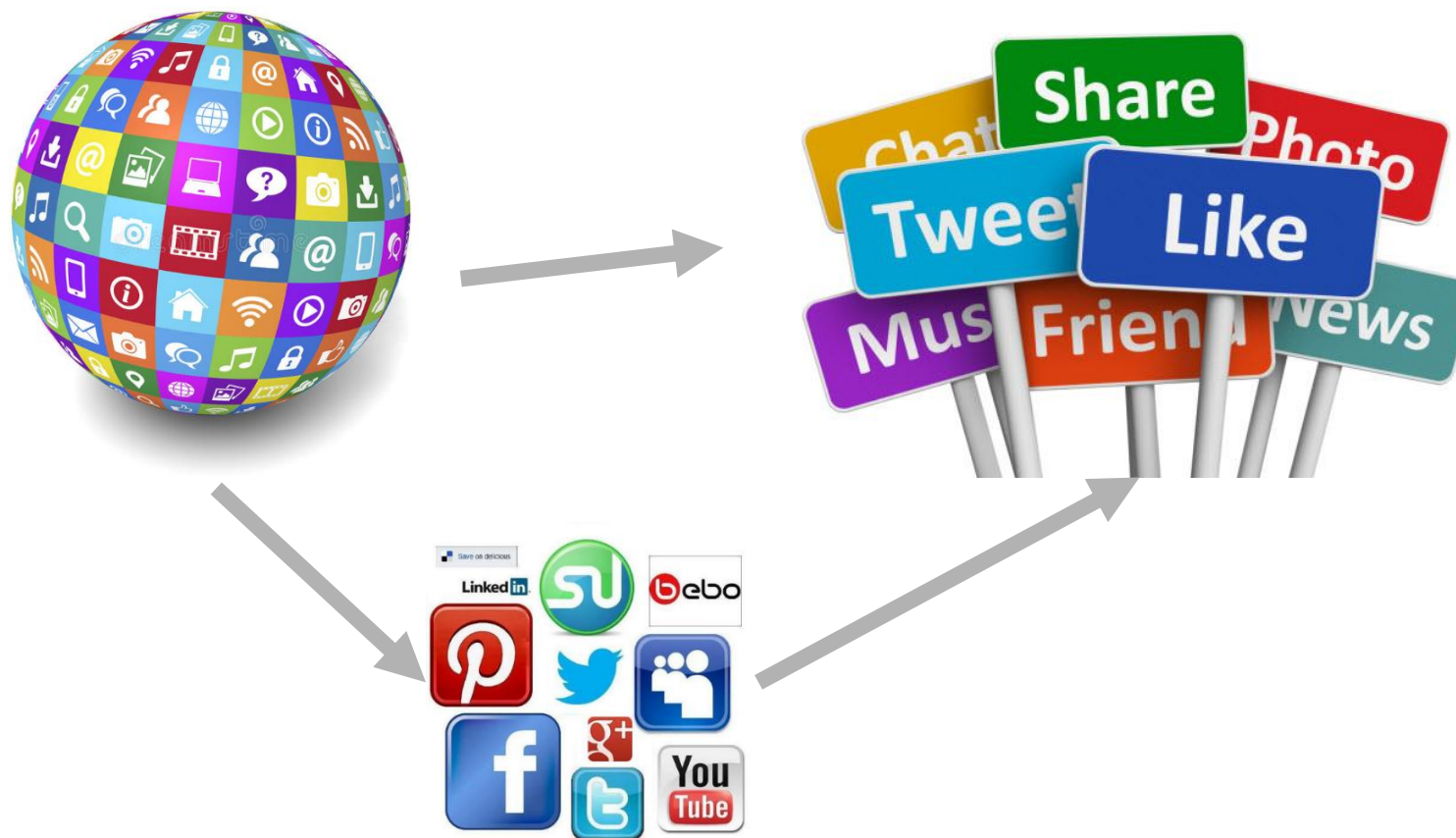
- Introduzir os conceitos de Análise de Texto e Imagens
- Formas de representação;
- Exemplos e exercícios;
- Conclusão;
- Perguntas.

