



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

“






# Zero to Hero Machine Learning na AWS

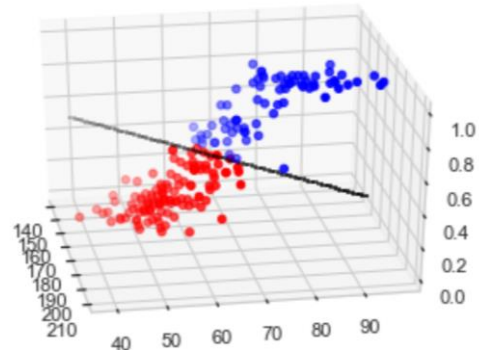
Parte 5/5

“

# Recapitulando...

# Regressão Linear

|   |          | 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    |



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

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

$$\beta_0 = 3$$

$$\beta_1 = 0,5$$

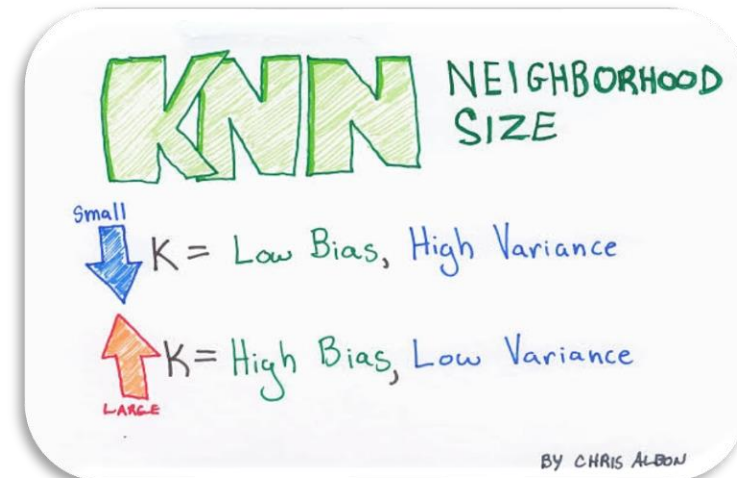
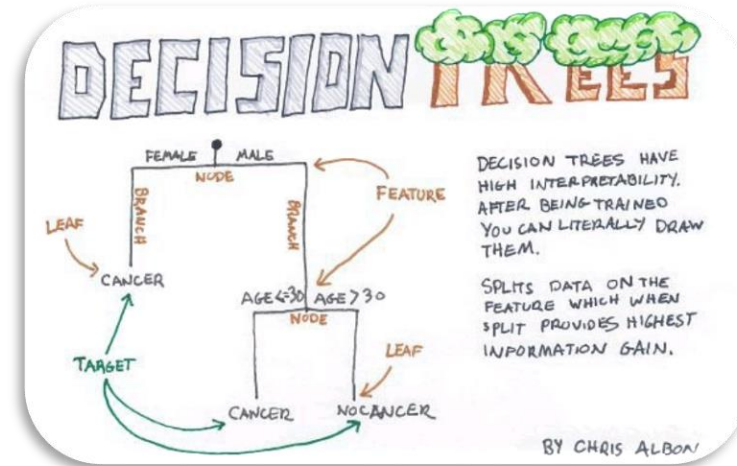
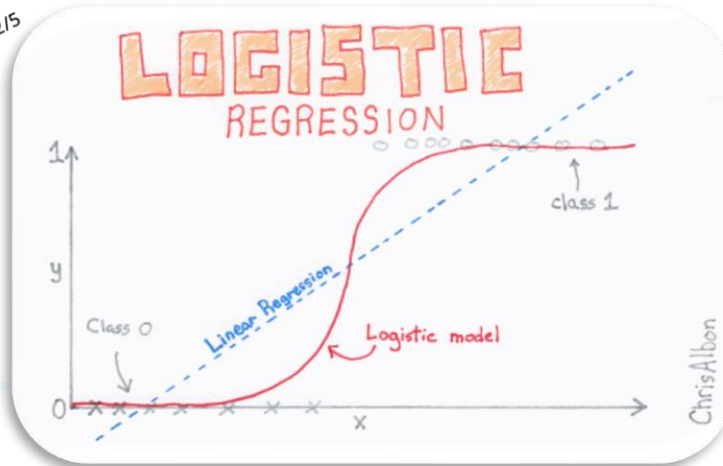
$$MSE = 7,125$$

