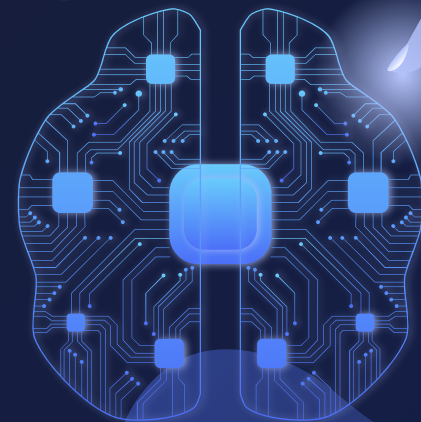


IA

Explicável:

Por trás dos panos da
Inteligência Artificial

Henrique Margotte



Quem sou eu?

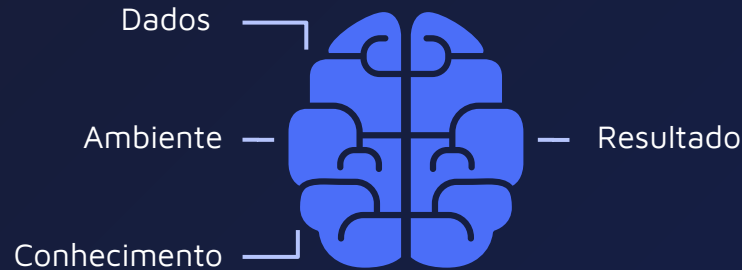
- Mestrando na linha de Inteligência Computacional pela UFPR
 - Sob orientação das profas. Aurora e Carmem
 - Laboratório de Computação Bioinspirada (C-Bio)
- Estagiário de IA no Estúdio de Inovação do TCE-PR
- Bacharel em Ciência da Computação pela UFPR
- Técnico em Programação de Jogos Digitais pelo IFPR
- Ex-petiano do grupo PET Computação UFPR
- Minhas pesquisas:
 - IC/TCC: Projeto Caravelas - RESMA - PrInt
 - Mestrado: XAI para caracterização de clusters



O que é
XAI?

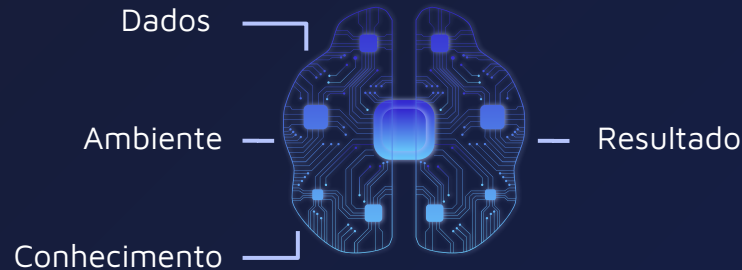
O que é Inteligência Artificial?

- Área da Computação em ascensão
- Tenta simular, através de modelos matemáticos, o funcionamento da cognição humana
- Lógica, aprendizado, reconhecimento de padrões, inferência, etc.

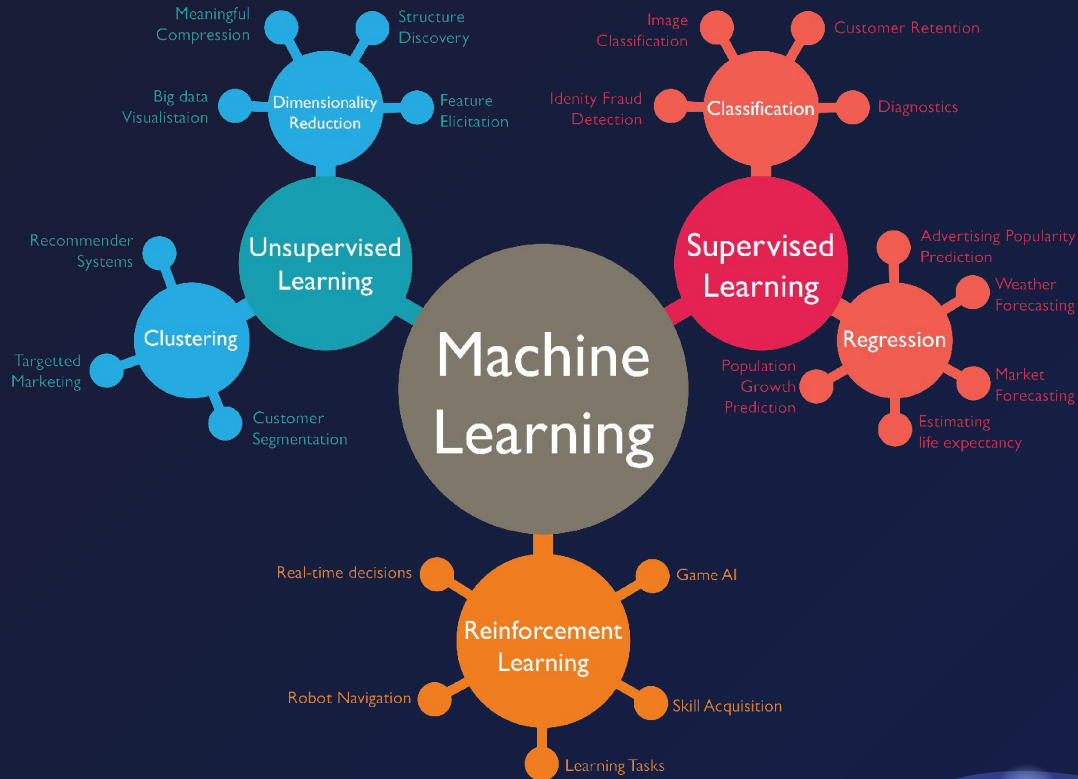


O que é Inteligência Artificial?

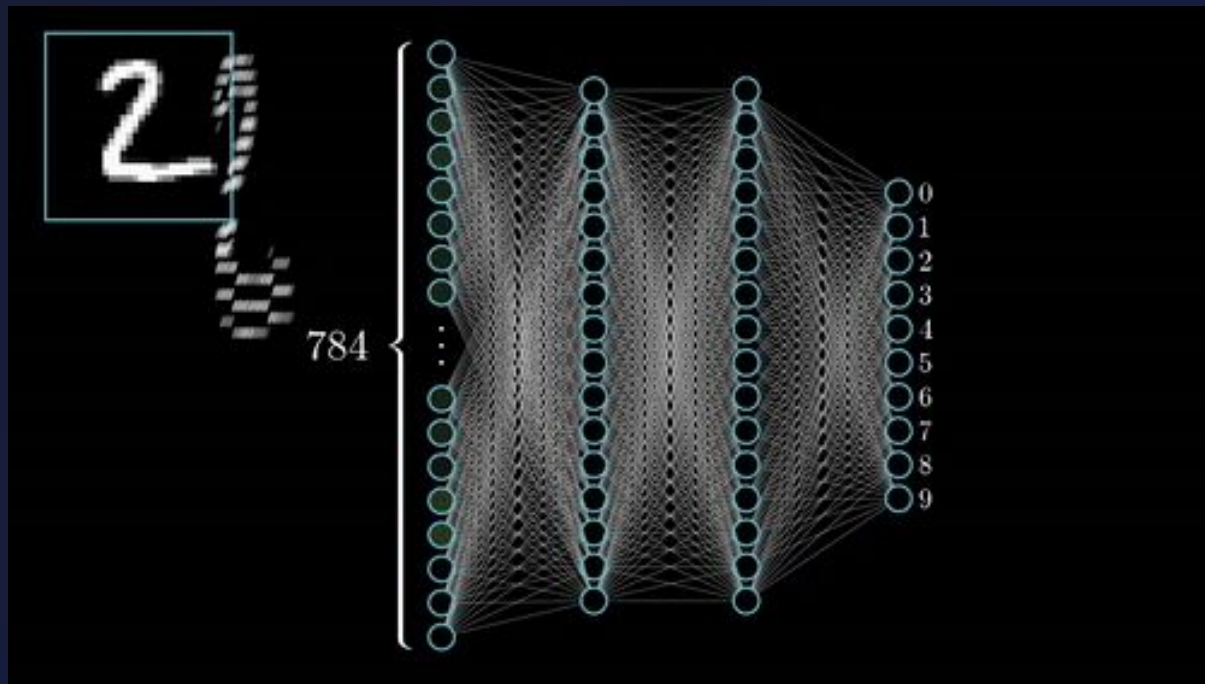
- Área da Computação em ascensão
- Tenta simular, através de modelos matemáticos, o funcionamento da cognição humana
- Lógica, aprendizado, reconhecimento de padrões, inferência, etc.
- **Mas como entender o que o modelo matemático está fazendo?!**



O que é Inteligência Artificial?



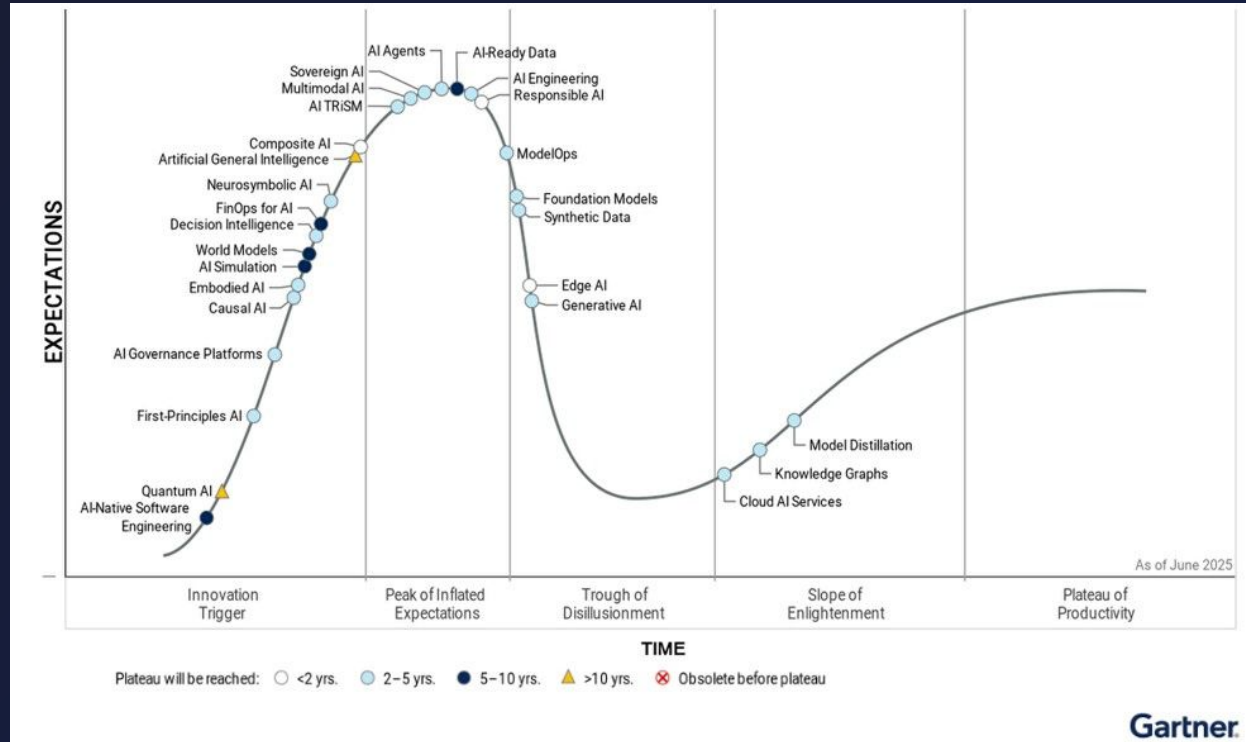
O que é Inteligência Artificial?

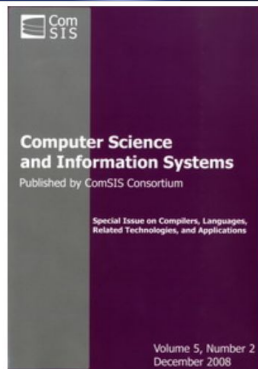


Por que XAI?

- **Confiança:** Saber se o que a IA está fazendo faz sentido ou não.
- **Justiça (Fairness):** Descobrir vieses nos dados e modelos.
- **Interpretabilidade:** Saber como a IA opera.
- **Explicabilidade:** Explicar porquê um modelo teve esse resultado.
- **Descoberta de Conhecimento:** Aprender com o que a IA aprendeu.

Por que entender a IA?





About the journal
Editorial policy
Instructions for authors



Computer Science and Information Systems 2024 Volume 21, Issue 4, Pages: 1887-1911
<https://doi.org/10.2298/CSIS240301064>

Full text (PDF) 2653 KB

Identifying Occurrences of the Cnidarian *Physalia physalis* in Social Media Data

Rocha Heloisa F. (Universidade Federal do Paraná, Departamento de Informática, Curitiba-PR, Brazil),
heloisarocha@ufpr.br

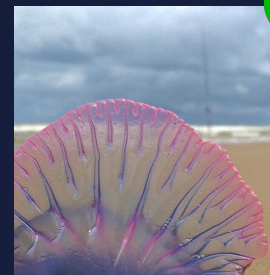
Nascimento Lorena S. (Universidade Federal do Paraná, Centro de Estudos do Mar, Pontal do Paraná PR, Brazil),
lorena.sn@ufpr.br

Camargo Leonardo (Universidade Federal do Paraná, Departamento de Informática, Curitiba-PR, Brazil),
camargo.s.leonardo@ufpr.br

Noernberg Mauricio A. (Universidade Federal do Paraná, Centro de Estudos do Mar, Pontal do Paraná PR, Brazil),
m.noernberg@ufpr.br

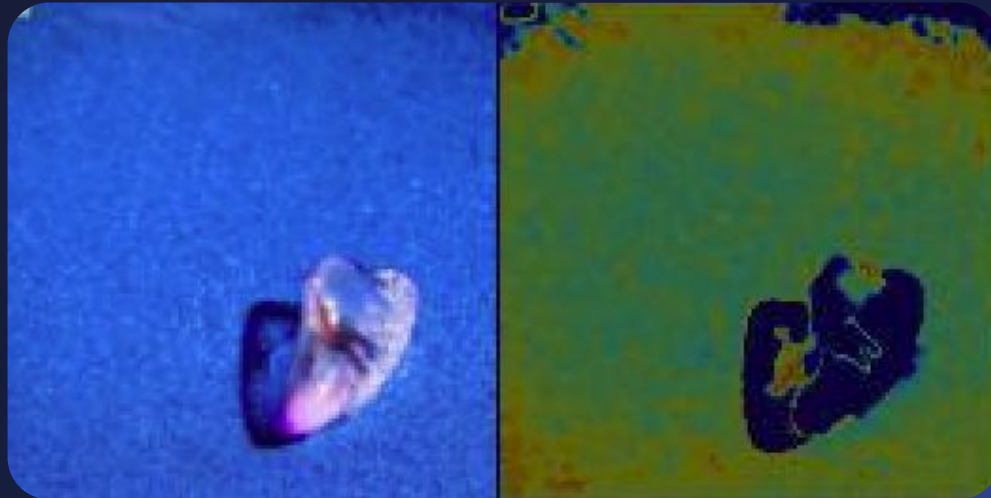
Ramirez Pozo Aurora T. (Universidade Federal do Paraná, Departamento de Informática, Curitiba-PR, Brazil),
aurora.pozo@ufpr.br

Hara Carmem S. (Universidade Federal do Paraná, Departamento de Informática, Curitiba-PR, Brazil),
carmhara@ufpr.br



Como explicar modelos de classificação
que usam AI?

