



RRAT

Reduzindo risco de acidentes com tubarões

Predições com IA



INTEGRANTES

Nome: Alan Andrade

Contato:alanandrade124@gmail.com

Formação: Ciência da computação

Nome: Leonardo Felipe Oliveira

Contato:leonardo.lustosa@hci.com

Formação: Ciência da computação

Nome: Caio Felipe

Contato:imcaiofelipe@gmail.com

Formação: Ciência da computação

Nome: Luiz Henrique Jose

Contato:contatoluizh8@gmail.com

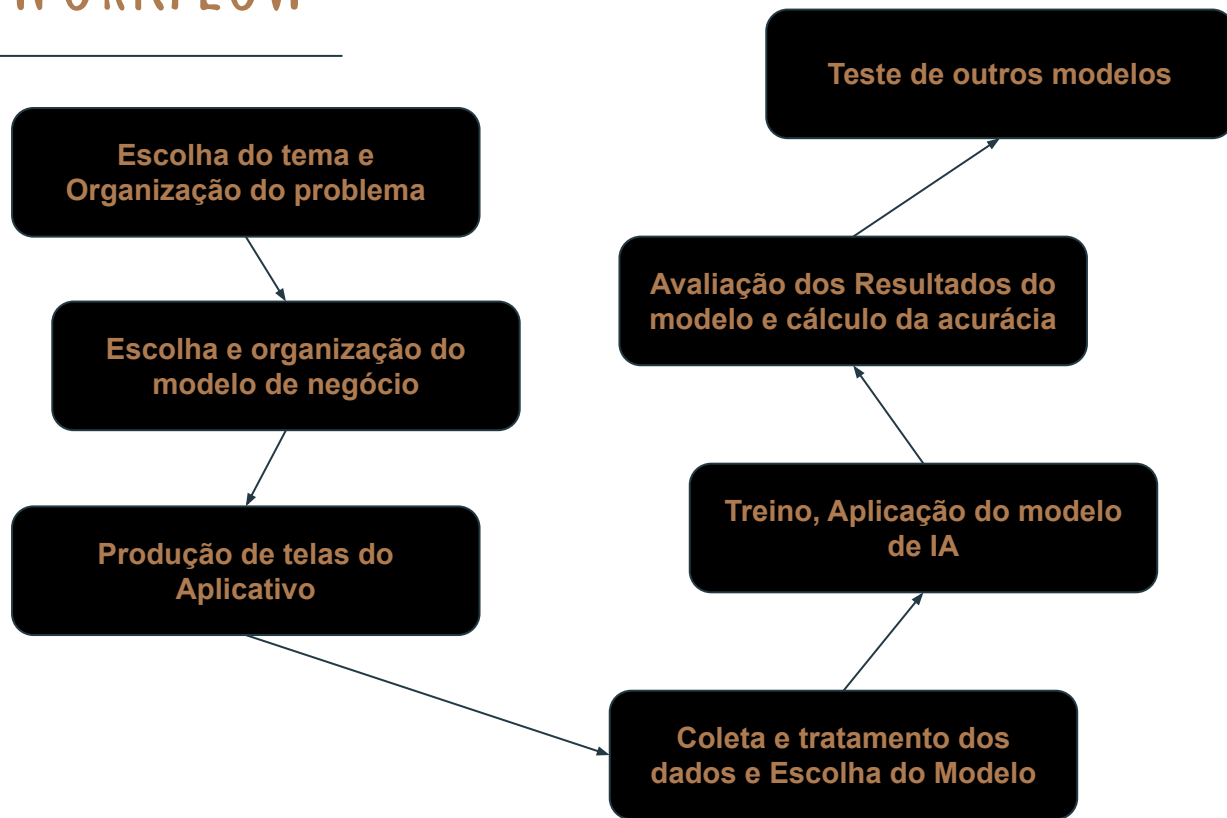
Formação: Ciência da computação

Nome: João Matheus

Contato:jm783311@gmail.com

Formação: Ciência da computação

WORKFLOW



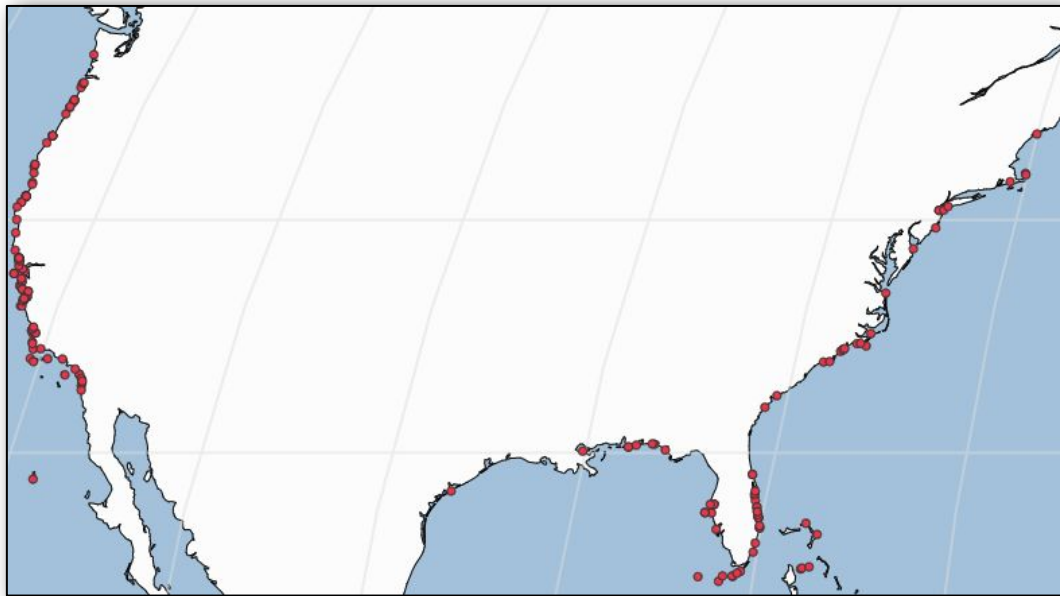
ENTENDIMENTO DO PROBLEMA E CONTEXTUALIZAÇÃO

Problema: Como podemos reduzir, com o uso de tecnologias inovadoras, os riscos de incidentes com tubarões?

Existem mais de **350 espécies** de tubarões, no entanto, apenas três delas são responsáveis pela maior parte das mortes: o **tubarão-branco**, o **tubarão-tigre** e o **tubarão-touro**. Segundo o dados do **ISAF** (International Shark Attack File), o EUA lidera o ranking de mais ataques de tubarões com 1640 ataques, Austrália em segundo com 706 e África do Sul em terceiro com 262.

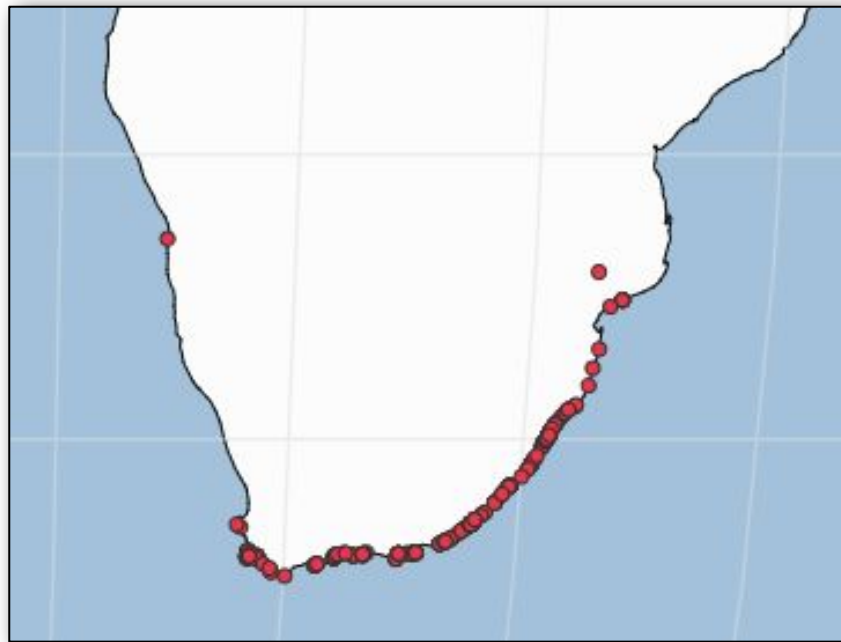
ENTENDIMENTO DO PROBLEMA E CONTEXTUALIZAÇÃO

