

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.

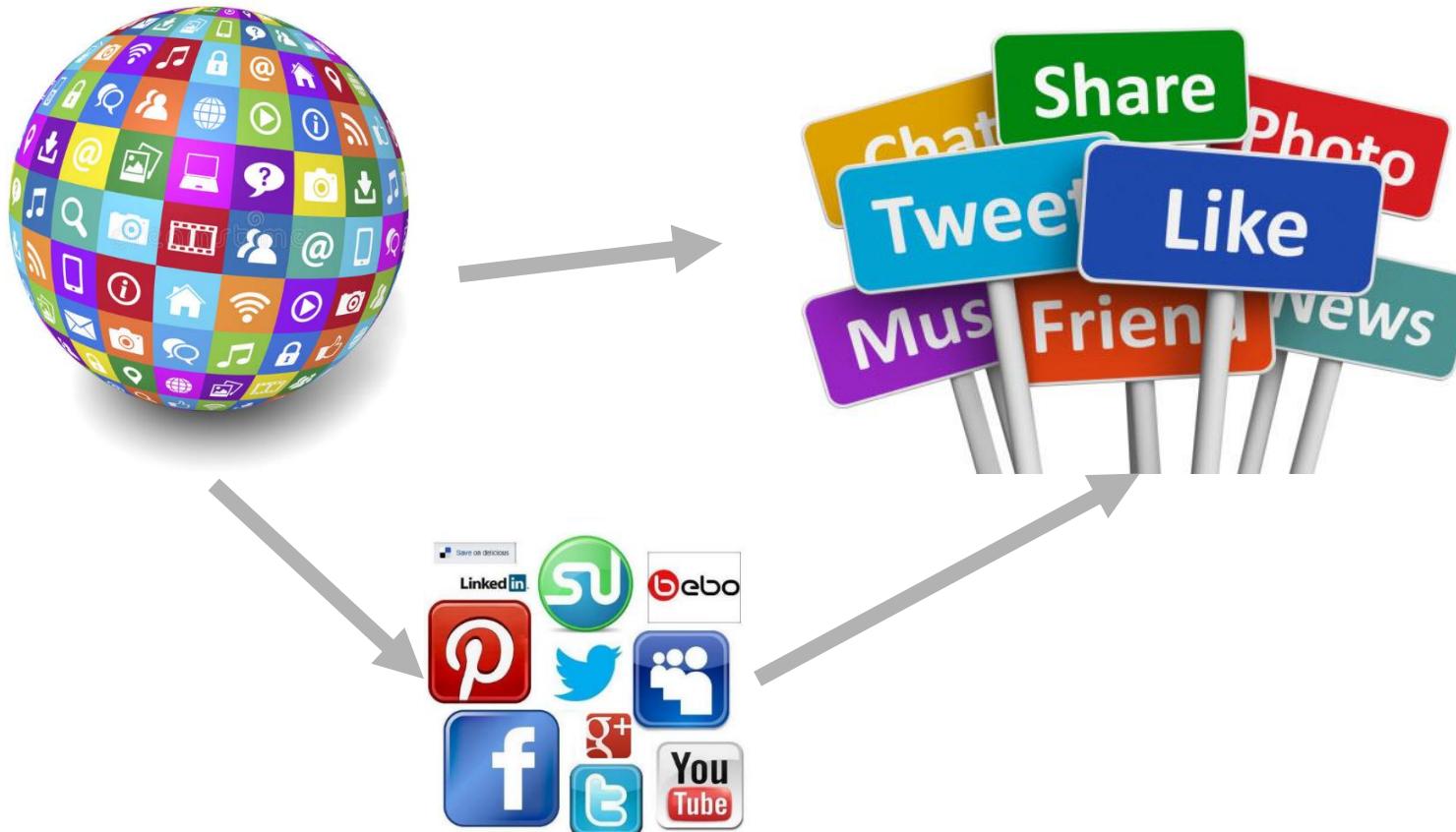
Conceitos

- 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

*NLP – Natural
Language
Processing*