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

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



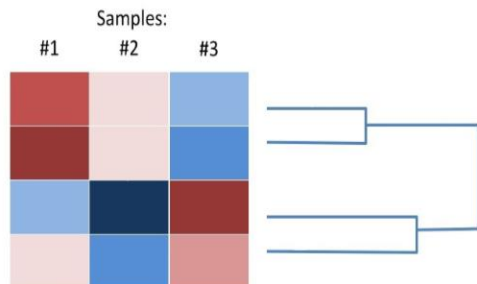
# Classificação





# Agrupamento

## Hierarchical Clustering, Clearly Explained!!!!



## 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.

ChrisAlbon



## 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

## DBSCAN

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

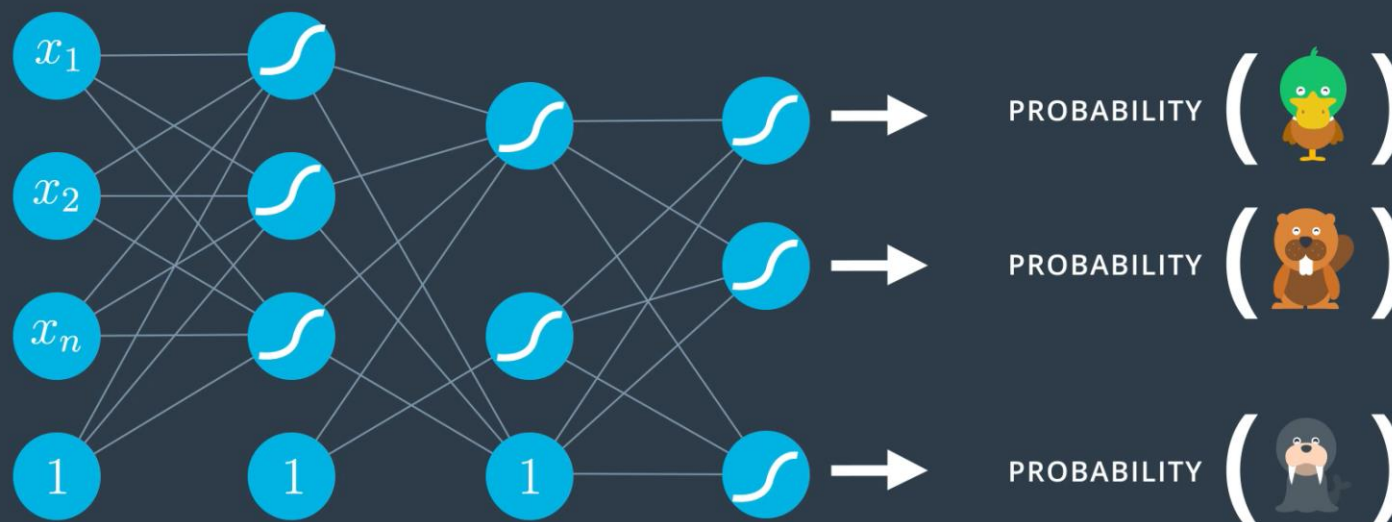
1. A random observation,  $x_i$ , is selected
2. If  $x_i$  has a minimum of close neighbors, we consider it part of a cluster.
3. Step 2 is repeated recursively for all of  $x_i$ 's neighbors, then neighbors' neighbors etc... These are the cluster's core members.
4. Once Step 3 runs out of observations,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

## Neural Network





“

# Classificação de Texto

end-to-end

# Embedding / NLP




★★★★★  
 **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

|          |   | tamanho do vocabulário |
|----------|---|------------------------|
| a        | → | 10000000000000000000   |
| abafar   | → | 01000000000000000000   |
| acampar  | → | 00100000000000000000   |
| acordo   | → | 00010000000000000000   |
| adeus    | → | 00001000000000000000   |
| aéreo    | → | 00000100000000000000   |
| afiado   | → | 00000010000000000000   |
| água     | → | 00000001000000000000   |
| águia    | → | 00000000100000000000   |
| .        |   | .                      |
| .        |   | .                      |
| .        |   | .                      |
| zimbábue | → | 00000000000000000001   |

