



Sistema de Monitoramento

Detecção de Derramamento de óleo em imagens

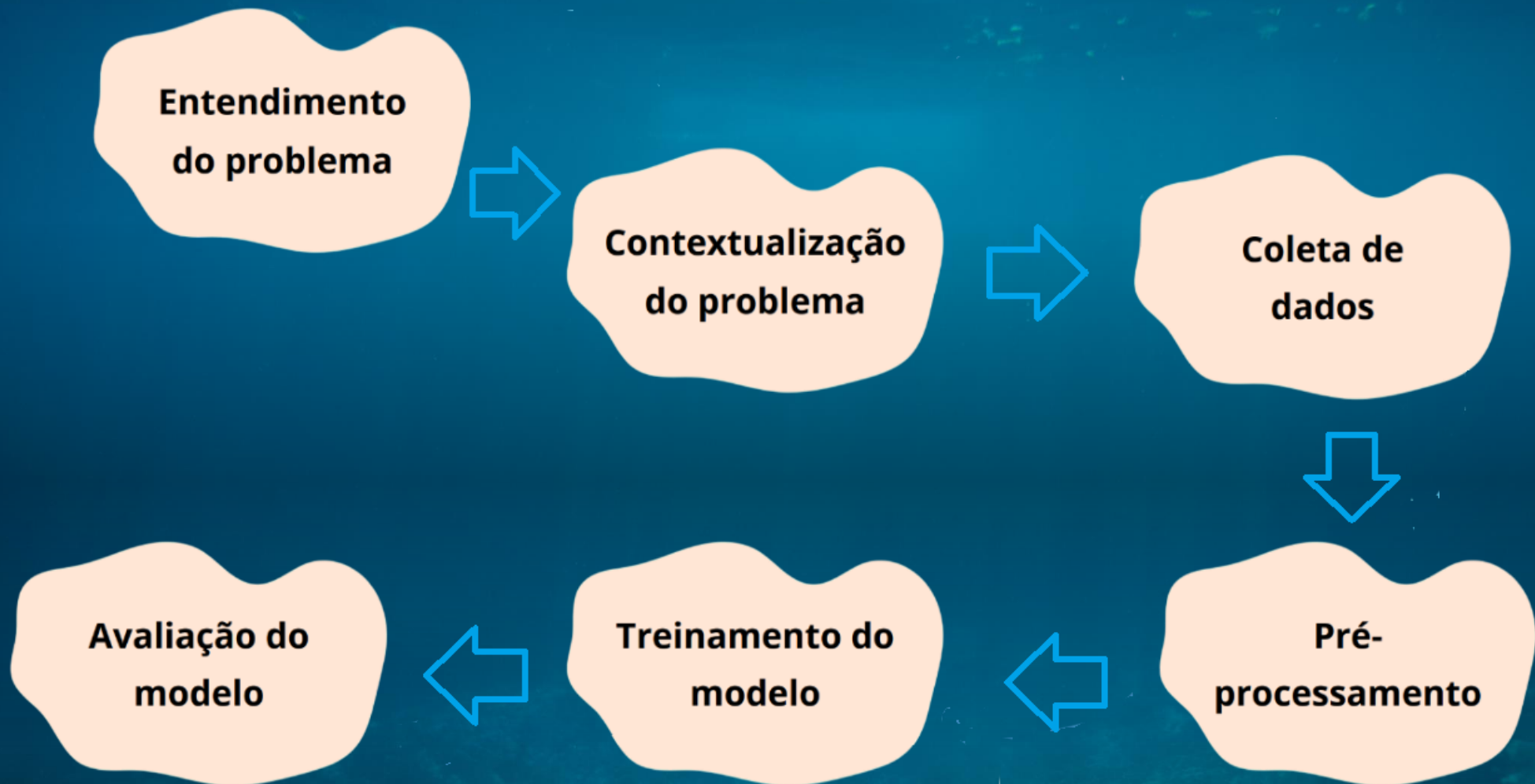
Integrantes

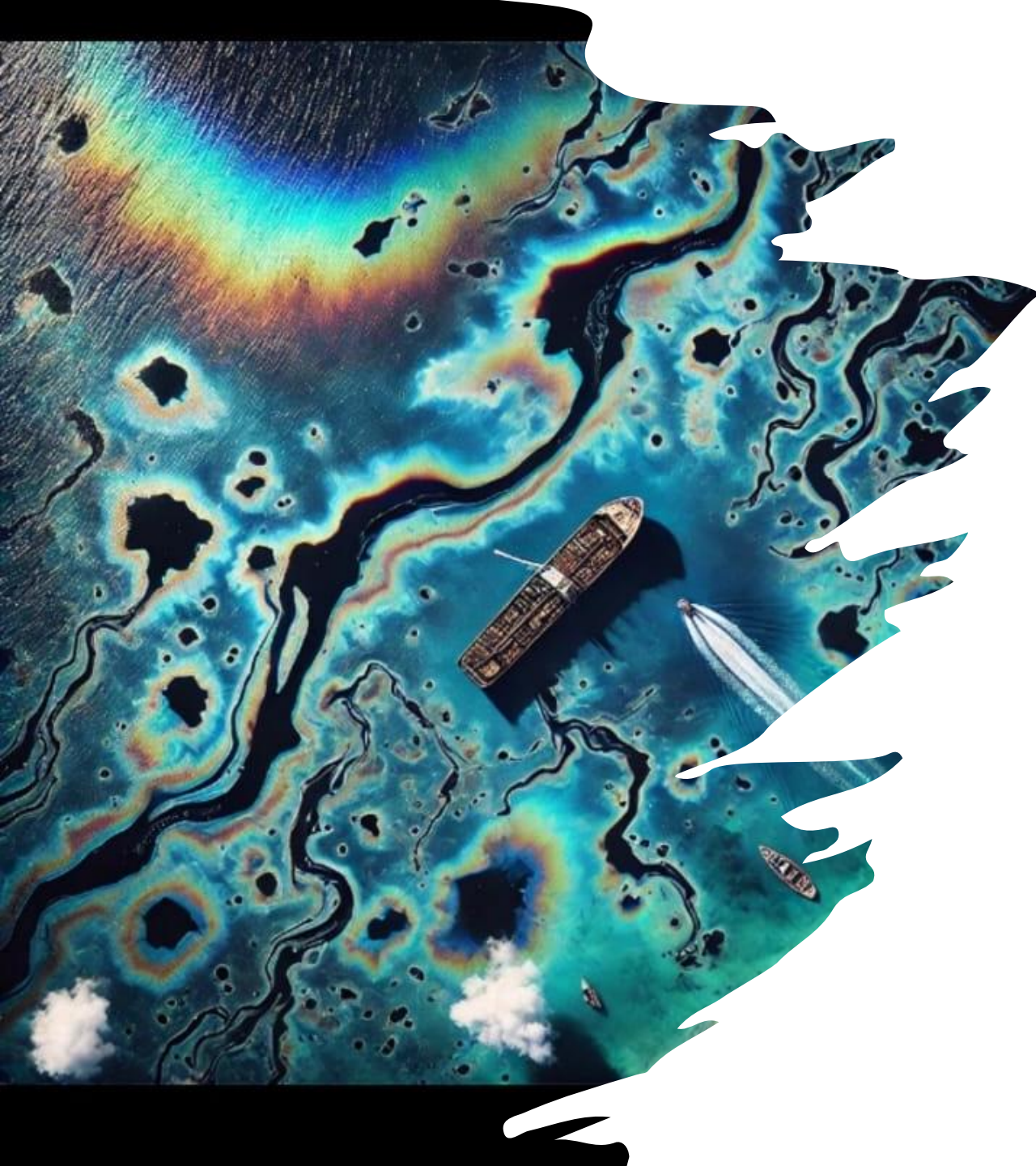


Guilherme Augusto



Marcos Neves





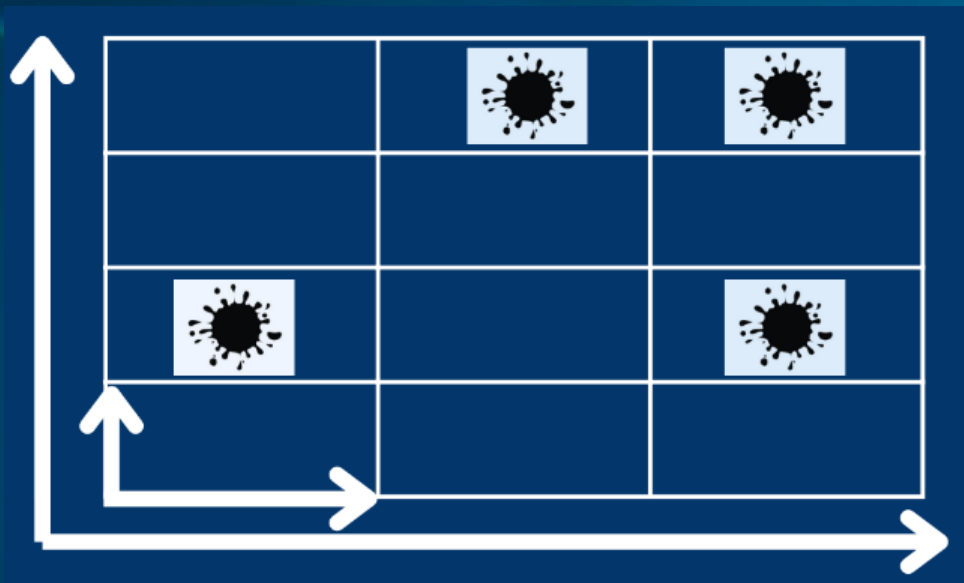
Entendimento do Problema

- Os vazamentos de petróleo são eventos ambientais desastrosos que tem impactos significativos no ecossistema marinho e nas comunidades dependentes. Tornando a recuperação desses ecossistemas um desafio complexo onde requer uma abordagem multidisciplinar.
