VINI CARIDÁ

Human-Centered Data & AI

Vinicius Caridá, Ph.D.

Head of Digital Customer Service Platforms, PCP, WFM, Data and AI - Itaú Unibanco

MBA Professor - FIAP

Google Developer Expert – Machine Learning

Co-organizer TFUGSP and AWSUGSP

@vinicius caridá @vfcarida @vinicius caridá @vfcarida @vinicius caridá @vfcarida

VINI CARIDÁ
Human-Centered Data & AI

Zero to Hero
Machine Learning na AWS

Parte 5/5

Vinicius F. Caridá

VINI CARIDÁ
Human-Centered Data & AI

“

Recapitulando...

Zero to Hero
Machine Learning na AWS
Parte 1/5

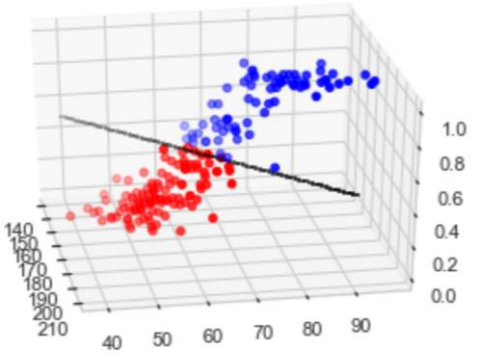
Regressão Linear

VINI CARIDÁ
Human-Centered Data & AI

	Peso	Altura
Pessoa 1	80 kg	163
Pessoa 2	85 kg	168
Pessoa 3	90 kg	175
Pessoa 4	95 kg	188
Pessoa 5	82 kg	158

$$\hat{y} = \beta_0 + \beta_1 X_1$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



$\beta_0 = 3$

$\beta_1 = 0,5$

MSE = 7,125

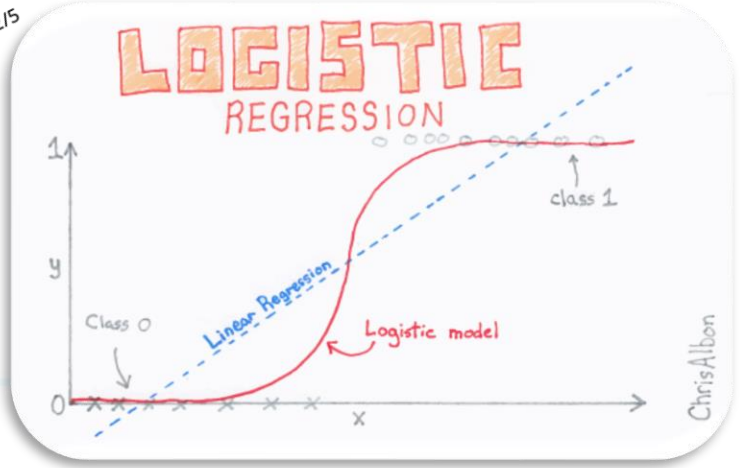
$$\hat{y} = 3 + 0,5\ X_1$$

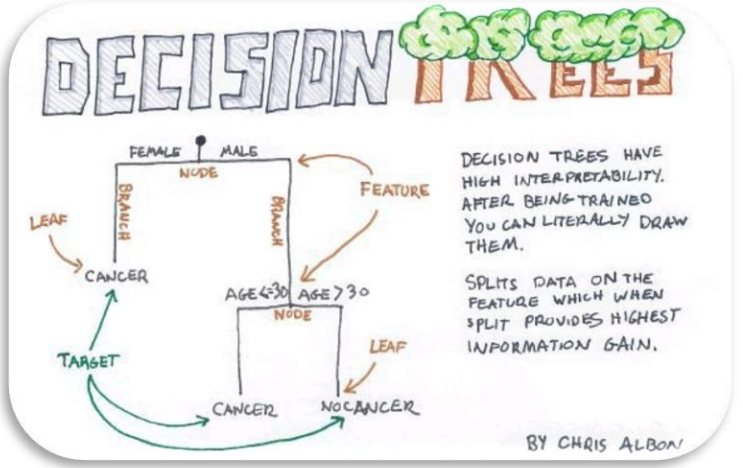
$$\hat{y} = 3 + 0,5\ x\ 158$$

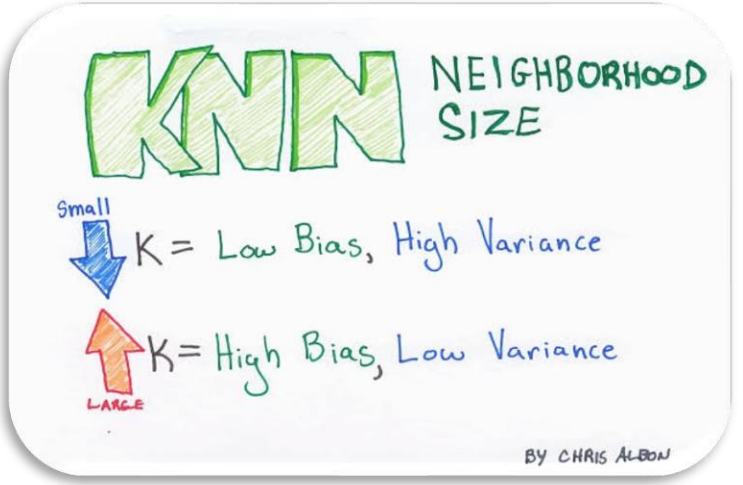
Amazon SageMaker Studio

Zero to Hero
Machine Learning na AWS
Parte 2/5

Classificação

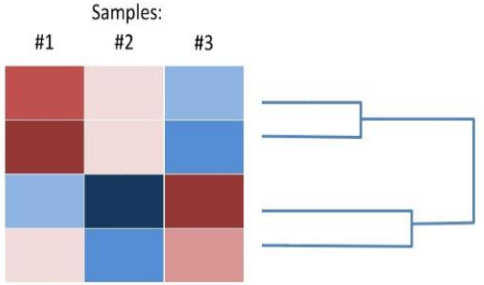






Agrupamento

Hierarchical Clustering, Clearly Explained!!!!



