

# GEAM:



# Grupo de Estudos de Aprendizado de Máquina

---

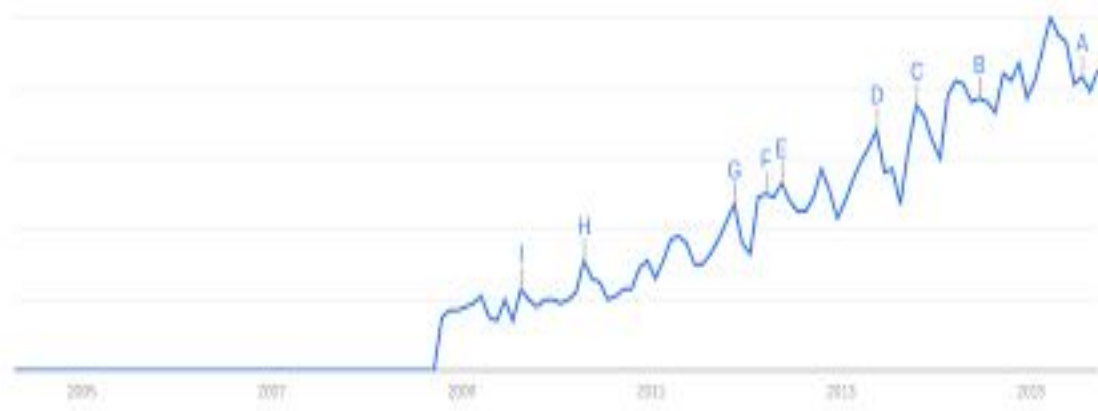
- Bolsista: Ítalo Lima Dantas
- Orientador: Régis Pires Magalhães

# Sumário

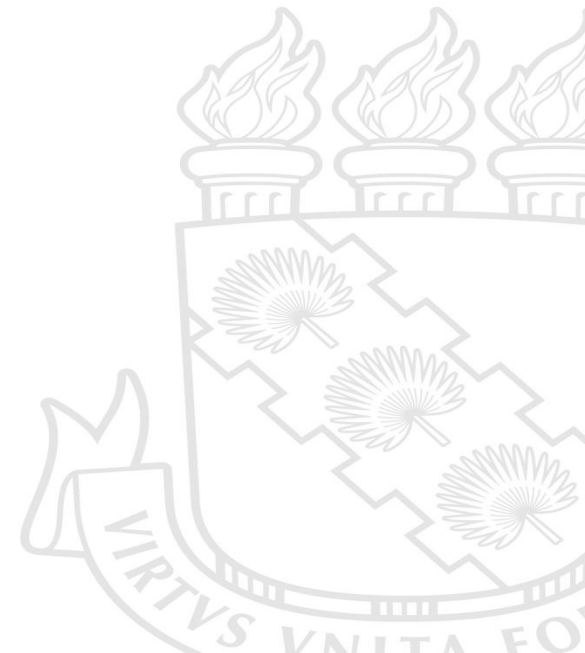


<b>Contextualização.....</b>	<b>3</b>
<b>Machine Learning.....</b>	<b>4</b>
<b>Twitter.....</b>	<b>5</b>
<b>Top Trends.....</b>	<b>6</b>
<b>Análise de Sentimentos.....</b>	<b>9</b>
<b>Dificuldades na Análise de Sentimentos.....</b>	<b>10</b>
<b>Prática no Orange.....</b>	<b>11</b>
<b>Resumo do Processo.....</b>	<b>15</b>
<b>Fluxo de Trabalho.....</b>	<b>16</b>
<b>Glossário.....</b>	<b>20</b>

## Contextualização do Problema



Pesquisa do termo “Sentiment Analysis” no Google Trends.