- Já no estudo de Impactos do Vazamento de óleo no Nordeste (Prof. Marcelo Soares e Prof. Emanuelle Fontenele Rabelo, 2021), destaca os impactos econômicos e sociais do derramamento de óleo, afetando principalmente a população litorânea e a economia local baseada no turismo e na pesca como duradouros, necessitando apoio financeiro e políticas públicas para ajudar a comunidade a se recuperar.

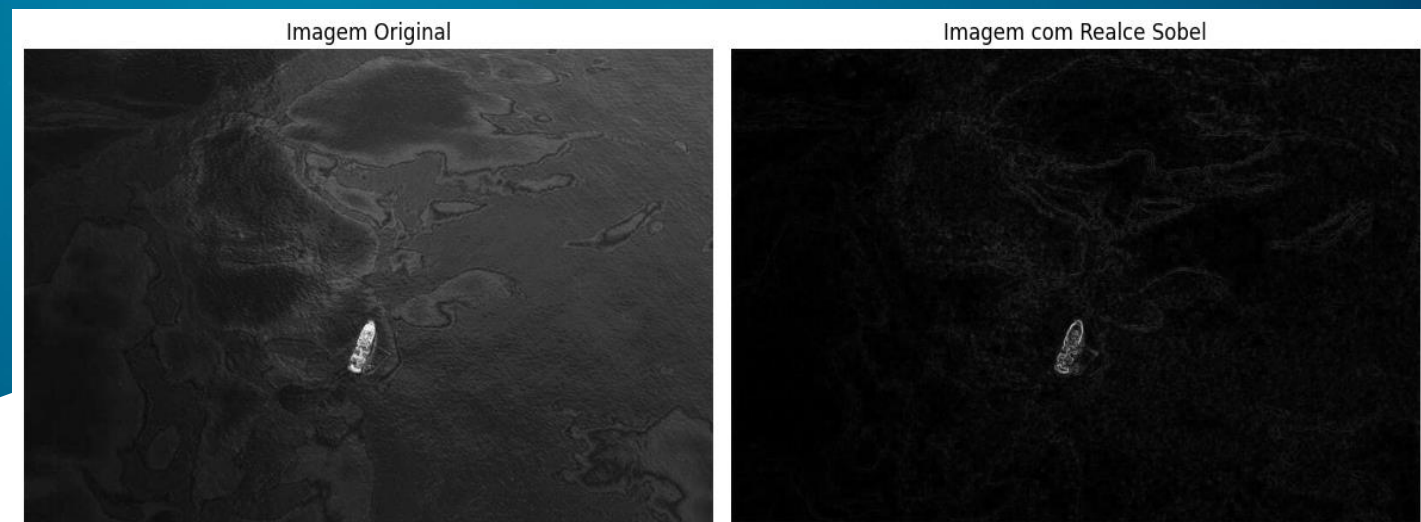


Contextualização do Problema

- O derramamento de óleo no oceano é um problema sério e representa uma ameaça significativa à vida marinha, alguns dos quais registrados historicamente como acontecimentos catastróficos em que cada evento causou um desafio contínuo para esse ecossistema e as comunidades litorâneas. Tendo como solução, a necessidade de uma abordagem dedicada para a prevenção e mitigação.
- O estudo sobre os Impactos Ambientais do Derramamento de Óleo na costa do Nordeste do Brasil, sugere que efeitos agudos, como no incidente de agosto de 2019, que afetou cerca de 1009 localidades em 130 municípios do Maranhão ao litoral norte do Rio de Janeiro, dura por vários anos até que o clima seja recuperado. O estudo prevê a recuperação completa em aproximadamente 10 anos; destacando a necessidade de medidas mais eficazes de resposta.