AGGLOMERATIVE CLUSTERING

All observations start as their own cluster. Clusters meeting some criteria are merged. This process is repeated, growing clusters until some end point is reached.

ChrisAlbon

K-MEANS CLUSTERING

1. k centerpoints are randomly initialized.
2. Observations are assigned to the closest centerpoint.
3. Centerpoints are moved to the center of their members.
4. Repeat steps 2 and 3 until no observation changes membership in step 2.

Chris Albon

DBSCAN

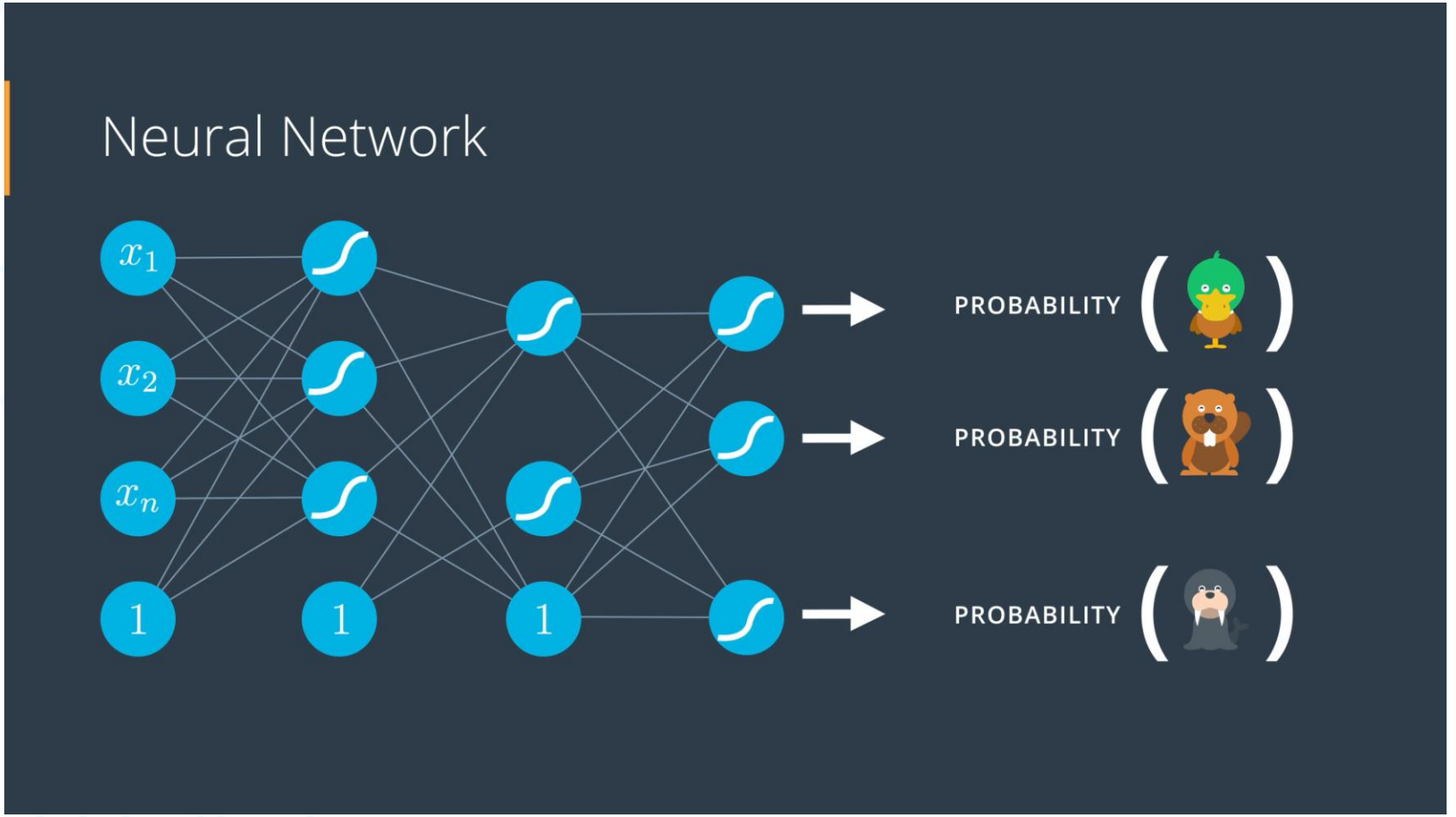
DBSCAN looks for densely packed observations and makes no assumptions about the number or shape of clusters.