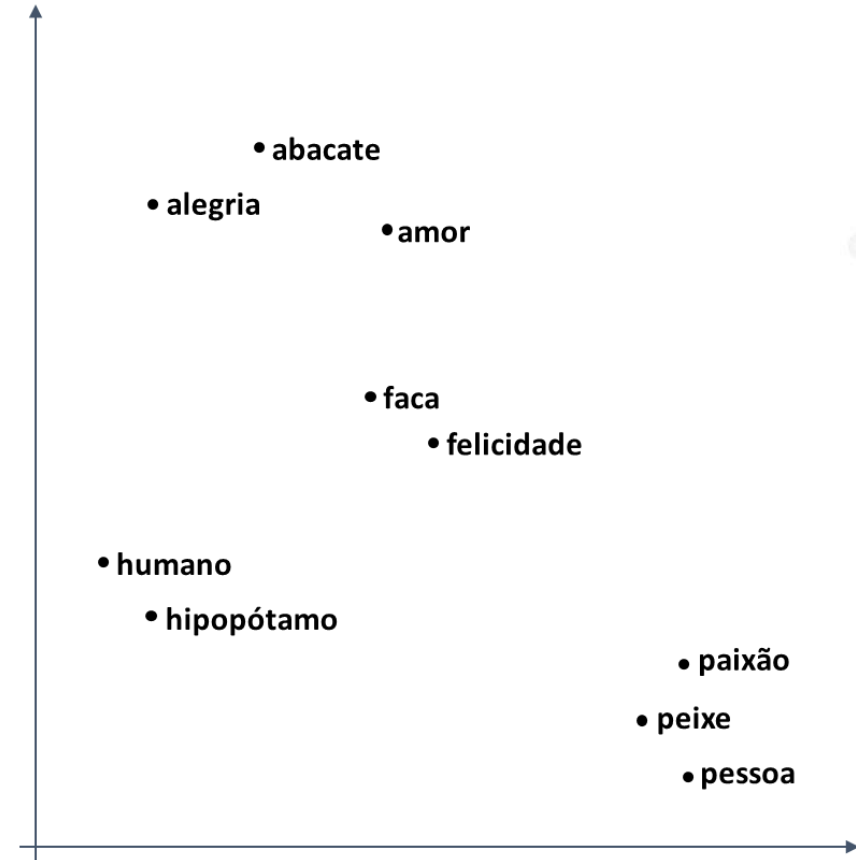


# Embedding / NLP

## Bag of Words

Representação muito grande e esparsa

Não há relação semântica entre palavras



# Embedding / NLP

## Word Embedding

Aprender a representação das palavras

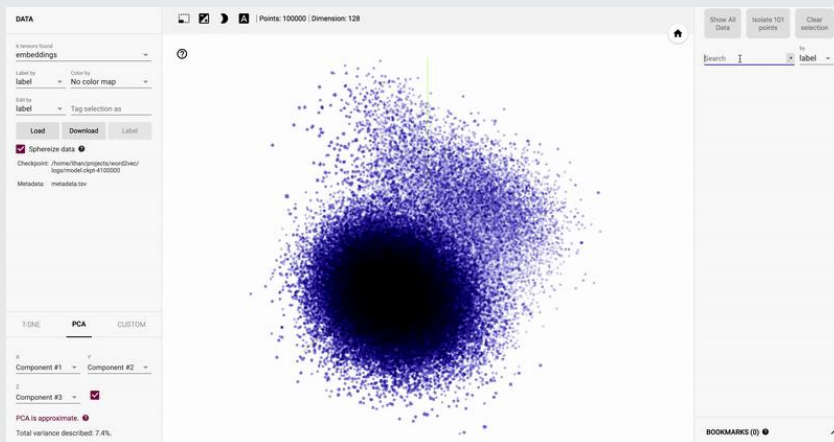
|          |   | tamanho do embedding |       |       |       |       |
|----------|---|----------------------|-------|-------|-------|-------|
|          |   |                      |       |       |       |       |
| a        | → | .1234                | .4763 | .0948 | .9384 | .7363 |
| abafar   | → | .6463                | .2347 | .9434 | .3232 | .5832 |
| acampar  | → | .5546                | .7293 | .1039 | .3220 | .1342 |
| acordo   | → | .9482                | .3293 | .4201 | .4403 | .8494 |
| adeus    | → | .9584                | .7564 | .9924 | .2134 | .1123 |
| aéreo    | → | .1111                | .4949 | .2847 | .2856 | .1345 |
| afiado   | → | .6754                | .4342 | .9683 | .3219 | .1103 |
| água     | → | .9831                | .3456 | .1234 | .9873 | .3435 |
| águia    | → | .4426                | .8464 | .9868 | .4458 | .3244 |
| .        |   | .                    | .     | .     | .     | .     |
| .        |   | .                    | .     | .     | .     | .     |
| .        |   | .                    | .     | .     | .     | .     |
| zimbábue | → | .9857                | .7746 | .2234 | .3342 | .9948 |