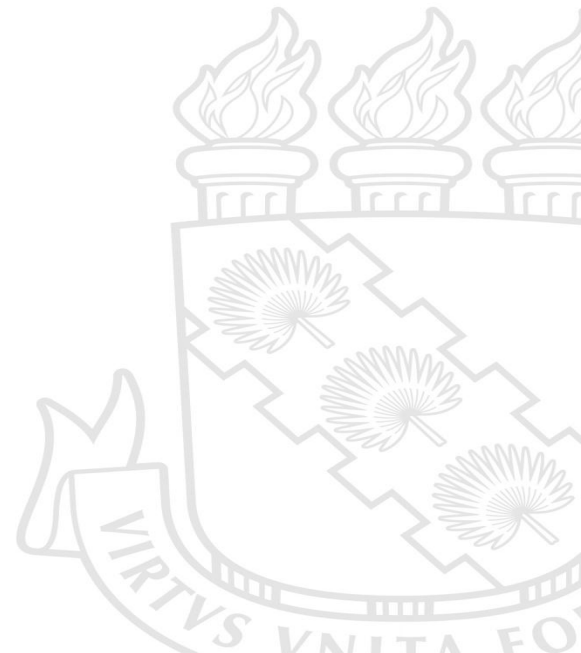


## O Machine Learning em Ação



- Atualmente, as mídias sociais se tornaram objetos de estudo, das mais diversas áreas: Empresarial, logística, psicologia, tecnologia, ciências sociais, etc. Nessas, as pessoas expressam suas opiniões de diversas formas, com posts claros e objetivos, imagens subjetivas, “indiretas”, tags, comentários, entre outras.
- Por isso, o Machine Learning cumpre papel muito importante para a análise, seja qual for a área interessada, algoritmos de processamento de texto e análise de sentimentos trabalham e são aperfeiçoados todos os dias, e isso pode ser observado através dos próprios anúncios, depois de visitarmos um produto na web, em empresas o capital é melhor direcionado, ao investir em serviços/produtos que apresentam um potencial maior de agrado ao cliente.

# Machine Learning no Twitter



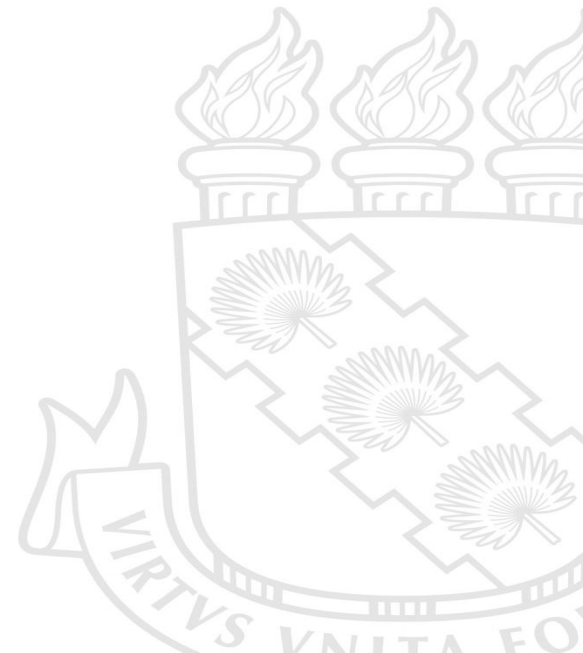


← Top trends 🔍 ⚙️

1	#StanLeeForever	31.5K Tweets	⌵
2	#TuesdayThoughts	15K Tweets	⌵
3	Dassault CEO Eric Trappier	3,690 Tweets	⌵
4	#TuesdayMotivation	3,621 Tweets	⌵
5	#छठपूजा	11.2K Tweets	⌵
6	#WorldKindnessDay	4,126 Tweets	⌵
7	#RIPStantheMan	41.7K Tweets	⌵
8	Kyrsten Sinema	151K Tweets	⌵

➕

Top trends do Twitter, exemplo da tamanha consideração das mídias sociais, em caso de análise.





# Top Trends : Funcionamento



## Como os assuntos do momento são determinados?

Os assuntos do momento são determinados por um algoritmo e, por padrão, são personalizados com base em quem você segue, em seus interesses e em sua localização. Esse algoritmo identifica os tópicos populares da atualidade, em vez de tópicos que já foram populares por algum tempo ou diariamente. Assim, você pode descobrir os tópicos que estão em discussão no Twitter no momento.

**Observação:** o número de Tweets relacionados aos assuntos do momento é apenas um dos fatores que o algoritmo analisa ao classificá-los e determiná-los. Por meio de algoritmos, assuntos do momento e hashtags são agrupados quando estão relacionados ao mesmo tópico. Por exemplo, #ÂnimoParaSegunda e #SegundaÂnimo podem ser representados por #ÂnimoParaSegunda.