ESTADOS UNIDOS

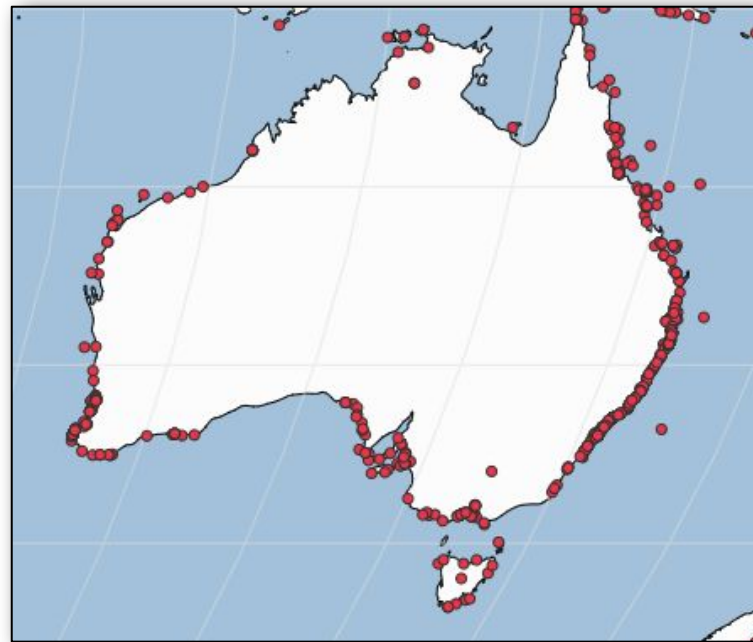


Florida representa 40% dos ataques
globais

ÁFRICA DO SUL



AUSTRÁLIA



Kwazulu, Cabo Oriental e Cabo Ocidental 40% dos ataques são fatais

O PROBLEMA

A ideia do projeto é, com ajuda de uma IA preditiva, **analisar dados de diferentes locais do mundo**, como, faixa etária das pessoas que sofreram o acidente, o tubarão que atacou, o tipo do ataque, a atividade que a pessoa estava fazendo, o sexo da pessoa, e com isso identificar os fatores que tornam um ataque mais propenso a ser fatal para que possamos criar estratégias de prevenção eficazes.

SOLUÇÃO DO PROBLEMA

Nós decidimos, depois de vários testes, usar o modelo de inteligência Artificial(IA), de Regressão Logística para prever a probabilidade de um ataque de tubarão ser fatal.

Com base no dataset analisado, nosso modelo foi treinado para identificar padrões em ataques fatais e não fatais. A solução proposta envolve:

- Análise dos dados:
 - Identificação de variáveis-chave, como atividade realizada, idade, e o tipo de ataque.
 - Uso de um modelo estatístico (Regressão logística) para separar fatalidades de ataques não fatais.

SOLUÇÃO DO PROBLEMA

- Recomendações Práticas:
 - Aplicar as previsões para criar sistemas de alerta baseados nas características de ataques mais propensos a serem fatais.
 - Por exemplo, alertar surfistas e banhistas em regiões e horários de maior risco ou durante atividades específicas.

Essa abordagem baseada em IA não só identifica os fatores de risco, mas também oferece suporte para intervenções preventivas, como campanhas de conscientização e alocação eficiente de recursos de segurança.

OUTROS TIPOS DE SOLUÇÕES PARA O PROBLEMA

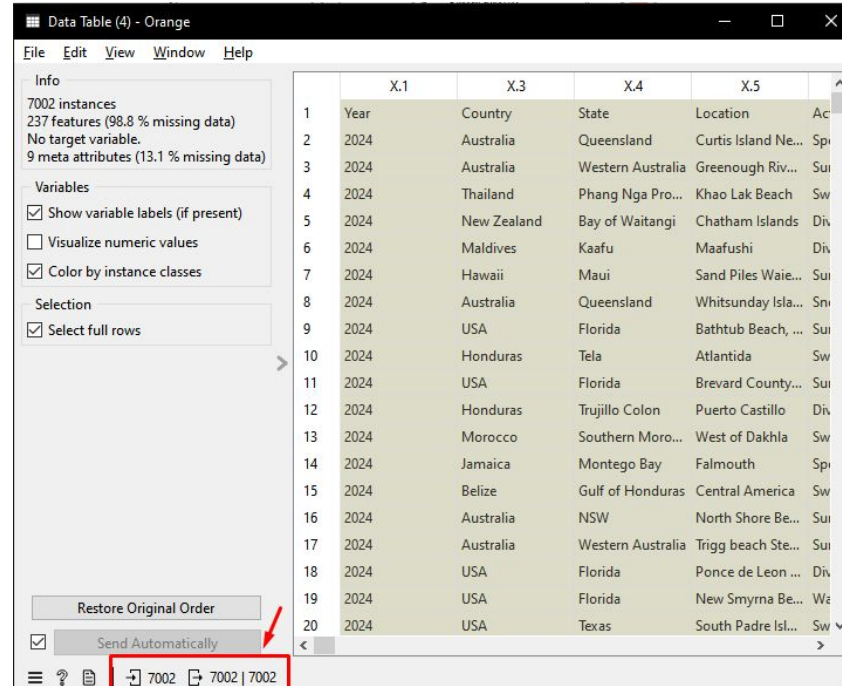
Uma outra forma de usar IA focado em ataques de tubarões é desenvolvendo um sistema de detecção inteligente integrado a drones! O governo de New South Wales, na Austrália, investiu milhões para que os drones ajudem a proteger os banhistas dos ataques de tubarões.

Pesquisadores criaram um modelo de aprendizado de máquina treinado para reconhecer 10 tipos de vidas marinhas, incluindo tubarões. Esse modelo foi incorporado em um aplicativo móvel que analisa imagens de drones em tempo real. O desafio desse projeto tem sido dimensionar esses sistemas para fazer uma diferença no mundo real, pois precisa de muita manutenção no software orientado a IA que foi usado.

DESENVOLVIMENTO DO PROJETO

- Procura do **Dataset:**
 - O nosso desenvolvimento no modelo, começou com a busca e tratamento dos dados. Tentamos procurar o dataset com mais informações possíveis para treinar o modelo de classificação com bastante informação, com o passar dos dias, vimos que seria muito complexo, e iria requerer muito poder computacional para tratar todos os acidentes registrados no dataset, eram mais de **7000 mil (seis mil)** dados para analisar, com isso reduzimos bastante o número de informação para o tratamento.