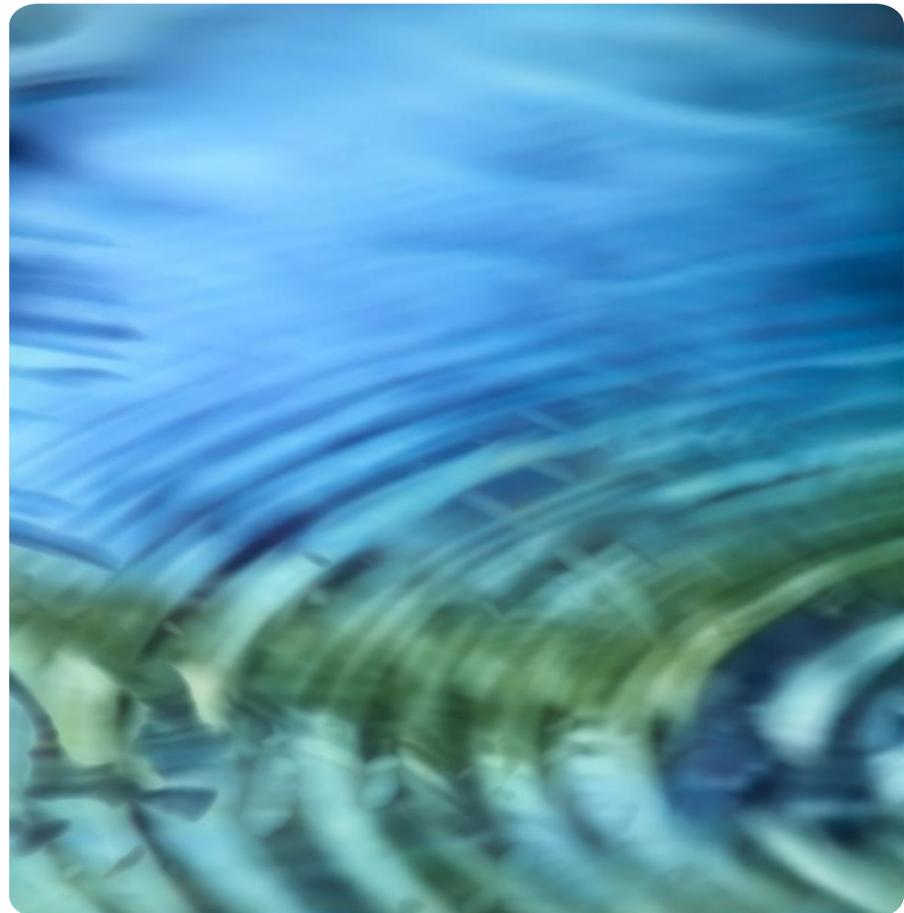
Manuseio de linguagem natural por computadores

Data Mining

Área focada em organizar e agregar dados visando encontrar padrões, anomalias e associações relevantes para um negócio