Conceitos

- Introdução Análise de Texto
 - *NLP, Data Mining e Text Mining*
 - Métodos de organização dos dados
 - Possíveis análises
-
- Introdução Análise de imagens
 - Processo de análise de imagens
 - Representação de imagens e tensores

Análise de Texto

- Importância:



Conceitos iniciais

Análise de Texto

<i>NLP – Natural Language Processing</i>	Data Mining	Text Mining
Manuseio de linguagem natural por computadores	Área focada em organizar e agregar dados visando encontrar padrões, anomalias e associações relevantes para um negócio	objetivo transformar textos (dados não estruturados) em informações, utilizando-se métodos estatísticos e processamento de linguagem natural



Análise de Texto

Corpus

Pré-processamento

Tabela de sinônimos

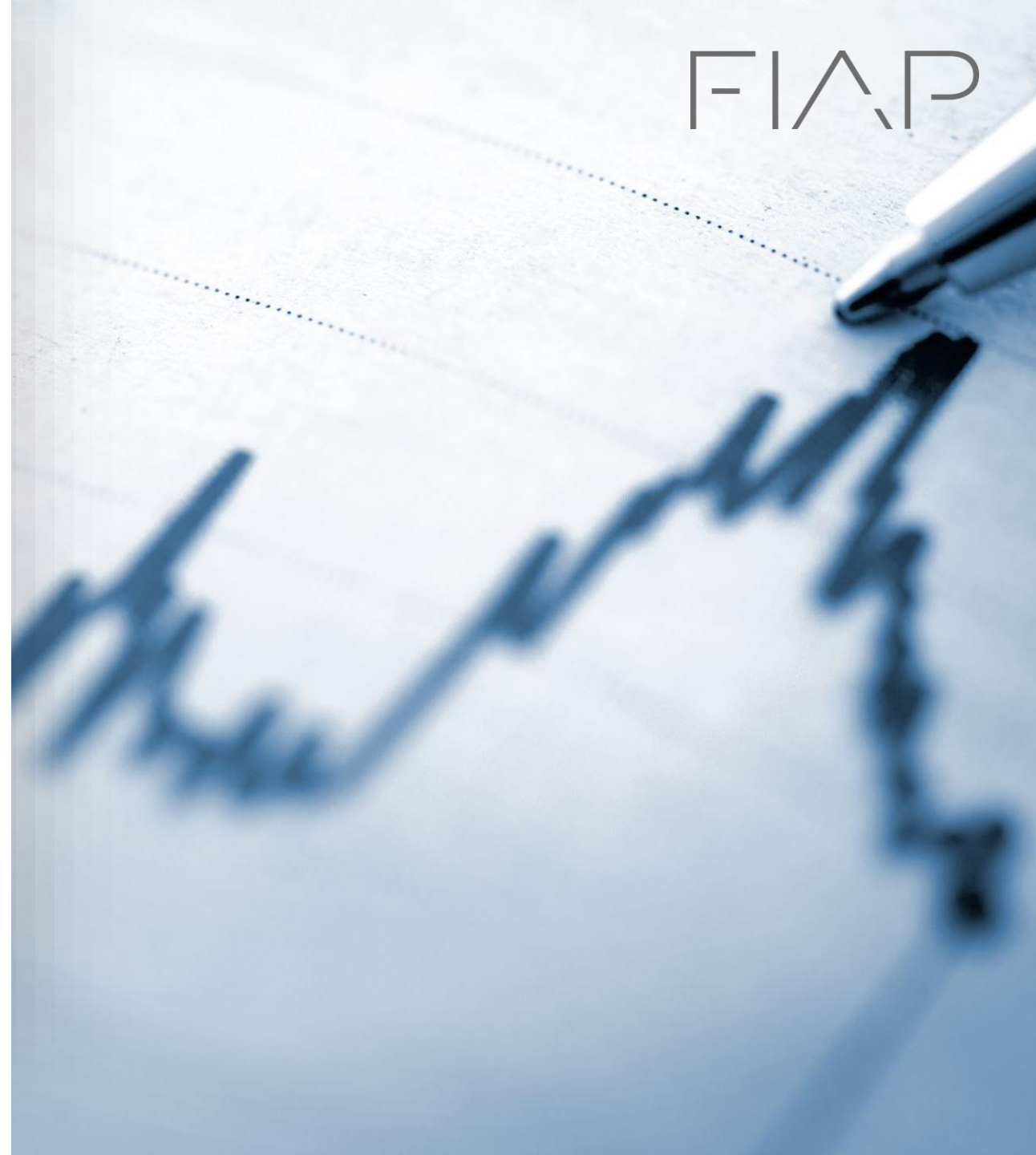
Tokens

Matriz de termos

Análise de Texto

Organização dos dados (matriz de termos e documentos)