DESENVOLVIMENTO DO PROJETO



Data Table (4) - Orange

File Edit View Window Help

Info
7002 instances
237 features (98.8 % missing data)
No target variable.
9 meta attributes (13.1 % missing data)

Variables
☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection
☒ Select full rows

Restore Original Order

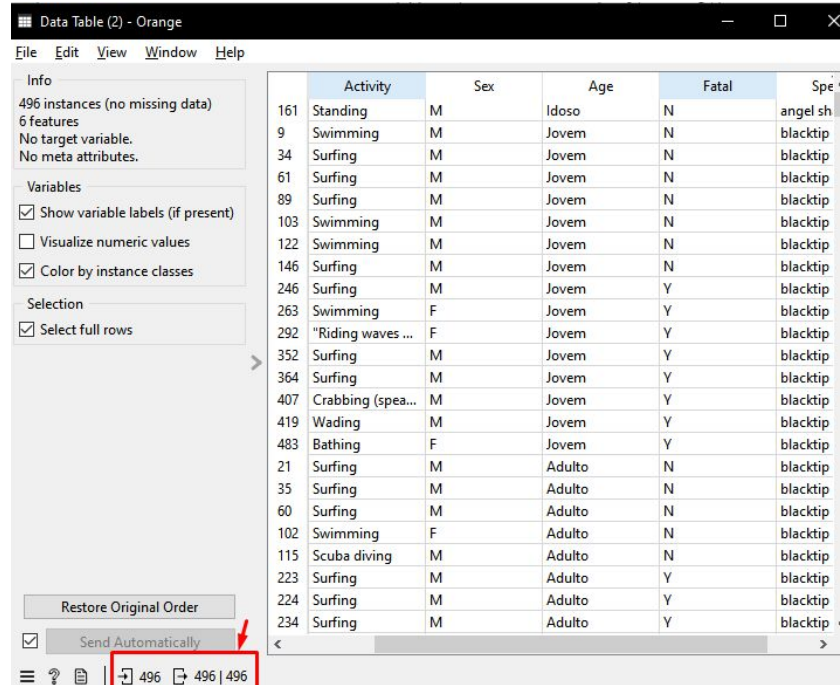
☒ Send Automatically

7002 | 7002

	X.1	X.3	X.4	X.5	
1	Year	Country	State	Location	Ac
2	2024	Australia	Queensland	Curtis Island Ne...	Sp
3	2024	Australia	Western Australia	Greenough Riv...	Su
4	2024	Thailand	Phang Nga Pro...	Khao Lak Beach	Sw
5	2024	New Zealand	Bay of Waitangi	Chatham Islands	Div
6	2024	Maldives	Kaafu	Maafushi	Div
7	2024	Hawaii	Maui	Sand Piles Waie...	Su
8	2024	Australia	Queensland	Whitsunday Isla...	Sn
9	2024	USA	Florida	Bathtub Beach, ...	Su
10	2024	Honduras	Tela	Atlantida	Sw
11	2024	USA	Florida	Brevard County...	Su
12	2024	Honduras	Trujillo Colon	Puerto Castillo	Div
13	2024	Morocco	Southern Moro...	West of Dakhla	Sw
14	2024	Jamaica	Montego Bay	Falmouth	Sp
15	2024	Belize	Gulf of Honduras	Central America	Sw
16	2024	Australia	NSW	North Shore Be...	Su
17	2024	Australia	Western Australia	Trigg beach Ste...	Su
18	2024	USA	Florida	Ponce de Leon ...	Div
19	2024	USA	Florida	New Smyrna Be...	Wa
20	2024	USA	Texas	South Padre Isl...	Sw

Dataset Antigo sem
tratamento

DESENVOLVIMENTO DO PROJETO



Data Table (2) - Orange

File Edit View Window Help

Info

496 instances (no missing data)
6 features
No target variable.
No meta attributes.

Variables

☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

Selection

☒ Select full rows

Restore Original Order

☒ Send Automatically

496 | 496

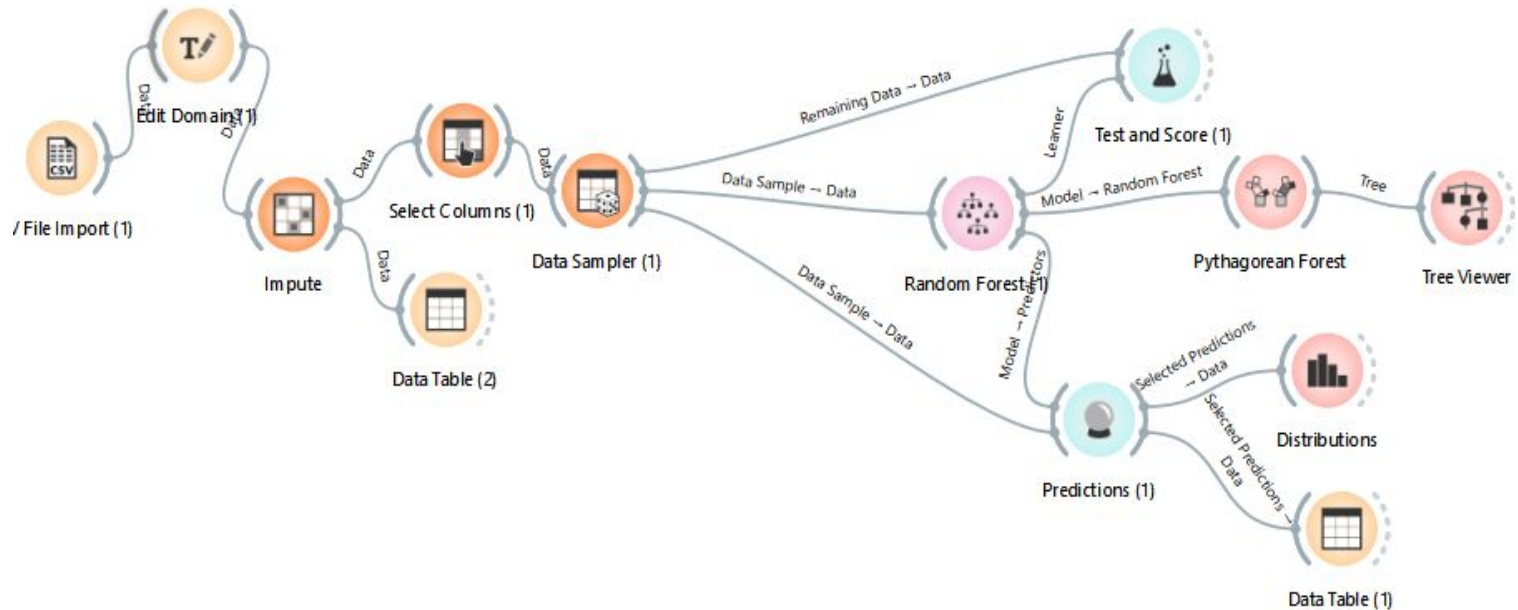
	Activity	Sex	Age	Fatal	Spe
161	Standing	M	Idoso	N	angel sh
9	Swimming	M	Jovem	N	blacktip
34	Surfing	M	Jovem	N	blacktip
61	Surfing	M	Jovem	N	blacktip
89	Surfing	M	Jovem	N	blacktip
103	Swimming	M	Jovem	N	blacktip
122	Swimming	M	Jovem	N	blacktip
146	Surfing	M	Jovem	N	blacktip
246	Surfing	M	Jovem	Y	blacktip
263	Swimming	F	Jovem	Y	blacktip
292	"Riding waves ...	F	Jovem	Y	blacktip
352	Surfing	M	Jovem	Y	blacktip
364	Surfing	M	Jovem	Y	blacktip
407	Crabbing (spea...	M	Jovem	Y	blacktip
419	Wading	M	Jovem	Y	blacktip
483	Bathing	F	Jovem	Y	blacktip
21	Surfing	M	Adulto	N	blacktip
35	Surfing	M	Adulto	N	blacktip
60	Surfing	M	Adulto	N	blacktip
102	Swimming	F	Adulto	N	blacktip
115	Scuba diving	M	Adulto	N	blacktip
223	Surfing	M	Adulto	Y	blacktip
224	Surfing	M	Adulto	Y	blacktip
234	Surfing	M	Adulto	Y	blacktip