# Embedding / NLP

## Word Embedding

Representação densa

Palavras com significado similar próximas



- alegria
- felicidade
- amor
- paixão

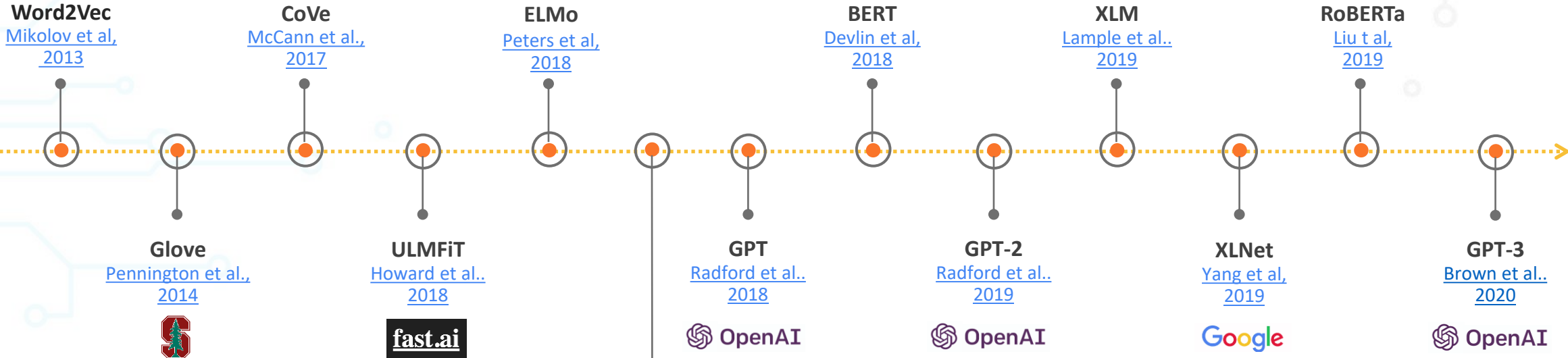
- pessoa
- humano

- hipopótamo
- peixe

- abacate

- faca

# Embedding / NLP



Google