- *Bag of words*
- *One hot encoding*
- TF-IDF (*Inverse Document Frequency*)



	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	1	1	1							
Doc 2	1		1	1	1	1				
Doc 3					1	1	1	2	1	1

Análise de Texto

- *Bag of words*

ications	creates	business	processes	bots	are	
1	1	1	1	0	0	
1	0	0	0	1	1	
0	0	1	0	0	0	

Análise de Texto

One hot encoding



Análise de Texto

TF-IDF (*Inverse Document Frequency*)

$$\text{TF} - \text{IDF} = \log \frac{N}{n}$$

	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	0.18	0.48	0.18							
Doc 2	0.18		0.18	0.48	0.18	0.18				
Doc 3					0.18	0.18	0.48	0.95	0.48	0.48



Review Id	nice	super	bad	average	shame	Label
R1	4	2	0	2	0	positive
R2	1	0	3	0	5	negative
R3	2	0	1	1	0	positive
R4	5	0	0	0	0	positive

Análise de Imagens

Importância



Vigilância visual
[video]



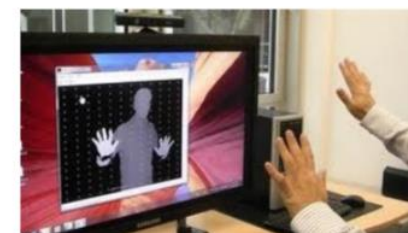
Monitoramento de tráfego



Automação industrial



Aplicativos comerciais



Interação homem/máquina



Games

Desafios

Desafios

- Imagens em diferentes posições.
- Imagens com fundos e cores diferentes.
- Natureza das imagens (a imagem de um coração na medicina e num romance).
- Problemas na captação.
- Qualidade da imagem.
- Analisar imagens em movimento.

Análise de Imagens

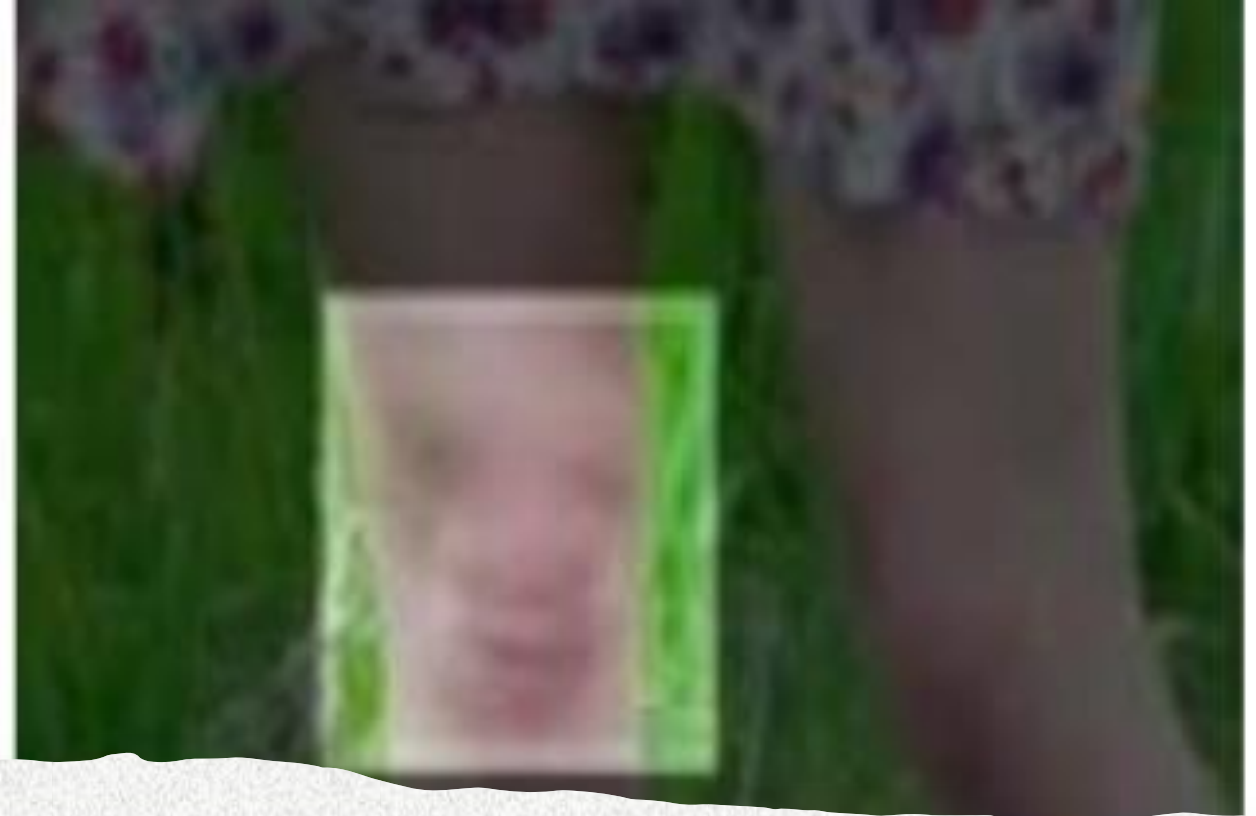
- Sistema de visão computacional
 - Sensores de visão (Captação)
 - Digitalização (Armazenamento)
 - Computador (Processamento)