Dataset novo e tratado

DESENVOLVIMENTO DO PROJETO

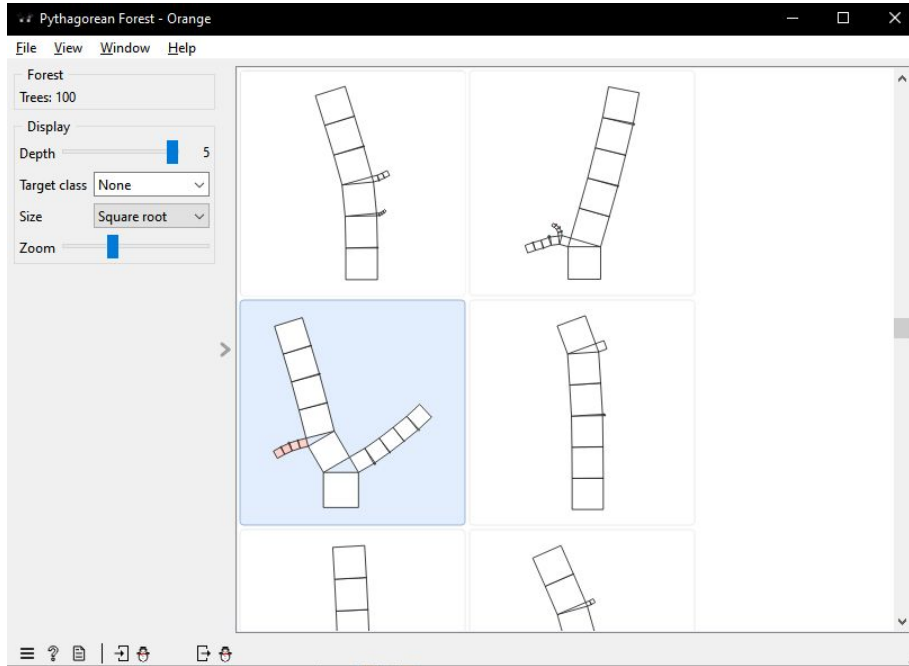
- Treinamento do modelo:
 - Após o tratamento dos dados, a escolha do modelo foi o próximo passo. Realizamos diversos testes com diferentes modelos até chegarmos à Regressão Logística, sempre buscando a melhor adequação ao nosso problema. Optamos por esse modelo porque é simples, eficiente e fácil de interpretar, ao contrário do Random Forest e do Gradient Boosting, que consideramos mais difíceis de interpretar. Utilizamos o Orange para realizar o treinamento e a análise dos resultados.