# Top Trends: Regras



## Há regras para os assuntos do momento?

Sim. Queremos que os assuntos promovam discussões saudáveis no Twitter. Isso significa que, às vezes, podemos impedir que determinado conteúdo entre nos assuntos do momento. Isso inclui os assuntos que:

- Contenham palavrões ou referências a conteúdo adulto/explicito.
- Incitem ódio com base em raça, etnia, nacionalidade, orientação sexual, sexo, identidade de gênero, religião, idade, deficiência ou doença.
- Violam as [Regras do Twitter](#).

Em alguns casos, podemos também considerar a importância do conteúdo, ou se ele é de interesse público, ao avaliar possíveis violações. Nesses casos, o conteúdo pode continuar nos assuntos do momento em nossa plataforma.

Observe que, mesmo que evitemos que a hashtag ou o conteúdo apareçam na lista de assuntos do momento, você ainda poderá acessar as conversas sobre esse conteúdo no Twitter.



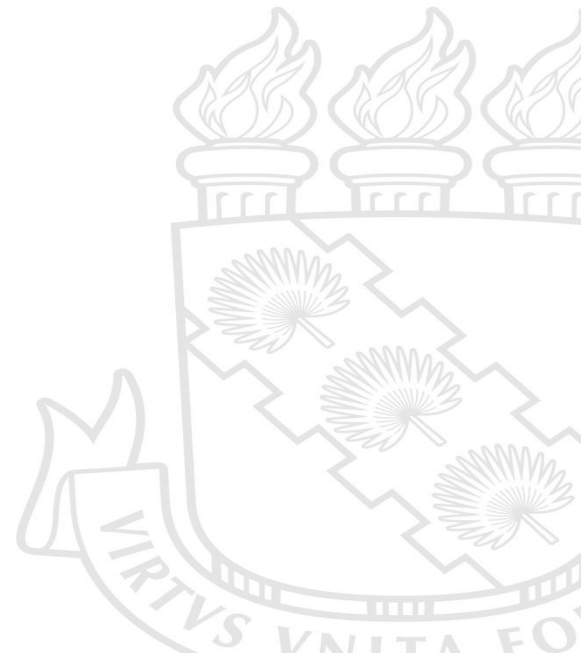
## Análise de Sentimentos



- Através de posts no Twitter, pode-se analisar e chegar a conclusão do sentimento que o mesmo expressa, antes de adentrar mais no assunto, precisamos conhecer alguns termos essenciais nesse tipo de estudo.
- **Polaridade:** Grau de Sentimento de um texto, podendo ser positivo, negativo ou neutro.
- **Força do Sentimento:** Representa a intensidade do sentimento, ou da polaridade (normalmente classificado entre -1 e 1).
- **Sentimento / Emoção:** Indica a conclusão de um sentimento em específico. Ex: Raiva, amor, etc.
- **Subjetividade Vs Objetividade:** Uma sentença subjetiva geralmente expressa opiniões ou sentimentos, já uma objetiva um fato ou informação. Pode criar variáveis de estudo (Ex: Análise de Fake News) .

## Dificuldades na Análise de Sentimentos

- Ironia
- Sarcasmo
- Metáforas
- Múltiplos sentimentos em uma única frase.
- Negações empregadas.



# O Processo



## PASSOS INICIAIS:

- Baixar add-on de Text no Orange.
- Utilizar o widget Twitter, e usar a chave de acesso : da API
- A chave de acesso pode ser obtida no link : <http://dontpad.com/grupogeam15/10>
- Filtrar o tipo de busca que se deseja.
- Nesse caso, faremos uma busca por conteúdo, sem destacar autor e informações sobre autores, primeiramente, destacaremos tweets em inglês.

## Passos



- Abrir o widget de Corpus Viewer, para visualizar os tweets sobre o tema pesquisado.
- Fazer o pré processamento do texto, transformando todas as letras em minúsculas, marcando a opção “Lower-Case”, utilizando “Tweet”, como forma de tokenização, removendo links, html e filtrando palavras e caracteres de forma padrão.
- Antes de partir para a análise de sentimentos, podemos visualizar algumas informações interessantes, com o widget “Word Cloud”, que vai mostrar as palavras que aparecem com maior frequência nos tweets buscados.
- Podemos visualizar o “Word Cloud” antes e após o pré processamento de texto, verificando a diferença entre os mesmos.
- Feito isso, é hora de partir para a Análise, ligando o widget “Sentiment Analysis”, e usando o modelo “Vader”, que é uma biblioteca do python, baseada em léxicos muito famosos, além de microblogs, para obter gírias e emoções.