- As imagens de satélite são divididas em menores chamadas patches
- Suponha que uma imagem de satélite tenha uma resolução de 4000 x 4000 pixels.
- Patches de $(4000/100 = 40)$. Se quisermos dividir essa imagem em patches de 100 x 100 pixels, a imagem será dividida em 40 x 40 patches $(4000/100 = 40)$



Coleta dos Dados

Pre-processamento dos Dados

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_10	...	f_40	f_41	f_42	f_43	f_44	f_45	f_46	f_47	f_48	f_49
0	-1.240922	1.152390	1.346434	-0.793007	0.129657	1.469091	-0.185871	-0.345107	3.165389	-0.343460	...	0.611105	1.913877	0.800597	0.950757	0.216514	-0.255448	-0.383248	3.686767	0.388730	-0.058377
1	-1.225524	11.389546	-1.033273	-0.057342	2.114766	14.374844	0.618905	-2.207407	7.100184	-2.226754	...	0.611105	4.810555	15.485710	17.486286	11.381341	-0.878152	-0.383248	6.362181	0.387769	-0.639664
2	-1.210126	-0.112818	1.252645	-0.502492	0.085544	-0.125929	-0.222058	-0.498440	-0.073589	-0.454242	...	0.611105	0.465538	-0.248340	-0.197438	-0.449905	0.858654	2.609278	3.460466	0.395456	-0.096212
3	-1.194727	0.449611	1.440556	-1.101091	-0.399705	0.583114	-0.066295	-0.322804	1.725979	-0.343460	...	0.611105	5.101741	0.467147	0.370349	0.286675	1.654442	2.609278	4.336762	0.382004	-0.017102
4	-1.179329	-0.010794	0.419520	-0.823188	-1.039352	0.002691	-0.142604	-0.584864	-0.072364	-0.565024	...	0.611105	0.385669	0.395889	0.481449	0.022483	-0.484237	-0.383248	3.072971	0.381043	-0.264751

5 rows x 49 columns

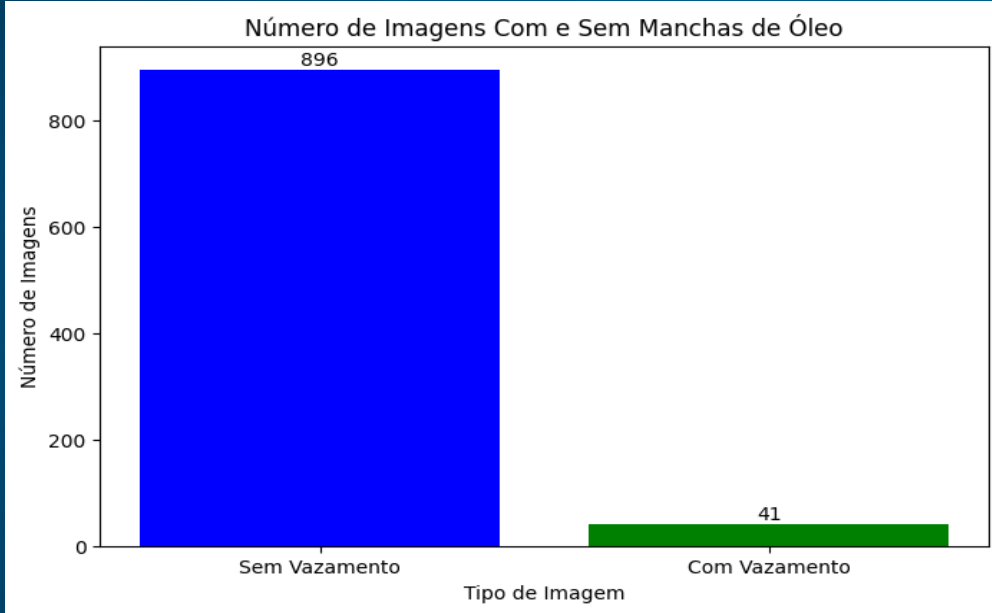
Legenda:

F_1 : Reflexão de Radar (SAR);
 F_2 : Intensidade de Pixel;
 F_3 : Cor (Banda RGB 1);
 F_4 : Cor (Banda RGB 2);
 F_5 : Cor (Banda RGB 3);
 F_6 : Índice de Vegetação por Diferença Normalizada (NDVI);
 F_7 : Índice de Água por Diferença Normalizada (NDWI);
 F_8 : Média da Intensidade de Pixel;
 F_9 : Desvio Padrão da Intensidade de Pixel;
 F_10 : Curtose;
 F_11 : Assimetria;
 F_12 : Média da Reflexão de Radar;
 F_13 : Desvio Padrão da Reflexão de Radar;
 F_14 : Curto da Reflexão de Radar;
 F_15 : Assimetria da Reflexão de Radar;
 F_16 : Textura (Características de Haralick - Contraste);
 F_17 : Textura (Características de Haralick - Dissimilaridade);