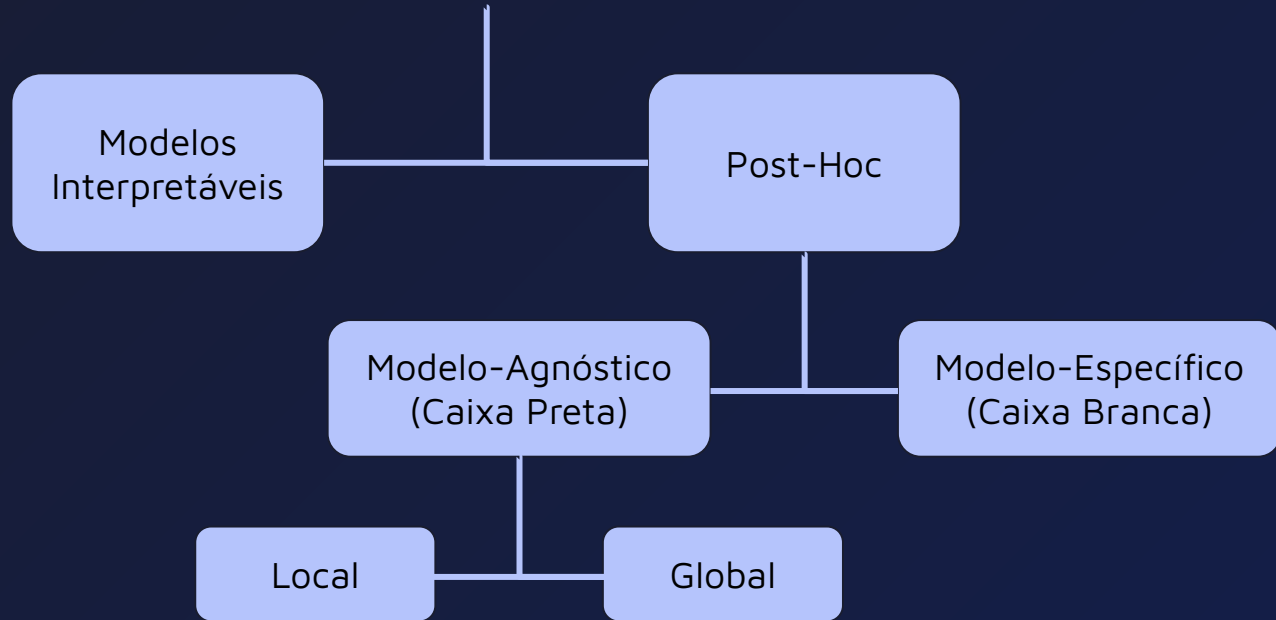
Explainable AI



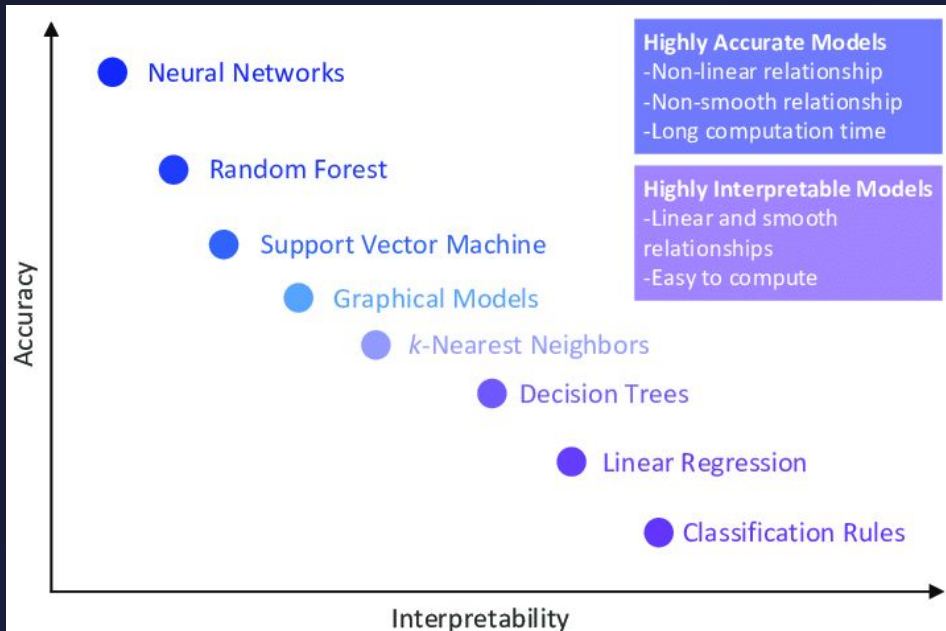
Conceitos principais

- **Interpretabilidade (*Interpretability*):**
 - O quanto uma pessoa consegue entender como uma decisão foi tomada.
- **Transparência (*Transparency*):**
 - O nível de acesso e abertura das estruturas internas e dados de um modelo.
- **Justiça (*Fairness*):**
 - A garantia que um sistema de IA não está enviesado, com base nos dados.
- **Explicabilidade (*Explainability*):**
 - Explica o porquê uma decisão foi tomada.

Métodos de XAI

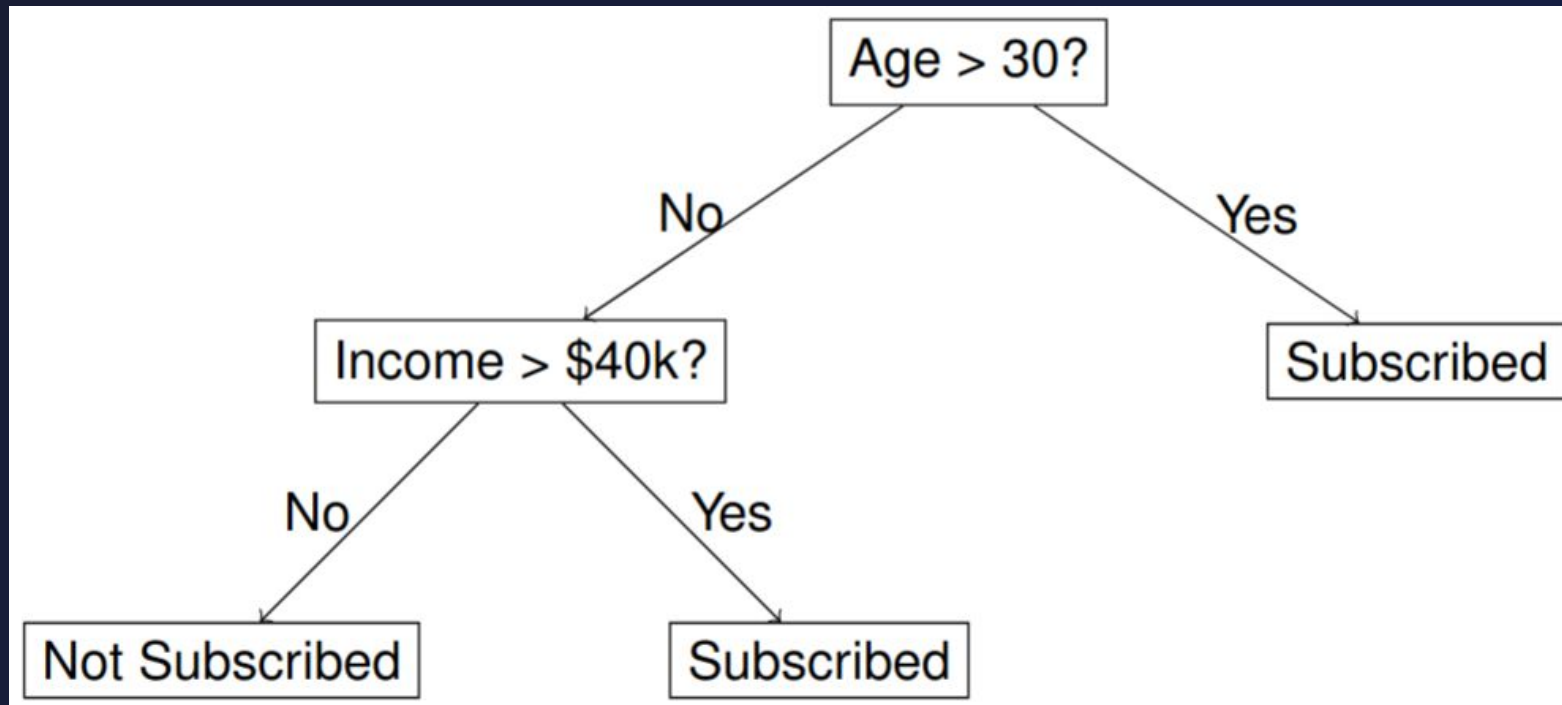


Modelos Interpretáveis

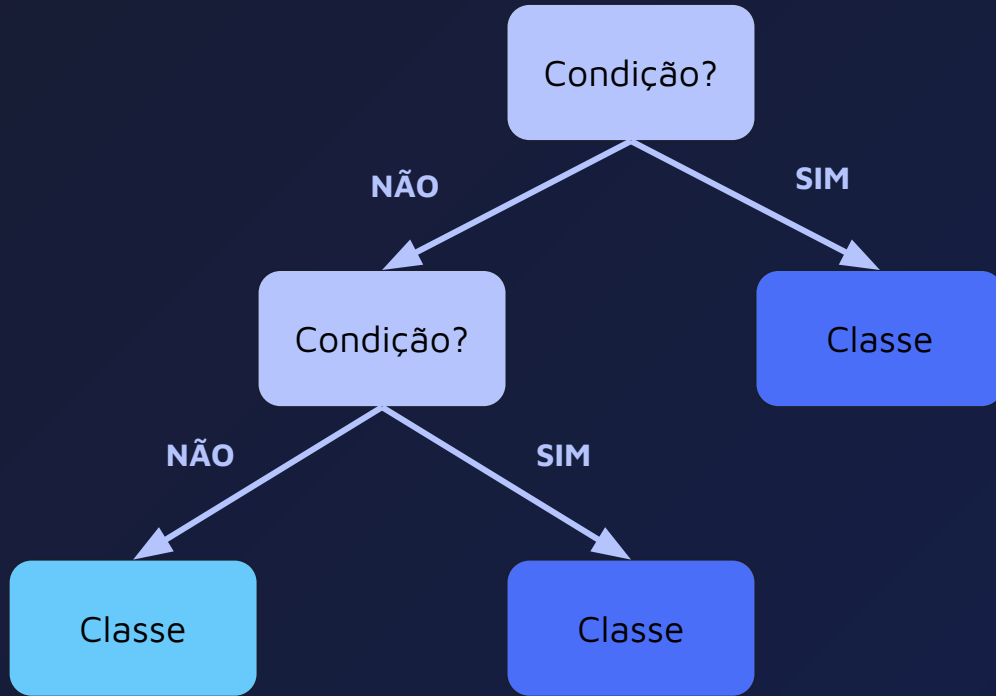


Morocho-Cayamcela, Manuel Eugenio & Lee, Haeyoung & Lim, Wansu. (2019). Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions. IEEE Access. 7. 137184-137206. 10.1109/ACCESS.2019.2942390.

