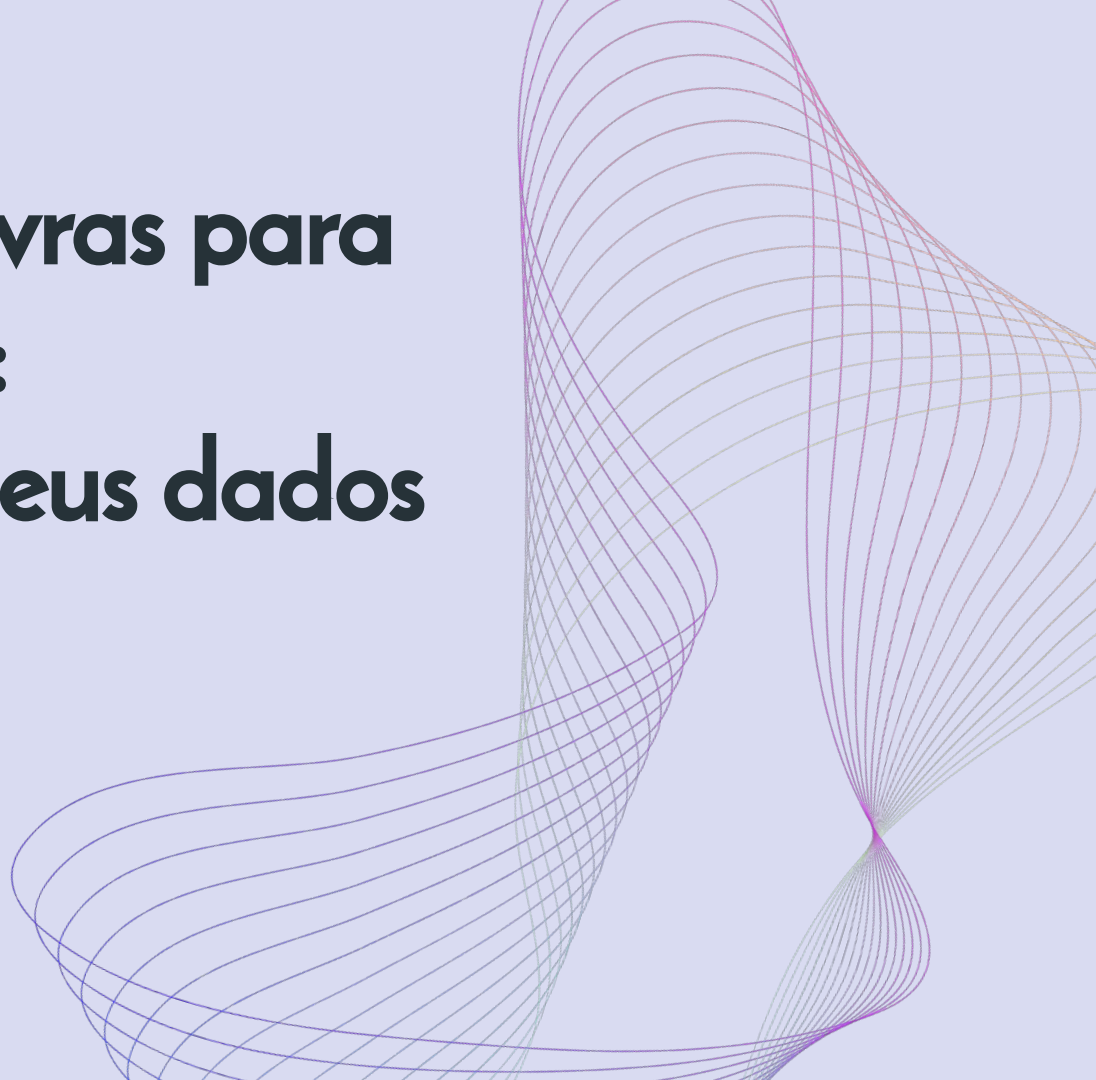


# **Explicando palavras para um computador: Como modelar seus dados**





Olá, meu nome é Giulia Falcão.