Text Mining

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

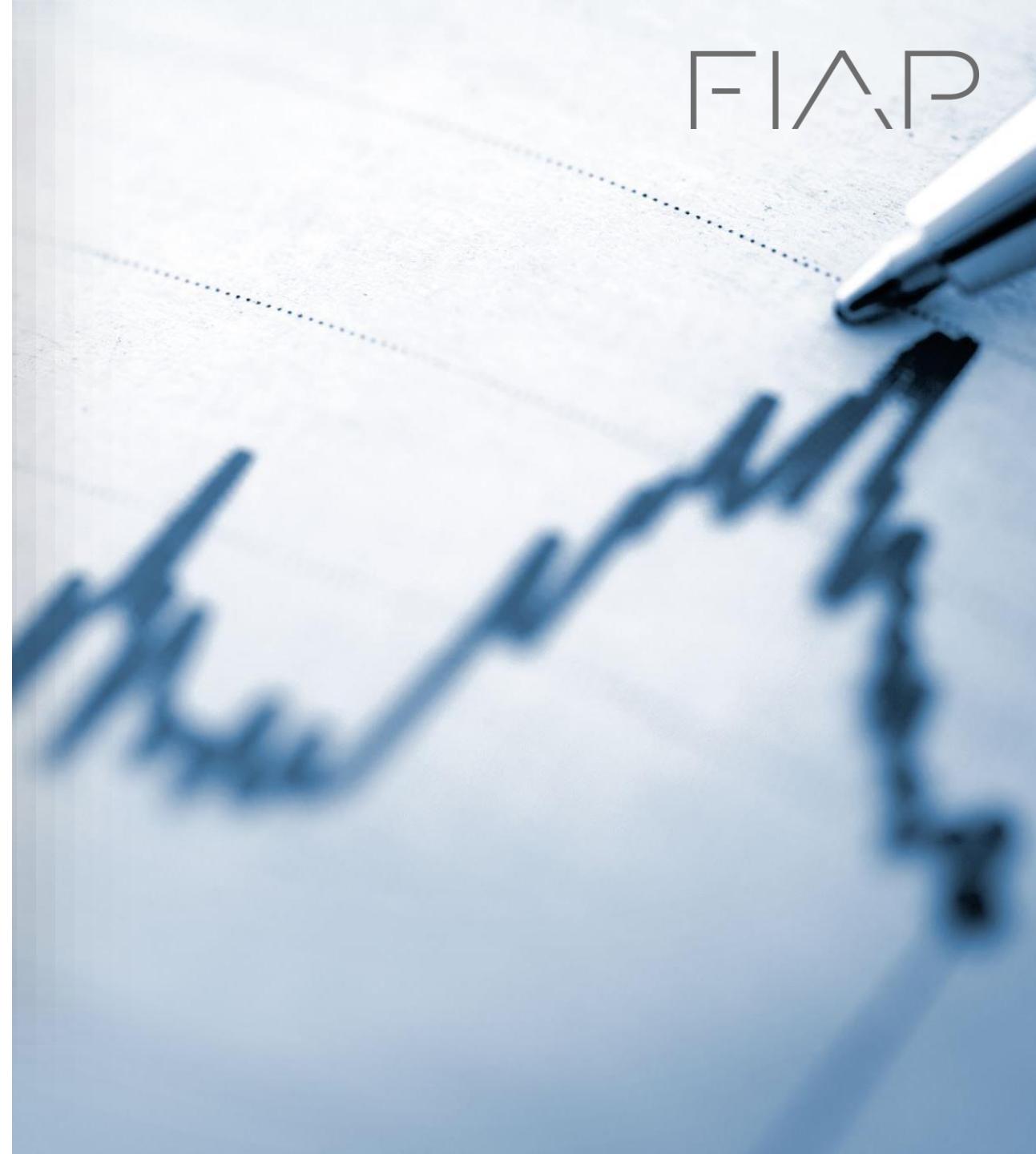
Tokens

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				
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

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

TF-IDF (*Inverse Document Frequency*)

	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

Análise de Texto

Possíveis análises:

FIAP

Nuvem de palavras



Análise de sentimentos

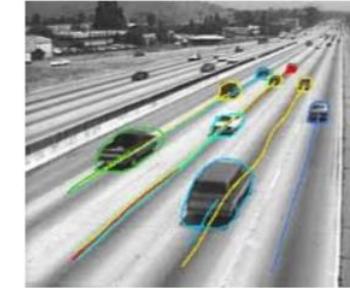
Review Id	nice	super	bad	average	shame	Label
R1	4	2	0	2	0	positive
R2	1	0	3	0	5	negative
R3	2	0	1	1	0	positive
R4	5	0	0	0	0	positive

Análise de Imagens

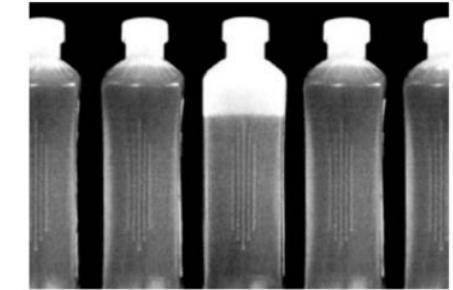
Importância



Vigilância visual
[[vídeo](#)]