DESENVOLVIMENTO DO PROJETO

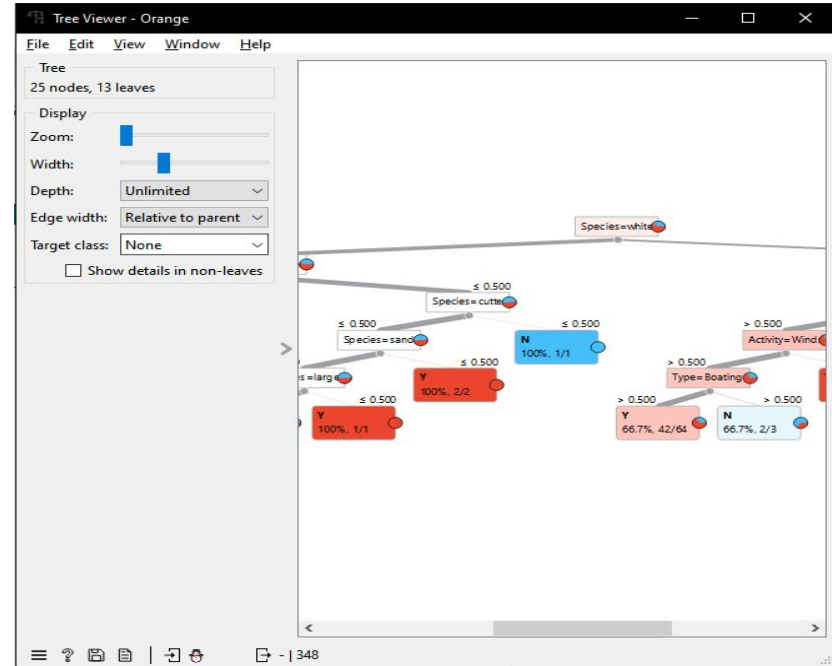


Random Forest no Orange

DESENVOLVIMENTO DO PROJETO

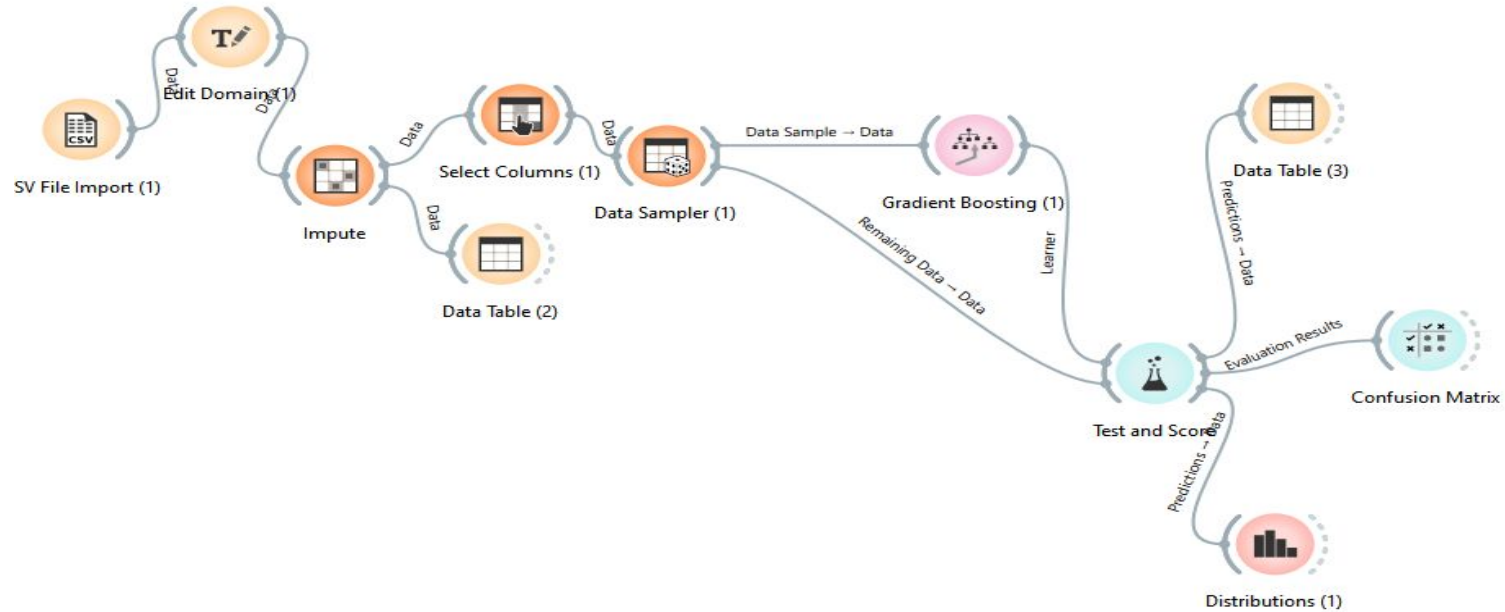


Random Forest e visualizador
Pythagorean Forest



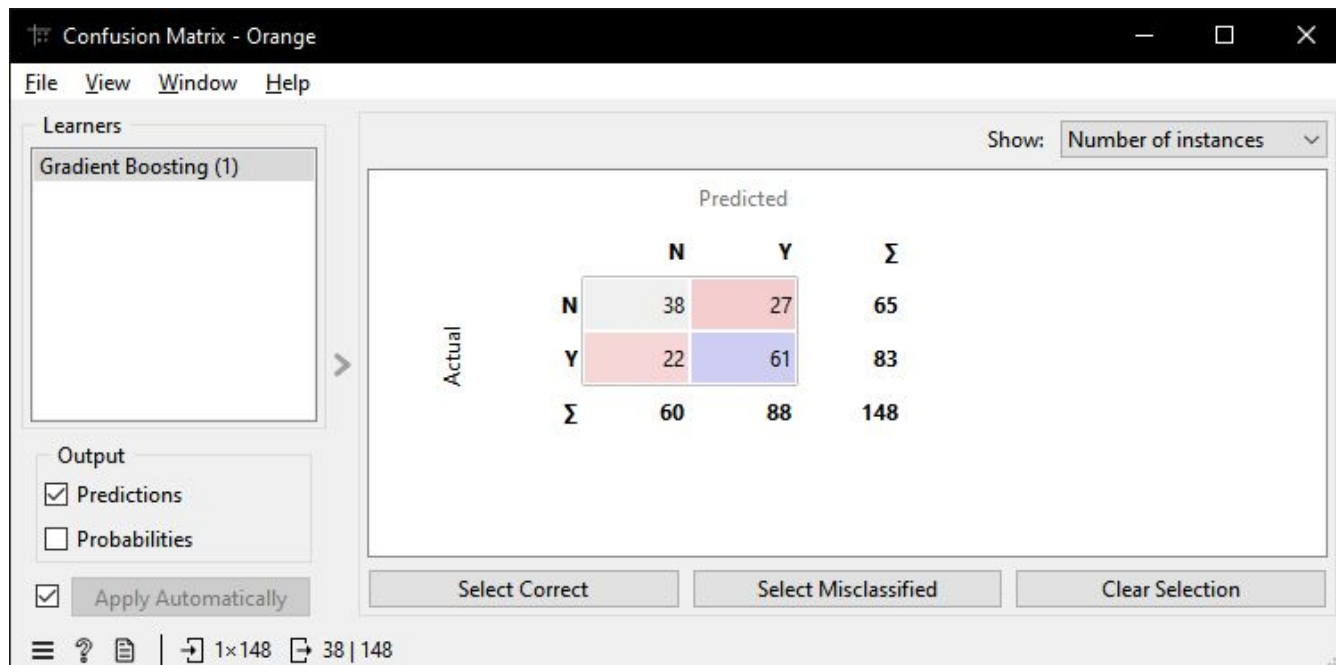
Pythagorean Forest e
visualizador Tree Viewer

DESENVOLVIMENTO DO PROJETO



Gradient Boosting no
Orange

DESENVOLVIMENTO DO PROJETO

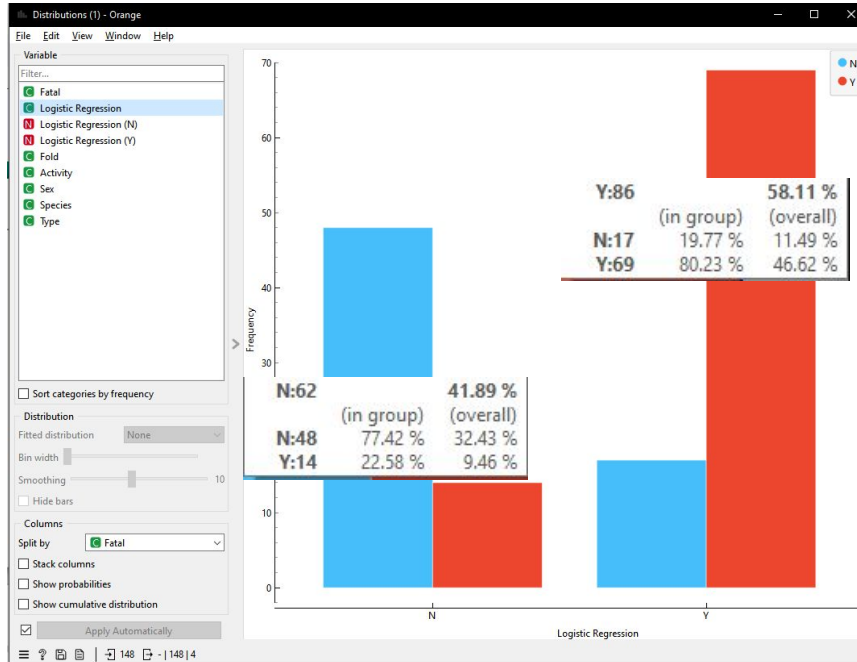


Gradient Boosting na matriz de confusão

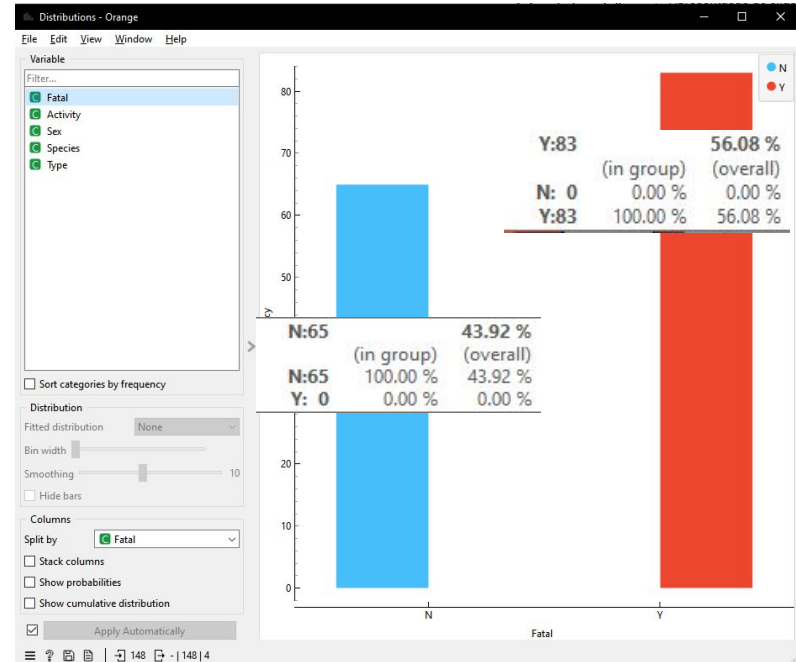
DESENVOLVIMENTO DO PROJETO

- Resultados do Modelo:
 - A seguir está o que conseguimos fazer com o modelo, após muitos testes, reajustes no banco de dados, troca da variável target, merged com variáveis que não correspondem ao tipo, remoção de valores nulos e em parâmetros do modelo como o uso do Ridge (R2), conseguimos chegar a um ponto satisfatório para a solução do problema.