## Continuação



- Depois dessa configuração, deve-se escolher apenas as colunas mais importantes para a análise, utilizando o widget “Select Columns”.
- Podemos visualizar a análise de duas formas : “Data Table” e “Heat Map”.
- O Heat Map é um mapa de calor que irá mostrar a polaridade dos tweets analisados, baseado na tonalidade das cores apresentadas.
- Por fim, visualizaremos novamente a estrutura dos tweets, com as pontuações na escala de polaridade definidas.
- Aqui, a nossa análise básica de sentimentos através de uma base de dados da rede social Twitter está concluída, porém, podemos ainda verificar as emoções contidas nesses tweets.



# Emoções != Sentimentos



- Utilizaremos o widget “Tweet Profiler”, que será ligado ao texto pré processado.
- Selecionaremos novamente as colunas mais importantes a serem exibidas, com o Select Columns. Feito isso, ligaremos diferentes formas de visualização de dados, como Data Table, Distributions e Box Plot.
- A saída deverá ser uma emoção, que pode ser: Anger, Joy, Fear, Surprise, Sadness, Disgust.
- Exemplo de uso: Debate Eleitoral 2018. Reconhecimento Facial + Análise de Emoções.
- <https://www.estadao.com.br/infograficos/politica,o-que-revela-uma-analise-das-emocoes-dos-candidatos-durante-o-debate.923037>



## Resumo do Processo



- Usar a base de dados do Twitter a partir de um tema a ser filtrado → Visualizar a estrutura desses tweets → Pré Processar os textos, retirando stopwords dos mesmos → Visualizar a nuvem de palavras de todos os tweets filtrados, verificando quais palavras estão relacionadas → Fazer a análise de sentimentos, utilizando o modelo Vader → Selecionar as colunas de maior importância para a exibição dos dados → Visualizar os dados e suas respectivas polaridades → Visualizar tweets específicos e suas polaridades.
- Usar o Tweet Profiler com o texto já pré processado, atribuir a relação de conteúdo, utilizar o modelo Ekman, com saída em classes → Visualizar em tabela as emoções dos respectivos tweets filtrados → Selecionar as colunas mais importantes → Visualizar os dados em formas de gráfico, com as saídas sendo as respectivas emoções.

**Data**

File CSV File Import Datasets SQL Table

Data Table Paint Data Data Info Data Sampler

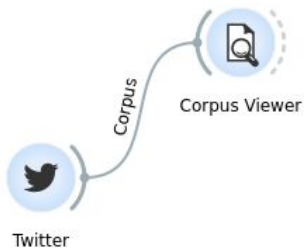
Select Columns Select Rows Pivot Table Rank

Correlat... Merge Data Concate... Select by Data In...

Transpose Random... Preproc... Apply Domain

**Twitter**

Load tweets from the Twitter API.

[more...](#)

**Twitter**

Twitter API Key

Query

Query word list: Machine Learning

Search by: Content

Language: English

Max tweets: ☒ 250

Allow retweets: ☐

Collect results: ☐

Text includes

☒ Content ☐ Author Description

Info

Tweets on output: 250

Search

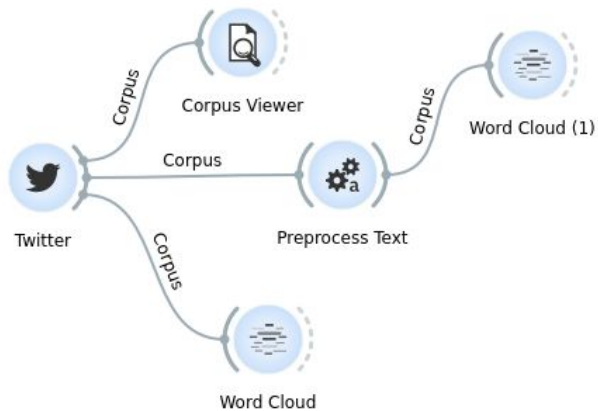
?

