



Universidade Federal do Piauí - UFPI  
Departamento de Computação/CNN

Trabalho da Disciplina de Aprendizado de Máquina

# **IMPLEMENTAÇÃO DE INTELIGÊNCIA ARTIFICIAL UTILIZANDO MÁQUINAS DE SUPORTE DE VETORES (SVM) NA CLASSIFICAÇÃO DE VIDROS**

Professor: Vinicius Ponte Machado

Alunos: Marcelo Pontes Rodrigues

Marcelo de Oliveira Rego

Novembro/2024

# Explicando os dados

## Conjunto de Dados de Identificação de Vidro

- Fonte: UCI (Universidade da Califórnia, Irvine)

## Descrição dos Atributos

- Número de Instâncias: 214
- Contém 10 atributos, incluindo o id, todos os atributos possuem valores contínuos.
- A resposta é o tipo de vidro (discreto, com 7 valores).
- Sem valores ausentes.

# Explicando os dados

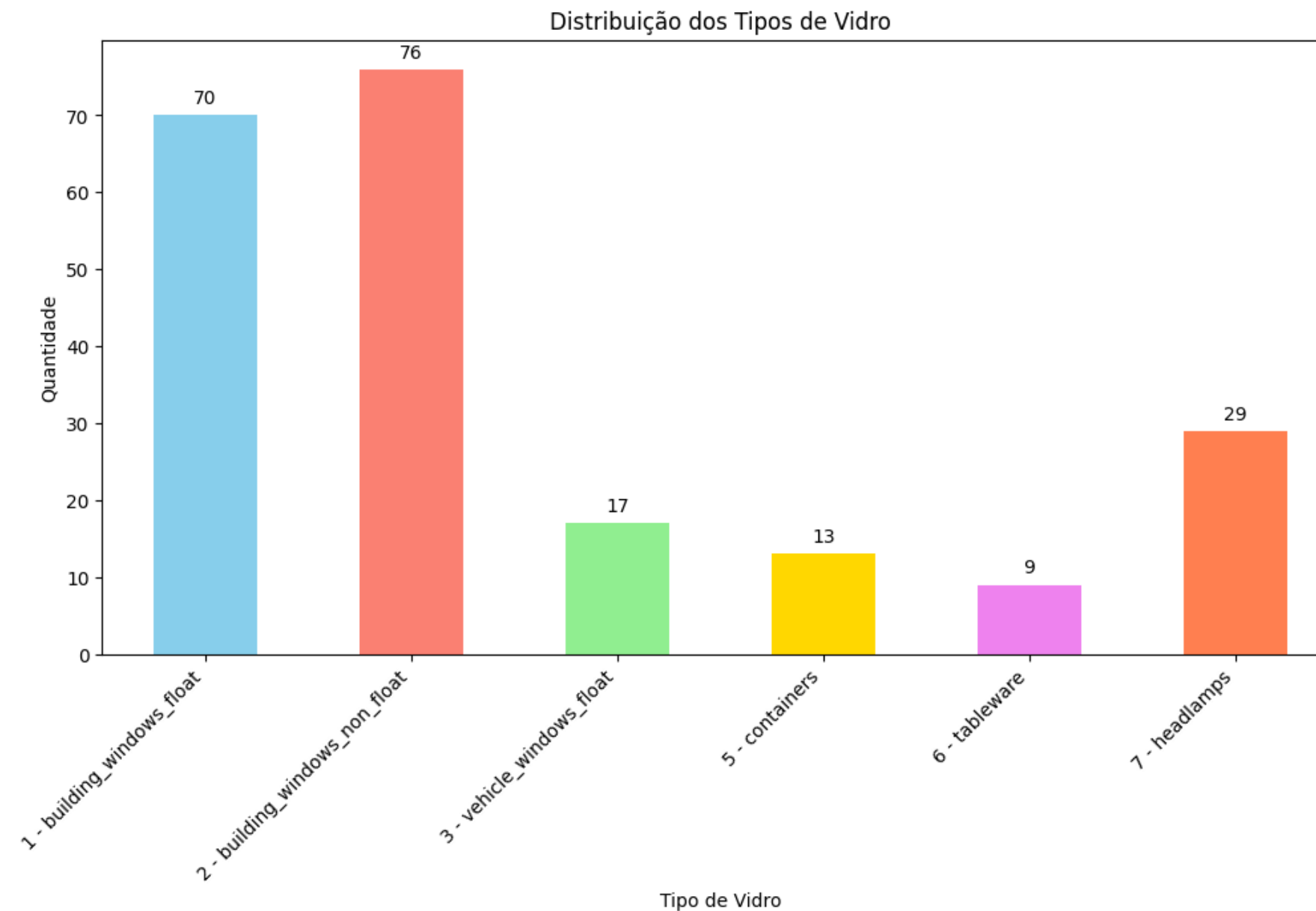
## Motivação do Estudo

- Investigações criminológicas.
- Na cena do crime, o vidro deixado pode ser usado como **evidência**, se for **corretamente identificado**.