Atualmente trabalho como Engenheira de Software no CESAR, mas estou buscando fazer transição de área para Data Science/Engenheira Machine Learning.

Sou formada em Ciência da computação pela Universidade Católica de Pernambuco, tendo ainda recebido a láurea acadêmica de minha turma. Minha carreira acadêmica consistiu em processamento de linguagem natural e manipulação de dados desbalanceados.

Eu sou uma grande viciada banoffeee, pedalo e faço academia pra compensar o tanto que como, também viciada em café ao ponto de ser uma barista em formação.

# Conteúdo da palestra

## BAG OF WORDS

O que é Bag Of Words, como fazê-lo e suas problemáticas

01

## WORD EMBEDDING

O que é Word Embedding, como fazê-lo e suas problemáticas

02

## LONG SHORT-TERM MEMORY (LSTM)

O que é LSTM como fazê-lo e suas problemáticas

03

## BERT

O que é BERT como fazê-lo e suas problemáticas

04

# INTRODUÇÃO

- **O que é modelar dados?**
  - Estruturar os dados de uma maneira que sejam utilizados pelos algoritmos de aprendizado de máquina
- **Qual a problemática de texto?**
  - Algoritmos de aprendizado de máquina não podem funcionar com texto bruto diretamente; o texto deve ser convertido em números. Especificamente, vetores de números.
- **Qual o ponto de relevância com relação a modelagem de dados?**
  - Modelagem de dados é importante porque pode afetar diretamente na performance do algoritmo de aprendizado de máquina



# Bag of Words

01



# BAG OF WORDS

01. O que é Bag of Words?

02. Como realizar o Bag of Words?

02.1. CountVectorizer

02.1.1 Problemática

02.2. TF-IDF

02.2.1 Problemática

03. Problemática do Bag of Words?



# Word Embedding

02



# WORD EMBEDDING

01. O que é Word Embedding?

02. Como funciona o Word Embedding?

03. Problemática do Word Embedding?

02.1. Word2Vec

02.1.1 CBOW

([palavras de contexto],  
palavra alvo)

02.1.2 Skip-gram

(palavra central,  
[palavras de contexto])





# LONG SHORT-TERM MEMORY (LSTM)

03



# LSTM

01. O que é o LSTM?

02. Como funciona o LSTM?

03. Problemática do LSTM?



# BERT

# 04



# BERT

01. O que é o BERT?

02. Como funciona o BERT?

03. Problemática do BERT?

# OBRIGADA!

Perguntas?

Redes Sociais:



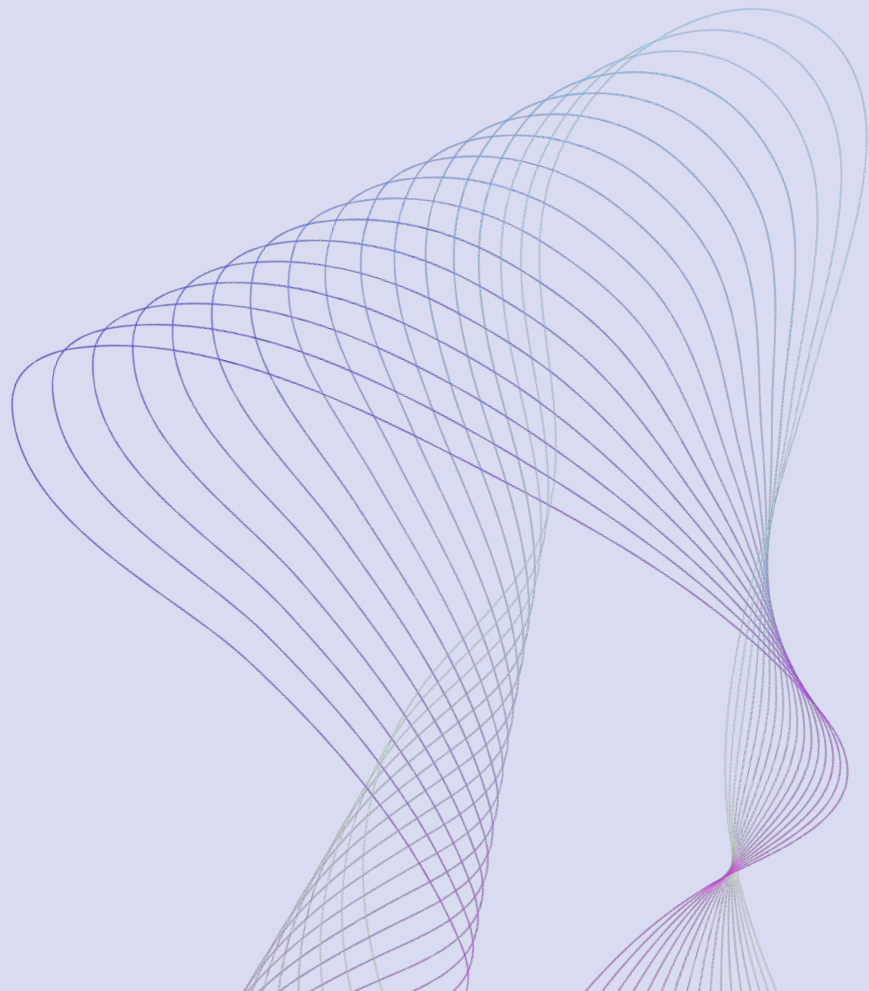
@giuufalcao



Giulia Falcão



@giufalcao



# REFERÊNCIAS

Did you like the resources on this template? Get them for free at our other websites.

## VECTORS

- Sound wave music poster template

## PHOTOS

- Close-up of beautiful young woman suffering from headache
- Portrait of a woman suffering from headache
- Portrait of a beautiful woman suffering from headache
- Senior man wearing eyeglasses with white bluetooth earphone in his ear
- Handsome man covering his ear with hand standing in front of modern building
- Portrait of a distraught girl covering her ears with two hands
- Portrait of a sick woman sitting on sofa with glass of water
- Medical team in a doctor's office

# Instructions for use

In order to use this template, you must credit [Slidesgo](#) by keeping the **Thanks** slide.

## You are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

## You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

# Fonts & colors used

This presentation has been made using the following fonts:

## **Reem Kufi**

(<https://fonts.google.com/specimen/Reem+Kufi>)

## **Montserrat**

(<https://fonts.google.com/specimen/Montserrat>)

#ffffff

#000000

#d9dbf1

#a7a3e2

#f3f4fa



# Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out [How it Works](#).