**Data**

File	CSV File Import	Datasets	SQL Table
Data Table	Paint Data	Data Info	Data Sampler
Select Columns	Select Rows	Pivot Table	Rank
Correlat...	Merge Data	Concate...	Select by Data In...
Transpose	Random...	Preproc...	Apply Domain

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

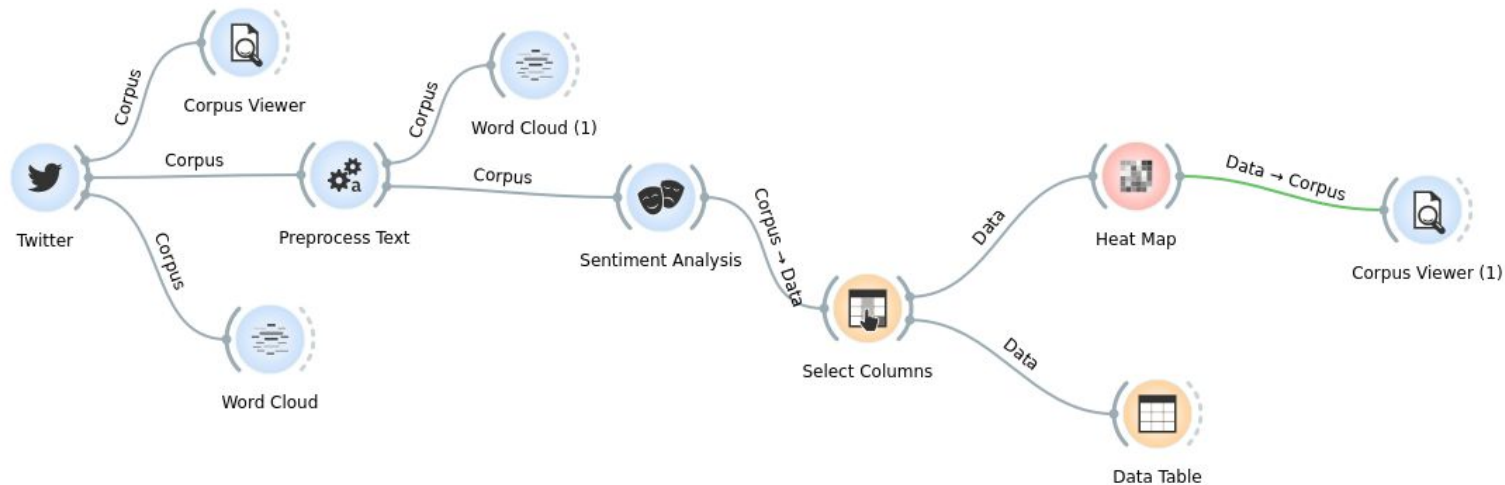


**Data**

File	CSV File Import	Datasets	SQL Table
Data Table	Paint Data	Data Info	Data Sampler
Select Columns	Select Rows	Pivot Table	Rank
Correlat...	Merge Data	Concat...	Select by Data In...
Transpose	Random...	Preproc...	Apply Domain

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).