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

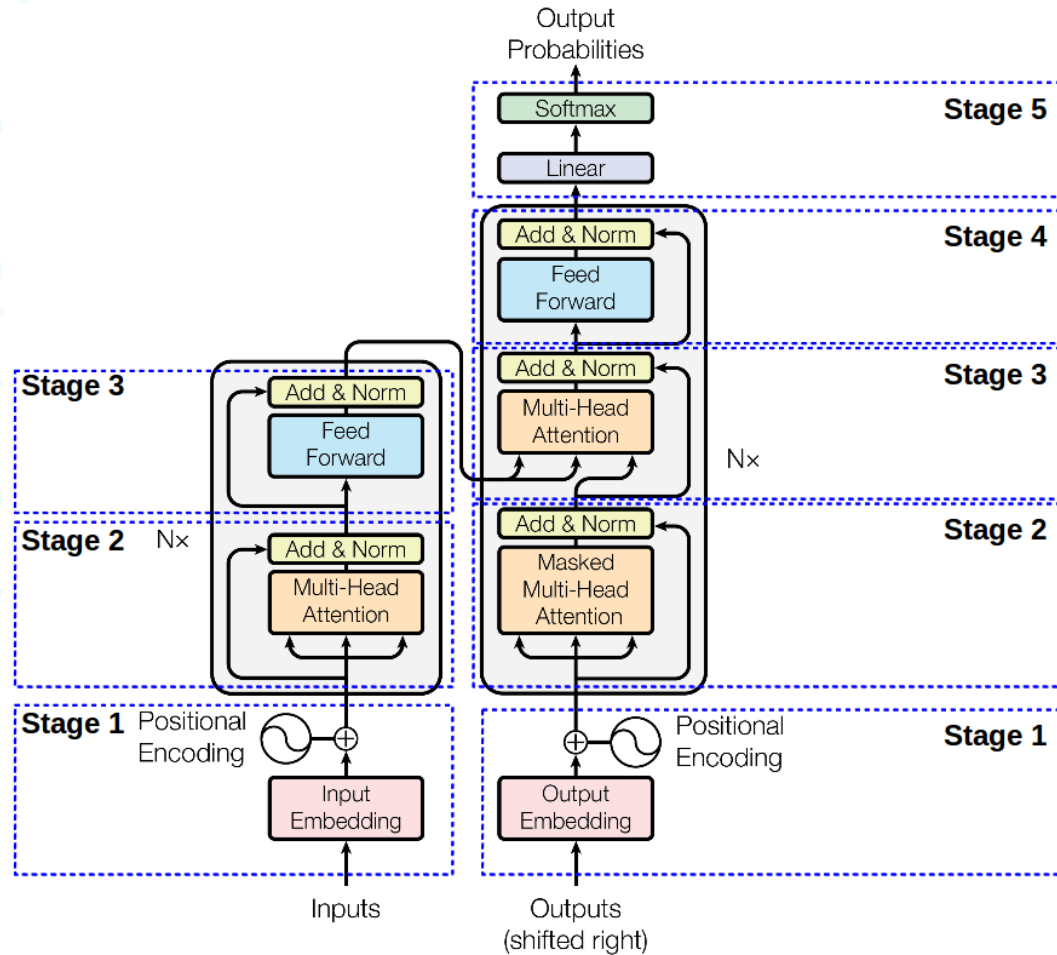
Llion Jones\*  
Google Research  
llion@google.com

Aidan N. Gomez\*<sup>†</sup>  
University of Toronto  
aidan@cs.toronto.edu

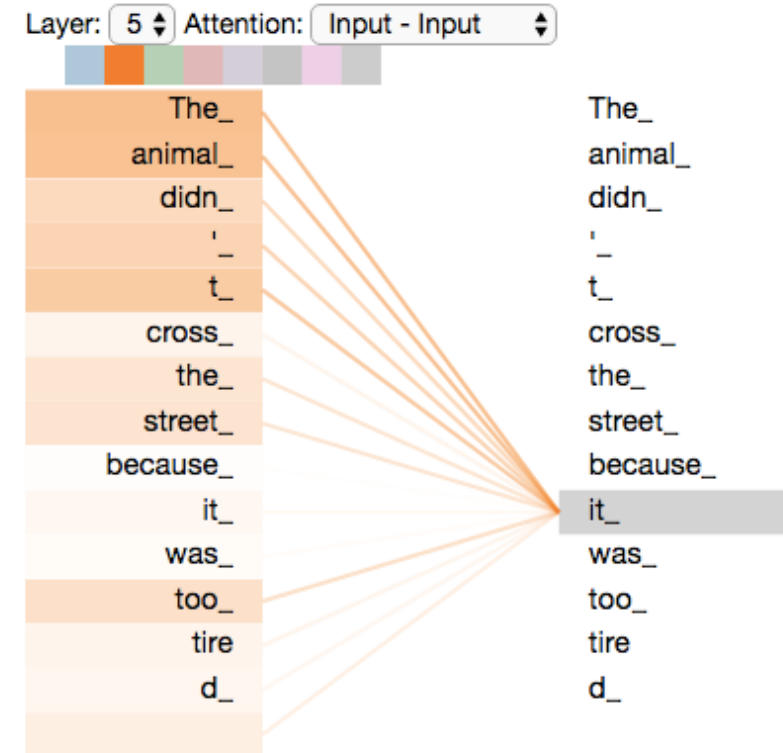
Lukasz Kaiser\*  
Google Brain  
lukaszkaiser@google.com