DESENVOLVIMENTO DO PROJETO

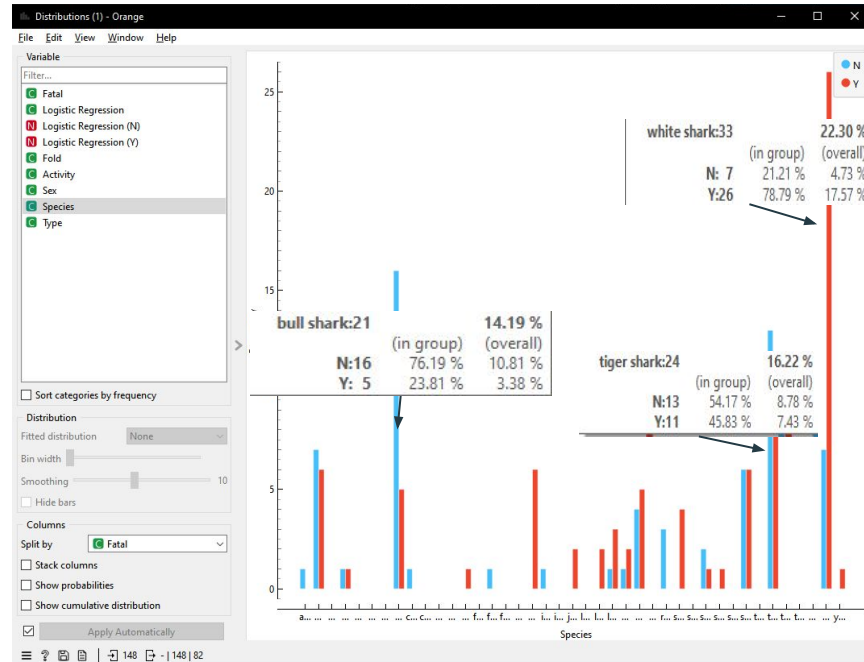


Casos Fatais **errados** e **acertados** pelo modelo



Banco de dados com os **valores reais**

DESENVOLVIMENTO DO PROJETO



Espécies que mais atacam segundo modelo que treinamos

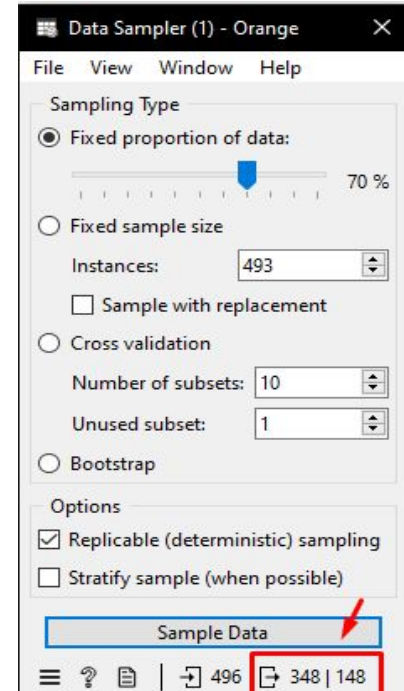
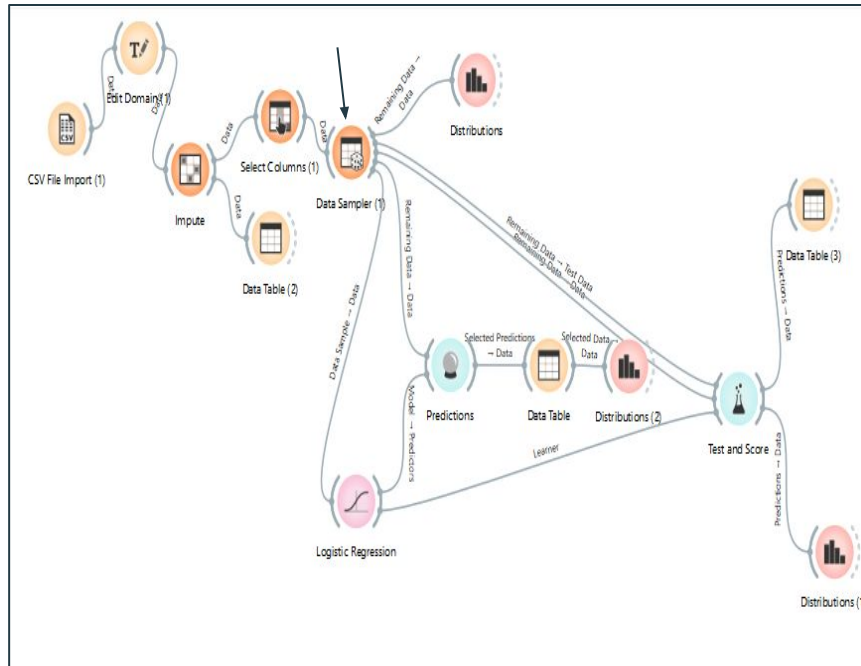
DESENVOLVIMENTO DO PROJETO

- “Teste e Score”:
 - No teste de desempenho do modelo que fizemos, depois de muitos testes e reajustes nos dados, tivemos uma nota que consideramos aceitável pelo tempo que restava.
 - No AUC conseguimos a maior nota do modelo, com isso ele pode separar melhor ataques fatais e não fatais.
 - No CA (Classification Accuracy), F1-Score, Precision, e Recall, conseguimos notas bem parecidas, e aceitáveis de quase **80%**, mas ainda com abertura para melhorias.

DESENVOLVIMENTO DO PROJETO

- “Teste e Score”:
 - E por fim o MCC (Matthews correlation coefficient) que leva em consideração verdadeiros e falsos positivos, verdadeiros e falsos negativos, e geralmente é considerado uma medida equilibrada para a performance do modelo.
 - Com essas notas e relacionando ao nosso problema, podemos dizer que é um modelo aceitável, que acerta uma boa parte das vezes os casos fatais e não fatais.

DESENVOLVIMENTO DO PROJETO



Separamos **70%** dos dados para treinamento e **30 %** para teste do modelo usando o **Data Sampler**

DESENVOLVIMENTO DO PROJETO

File Edit View Window Help

☐ Cross validation
Number of folds: 5
☐ Stratified

☐ Cross validation by feature
[Dropdown]

☐ Random sampling
Repeat train/test: 5
Training set size: 70 %
☒ Stratified

☐ Leave one out

☐ Test on train data

☒ Test on test data

Evaluation results for target (None, show average over classes) [Dropdown]

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.900	0.791	0.790	0.790	0.791	0.573

Compare models by: Area under RO [Dropdown] ☐ Negligible diff.: 0.1

Logistic Reg...	
Logistic Regression	

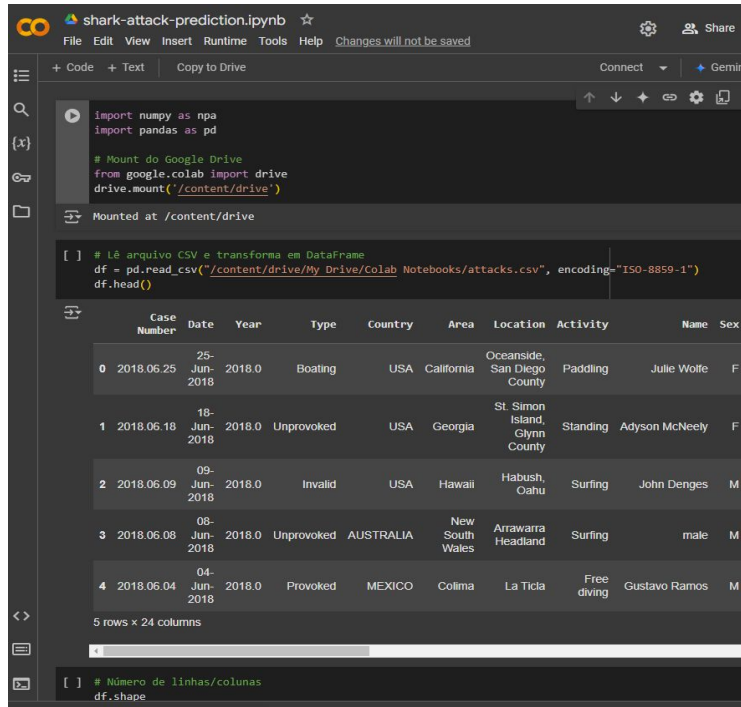
Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

Test and Score

DESENVOLVIMENTO DO PROJETO

- Alternativa com python:
 - A partir de um [notebook no Google Colab](#) que aborda o tratamento e a análise de um conjunto de dados de ataques de tubarões, bem como a aplicação de um modelo de classificação para prever se um ataque resultaria em um desfecho fatal ou não, conseguimos desde tratar os dados (Remoção de colunas irrelevantes, tratamento de valores nulos, padronização das informações, ajustes de inconsistências), até o treinamento de um modelo Random Forest, avaliação dos resultados e Cálculo de acurácia do modelo.