Análise de Imagens

Processo de análise de imagens

- Aquisição.
- Pré-Processamento.
- Segmentação
- Representação e descrição.
- Reconhecimento de padrões e Interpretação.
- Base de conhecimento.



Desafios

Desafios

Golpe da cara falsa: homem usava boneco e foto para burlar biometria facial



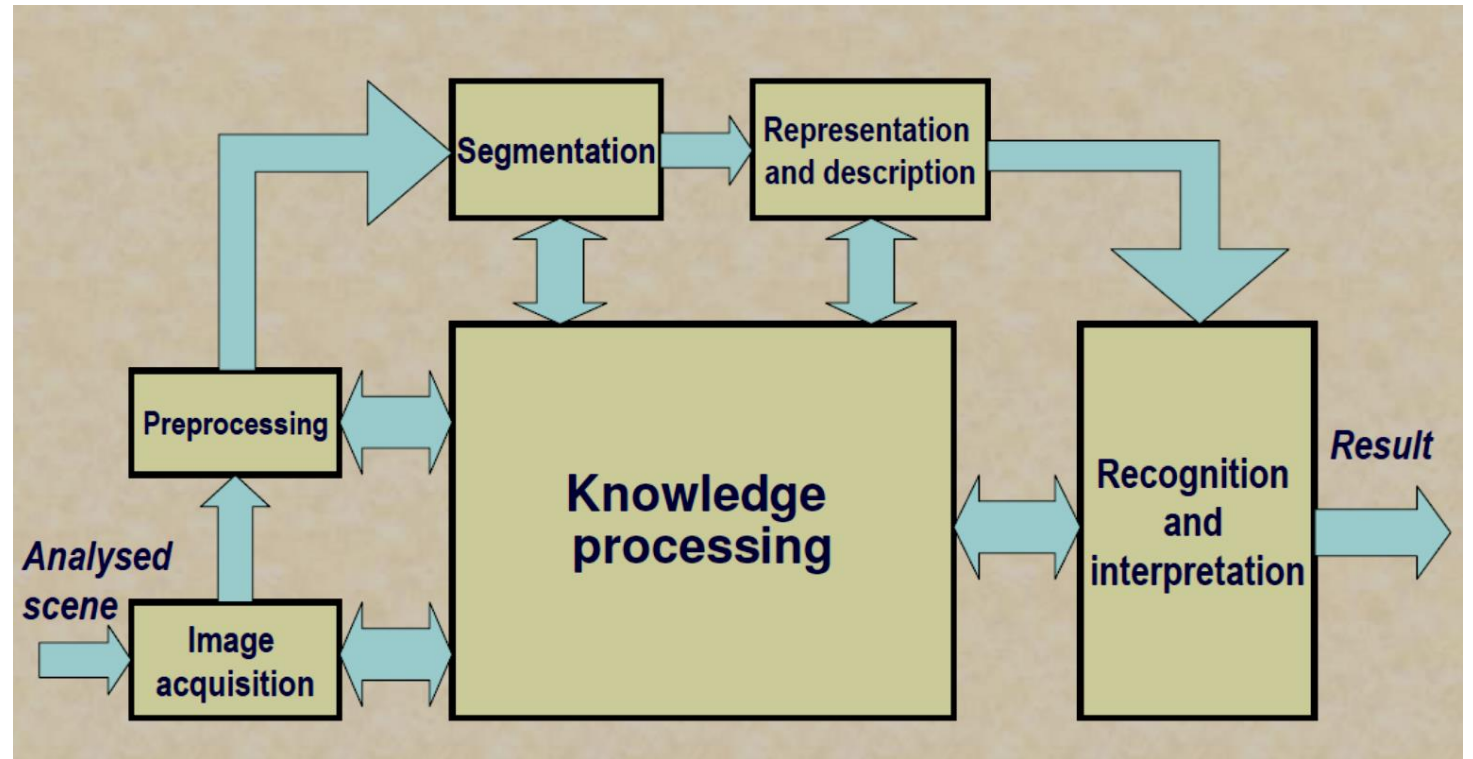
PUBLICIDADE

Oficina®

-40%	Calça Jeans Selvedge	-30%	Calça Jeans Selvedge Dy...
	R\$ 1.199,40		R\$ 1.399,30
-40%	Calça Jeans 5 Pockets		Tênis Chelsea
	R\$ 299,40		R\$ 749

Análise de Imagens

- Processo de análise de imagens

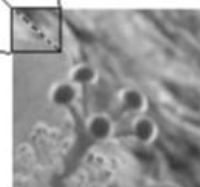


Análise de Imagens