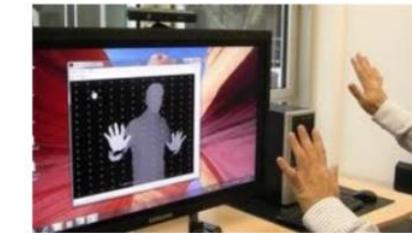
Monitoramento de tráfego



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)



**TESLA DESVIA SOZINHO DE
ACIDENTE IMPRESSIONANTE
NOS EUA**



Análise de Imagens

Processo de análise de imagens

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



Desafios

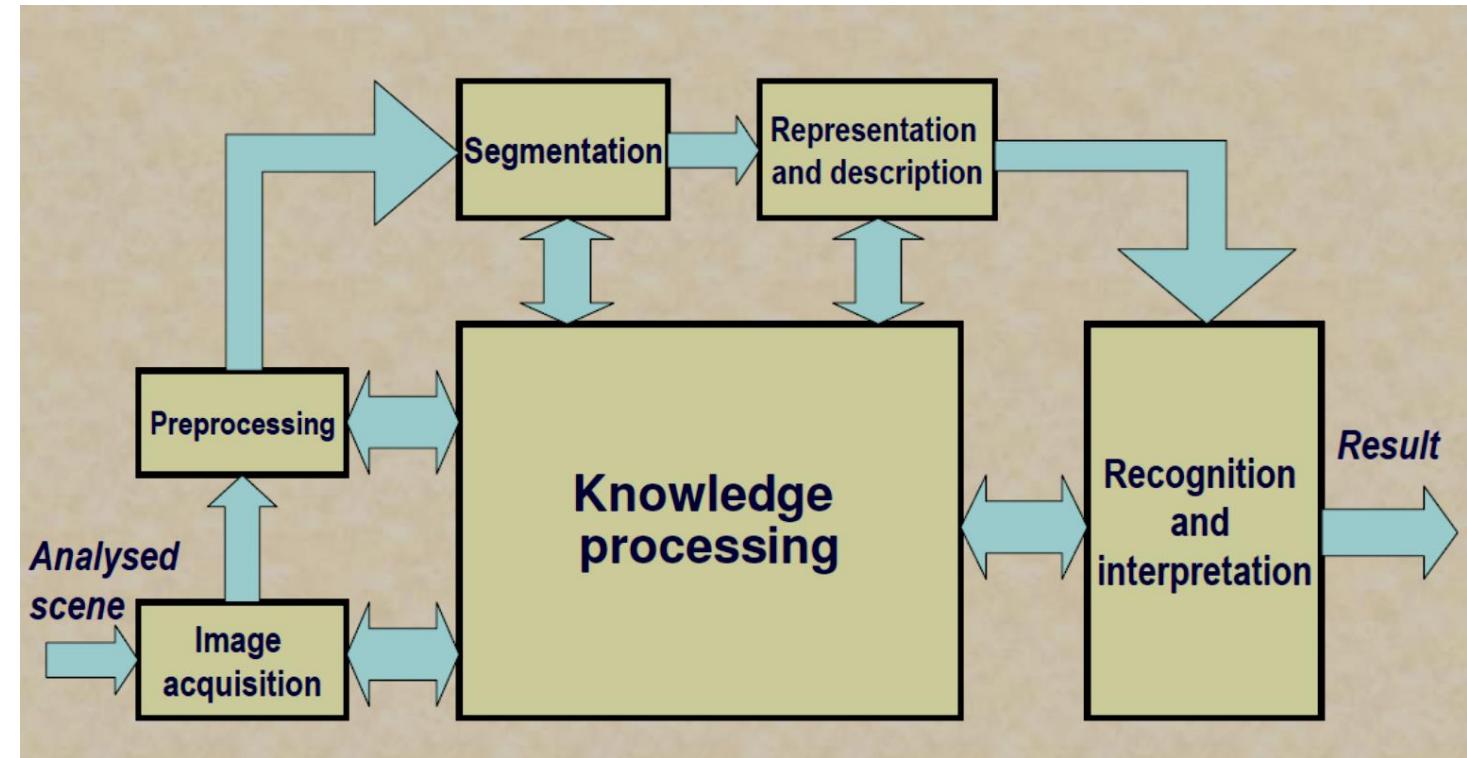
Desafios