Árvores de Decisão

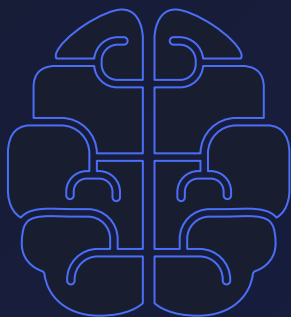


Árvores de Decisão

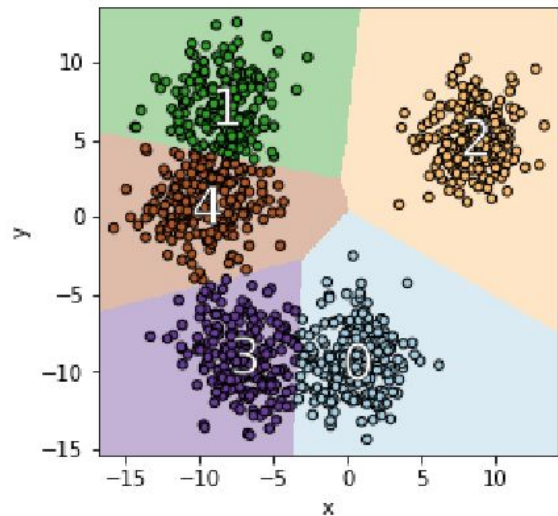


Métodos Post-Hoc

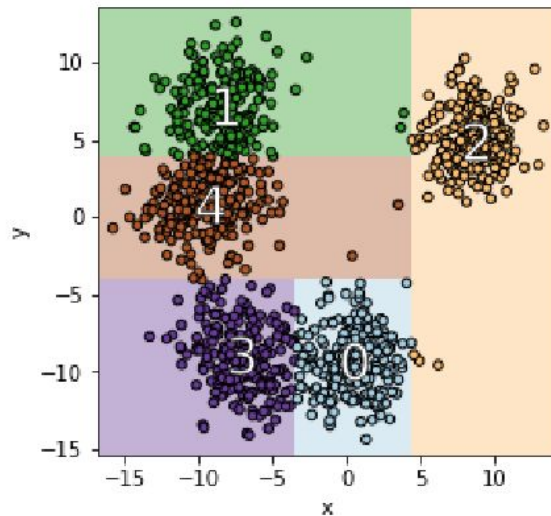
Modelo-Agnóstico
Global



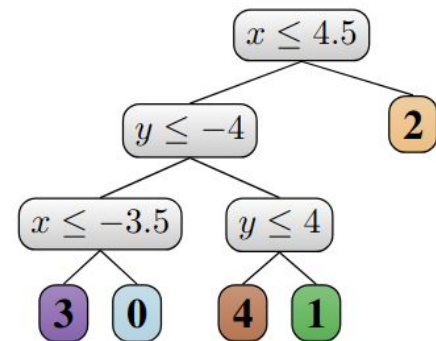
Modelos Substitutos



(a) Optimal 5-means clusters



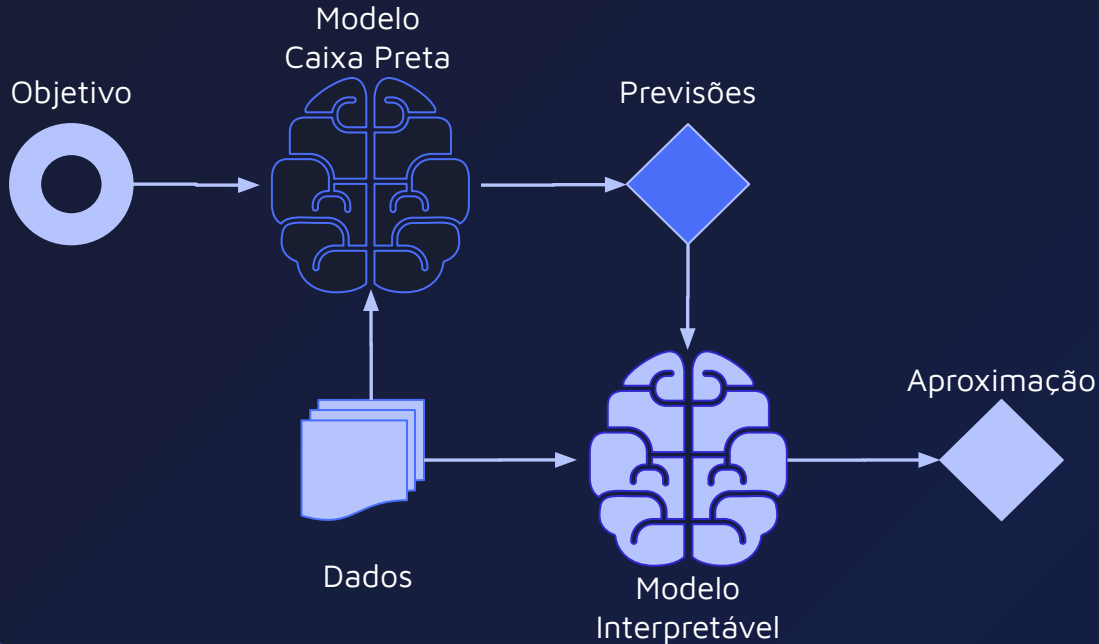
(b) Tree based 5-means clusters



(c) Threshold tree

Modelos Substitutos

- Treinar um modelo interpretável para aproximar um modelo mais complexo (caixa preta)
- O treinamento é feito a partir das previsões do modelo caixa preta, em vez das classes



Permutation Importance

- Faz a permutação aleatória de valores para cada atributo, medindo o impacto de cada um na precisão do modelo
- Identifica os atributos mais usados pelo modelo e os que prejudicam

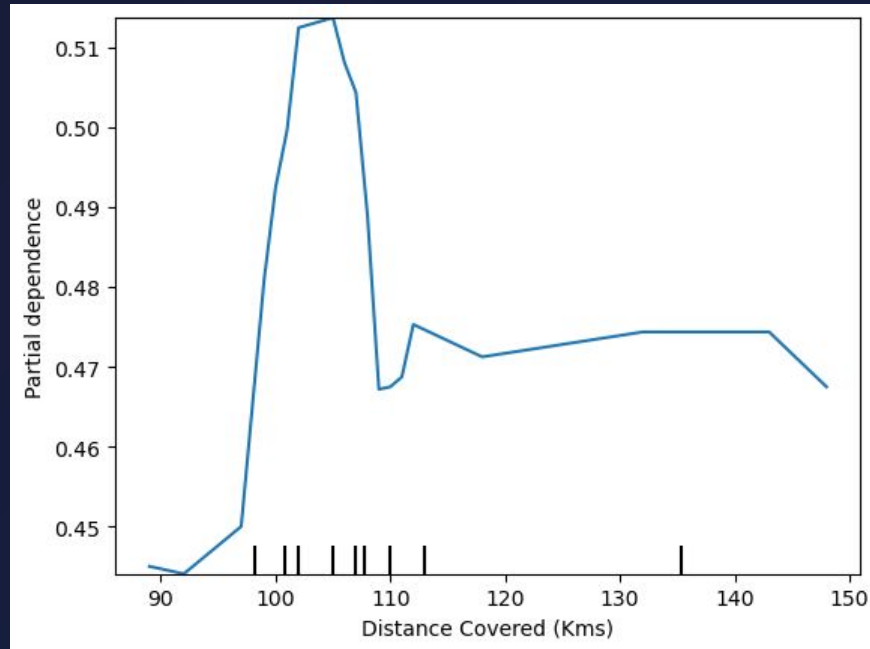
Height at age 20 (cm)	Height at age 10 (cm)	...	Socks owned at age 10
182	155	...	20
175	147	...	10
...	...	...	...
156	142	...	8
153	130	...	24



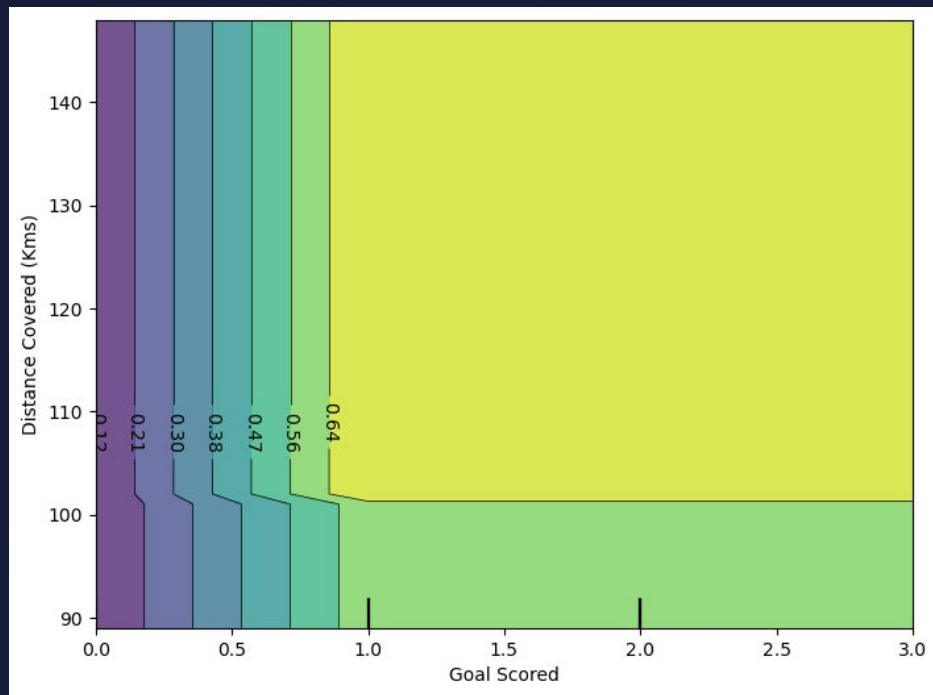
Weight	Feature
0.1750 ± 0.0848	Goal Scored
0.0500 ± 0.0637	Distance Covered (Kms)
0.0437 ± 0.0637	Yellow Card
0.0187 ± 0.0500	Off-Target
0.0187 ± 0.0637	Free Kicks
0.0187 ± 0.0637	Fouls Committed
0.0125 ± 0.0637	Pass Accuracy %
0.0125 ± 0.0306	Blocked
0.0063 ± 0.0612	Saves
0.0063 ± 0.0250	Ball Possession %
0 ± 0.0000	Red
0 ± 0.0000	Yellow & Red
0.0000 ± 0.0559	On-Target
-0.0063 ± 0.0729	Offsides
-0.0063 ± 0.0919	Corners
-0.0063 ± 0.0250	Goals in PSO
-0.0187 ± 0.0306	Attempts
-0.0500 ± 0.0637	Passes

Partial Dependence Plots (PDP)

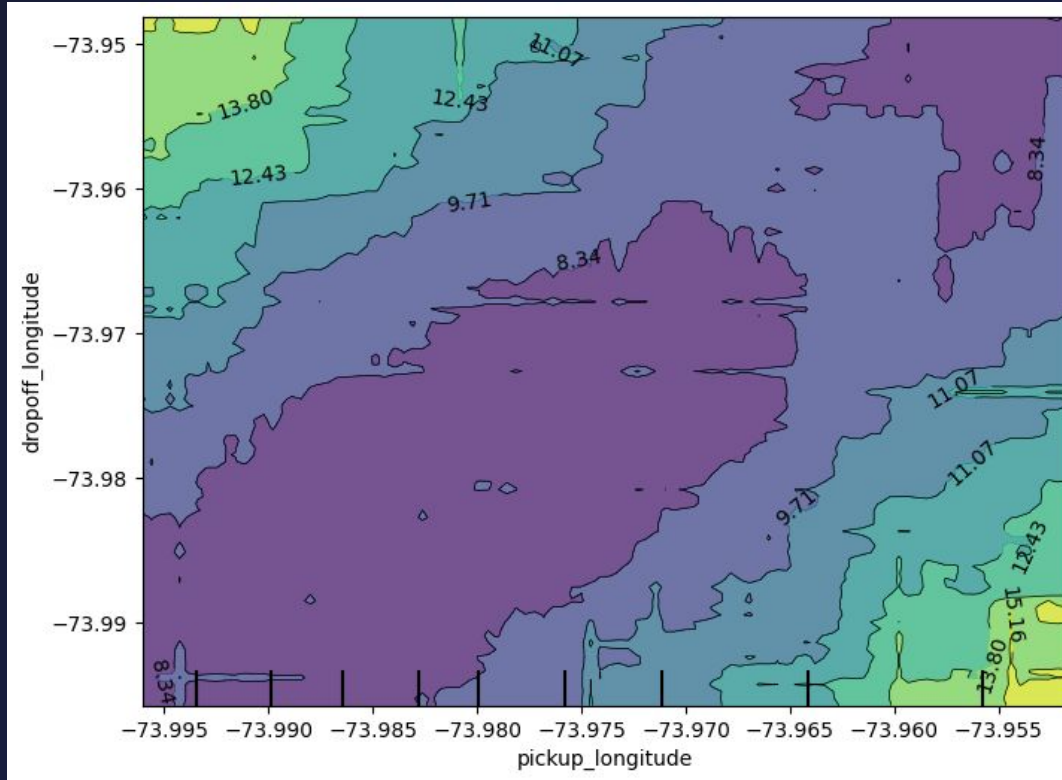
- Varia os valores de um único atributo em várias instâncias dos dados, medindo o impacto
- Gera gráficos que mostram o impacto de cada variável no modelo



Partial Dependence Plots (PDP)

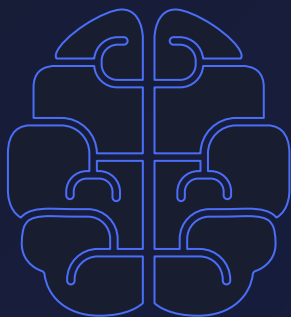


Partial Dependence Plots (PDP)