1. A random observation, x_i, is selected
2. If x_i has a minimum of close neighbors, we consider it part of a cluster.
3. Step 2 is repeated recursively for all of x_i's neighbors, then neighbors' neighbors etc... These are the cluster's core members.
4. Once Step 3 runs out of observations, a new random point is chosen

Afterwards, observations not part of a core are assigned to a nearby cluster or marked as outliers.

ChrisAlbon

Redes Neurais



VINI CARIDÁ
Human-Centered Data & AI

Classificação de Texto

end-to-end

Vinicius F. Caridá

Embedding / NLP

VINI CARIDÁ
Human-Centered Data & AI

amazon

★★★★★

Good price! Quality not bad! I'm happy I bought it.

★★★★★

Bad quality! I'm sad! I bought it I will return it.

Embedding / NLP

amazon

★★★★★

Good price! Quality not bad! I'm happy I bought it.

★☆☆☆☆

Bad quality! I'm sad! I bought it I will return it.

fail	good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
0	1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	0	1	1	0	1	1	1	1	1	0	1	1

Embedding / NLP

amazon

★★★★★

Good price! Quality not bad! I'm happy I bought it.

★★★★★

Bad quality! I'm sad! I bought it I will return it.

fail	good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
0	1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	0	1	1	0	1	1	1	1	1	0	1	1

★★★★★

Price not good. Quality bad! I'm not happy I bought it.

Embedding / NLP

amazon

★★★★★

Good price! Quality not bad! I'm happy I bought it.

★★★★★

Bad quality! I'm sad! I bought it I will return it.

★★★★★

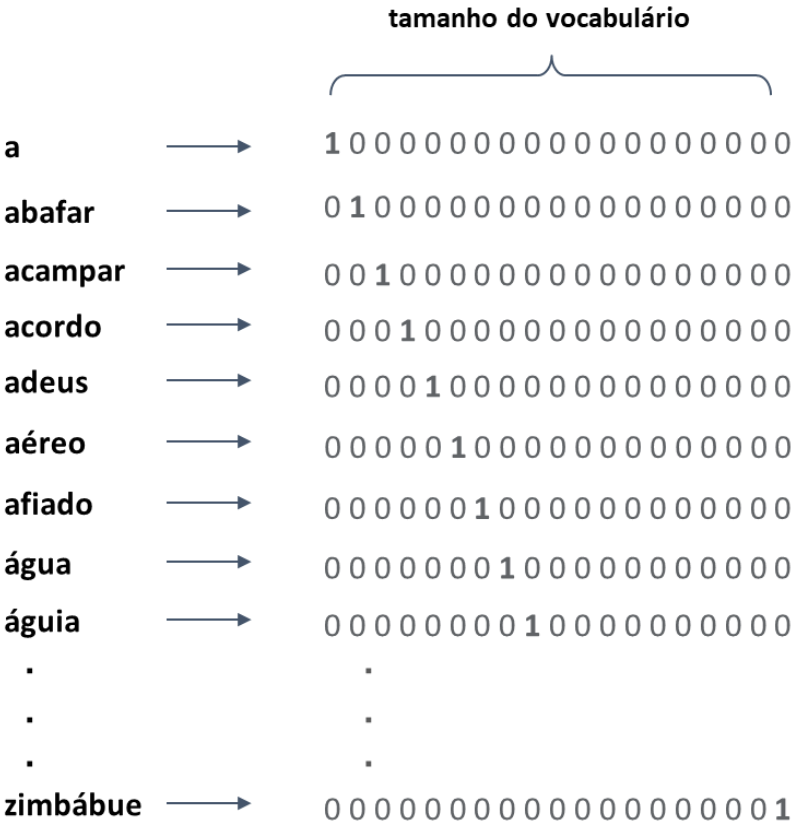
Price not good. Quality bad! I'm not happy I bought it.

fail	good	card	price	quality	bad	not	I	am	it	bought	return	happy	sad	will
0	1	0	1	1	1	1	1	1	1	1	0	1	0	0
0	0	0	0	1	1	0	1	1	1	1	1	0	1	1
0	1	0	1	1	1	1	1	1	1	1	0	1	0	0