F_18 : Textura (Características de Haralick - Homogeneidade);
 F_19 : Textura (Características de Haralick - Energia);
 F_20 : Textura (Características de Haralick - Correlação);
 F_21 : Variância;
 F_22 : Entropia;
 F_23 : Área da Mancha;
 F_24 : Perímetro da Mancha;
 F_25 : Raio Médio da Mancha;
 F_26 : Diâmetro Equivalente;
 F_27 : Excentricidade;
 F_28 : Circularidade;
 F_29 : Rugosidade;
 F_30 : Albedo;
 F_31 : Reflexão de Luz Visível;
 F_32 : Reflexão de Infravermelho;
 F_33 : Reflexão de Microondas;

F_34 : Reflexão de Ultravioleta;
 F_35 : Reflexão de Radar de Alta Resolução;
 F_36 : Reflexão de Radar de Baixa Resolução;
 F_37 : Reflexão de Radar de Média Resolução;
 F_38 : Reflexão de Radar de Alta Frequência;
 F_39 : Reflexão de Radar de Baixa Frequência;
 F_40 : Reflexão de Radar de Frequência Média;
 F_41 : Reflexão de Radar de Alta Potência;
 F_42 : Reflexão de Radar de Baixa Potência;
 F_43 : Reflexão de Radar de Potência Média;
 F_44 : Reflexão de Radar de Alta Sensibilidade;
 F_45 : Reflexão de Radar de Baixa Sensibilidade;
 F_46 : Reflexão de Radar de Sensibilidade Média;
 F_47 : Reflexão de Radar de Alta Precisão;
 F_48 : Reflexão de Radar de Baixa Precisão;
 F_49 : Reflexão de Radar de Precisão Média.

Pre-processamento dos Dados



80%

Dado de treino

20%

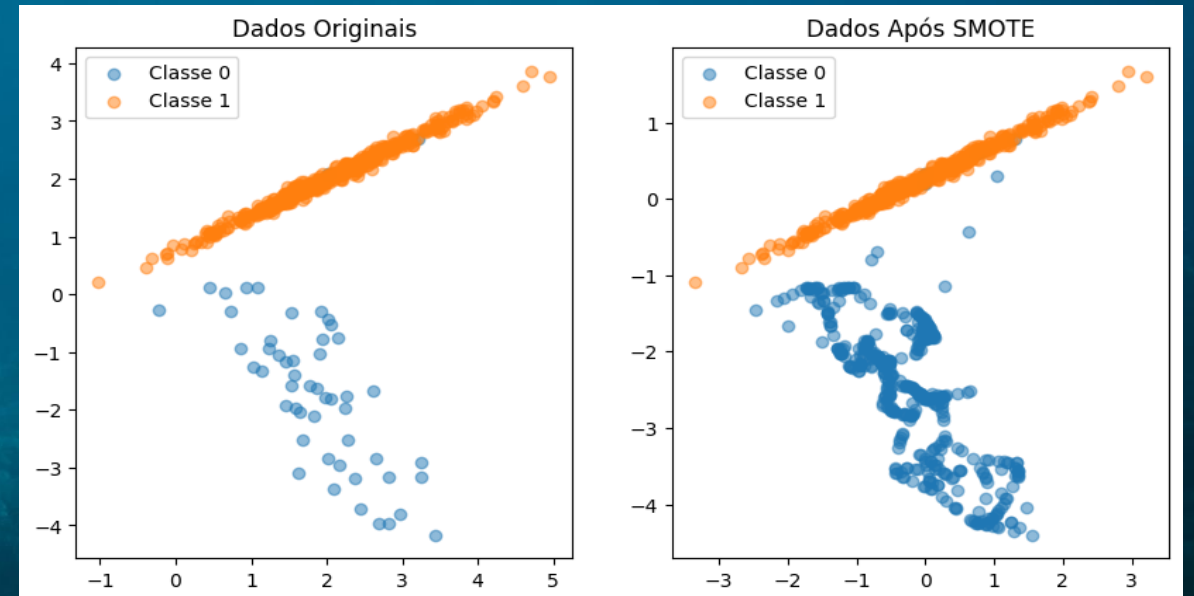
Dado de Teste

StandardScaler: Normaliza as variáveis independentes para que todas estejam na mesma escala (média 0 e desvio padrão 1).

Criar uma instância de StandardScaler
scaler = StandardScaler()

Ajustar e transformar os dados
X_normalized = scaler.fit_transform(X)

Converter de volta para um DataFrame com os nomes das colunas originais
X = pd.DataFrame(data=X_normalized, columns=X.columns)



Modelos Usados

Rede Neural

Projetada para conter rótulos de dados de entrada com base em suas características. Esse tipo de tarefa é comum no reconhecimento de imagem, classificação de texto, entre outros.