**Será que conseguiremos incriminar o Malfeitor?!**



<b>nº</b>	<b>Tipos</b>	<b>Processo</b>	<b>Usados em/para:</b>
<b>1</b>	<b>Building_windows</b>	<b>float</b>	<b>Janelas de edifícios</b>
<b>2</b>	<b>Building_windows</b>	<b>não "float"</b>	<b>Janelas de edifícios</b>
<b>3</b>	<b>Vehicle_windows</b>	<b>float</b>	<b>Janelas de veículos</b>
<b>4</b>	<b>Vehicle_windows</b>	<b>não "float"</b>	<b>Janelas de veículos(<b>sem amostras</b>)</b>
<b>5</b>	<b>Containers</b>	<b>não "float"</b>	<b>Fabricação de recipientes, como garrafas e frascos</b>
<b>6</b>	<b>Tableware</b>	<b>não "float"</b>	<b>Utensílios domésticos, como copos, pratos e tigelas</b>
<b>7</b>	<b>Headlamps</b>	<b>não "float"</b>	<b>Específicos para faróis, feitos para resistir a impactos e altas temperaturas</b>



**Vidro float** é um tipo de vidro plano, transparente ou colorido, que se caracteriza por ser **uniforme e homogêneo**. O nome "float" vem do processo de fabricação, em que o vidro ainda não derretido flutua em estanho derretido.