Embedding / NLP

Bag of Words

Marcar apenas a posição correspondente à palavra em um vetor



Embedding / NLP

Bag of Words

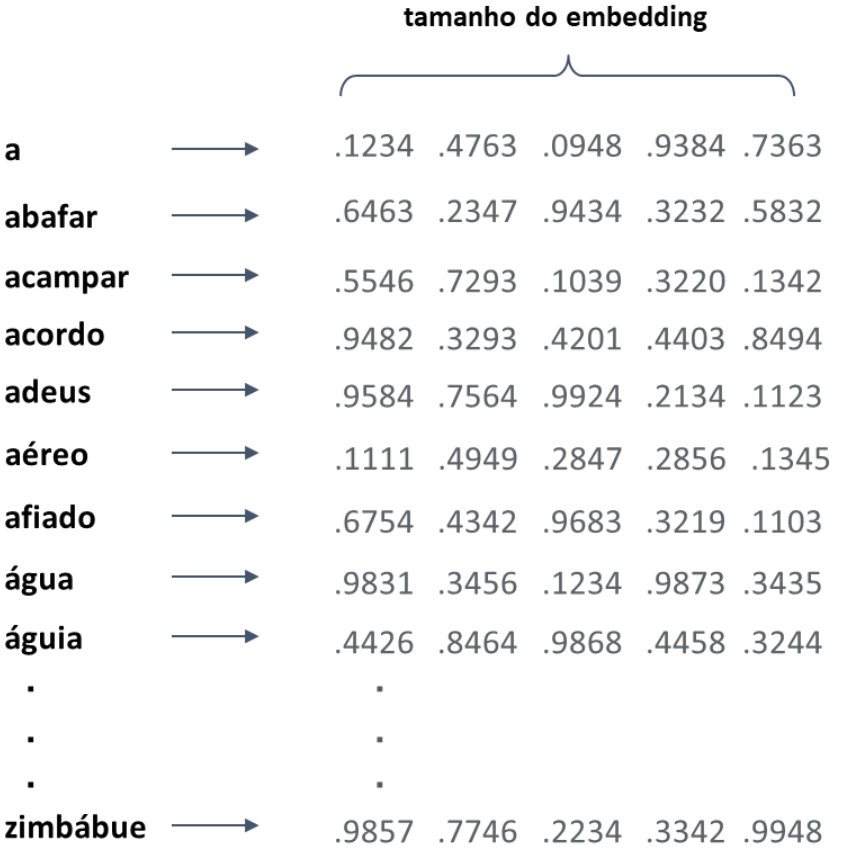
Representação muito grande e esparsa

Não há relação semântica entre palavras

Embedding / NLP

Word Embedding

Aprender a representação das palavras

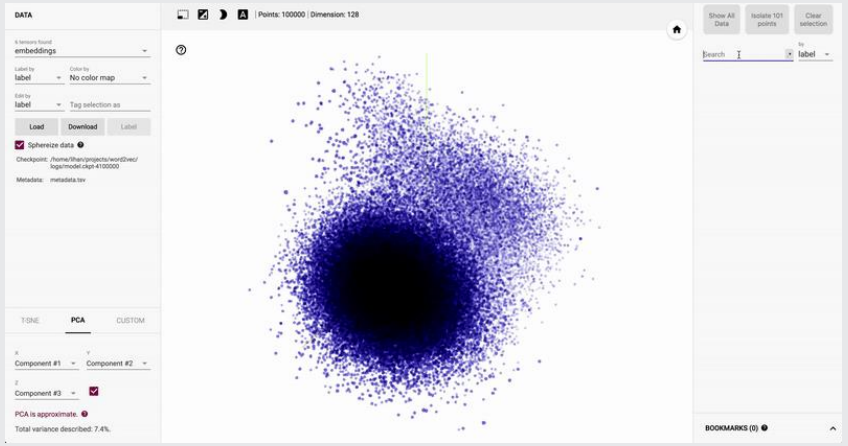


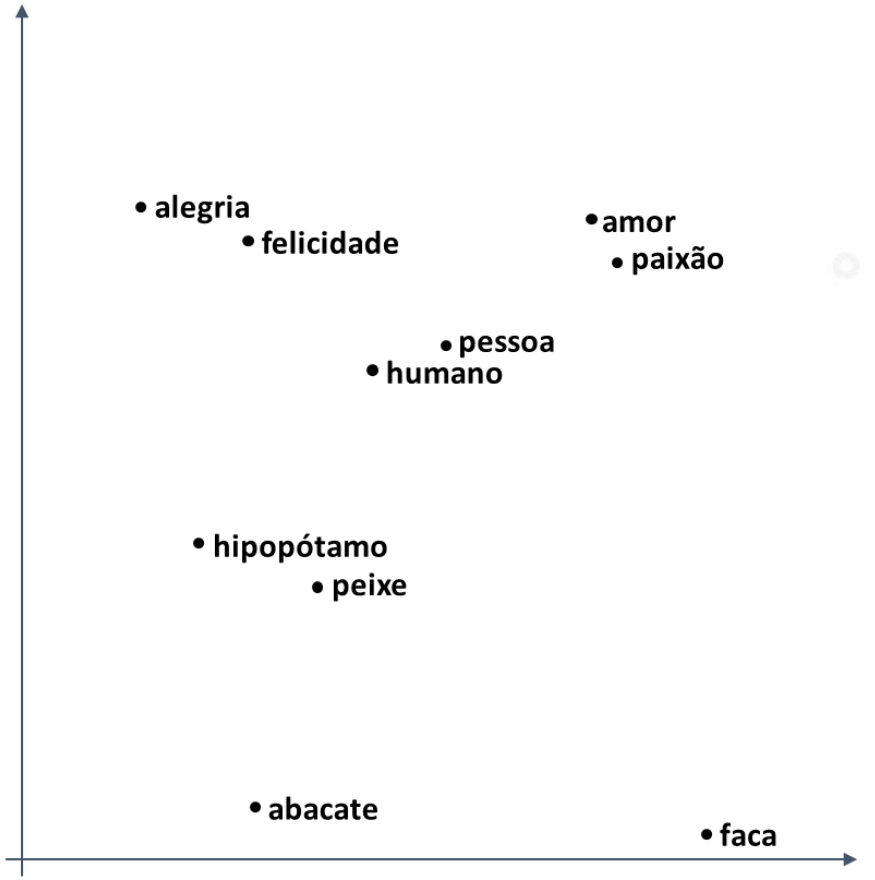
Embedding / NLP

Word Embedding

Representação densa

Palavras com significado similar próximas





Embedding / NLP

Word2Vec
Mikolov et al, 2013