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



Análise de Imagens

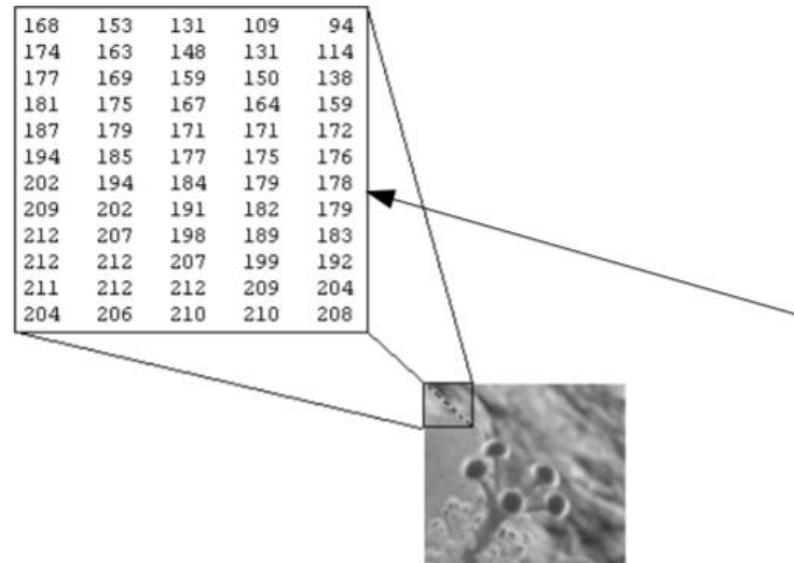
- Processo de análise de imagens



Análise de Imagens

- Representação de uma imagem

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



Preto: valores baixos (0)

Cinza: valores intermediários

Branco: valores altos (255)