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

# NLP



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



# Embedding / NLP

qual o valor da minha fratura

All Shopping News Images Videos More Settings Tools

About 1,770,000 results (0.52 seconds)

www.doctoralia.com.br › em-media-... · [Translate this page](#)

## Em media quanto custa a cirurgia de correção da tibia? Boa ...

é normal essa dor e inchaço? Você quer enviar sua pergunta? Nossos especialistas responderam a 481 perguntas sobre Fraturas Da Tíbia.

diegoariel.com.br › fratura-da-tibia · [Translate this page](#)

## Fratura da Tíbia - Dr. Diego Ariel

Basicamente são 5 causas principais de **fratura** da tíbia: Quedas (própria altura, escada ou lugares altos); Lesões ... Após a cirurgia, em quanto tempo retorno às **minhas** atividades? ... Eu gostaria de saber o **preço** da cirurgia tíbia fibula ...

www.scielo.br › pdf › rbgg · PDF

## O Custo Direto da fratura de fêmur por quedas em ... - SciELO

by ÂBM Arndt · 2011 · Cited by 37 — 17 O tratamento da **fratura** tem por finalidade restaurar a anatomia óssea, a função do membro e reabilitar o paciente efetivamente, devolvendo sua ...

www.aig.com.br › brazil › documents › brochure · PDF

## Fratura de Ossos - CONDIÇÕES GERAIS

Indenização desta cláusula, aplicados sobre o **valor** do capital segurado ... Contratuais. A **fratura** de ossos, deve estar relacionada exclusivamente à ocorrência de ... seguro, ainda que sua manifestação ocorra durante a vigência do seguro;

qual o valor da minha fratura do cartão de credito

All News Shopping Videos Images More Settings Tools

About 5,570,000 results (0.88 seconds)

## Showing results for qual o valor da minha **fatura** do cartão de credito

Search instead for qual o valor da minha fratura do cartão de credito

pagoquandopuder.com.br › fatura-ca... · [Translate this page](#)

## [GUIA COMPLETO] Entenda A Fatura Do Seu Cartão De Crédito

Uma opção é escolher o parcelamento proposto pelo banco. Funciona assim: o **valor** da **fatura** atual é parcelado em algumas vezes (à sua escolha) e você paga ...

## People also ask

Quanto é a fatura do cartão de crédito?

O que acontece se não pagar o valor total da fatura do cartão de crédito?

Como faço para ver a fatura do cartão de crédito?

Como funciona o pagamento de fatura de cartão de crédito?

Feedback

jurosbaixos.com.br › conteudo › apr... · [Translate this page](#)

## Aprenda a ler a sua Fatura do Cartão de Crédito - Juros Baixos

Total da **Fatura**. É o **valor** total de todas as compras realizadas no **cartão de crédito**, no período da data de fechamento do cartão. Em suma, é o ...




# Embedding / NLP

Google


me mostre fotos de raças cachorro exceto beagles

Tudo Imagens Compras Vídeos Notícias Mais Definições Ferramentas


filhote raças caninas




Beagle (Beagle Inglês ou famoso Snoopy...  
cachorrogato.com.br




Beagle – Raças Caninas - Raças de Cac...  
petvale.com.br




476 × 238  
Beagle (Beagle Inglês ou famoso Snoopy) - Raças de...  
cachorrogato.com.br




Beagle | Raças de cães | Royal Canin  
royalcanin.pt



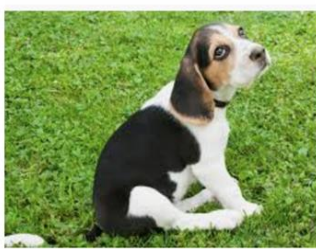
Beagle - Blog do Cachorro  
blogdocachorro.com.br




Nós, os Cachorros - N...  
nososcachorros.blogspot...