Glove
Pennington et al., 2014

CoVe
McCann et al., 2017

ULMFiT
Howard et al.. 2018

ELMo
Peters et al, 2018

GPT
Radford et al.. 2018

BERT
Devlin et al, 2018

GPT-2
Radford et al.. 2019

XLM
Lample et al.. 2019

XLNet
Yang et al, 2019

RoBERTa
Liu t al, 2019

GPT-3
Brown et al.. 2020

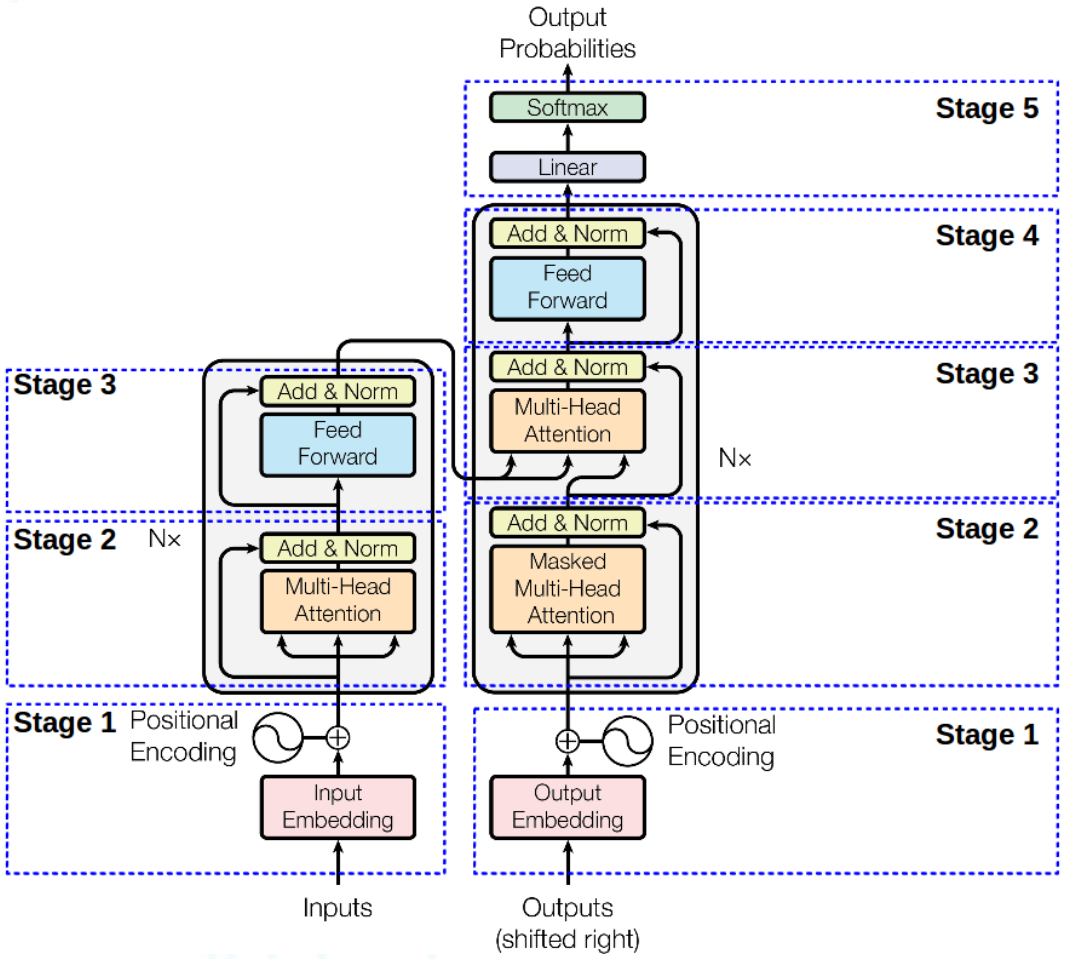
Transformer / Attention - 2017
https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Attention Is All You Need

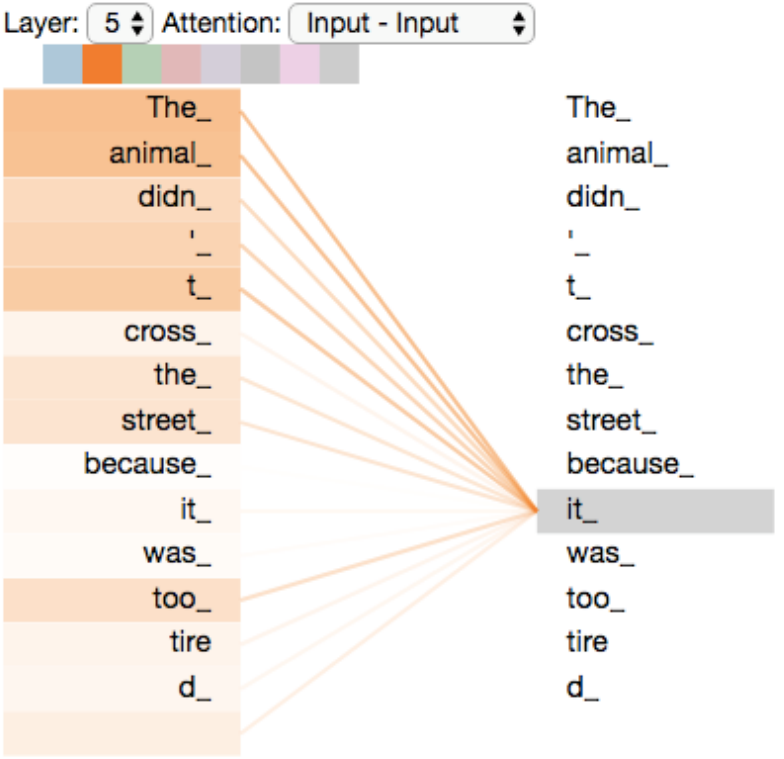
Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com