Análise de Imagens

- Representação de uma imagem colorida

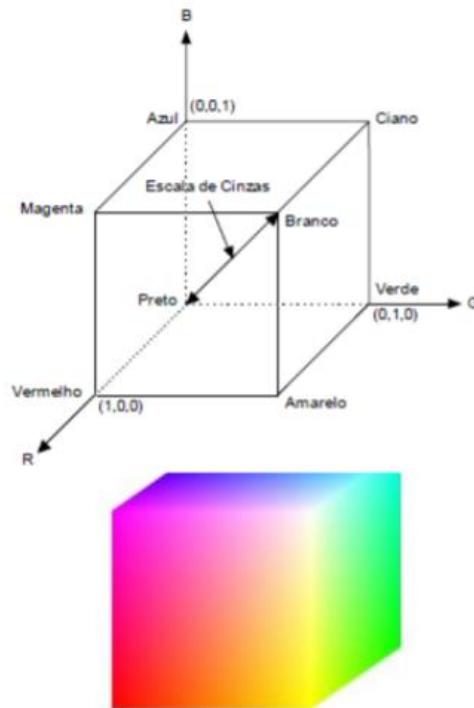


Imagen colorida



Camada R



Camada G



Camada B

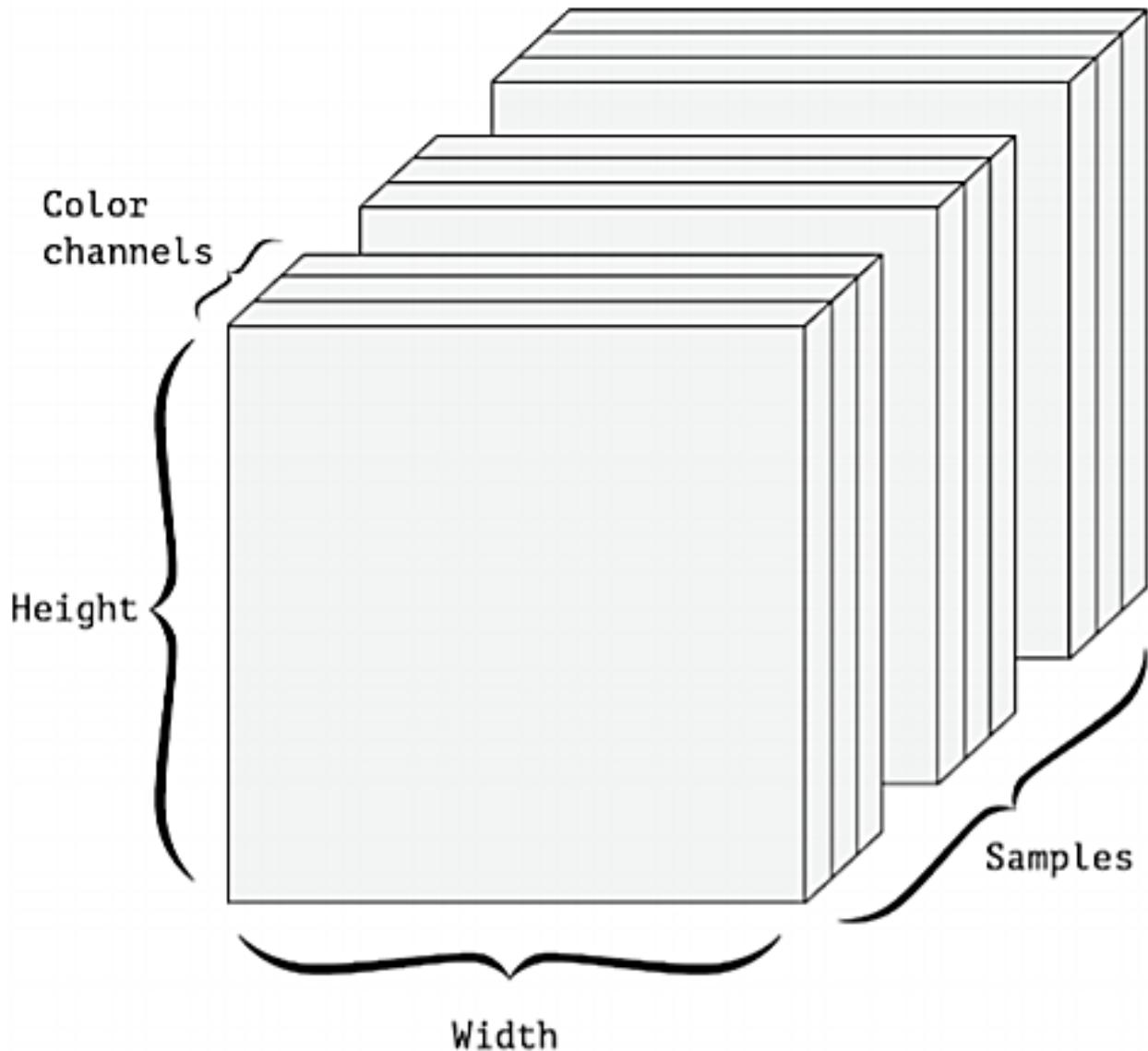
Análise de Imagens

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

	Scalar	Vector	Matrix	Tensor
1	$[1]$	$[1 \ 2]$	$[1 \ 2]$	$[1 \ 2] \ [3 \ 2]$ $[1 \ 7] \ [5 \ 4]$

Análise de Imagens

- Imagens como tensor



Onde estudar mais!!