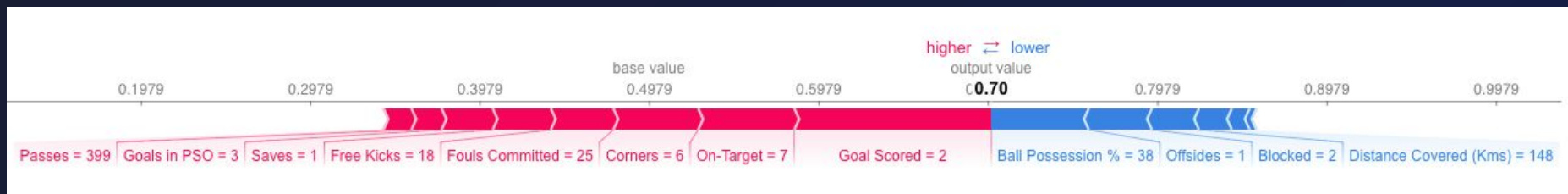
Métodos Post-Hoc

Modelo-Agnóstico
Local

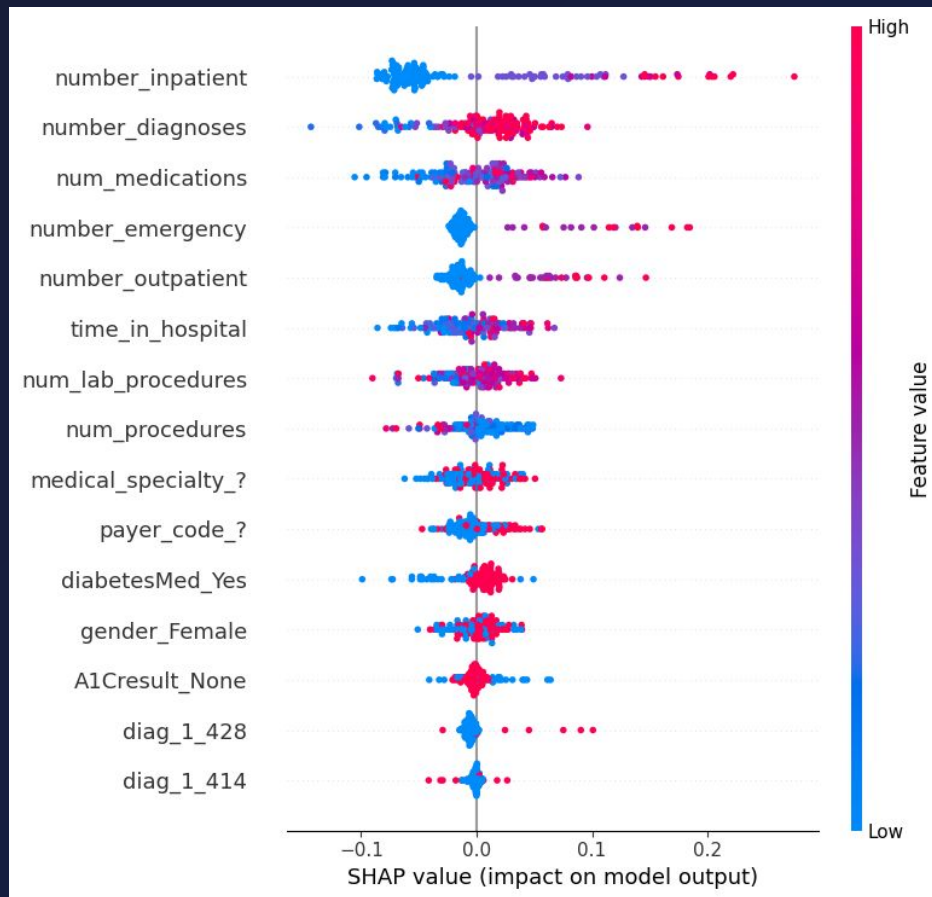


SHAP: Shapley Additive Explanations

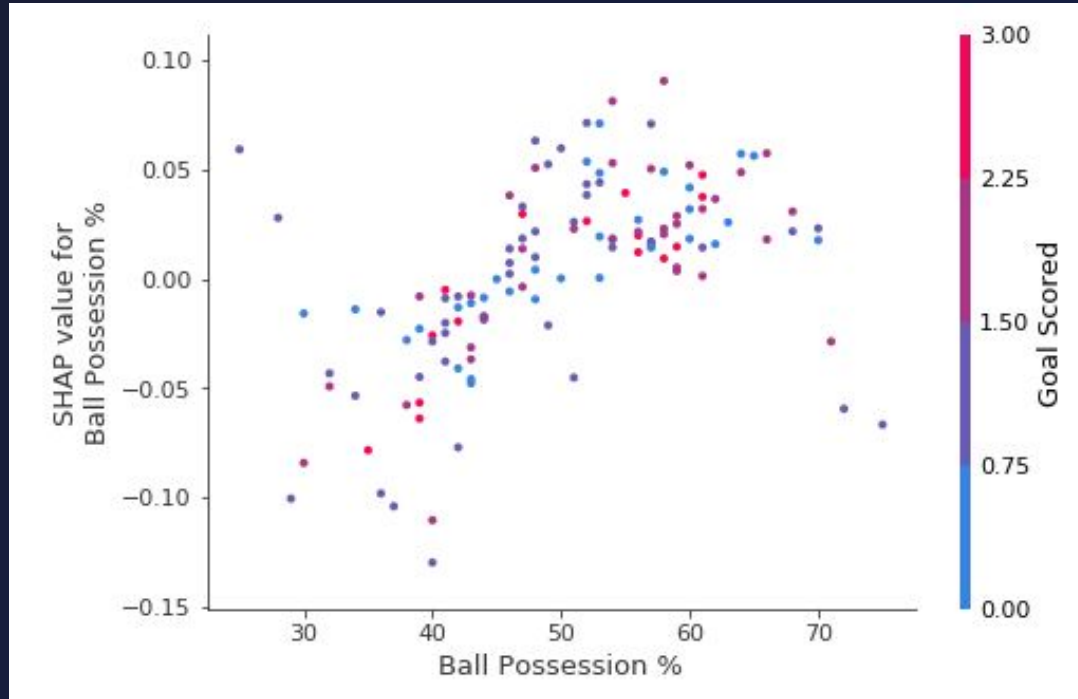
- Mede o impacto de um atributo comparando com um valor base
- Permite medir como cada atributo contribuiu para o resultado de cada instância (local)
- Se calculado para todos os elementos, pode gerar um panorama global
- Pode também ser usado para mostrar a dependência de atributos, associado com a contribuição

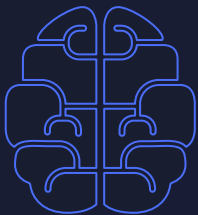


SHAP: Shapley Additive Explanations



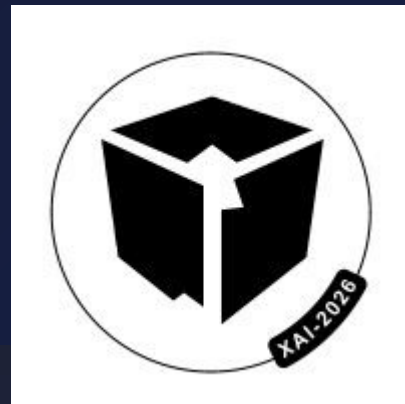
SHAP: Shapley Additive Explanations





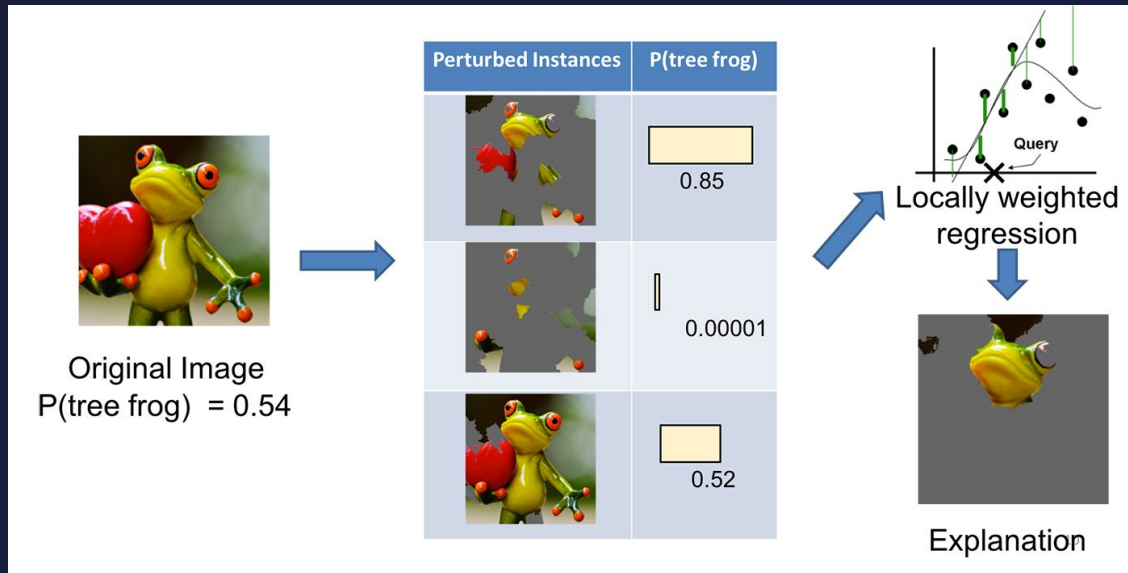
GPUPTextSHAP

- Implementação SHAP em GPU para texto
- 118x mais rápido em testes preliminares (10min em 5s)



LIME: Local Interpretable Model-Agnostic Explanations

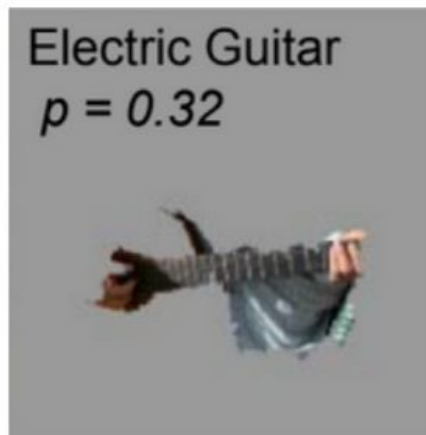
- Define um conjunto de dados locais através de perturbações na instância a ser explicada
- Utiliza de um modelo mais interpretável para aproximar o modelo nesse conjunto de dados



LIME: Local Interpretable Model-Agnostic Explanations



(a) Original image



(b) Explaining *electric guitar*



(c) Explaining *acoustic guitar*

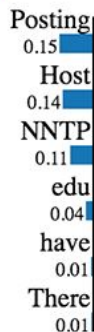
LIME: Local Interpretable Model-Agnostic Explanations

Prediction probabilities



atheism

christian



Text with highlighted words

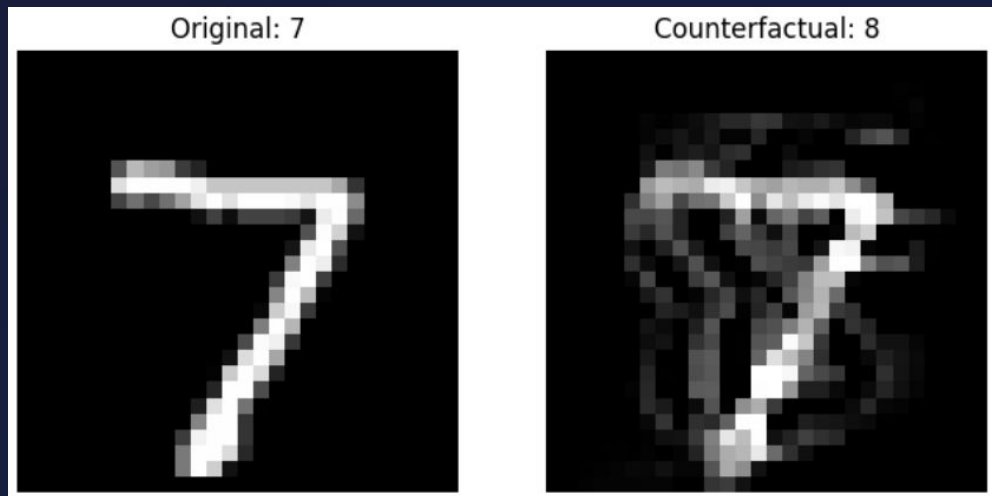
From: johnchad@triton.unm.edu (jchadwic)
Subject: Another request for Darwin Fish
Organization: University of New Mexico, Albuquerque
Lines: 11
NNTP-Posting-Host: triton.unm.edu

Hello Gang,

There have been some notes recently asking where to obtain the DARWIN fish.
This is the same question I have and I have not seen an answer on the net. If anyone has a contact please post on the net or email me.

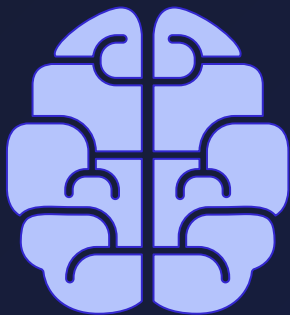
Counterfactual Explanations

- Encontra a mudança mínima na instância que produz uma saída esperada
- Várias técnicas para isso
- Boa parte define uma função de custo (loss) e utiliza um algoritmo de otimização para minimizar essa função
- Aplicações de *prompt injection* e *one pixel attack*



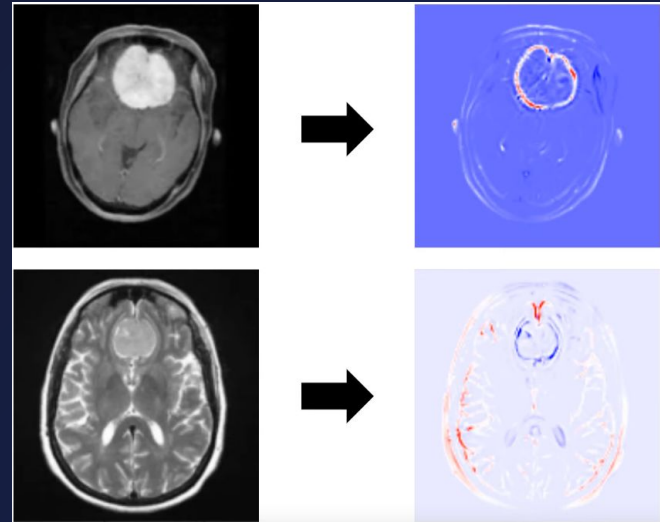
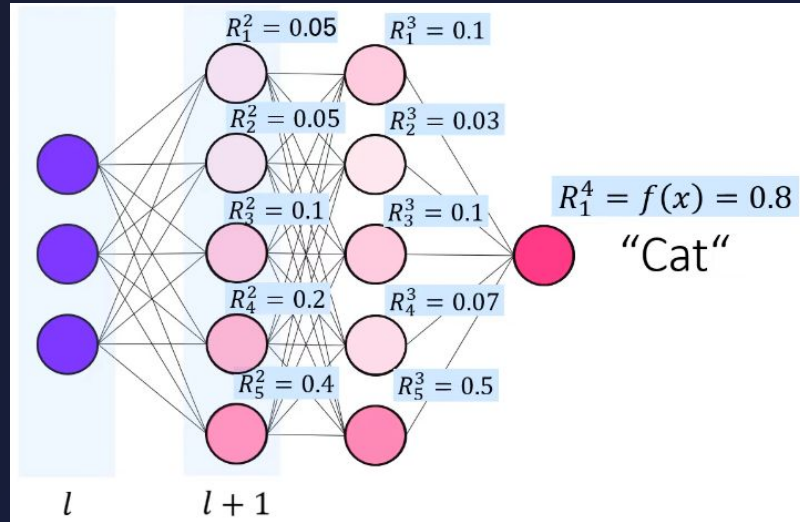
Métodos Post-Hoc