Regressão Logística

Modelo de classificação estatístico que prevê a probabilidade de um evento acontecer.

Árvore de decisão

Algoritmo de aprendizado de máquina supervisionado que é utilizado para classificação e para regressão.

Árvore de Decisão

Impureza de Gini

Ela fornece mais informações sobre a distribuição de dados por nó do que a precisão de classificação usada para relatar a precisão da árvore.

Ganho de informação

Mede a redução de entropia após uma divisão

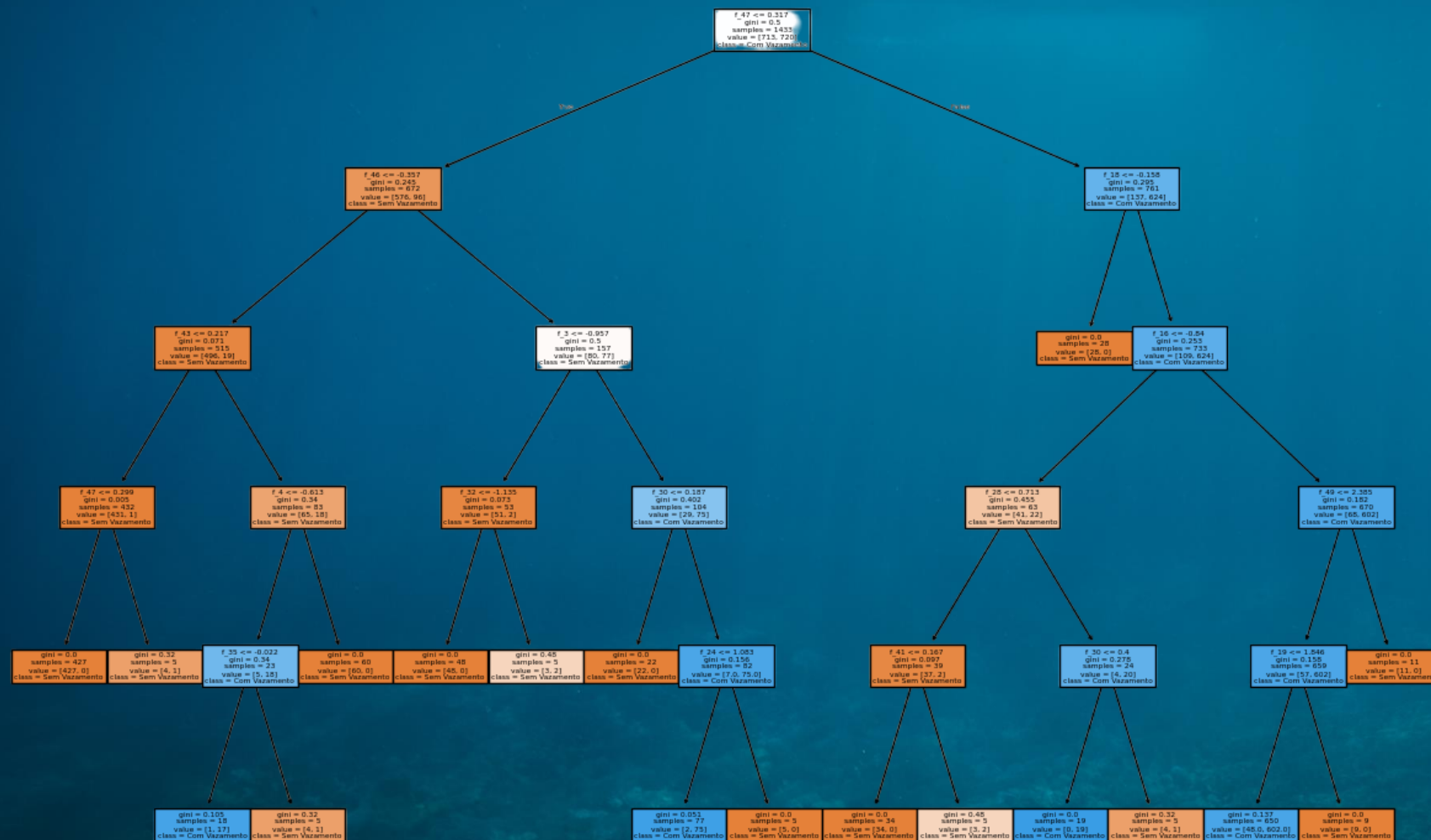
Entropia

A entropia de Shannon mede a incerteza ou a quantidade de informação contida em uma variável aleatória.