- Leitura



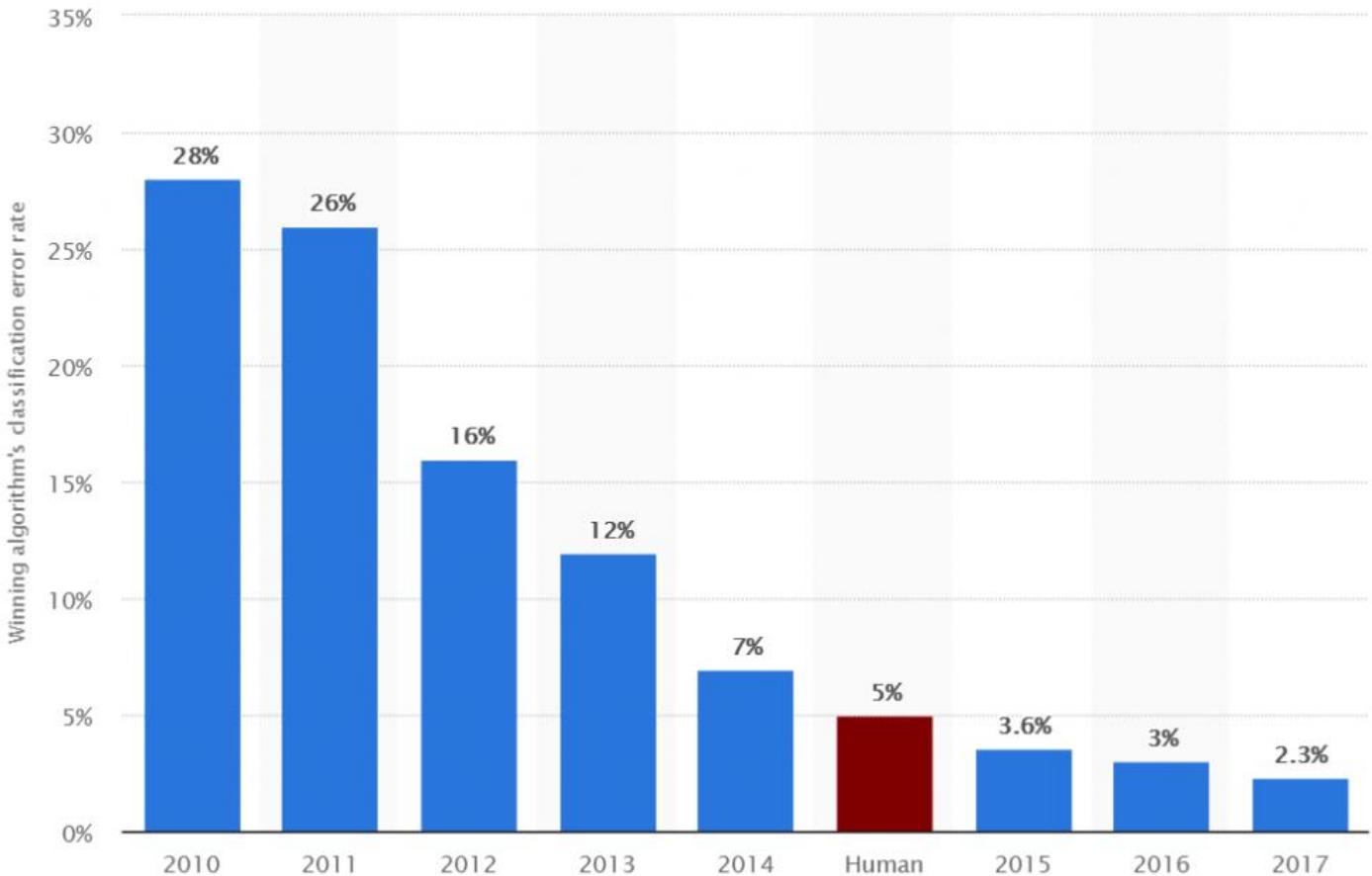
- <https://image-net.org/challenges/LSVRC/>
- <https://www.kaggle.com/getting-started/149448>
- <https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/>
- <http://www.nilc.icmc.usp.br/nilc/index.php>

- Vídeos

- <https://pt.coursera.org/learn/introduction-computer-vision-watson-opencv?#syllabus>
- <https://pt.coursera.org/learn/getting-started-with-tensorflow2?#syllabus>
- <https://pt.coursera.org/specializations/natural-language-processing?#courses>

ImageNet

- ImageNet - evolução



https://github.com/jegydio/PhD/blob/main/Exerc%C3%ADo_Programa_v2_com_Triplet.ipynb

Obrigado!