Modelo-Específico



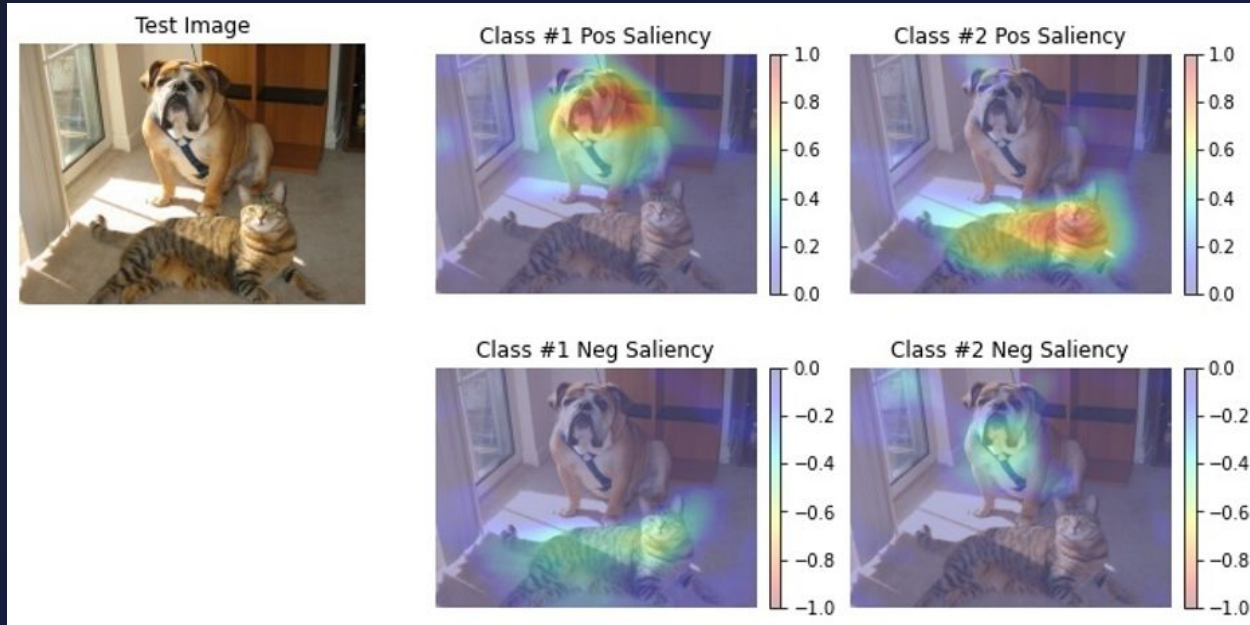
Layerwise Relevance Propagation

- Calcula a relevância de cada atributo através dos pesos da rede, de “trás pra frente”
- Capaz de captar como cada camada da rede contribuiu para a decisão



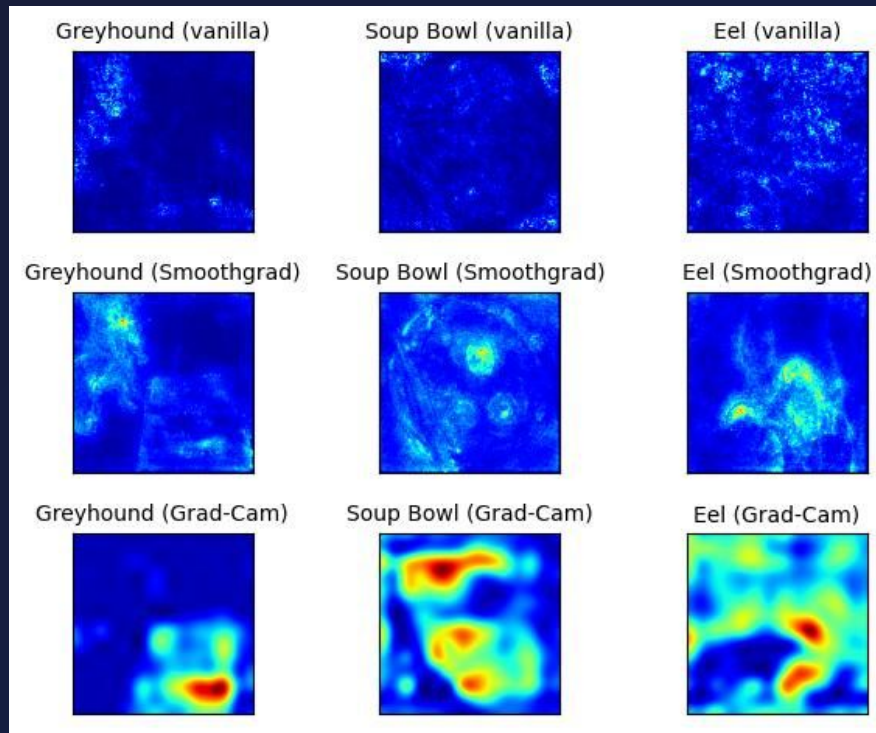
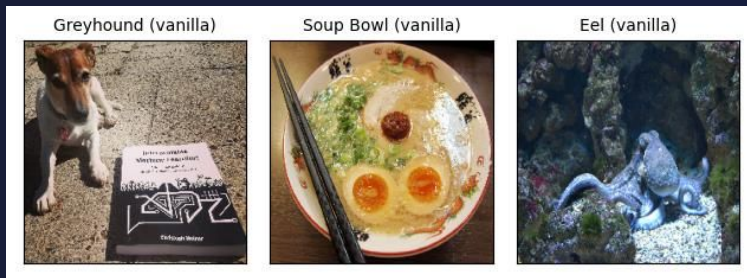
Saliency Maps

- Analisa a importância dos pixels em imagens
- Usa técnicas semelhantes às vistas, descida de gradiente, entre outras



Saliency Maps

- Analisa a importância dos pixels em imagens
- Usa técnicas semelhantes às vistas, descida de gradiente, entre outras



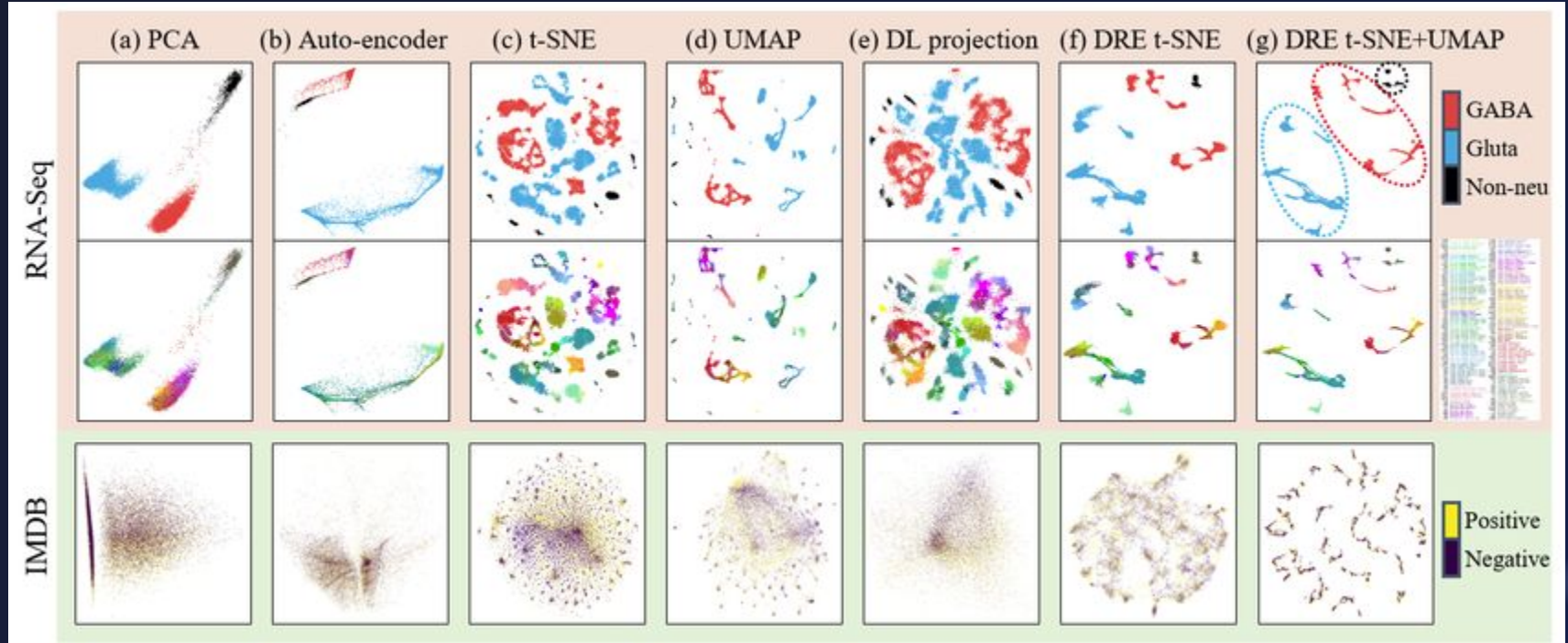
Outros Modelo-Específico...

- Muitas outras técnicas podem ser usadas para diferentes tipos de modelos
- Extração de valores importantes, do Gini em Random Forests até calcular razões de probabilidades em Regressão Logística;
- Verificação de parâmetros internos, testes com partes ou estados do modelo, etc.

E quanto às LLMs?

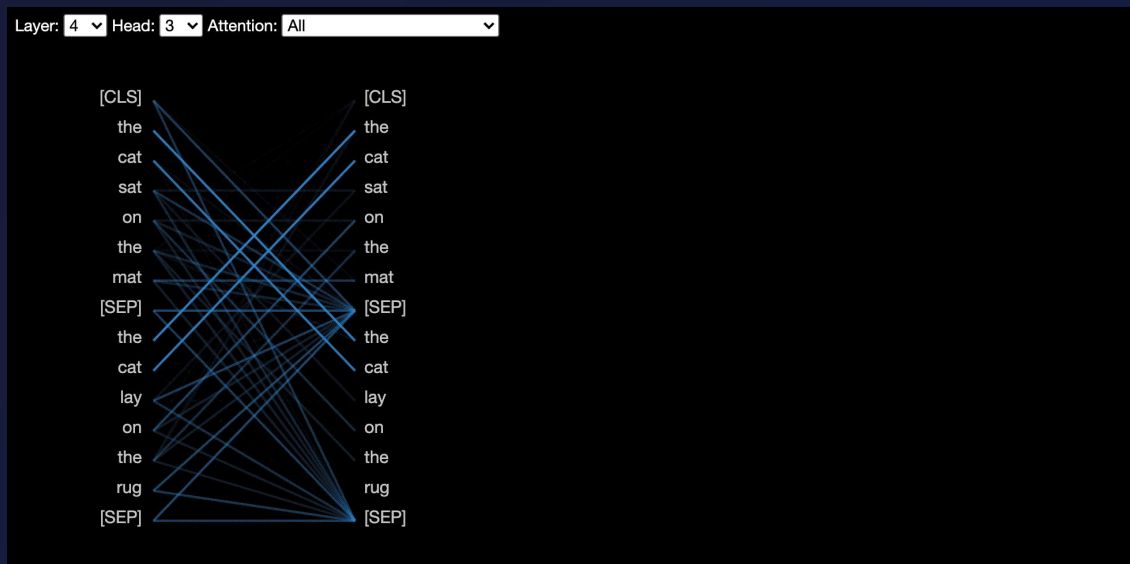
- Podem ser amplamente integradas na Explicabilidade de outros modelos
- Capazes de transformar os resultados das técnicas de XAI em linguagem natural (texto), mais acessível para humanos
- **Mas ainda são um desafio para a XAI!**
- Várias técnicas podem ser aplicadas, mas ainda há muito a ser explorado nesses modelos de bilhões de parâmetros!

E quanto às LLMs?

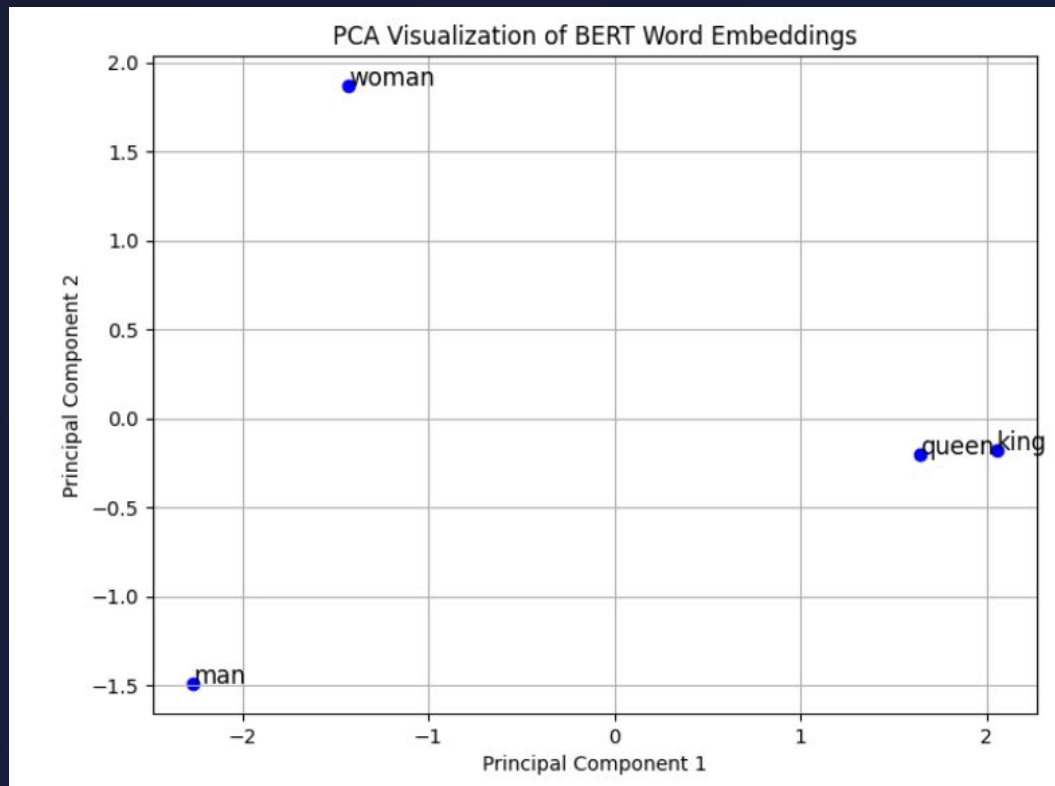


Como explicar modelos generativos que usan AI?

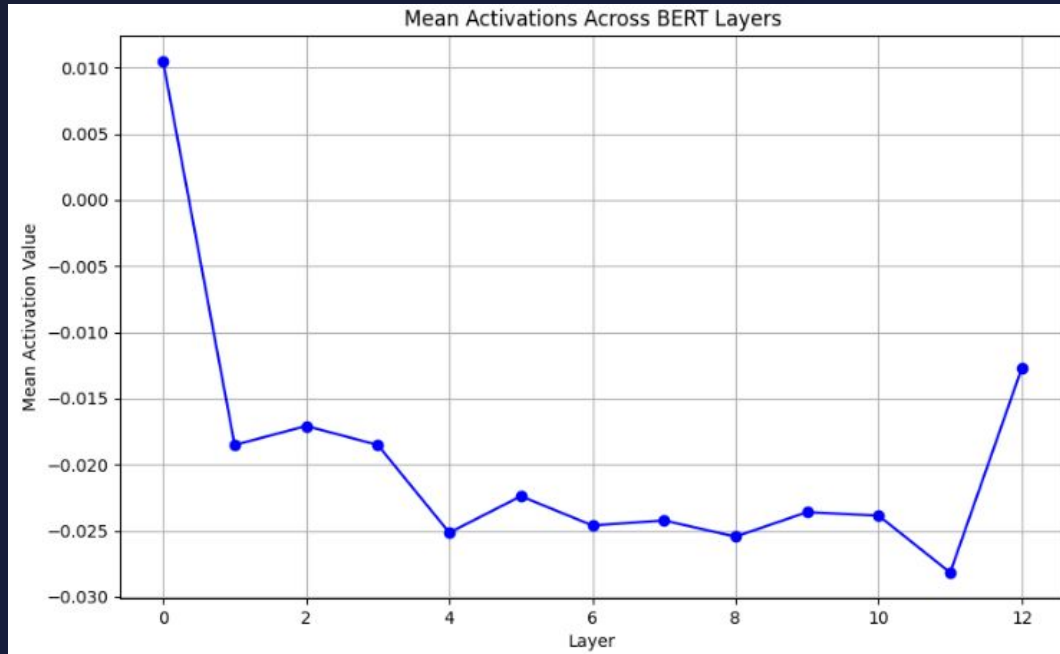
Explainable AI



E quanto às LLMs?