https://arxiv.org/pdf/1706.03762.pdf

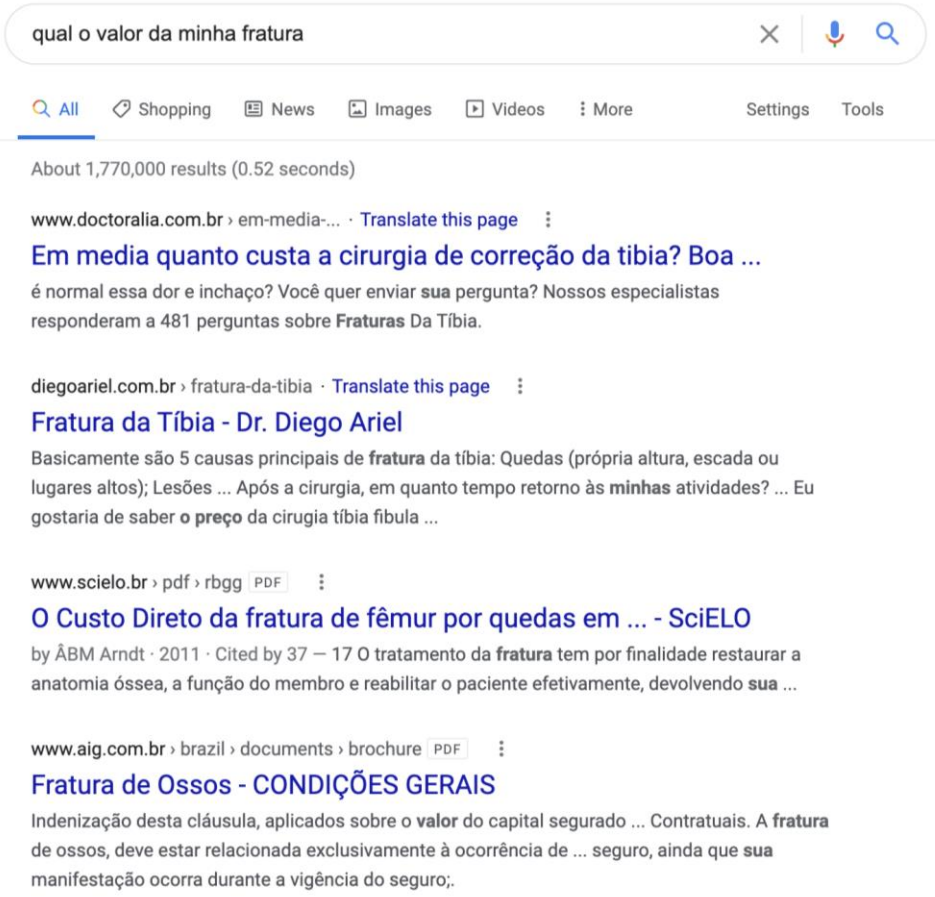
NLP

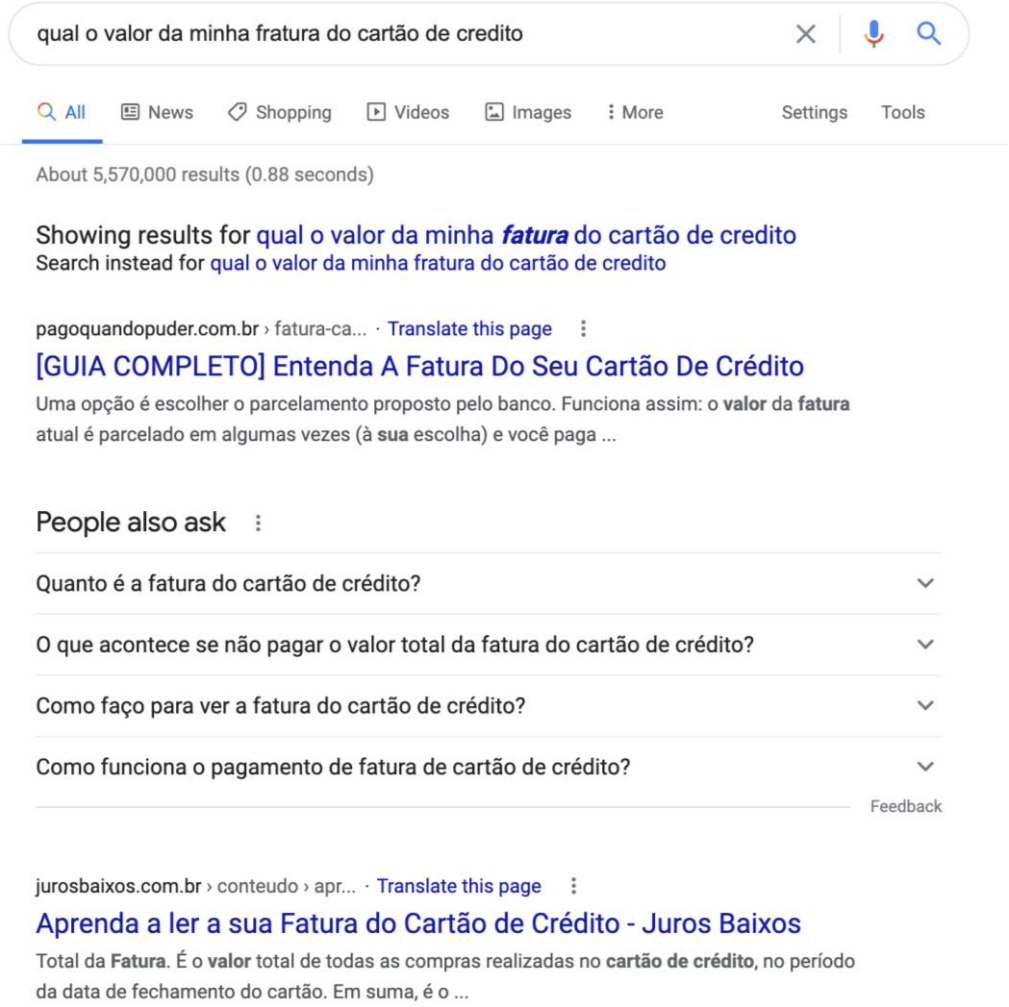


The animal didn't cross the street because it was too tired

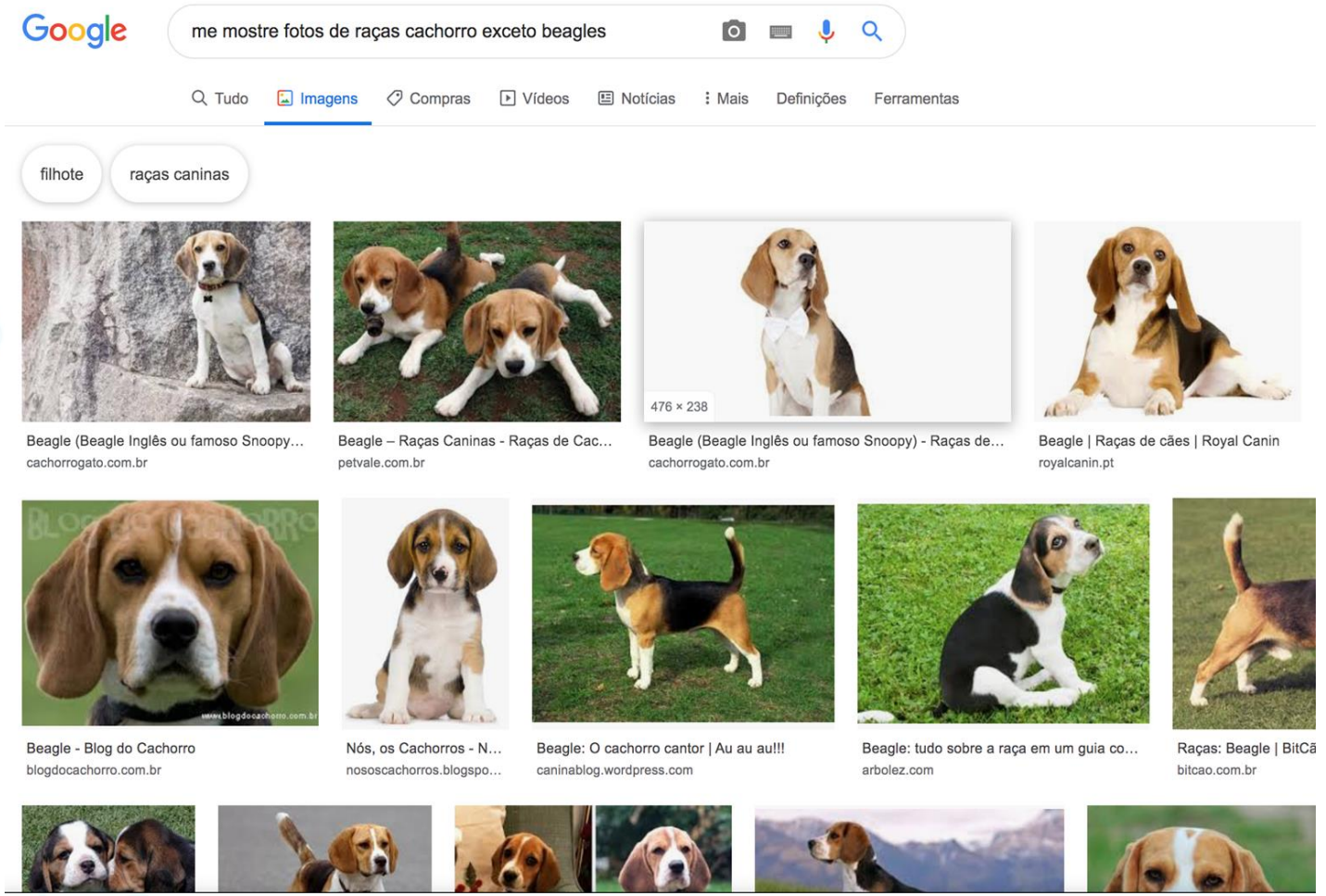


Embedding / NLP





Embedding / NLP



VINI CARIDÁ
Human-Centered Data & AI

Classificação de Texto

end-to-end

Vinicius F. Caridá

Website

templates