- Representação de uma imagem

Escala de intensidade: [0, ..., 255]

168	153	131	109	94
174	163	148	131	114
177	169	159	150	138
181	175	167	164	159
187	179	171	171	172
194	185	177	175	176
202	194	184	179	178
209	202	191	182	179
212	207	198	189	183
212	212	207	199	192
211	212	212	209	204
204	206	210	210	208



Preto: valores baixos (0)

Cinza: valores intermediários

Branco: valores altos (255)



Análise de Imagens

- Representação de uma imagem colorida

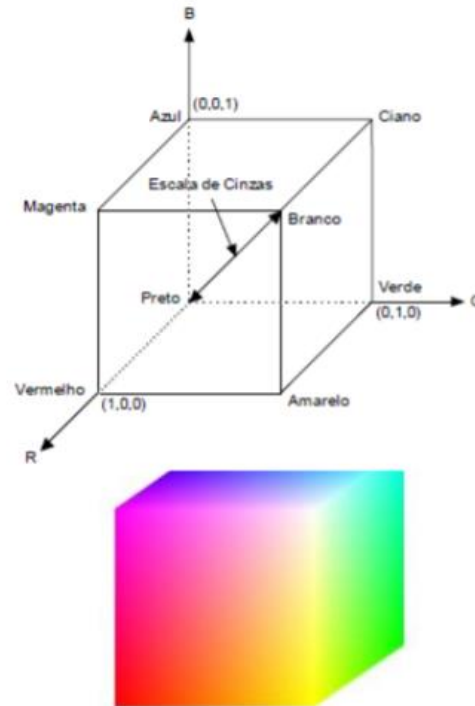


Imagem colorida



Camada R



Camada G



Camada B

Análise de Imagens

- Tensores
- Representam vetores e matrizes com número arbitrário de dimensões.