E quanto às LLMs?



Como começar em XAI?

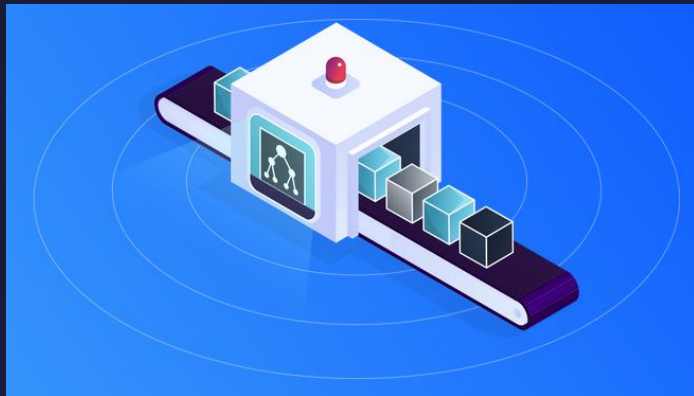
- **Cursos e Aulas:**
 - [Learn Machine Learning Explainability Tutorials | Kaggle](#)
 - [Intrepreting Machine Learning models | Kaggle](#)
 - [Explainable AI - YouTube](#)
- **Livros e Artigos:**
 - [Interpretable Machine Learning](#)
 - [\(PDF\) A Comprehensive Guide to Explainable AI: From Classical Models to LLMs](#)
- **Disciplina do PPGINF**

Algumas Ferramentas

InterpretML: <https://interpret.ml/>

Vertex AI (Google):

<https://cloud.google.com/vertex-ai/docs/explainable-ai/overview?hl=pt-br>



Obrigado!

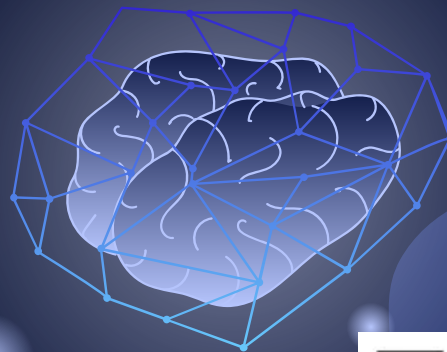
Dúvidas?

hmargotte@inf.ufpr.br

www.linkedin.com/in/henrique-margotte



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#) and infographics & images by [Freepik](#)



Referências

Local Interpretable Model-Agnostic Explanations (LIME): An Introduction – O'Reilly

Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, Volume 58, 2020, Pages 82-115, ISSN 1566-2535. <https://doi.org/10.1016/j.inffus.2019.12.012>. (<https://www.sciencedirect.com/science/article/pii/S1566253519308103>)

Moshkovitz, M., Dasgupta, S., Rashtchian, C. & Frost, N.. (2020). Explainable k-Means and k-Medians Clustering. *Proceedings of the 37th International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 119:7055-7065 Available from <https://proceedings.mlr.press/v119/moshkovitz20a.html>.

Content of this template_

You can delete this slide when you're done editing the presentation

<u>Fonts</u>	To view this template correctly in PowerPoint, download and install the fonts we used
<u>Used</u> and <u>alternative resources</u>	An assortment of graphic resources that are suitable for use in this presentation
<u>Thanks slide</u>	You must keep it so that proper credits for our design are given
<u>Colors</u>	All the colors used in this presentation
<u>Icons</u> and <u>infographic resources</u>	These can be used in the template, and their size and color can be edited
Editable presentation theme	You can edit the master slides easily. For more info, click <u>here</u>

For more info:

[SLIDESGO](#) | **[BLOG](#)** | **[FAQs](#)**

You can visit our sister projects:

[FREEPIK](#) | **[FLATICON](#)** | **[STORYSET](#)** | **[WEPIK](#)** | **[VIDEVO](#)**

Table of contents_

01

Introduction

03

Problem &
solution

02

Market
opportunity

04

Product
demo

What is a pitch deck?_

A pitch deck is a short presentation that entrepreneurs use to describe their business and present it to potential investors. It usually consists of about 10 slides that provide:

- An overview of the company's mission, product and service
- Market opportunity
- Team, financials, etc.

The goal is to garner enough interest from investors so that they are motivated to invest in the business





01

Introduction_

Introduction_



Our company

The introduction of your company in the pitch deck should include a brief overview of who you are and what you do. It should also include a brief description of why your product or service is unique and the value it provides to customers



Our idea

The “our idea” part of a pitch deck should include an explanation of your product/service, how it works, and what the customer value is. You should also explain the market opportunity for your product/service

Main ideas of a pitch deck_



Presentation

A great presentation for a pitch deck should be concise, clear and easy to understand. Visual elements such as charts, graphs, images, and videos can help to make the presentation more engaging



Investors

When it comes to picking investors for a pitch deck, it's important to research potential investors and understand their preferences, goals and interests



Value

The value of a product is determined by how much customers are willing to pay for it. It's a combination of factors, including the quality of the product, its features, and how well it meets customer needs

Key ideas in market opportunity_



Research new markets

Analyze existing customer needs and identify gaps in the market to find new opportunities



Innovative solutions

Utilize technology and customer feedback to come up with novel solutions that meet your customer's demands or needs



Stay ahead of trends

Monitor changes in the marketplace, such as shifts in consumer behavior or emerging technologies, to stay competitive and capitalize on new opportunities and trends



Seasonal spikes

Identify times of year when demand for certain products may spike, such as holidays or special occasions, and use these times to target customers or maximize sales

Six recommendations_

Keep it simple

Structure your pitch deck clearly and make sure all the important points are easy to understand

Use visuals

Incorporate visuals or graphics to illustrate your message and keep the audience engaged

Keep it short

A good pitch deck should be concise and clear; avoid trying to cram in information in excess

Make it memorable

Include impactful stories, statistics, or facts that will help your audience remember your message after the presentation

Test & iterate

Practice presenting your deck beforehand with friends or colleagues for feedback, then use this input to refine it further

Main points

A pitch deck is an overview of your business; focus on key points that get the most important ideas across



333,000_

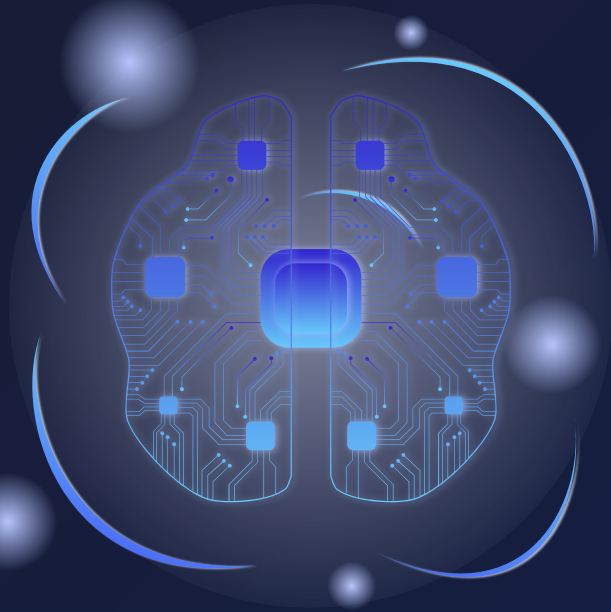
Users bought our product

9h 55m 23s_

Estimated delivery time per unit

386,000 km_

Avg. Distance travelled by logistics team



4,498,300,000_

Number of users analyzed in
our market research

Revenue by quarter_

20%

1st quarter

Highlight the main
sources of revenue

20%

2nd quarter

Highlight the main
sources of revenue

35%

3rd quarter

Highlight the main
sources of revenue

25%

4th quarter

Highlight the main
sources of revenue



Follow the link in the graph to modify its data and then paste the new one here. [For more info, click here](#)

Competition comparison_

Team A Team B Team C Team D Team E

Mercury	Yes	Yes	No	Yes	Yes
Mars	Yes	No	Yes	Yes	No
Saturn	No	Yes	Yes	No	No
Venus	Yes	Yes	Yes	Yes	Yes
Jupiter	No	No	Yes	Yes	Yes
Earth	Yes	Yes	No	No	Yes

Product infographics_

Visuals

Showcase the design of your product



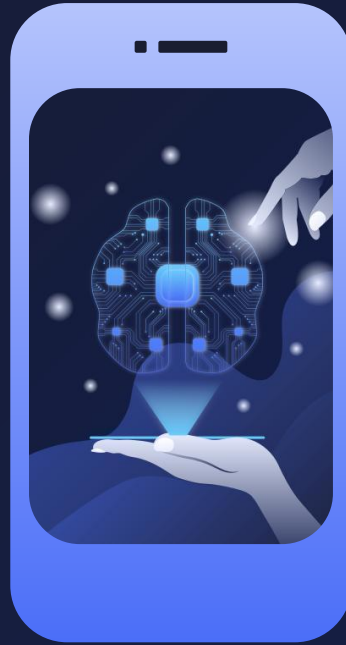
Price

Share the cost and price with investors



Availability

When is the product expected to be ready?



Features

What makes your product unique?



Users

Speak about the target audience



Updates

Do you plan on updating it?



Timeline of your presentation_

Provide a brief overview of the pitch deck content



Engage the audience with a concise and compelling company introduction

Identify the target customer pain points and challenges in a brief manner



Describe how your product or service can solve the problem

Highlight the unique value proposition and benefits of your solution



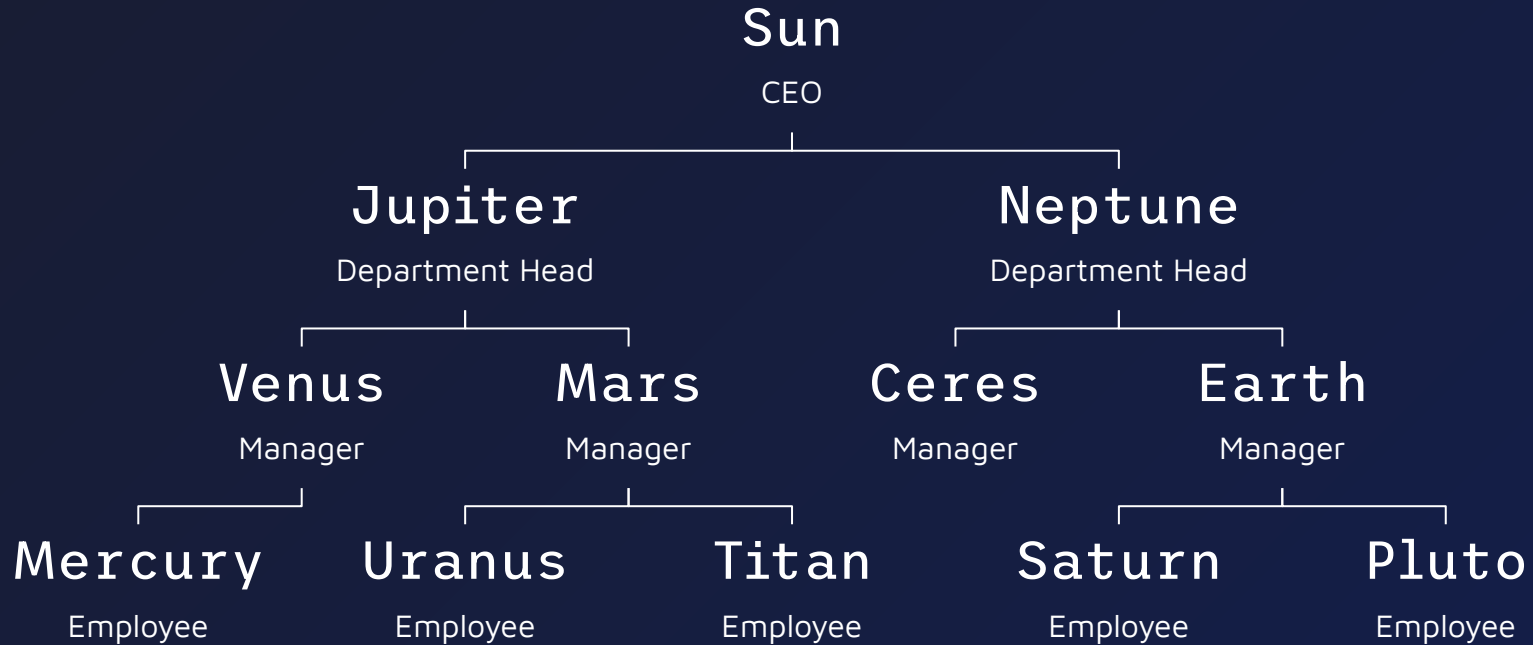
Analyze the target market size, growth potential, and competition briefly.