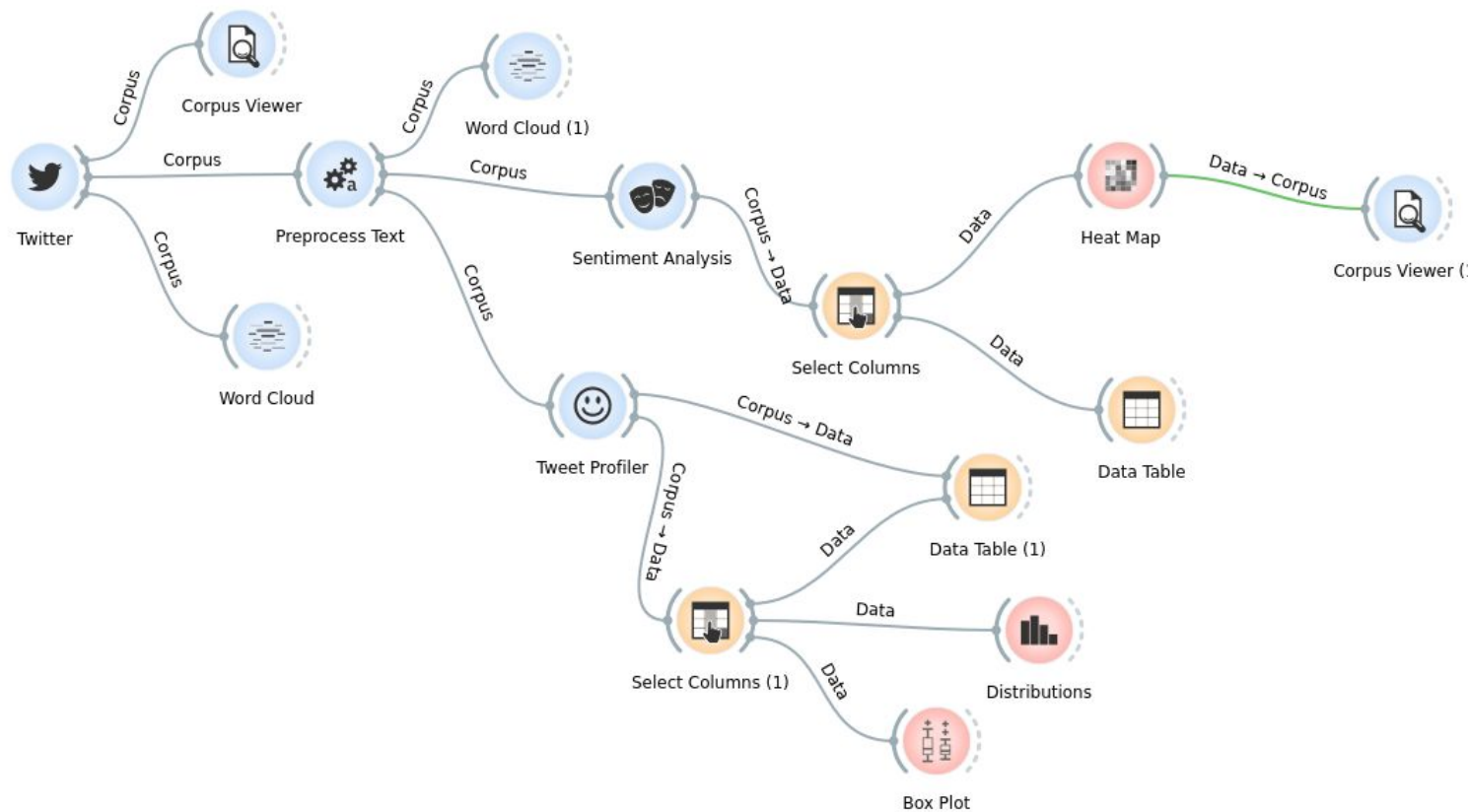
**Data**

File	CSV File Import	Datasets	SQL Table
Data Table	Paint Data	Data Info	Data Sampler
Select Columns	Select Rows	Pivot Table	Rank
Correlat...	Merge Data	Concat...	Select by Data In...
Transpose	Random...	Preproc...	Apply Domain

**Heat Map**

Plot a heat map for a pair of attributes.

[more...](#)



# Glossário



## Alguns termos utilizados durante a realização do problema:

- **Vader** - Biblioteca do python, feita com base nos principais léxicos internacionais, para a coleta de palavras, depois foi integrada a microblogs para a coleta de gírias e emoções. Base para o Sentiment Analysis. <https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f>
- **Ekman**: Classificação das 6 principais emoções que podem ser analisadas, são: Tristeza, Medo, Surpresa, Repulsa, Raiva e Alegria. Base para o Tweet Profiler.
- **StopWords**: Palavras/caracteres que não agregam valor para a análise de sentimentos, podem ser artigos, preposições, vírgulas e sinais de pontuação em geral. Utilizada no Preprocess Text.
- **Compound**: Espécie de peso, que exhibe uma pontuação sobre o sentimento total analisado. Onde o -1 é muito negativo, e o 1 é muito positivo.



**“O *Homo sapiens* é a espécie que adapta o mundo às suas necessidades e não uma espécie que se adapta ao mundo. O Machine Learning é o mais novo capítulo desta saga de milhões de anos.”**  
**Pedro Domingos, em seu livro o Algoritmo Mestre.**