DESENVOLVIMENTO DO PROJETO



```
import numpy as npa
import pandas as pd

# Mount do Google Drive
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

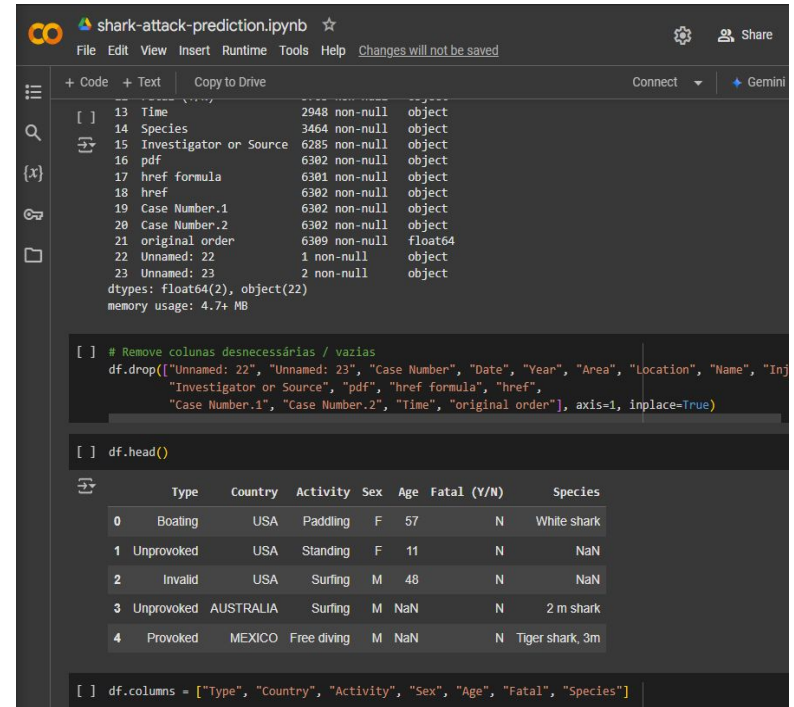
[ ] # Lê arquivo CSV e transforma em DataFrame
df = pd.read_csv("/content/drive/My Drive/Colab Notebooks/attacks.csv", encoding="ISO-8859-1")
df.head()
```

	Case Number	Date	Year	Type	Country	Area	Location	Activity	Name	Sex
0	2018.06.25	25-Jun-2018	2018.0	Boating	USA	California	Oceanside, San Diego County	Paddling	Julie Wolfe	F
1	2018.06.18	18-Jun-2018	2018.0	Unprovoked	USA	Georgia	St. Simon Island, Glynn County	Standing	Adyson McNeely	F
2	2018.06.09	09-Jun-2018	2018.0	Invalid	USA	Hawaii	Habush, Oahu	Surfing	John Denges	M
3	2018.06.08	08-Jun-2018	2018.0	Unprovoked	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	male	M
4	2018.06.04	04-Jun-2018	2018.0	Provoked	MEXICO	Colima	La Ticia	Free diving	Gustavo Ramos	M

5 rows x 24 columns

```
[ ] # Número de linhas/colunas
df.shape
```

Script no Google Colab



```
[ ] 13 Time 2948 non-null object
    14 Species 3464 non-null object
    15 Investigator or Source 6285 non-null object
    16 pdf 6302 non-null object
    17 href formula 6301 non-null object
    18 href 6302 non-null object
    19 Case Number.1 6302 non-null object
    20 Case Number.2 6302 non-null object
    21 original order 6309 non-null float64
    22 Unnamed: 22 1 non-null object
    23 Unnamed: 23 2 non-null object
dtypes: float64(2), object(22)
memory usage: 4.7+ MB

[ ] # Remove colunas desnecessárias / vazias
df.drop(["Unnamed: 22", "Unnamed: 23", "Case Number", "Date", "Year", "Area", "Location", "Name", "Investigator or Source", "pdf", "href formula", "href", "Case Number.1", "Case Number.2", "Time", "original order"], axis=1, inplace=True)

[ ] df.head()
```

	Type	Country	Activity	Sex	Age	Fatal (Y/N)	Species
0	Boating	USA	Paddling	F	57	N	White shark
1	Unprovoked	USA	Standing	F	11	N	NaN
2	Invalid	USA	Surfing	M	48	N	NaN
3	Unprovoked	AUSTRALIA	Surfing	M	NaN	N	2 m shark
4	Provoked	MEXICO	Free diving	M	NaN	N	Tiger shark, 3m

```
[ ] df.columns = ["Type", "Country", "Activity", "Sex", "Age", "Fatal", "Species"]
```

Removendo colunas desnecessárias

MODELO DE NEGÓCIO

Nosso negócio usará o Modelo Freemium, que consiste em disponibilizar o aplicativo gratuitamente, mas com funcionalidades premium pagas.

A versão gratuita incluirá informações básicas sobre riscos em áreas específicas e dicas gerais de prevenção.

Já a versão paga/Premium, irá oferecer análises personalizadas com base no perfil do usuário (idade, sexo, localidade, etc.) e alertas em tempo real sobre riscos locais.

MODELO DE NEGÓCIO

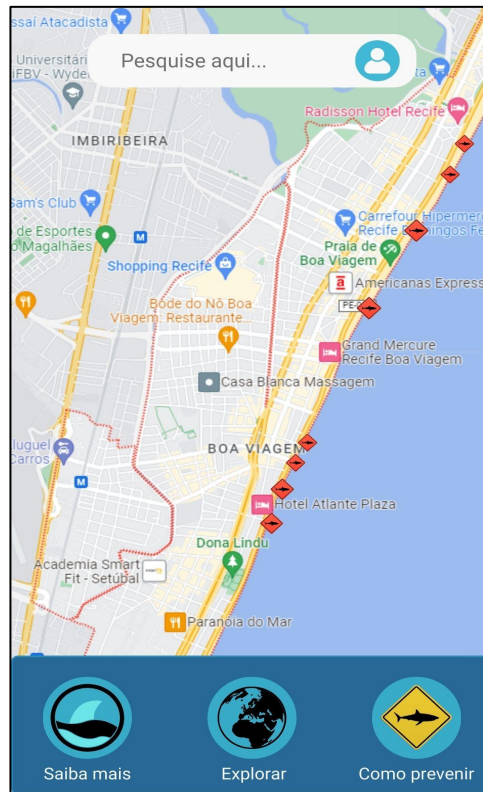
Outro tipo de modelo de negócio que também podemos usar com o modelo de classificação que treinamos é, desenvolver parcerias com prefeituras, órgãos governamentais e empresas de turismo que se beneficiam da redução de incidentes com tubarões.

Um exemplo é: Áreas costeiras que podem integrar o nosso sistema de IA como parte de políticas de segurança pública ou turismo sustentável.