Explain your financial needs and briefly outline your funding allocation



End with a clear and concise call to action

Organizational chart_



Market size overview_

Include the total size of the market, which represents the entire potential customer base for the product or service

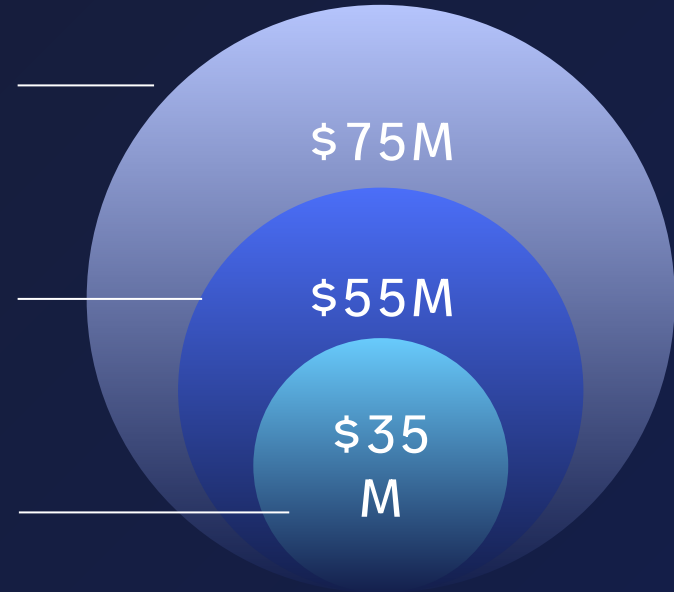
Identify the target market for the product or service, which may be a subset of the total market. This could be based on factors such as demographics, geography, or specific needs

Indicate the current market size, which represents the portion of the target market that the company has successfully captured

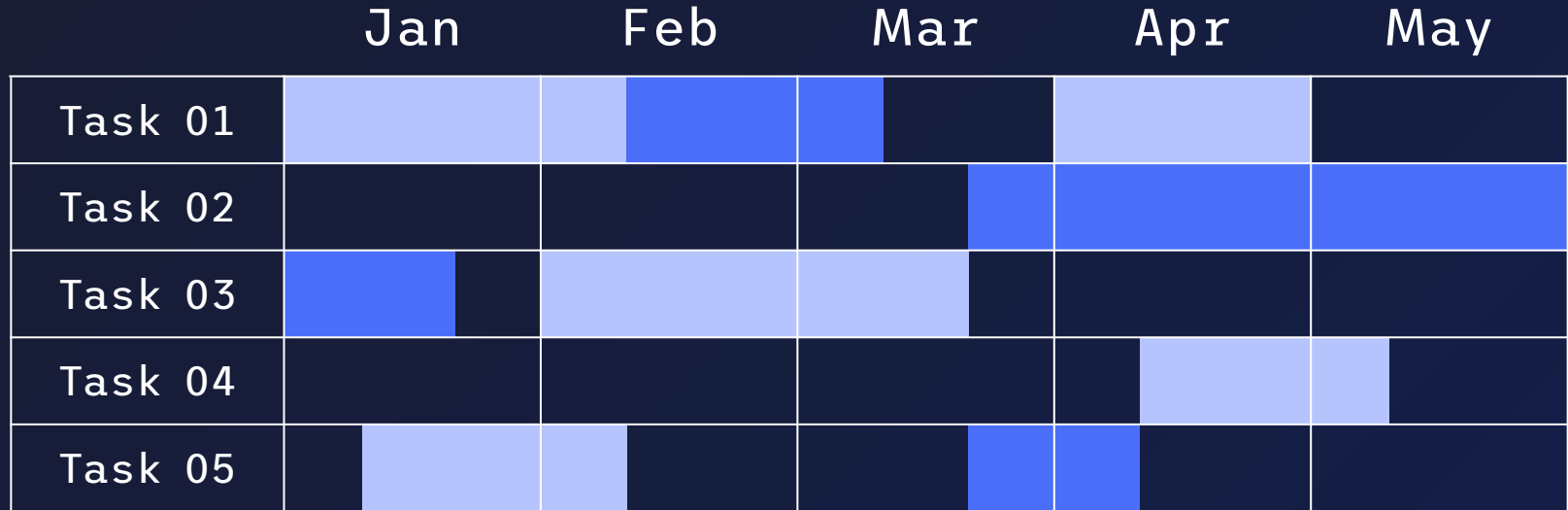
Outer
circle

Middle
circle

Inner
circle



Roadmap infographics_



● Team 01

Give a brief description of this team and their responsibilities

● Team 02

Give a brief description of this team and their responsibilities

KPI dashboard_

70%

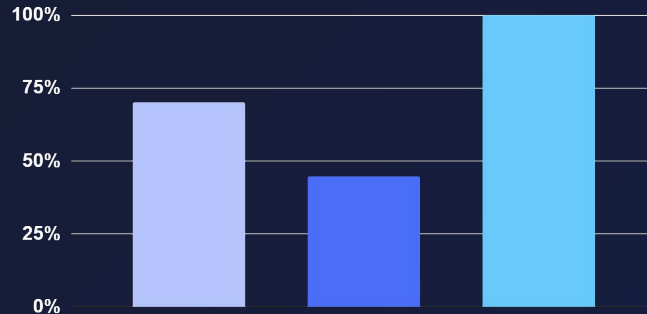
Development was mostly good in the beginning

45%

The company faced some difficulties halfway

100%

Bugs were fixed and the final product is ready



Product

Column 1

Column 2

Jupiter

50%

2,000,000

Saturn

20%

50,000

Mercury

100%

1,500,000

Follow the link in the graph to modify its data and then paste the new one here. [For more info, click here](#)

Thanks!_

Do you have any questions?

youremail@freepik.com

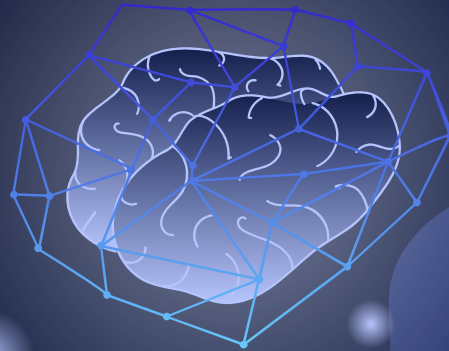
+91 620 421 838

yourwebsite.com



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#) and infographics & images by [Freepik](#)

Please keep this slide for attribution



Icon pack_



Alternative resources_

Here's an assortment of alternative resources whose style fits that of this template:

Vectors

- Landing page template artificial intelligence
- Landing page neon with smartphone



Resources_

Did you like the resources on this template? Get them at our other websites:

Photos

- Side view of man using smartphone
- Side view woman working on digital monitors

Vectors

- Artificial intelligence template landing page
- Artificial intelligence landing page
- Artificial intelligence landing page template

Icons

- Icon Pack: Artificial intelligence | Filled

Instructions for use

If you have a free account, in order to use this template, you must credit [Slidesgo](#) by keeping the [original slide](#)



As a Free user, 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



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 [Editing Slides](#) or visit our [FAQ](#)



Instructions for use (premium users)

As a Premium user, you can use this template without attributing or keeping the slide



You are allowed to:



Modify this template



Share this template in an editable format



Use it for both personal and commercial projects



Hide or delete the “Thanks” slide and the mention to Slidesgo in the credits



You are not allowed to:



Sublicense, sell or rent any of Slidesgo Content



Use any of the elements that are part in a separated way



Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo



Register any of the elements as a trademark or logo in an intellectual property registry

For more information about editing slides, please read our or visit our

Fonts

This presentation has been made using the following fonts:



Click on the button of the link to the fonts

[Trispace](#)

[Maven Pro](#)

To view this template correctly in PowerPoint, download and install the fonts we used

Colors

All the colors used in this presentation



#ffffff

#191d30

#b5c4fc

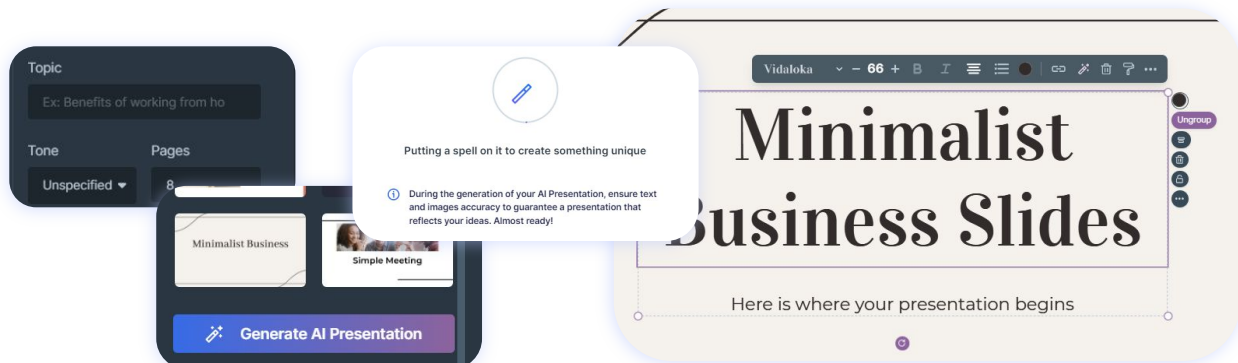
#4b6ef8

#131e4b

#3327cf

#68cafb

Presentation Maker



Slidesgo introduces its **latest feature**: the **Presentation Maker**. Enjoy **two main functionalities** - **firstly**, with a few clicks, **create marvelous presentations with Artificial Intelligence** that adapt to your needs. And it's completely free!

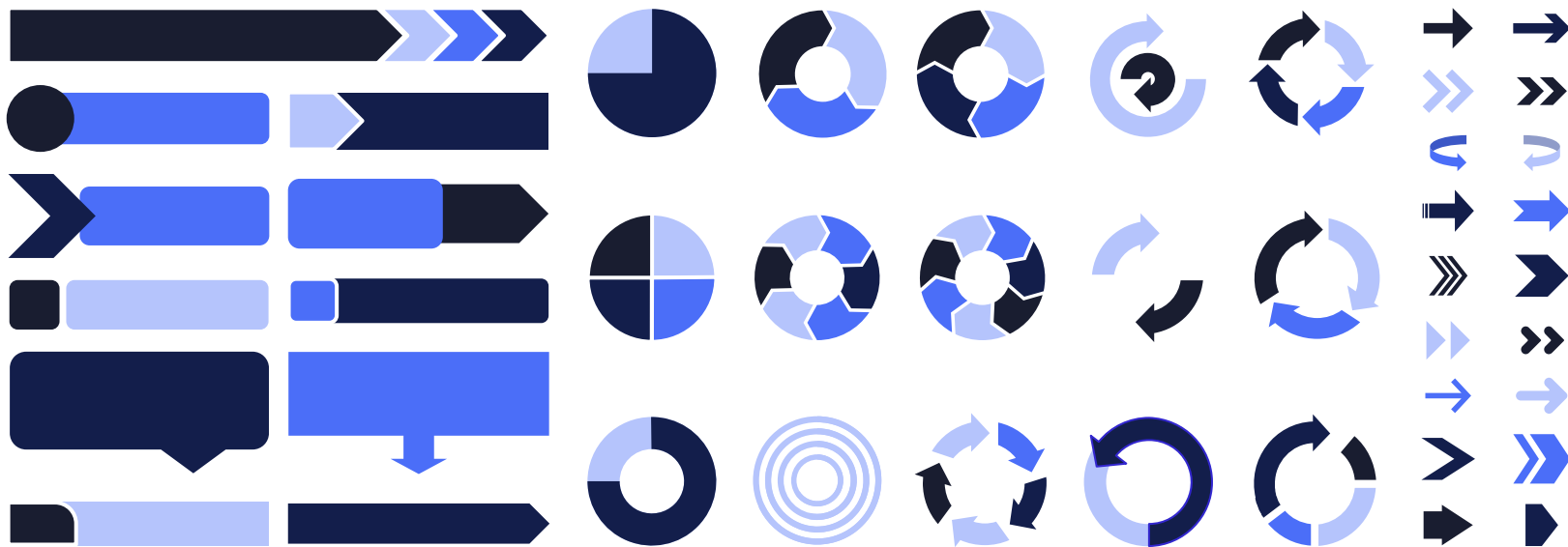
The **second functionality** of this tool is that **you can edit presentations through the online editor**. Create interactive resources easily, quickly and without the need for any software. Change everything or start from scratch

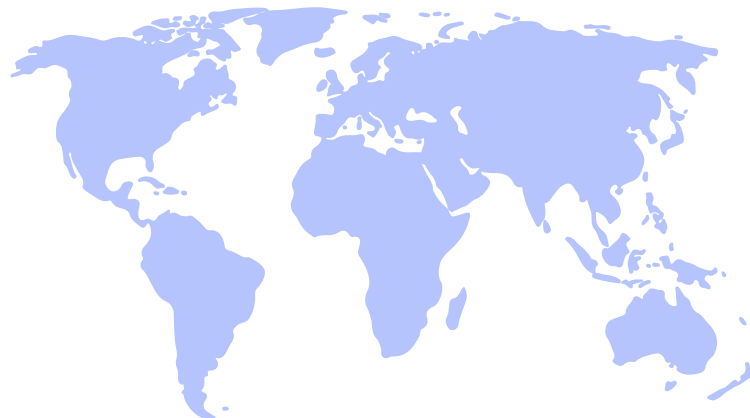
[Generate AI Presentation](#)