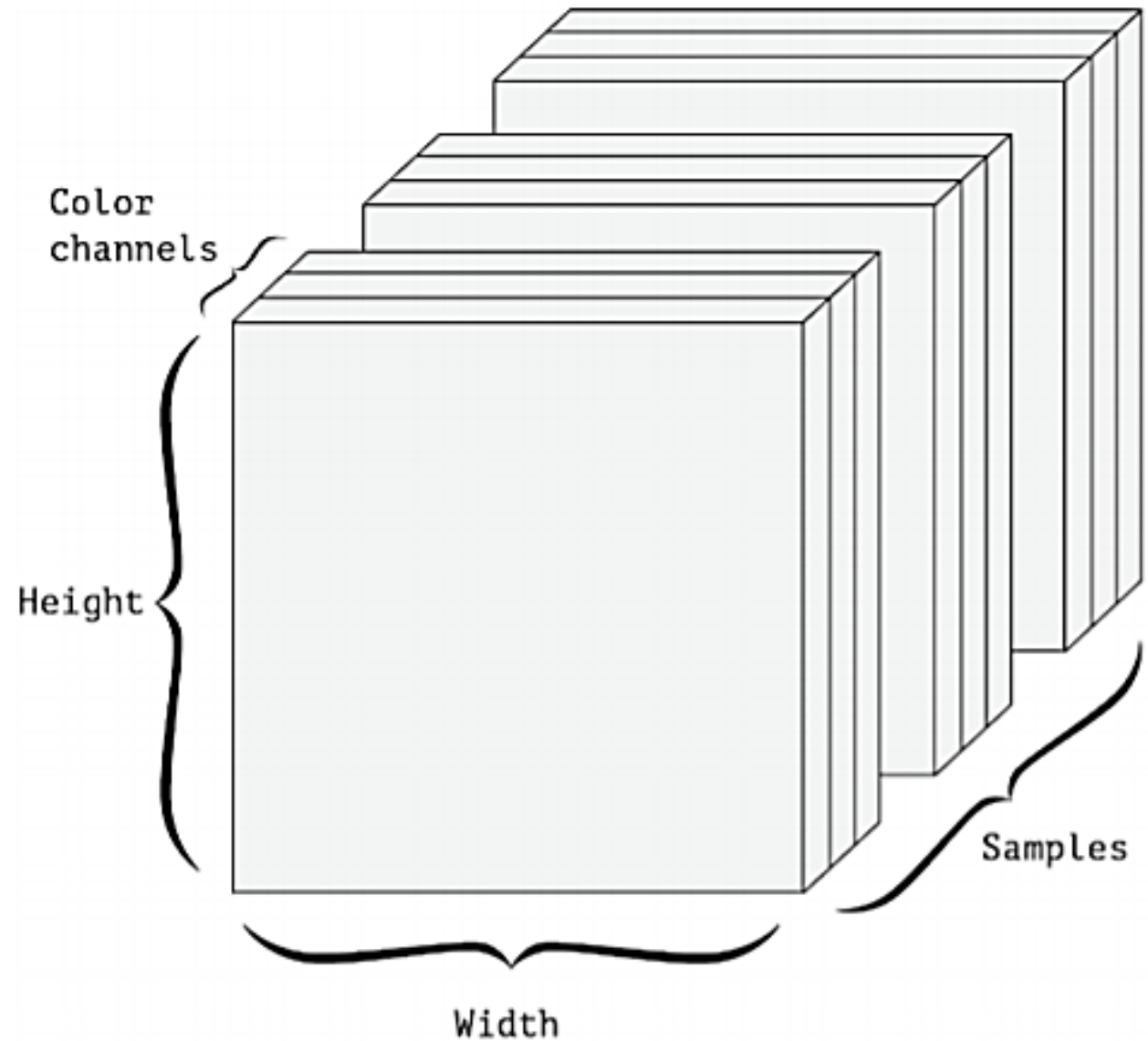
Scalar Vector Matrix Tensor

1

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$
$$\begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} & \begin{bmatrix} 3 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & 7 \end{bmatrix} & \begin{bmatrix} 5 & 4 \end{bmatrix} \end{bmatrix}$$

Análise de Imagens

- Imagens como tensor



Onde estudar mais!!