Cálculos de divisão

A árvore decide onde dividir com base em valores numéricos ou categorias para minimizar as impurezas

Diagrama da árvore e nós promissores



Nós com a Condição $F_3 \leq -0.957$:

- Índice de Gini: 0.0
- Amostras: 17
- Valores: [0, 17]
- Classe Predominante: Com Vazamento

Nós com a Condição $F_{38} \leq -0.138$:

- Índice de Gini: 0.0
- Amostras: 24
- Valores: [0, 24]
- Classe Predominante: Com Vazamento

Nós com a Condição $F_{43} \leq 0.217$:

- Índice de Gini: 0,215
- Amostras: 77
- Valores: [68, 9]
- Classe Predominante: Sem Vazamento

Legenda:

F_3 : Cor (Banda RGB 1);

F_{38} : Reflexão de Radar de Alta Frequência;

F_{43} : Reflexão de Radar de Potência Média.

Metricas para Avaliação

Precision

A precisão é a razão entre verdadeiros positivos (TP) e o total de predições positivas (TP + FP). Indica apenas as instâncias em que o modelo anterior é realmente positivo .

Fórmula: Precisão = $TP / (TP + FP)$

Recall

Também conhecido como sensibilidade, é a razão entre verdadeiros positivos (TP) e o total de instâncias reais positivas (TP + FN). Indica quantas instâncias positivas reais o modelo conseguiu capturar .

Fórmula :Recall = $TP / (TP + FN)$

F1-score

É a média harmônica entre a precisão e o recall. Ele fornece uma única métrica que combina ambos os aspectos, sendo útil quando há um desequilíbrio entre as classes.

Fórmula: F1-Score = $2 * (Precisão * Recall) / (Precisão + Recall)$

Accuracy

A acurácia é a razão entre o número de predições corretas e o total de predições.

Fórmula :Acurácia = $(TP + TN) / (TP + TN + FP + FN)$

AVALIAÇÃO DE MODELOS

1. Árvore de Decisão

Relatório de classificação do Árvore de Decisão:				
	precision	recall	f1-score	support
0	0.99	0.88	0.93	183
1	0.89	0.99	0.94	176
accuracy			0.93	359
macro avg	0.94	0.93	0.93	359
weighted avg	0.94	0.93	0.93	359