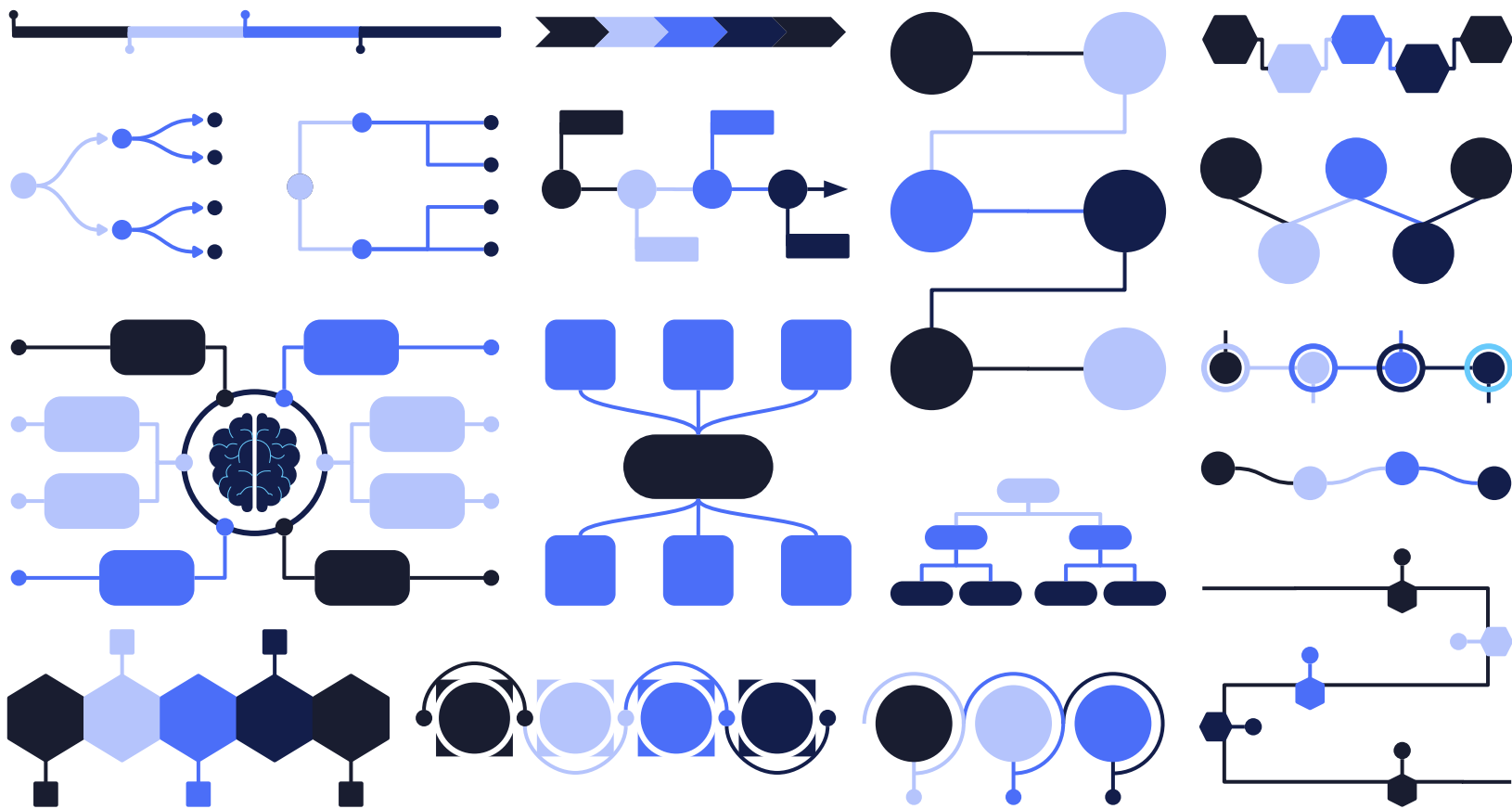
[Edit online](#)

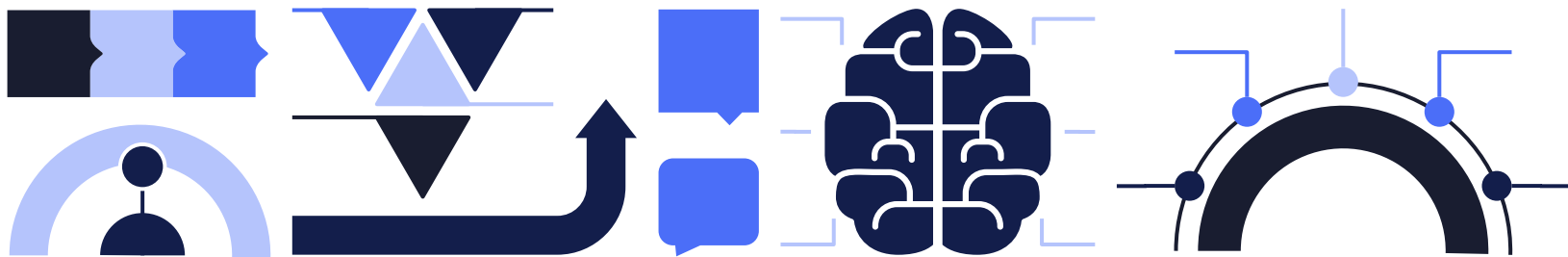
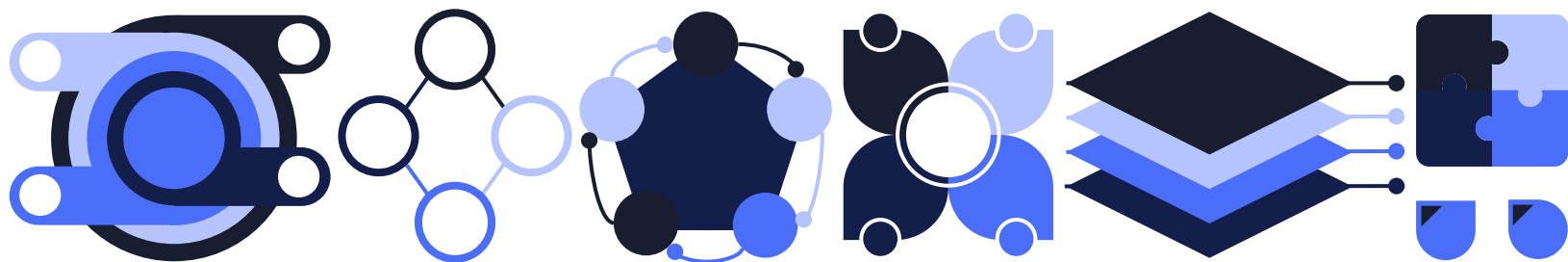
Use our editable graphic resources...

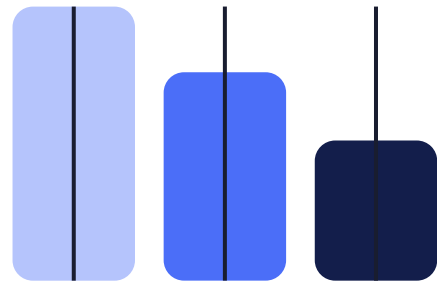
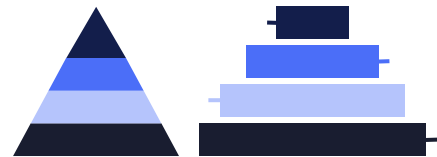
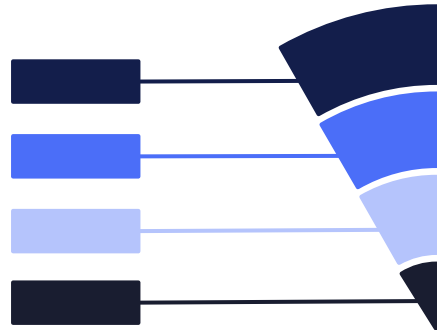
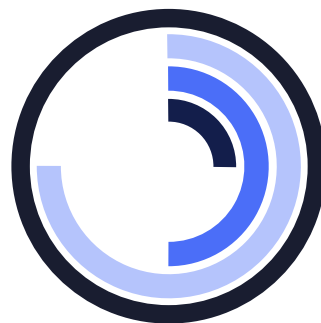
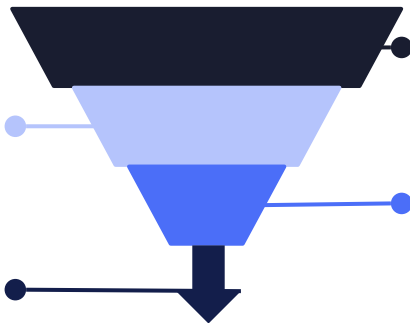
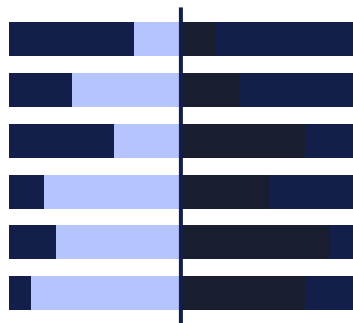
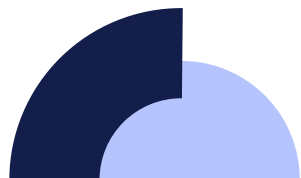
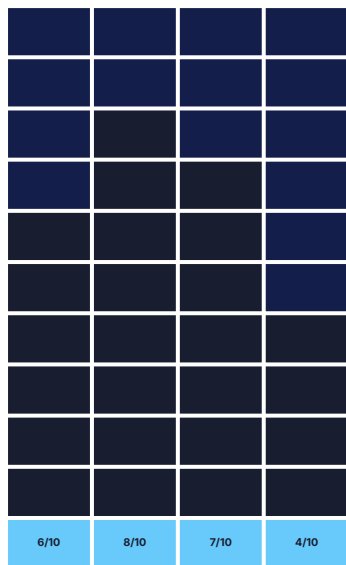
You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. You can also look for more on Slidesgo











...and our sets of editable icons

You can **resize** these icons without losing quality

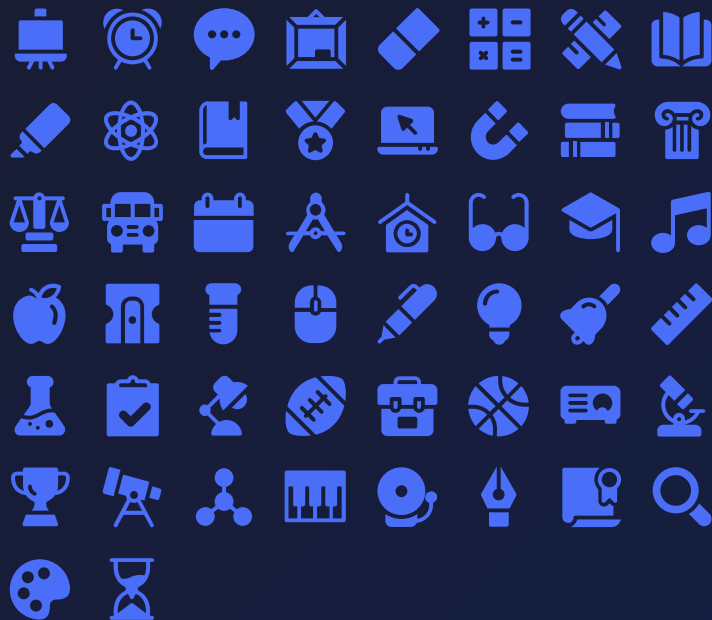
You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**

In Google Slides, you can also use

, allowing you to customize and add even more icons



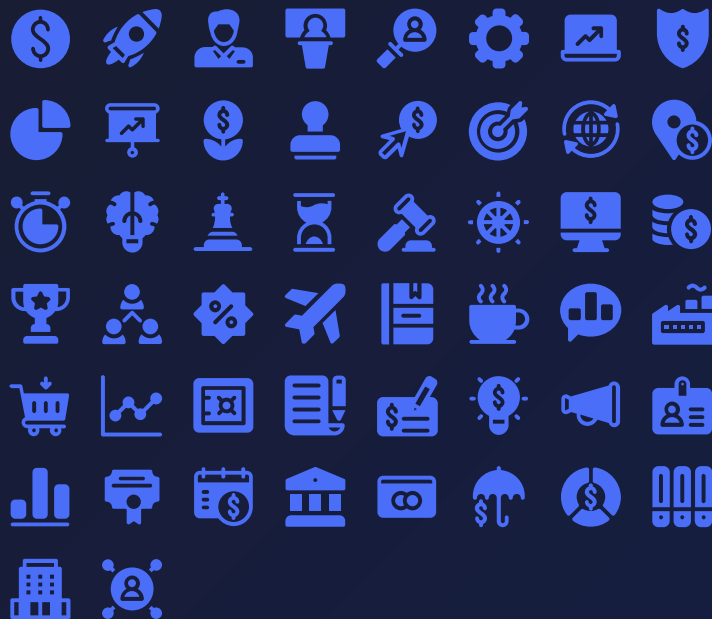
Educational icons



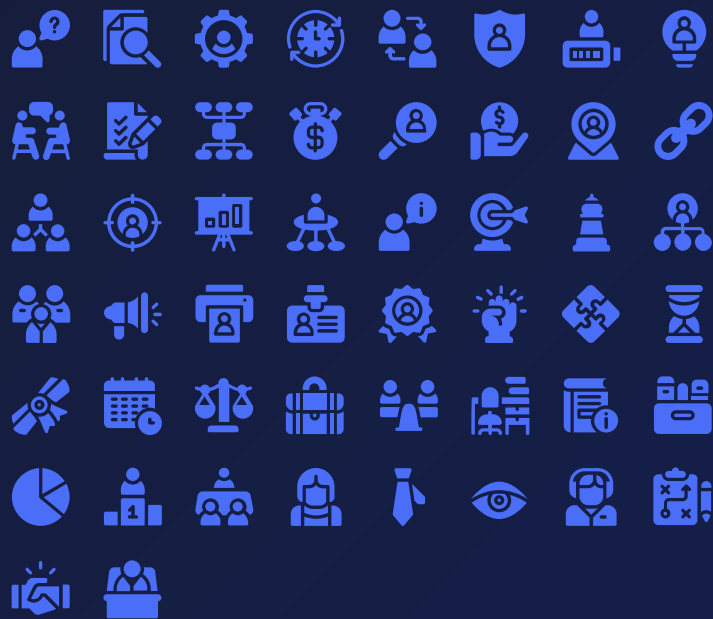
Medical icons



Business icons



Teamwork icons



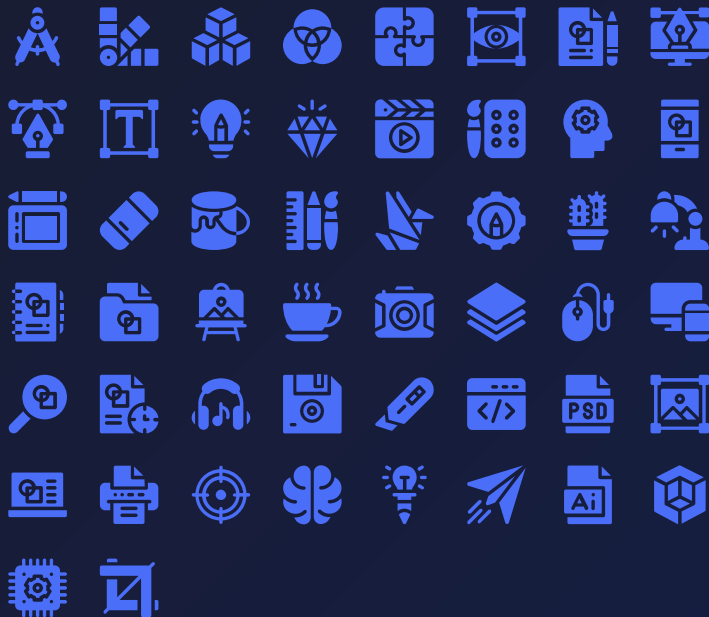
Help & support icons



Avatar icons



Creative process icons



Performing arts icons



Nature icons



SEO & marketing icons