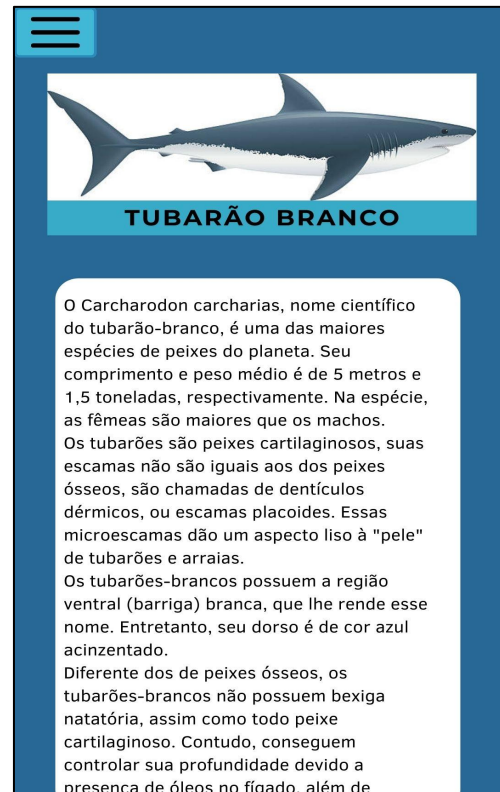
TELAS DO APP



INÍCIO

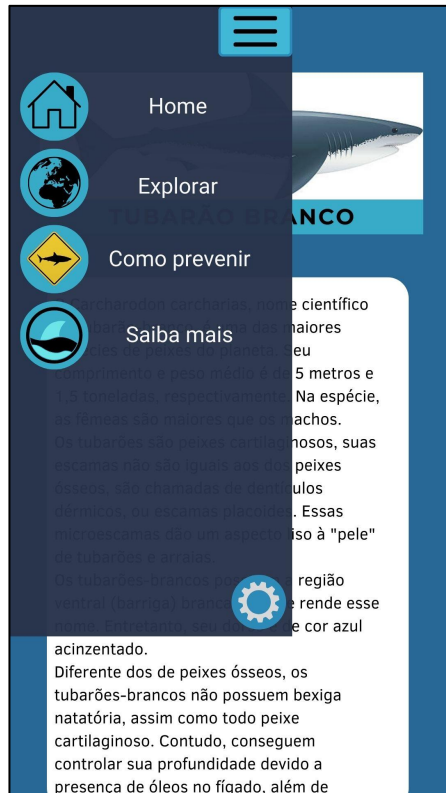


HOME

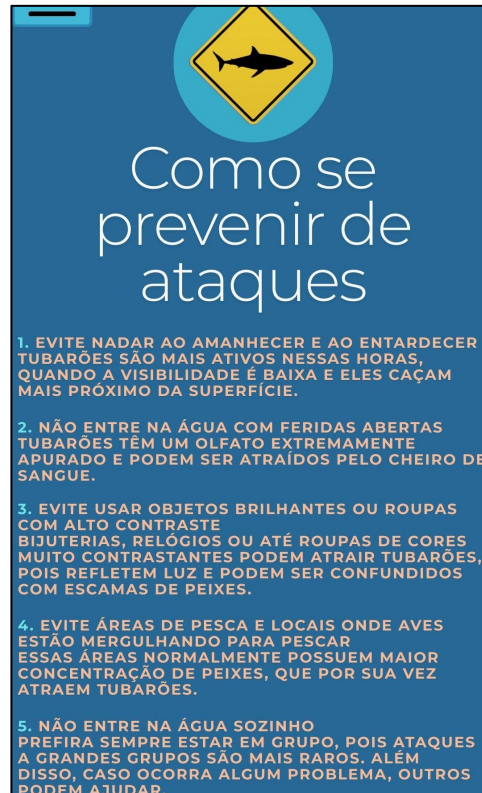


INFORMATIVO

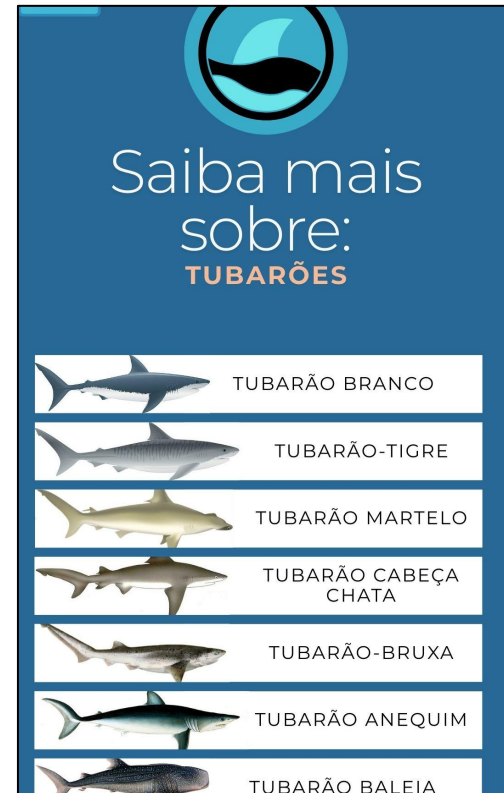
TELAS DO APP



MENU



PREVENÇÕES



TUBARÕES

TELAS DO APP

PREMIUM

Análise personalizada

IDADE:

Local atual:

Gênero

Análise Personalizada

PREMIUM

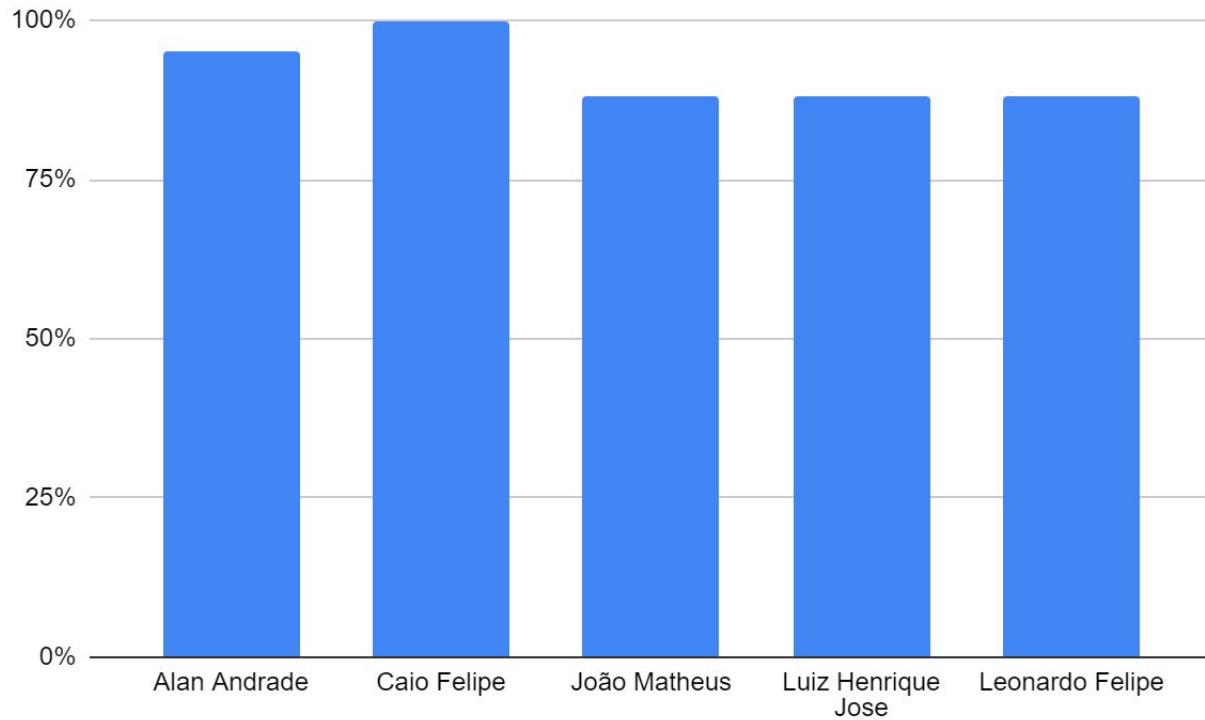
Notificações do dia

Acidente na praia de Boa Viagem

Acidente na praia de Boa Viagem

Notificações do dia

PORCENTAGEM DE PARTICIPAÇÃO



BIBLIOGRAFIA

Registros do ISAF:

<https://www.floridamuseum.ufl.edu/shark-attacks/yearly-worldwide-summary>

Estatísticas e banco de dados: <https://www.sharkattackfile.net/>

Sobre tubarões: <https://www.sharkattackfile.net/species.htm>

Drones e IA:

<https://forbes.com.br/forbes-tech/2023/01/os-drones-e-a-inteligencia-artificial-podem-nos-proteger-dos-tubaroos/>

Orange: <https://orangedatamining.com/widget-catalog/evaluate/testandscore/>

Github: <https://github.com/Tantataito/Shark-Attack-predictions>