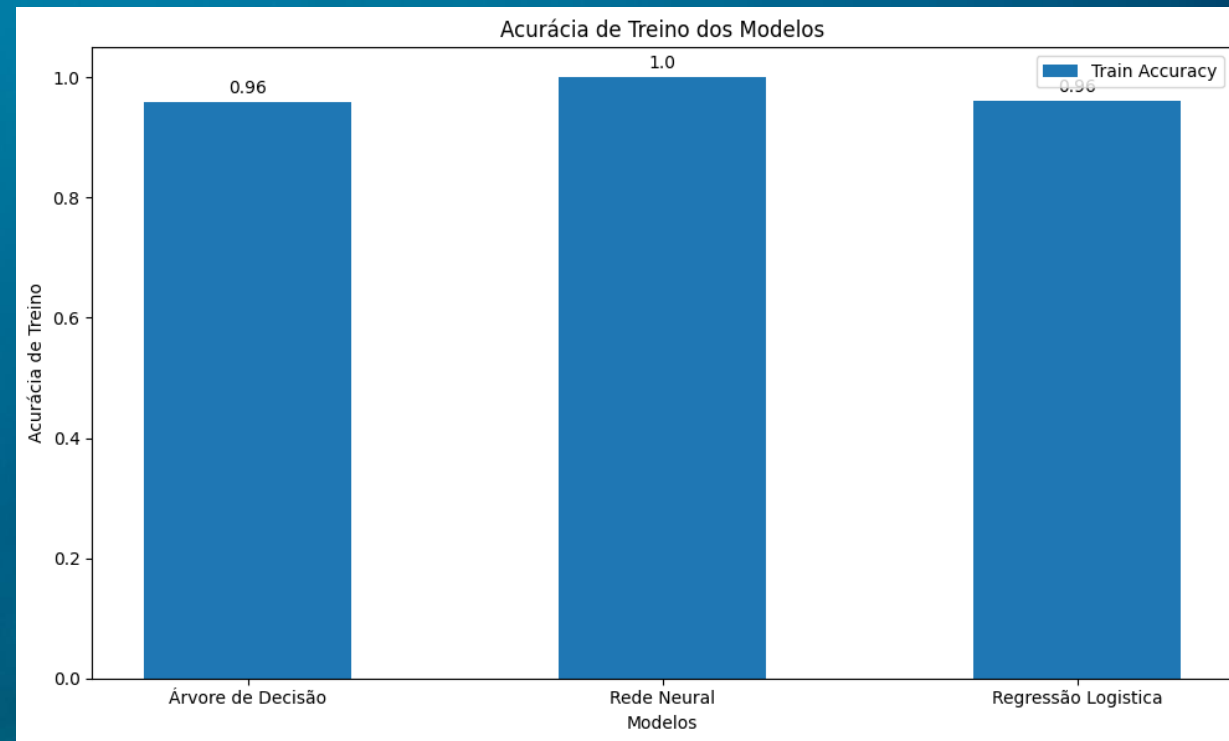
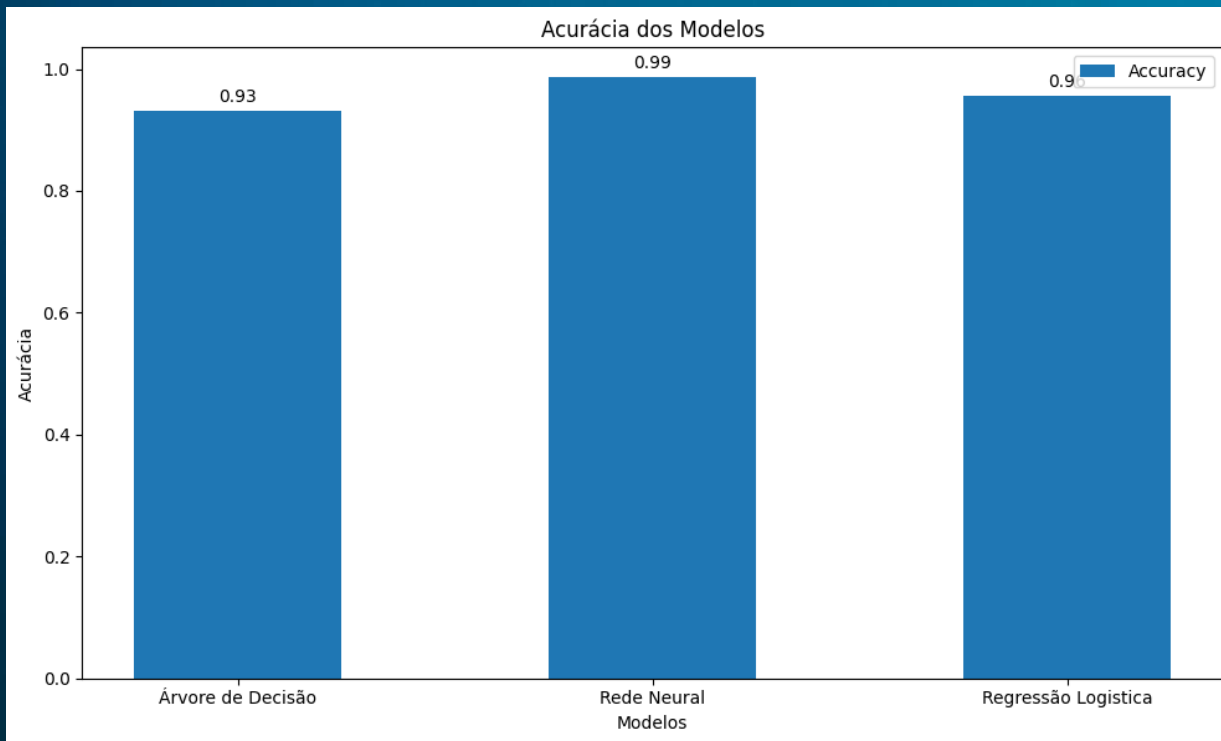
2. Rede neural

Relatório de classificação do Regressão Logística:				
	precision	recall	f1-score	support
0	0.99	0.92	0.95	183
1	0.93	0.99	0.96	176
accuracy			0.96	359
macro avg	0.96	0.96	0.96	359
weighted avg	0.96	0.96	0.96	359

3. Regressão logística

Relatório de classificação do Rede Neural:				
	precision	recall	f1-score	support
0	1.00	0.97	0.99	183
1	0.97	1.00	0.99	176
accuracy			0.99	359
macro avg	0.99	0.99	0.99	359
weighted avg	0.99	0.99	0.99	359

Acurácia dos modelos



Reflexão final

<https://colab.research.google.com/drive/1SL8Wg4w7nWkePKHYn3lLAyVWAbCKaJLp?usp=sharing>

- Com base nos resultados acima, a Rede Neural (MLP) parece ser o melhor modelo para este conjunto de dados devido à sua alta acurácia (99%) e excelente equilíbrio entre precisão e revogação para ambas as classes. No entanto, vale lembrar que as Redes Neurais são mais complexas e podem exigir mais recursos computacionais.

Melhor Modelo:

- *Se você busca interpretabilidade e um modelo mais simples, a Árvore de Decisão pode ser uma boa escolha. Já a Regressão Logística pode ser útil se você precisar de um modelo linear e menos complexo.*