App.py

Requeriments.txt

Sagemaker

SageMaker Project.ipynb

Train

Model.py

train.py

Requeriments.txt

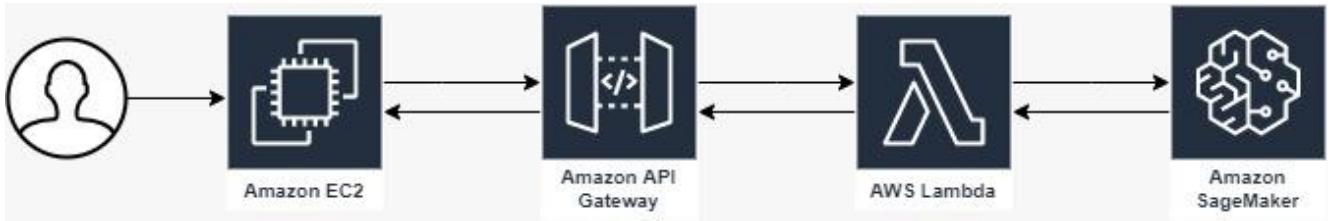
Serve

Model.py

Predict.py

Utils.py

Requeriments.txt



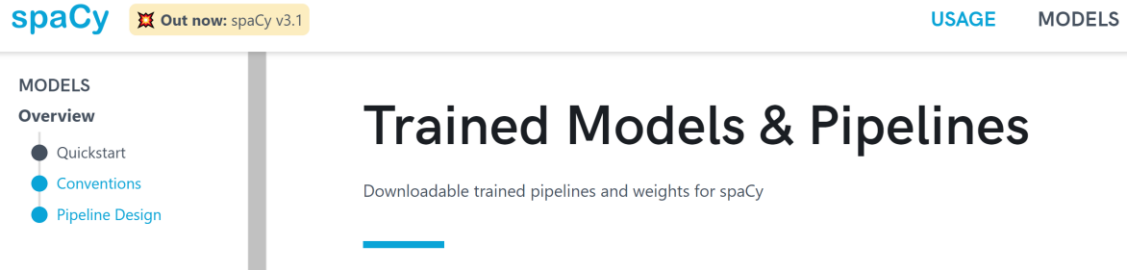
Embedding / NLP

VINI CARIDÁ
Human-Centered Data & AI

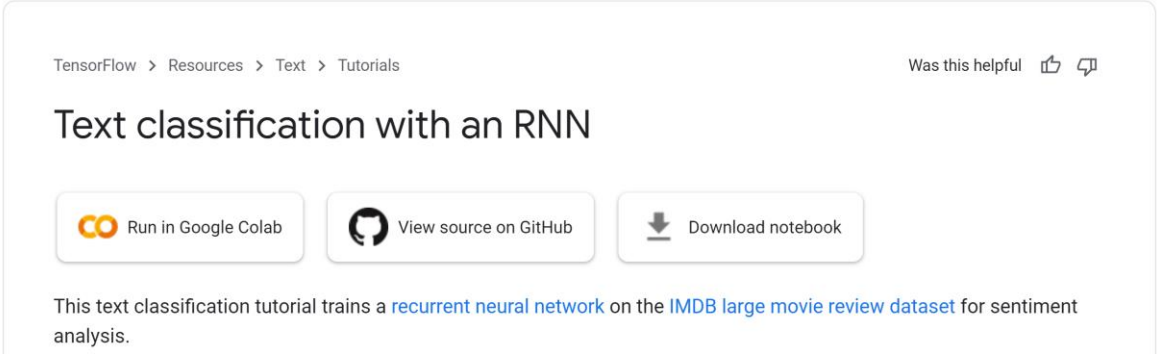
https://www.nltk.org/