Pana



Amico



Bro



Rafiki



Cuate

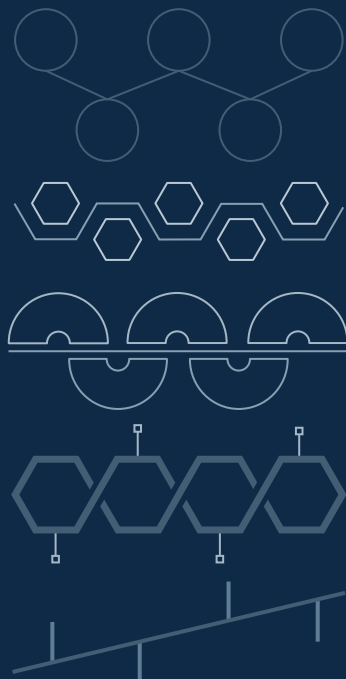
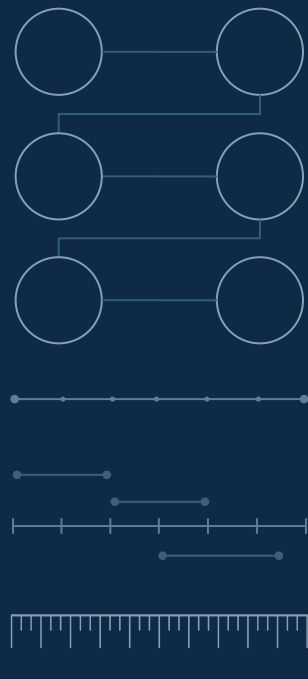
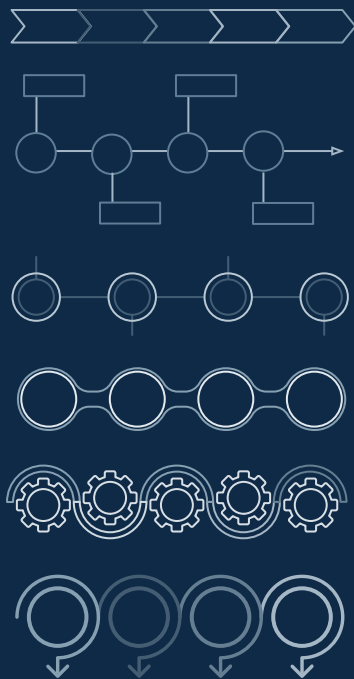
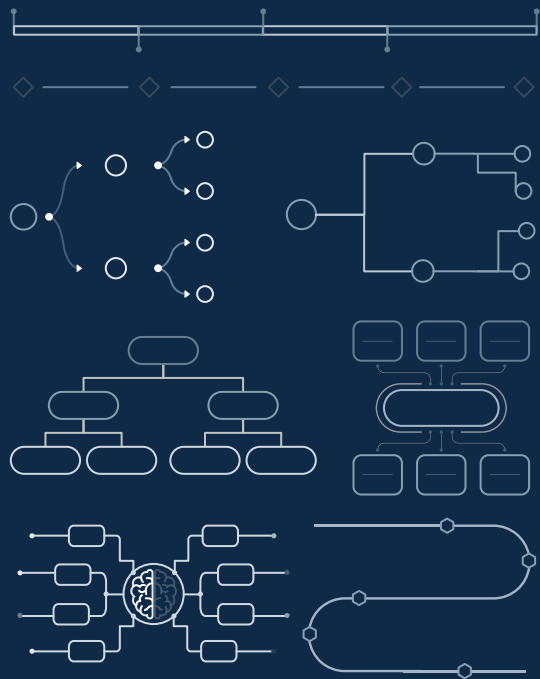
# Use our editable graphic resources...

You can easily [resize](#) these resources without losing quality. To [change the color](#), just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want.

Group the resource again when you're done. You can also look for more [infographics](#) on [Slidesgo](#).



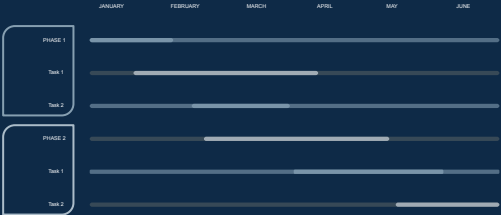




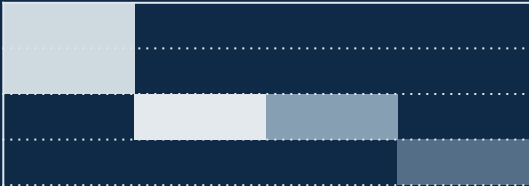
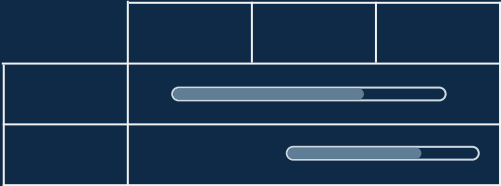
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |
|--|--|
|  |  |
|  |  |
|  |  |

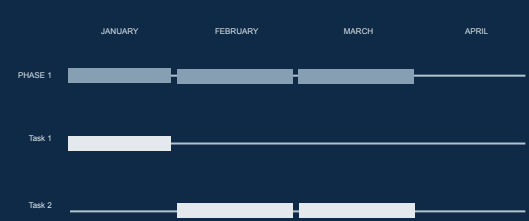
|  |  |  |  |  |
|--|--|--|--|--|
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |



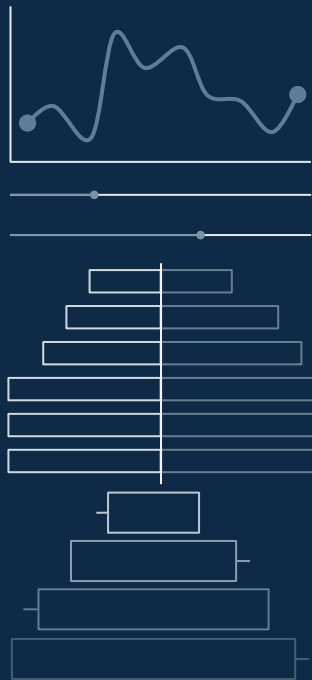
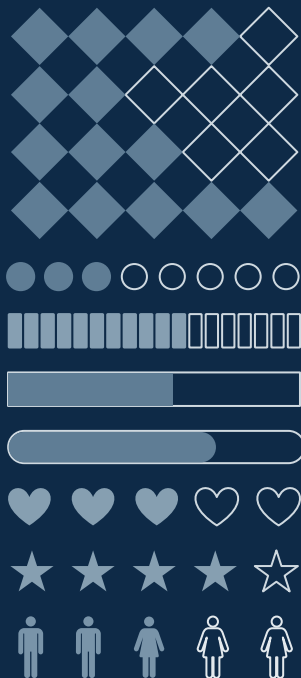
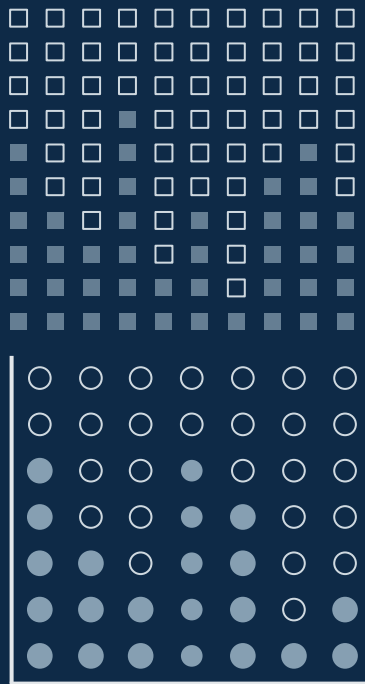
|  |  |  |  |
|--|--|--|--|
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |



|  |  |  |  |  |
|--|--|--|--|--|
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |







## ...and our sets of editable icons

You can resize these icons without losing quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.





## Educational Icons

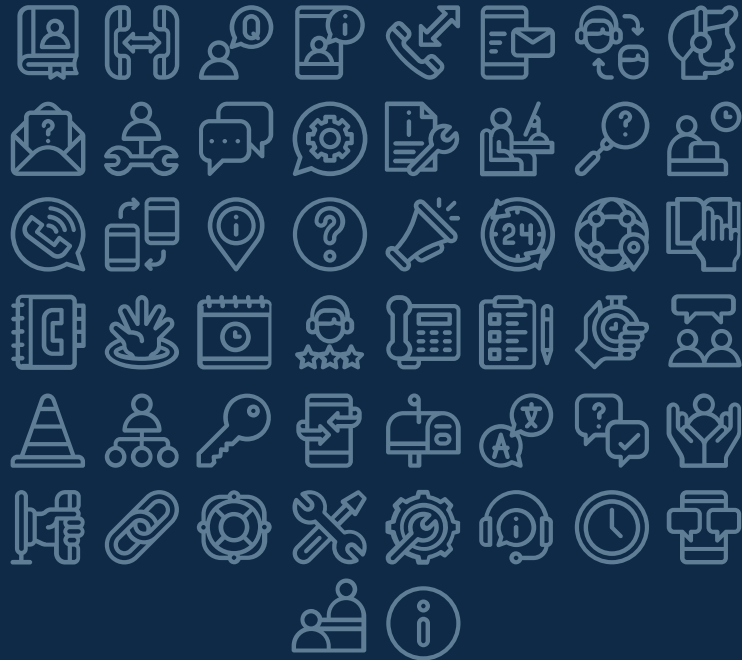


## Medical Icons





## Help & Support Icons



## Avatar Icons



## Creative Process Icons



## Performing Arts Icons



## Nature Icons



# SEO & Marketing Icons