Beagle: O cachorro cantor | Au au au!!!  
caninablog.wordpress.com





Beagle: tudo sobre a raça em um guia co...  
arbolez.com





Raças: Beagle | BitCã  
bitcao.com.br














“

# Classificação de Texto

end-to-end

Website

templates

App.py

Requeriments.txt

Sagemaker

SageMaker Project.ipynb

Train

Model.py

train.py

Requeriments.txt

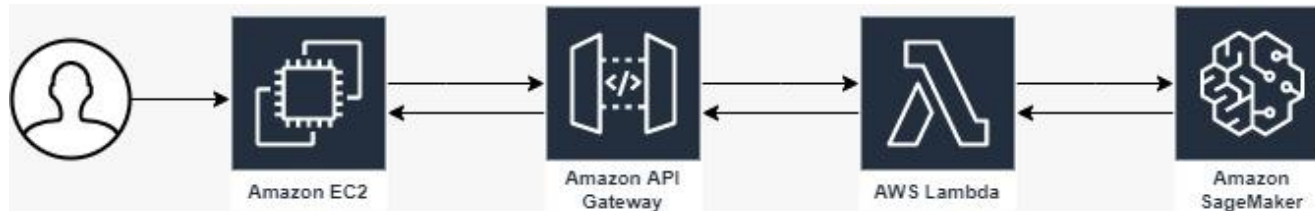
Serve

Model.py

Predict.py

Utils.py

Requeriments.txt



# Embedding / NLP

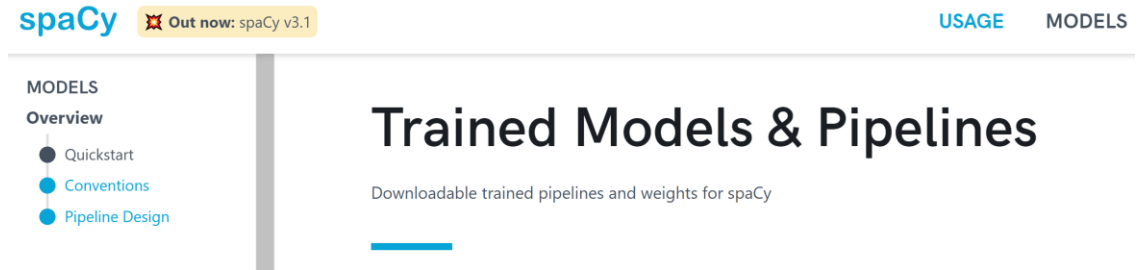
<https://www.nltk.org/>

NLTK 3.6.2 documentation

[NEXT](#) | [MODULES](#) | [INDEX](#)

Natural Language Toolkit

<https://spacy.io/models>



spaCy Out now: spaCy v3.1 USAGE MODELS

MODELS  
Overview

- Quickstart
- Conventions
- Pipeline Design

## Trained Models & Pipelines

Downloadable trained pipelines and weights for spaCy

[https://www.tensorflow.org/text/tutorials/classify\\_text\\_with\\_bert](https://www.tensorflow.org/text/tutorials/classify_text_with_bert)

[https://www.tensorflow.org/text/tutorials/classify\\_text\\_with\\_bert](https://www.tensorflow.org/text/tutorials/classify_text_with_bert)

TensorFlow > Resources > Text > Tutorials

Was this helpful  

## Text classification with an RNN



Run in Google Colab



View source on GitHub



Download notebook

This text classification tutorial trains a [recurrent neural network](#) on the [IMDB large movie review dataset](#) for sentiment analysis.

TensorFlow > Resources > Text > Tutorials

Rate and review  

## Classify text with BERT



Run in Google Colab



View on GitHub



Download notebook



See TF Hub model

This tutorial contains complete code to fine-tune BERT to perform sentiment analysis on a dataset of plain-text IMDB movie reviews. In addition to training a model, you will learn how to preprocess text into an appropriate format.

# Brindes



# Thanks !



Vinicius Fernandes Caridá

vfcarida@gmail.com



@vinicius caridá



@vfcarida



@vinicius caridá



@vfcarida



@vinicius caridá



@vfcarida