Natural Language Toolkit

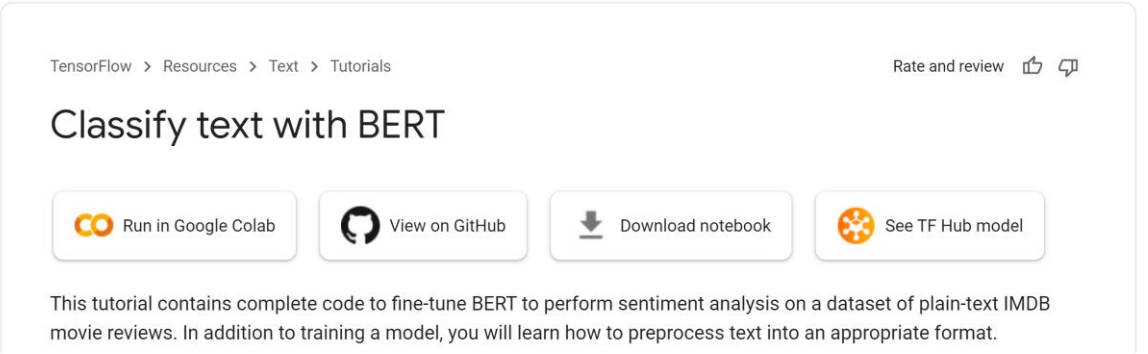
https://spacy.io/models



https://www.tensorflow.org/text/tutorials/classify_text_with_bert



https://www.tensorflow.org/text/tutorials/classify_text_with_bert



Brindes

Thanks !

Vinicius Fernandes Caridá

vfcarida@gmail.com


@vinicius caridá


@vfcarida


@vinicius caridá


@vfcarida


@vinicius caridá


@vfcarida