- Leitura



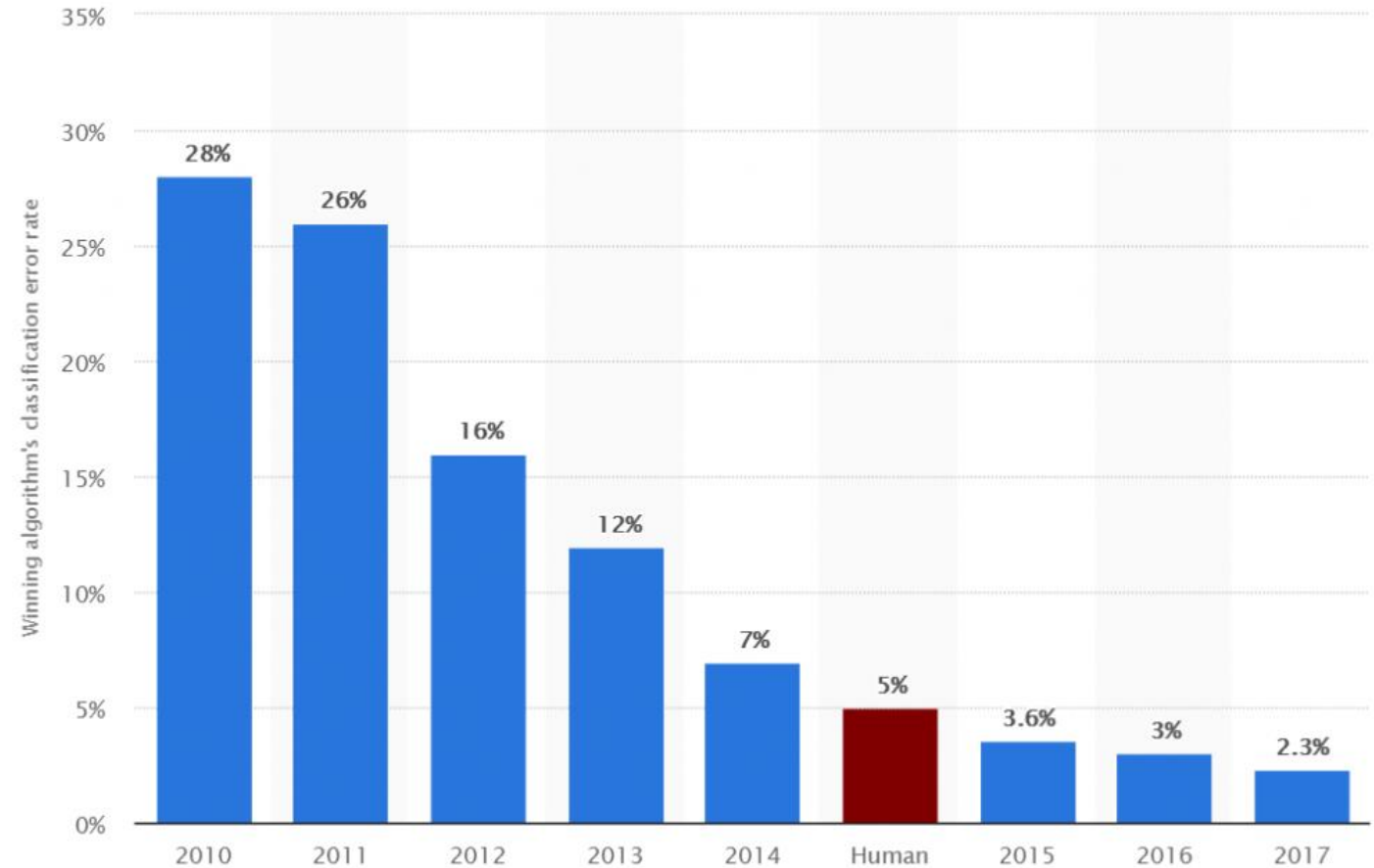
- <https://image-net.org/challenges/LSVRC/>
- <https://www.kaggle.com/getting-started/149448>
- <https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/>
- <http://www.nilc.icmc.usp.br/nilc/index.php>

- Vídeos

- <https://pt.coursera.org/learn/introduction-computer-vision-watson-opencv?#syllabus>
- <https://pt.coursera.org/learn/getting-started-with-tensor-flow2?#syllabus>
- <https://pt.coursera.org/specializations/natural-language-processing?#courses>

ImageNet

- ImageNet - evolução



Obrigado!