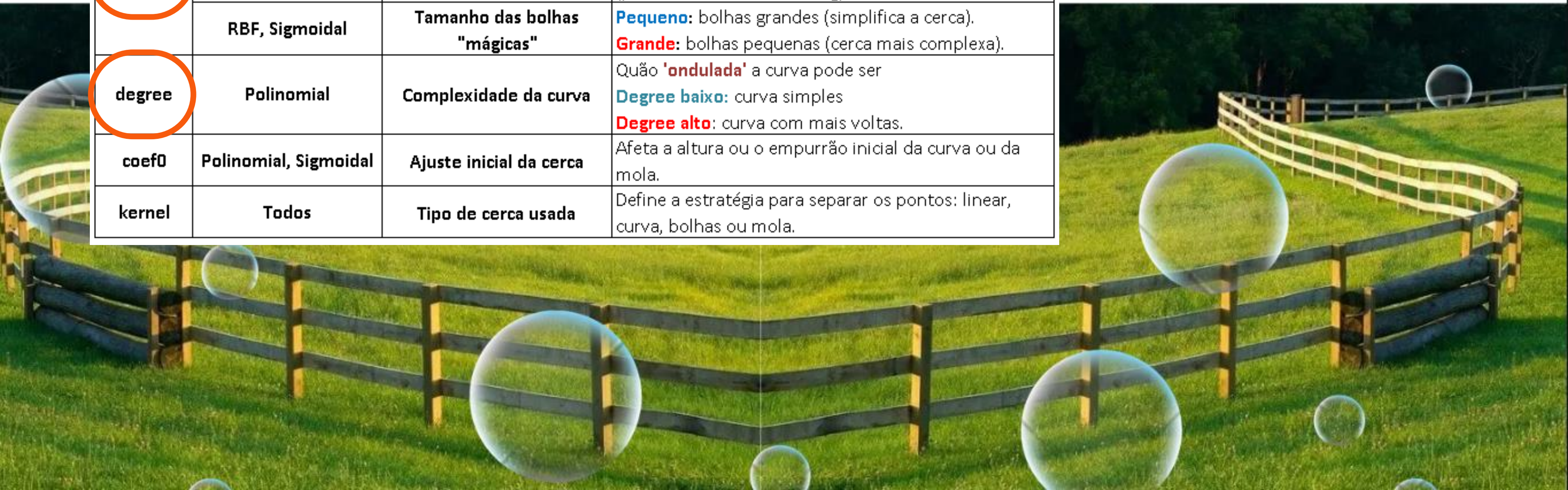
# Parâmetros SVM

Parâmetro	Kernel	Descrição/analogia	O que Controla?
C	Todos	Regularização (disciplina da cerca)	Quanto o modelo tenta separar os pontos perfeitamente. <b>C baixo:</b> mais flexível <b>C alto:</b> mais rígido.
gamma	Polinomial	Poder de "Lupa"	<b>Pequeno:</b> menos detalhes percebidos, curvas suaves(menos sensível) <b>Grande:</b> mais detalhes percebidos, curva complexa (possibilidade de overfitting)
	RBF, Sigmoidal	Tamanho das bolhas "mágicas"	<b>Pequeno:</b> bolhas grandes (simplifica a cerca). <b>Grande:</b> bolhas pequenas (cerca mais complexa).
degree	Polinomial	Complexidade da curva	Quão ' <b>ondulada</b> ' a curva pode ser <b>Degree baixo:</b> curva simples <b>Degree alto:</b> curva com mais voltas.
coef0	Polinomial, Sigmoidal	Ajuste inicial da cerca	Afeta a altura ou o empurrão inicial da curva ou da mola.
kernel	Todos	Tipo de cerca usada	Define a estratégia para separar os pontos: linear, curva, bolhas ou mola.



# Parâmetros SVM

Parâmetro	Kernel	Descrição/analogia	O que Controla?
C	Todos	Regularização (disciplina da cerca)	Quanto o modelo tenta separar os pontos perfeitamente. <b>C baixo:</b> mais flexível <b>C alto:</b> mais rígido.
gamma	Polinomial	Poder de "Lupa"	<b>Pequeno:</b> menos detalhes percebidos, curvas suaves(menos sensível) <b>Grande:</b> mais detalhes percebidos, curva complexa (possibilidade de overfitting)
	RBF, Sigmoidal	Tamanho das bolhas "mágicas"	<b>Pequeno:</b> bolhas grandes (simplifica a cerca). <b>Grande:</b> bolhas pequenas (cerca mais complexa).
degree	Polinomial	Complexidade da curva	Quão ' <b>ondulada</b> ' a curva pode ser <b>Degree baixo:</b> curva simples <b>Degree alto:</b> curva com mais voltas.
coef0	Polinomial, Sigmoidal	Ajuste inicial da cerca	Afeta a altura ou o empurrão inicial da curva ou da mola.
kernel	Todos	Tipo de cerca usada	Define a estratégia para separar os pontos: linear, curva, bolhas ou mola.



## Pré-processamento:

- Os dados estão normalizados.
  - SVM --> MinMaxScaler
  - RNA --> StandardScaler

Característica	StandardScaler	MinMaxScaler
O que faz	Centraliza e escala (média 0, desv. 1).	Reescalona para um intervalo.
Resultado típico	Valores negativos e positivos.	Valores entre 0 e 1 (ou outro intervalo).
Quando usar	Dados aproximadamente normais (Gaussianos).	Dados com escala arbitrária.



**Sabemos que as SVM irão confundir classes , sendo assim fazemos as:**

**Pergunta 1:** Seria melhor que fossem duas classes bem parecidas no formato e utilidade?

**Pergunta 2:** Seria mais interessante que ocorresse entre duas classes que possuem características distintas, como atributos visuais, curvatura e outros aspectos?



# Hipótese 1:Dados com as 7 classes e todos os atributos

SVM\_poly

RNA

		precision	recall	f1-score	support
Building_windows_float	1	0.7619	0.7619	0.7619	21
Building_windows	2	0.7600	0.8261	0.7917	23
Vehicle_windows_float	3	0.5000	0.4000	0.4444	5
Containers	5	1.0000	0.7500	0.8571	4
Tableware	6	0.7500	1.0000	0.8571	3
Headlamps	7	1.0000	0.8889	0.9412	9
accuracy				0.7846	65
macro avg		0.7953	0.7711	0.7756	65
weighted avg		0.7882	0.7846	0.7831	65

		precision	recall	f1-score	support
Building_windows_float	1	0.71	0.71	0.71	21
Building_windows	2	0.69	0.78	0.73	23
Vehicle_windows_float	3	0.33	0.20	0.25	5
Containers	5	1.00	0.25	0.40	4
Tableware	6	0.60	1.00	0.75	3
Headlamps	7	1.00	1.00	1.00	9
accuracy				0.72	65
macro avg		0.72	0.66	0.64	65
weighted avg		0.73	0.72	0.71	65

Parâmetros:

C=100; class\_weight= Balanced; gamma=1, Kernel=poly

# Hipótese 1:Dados com as 7 classes e todos os atributos

SVM\_rbf

RNA

		precision	recall	f1-score	support
Building_windows_float	1	0.7308	0.9048	0.8085	21
Building_windows	2	0.7391	0.7391	0.7391	23
Vehicle_windows_float	3	1.0000	0.4000	0.5714	5
Containers	5	1.0000	0.5000	0.6667	4
Tableware	6	0.7500	1.0000	0.8571	3
Headlamps	7	1.0000	0.8889	0.9412	9
accuracy				0.7846	65
macro avg		0.8700	0.7388	0.7640	65
weighted avg		0.8092	0.7846	0.7776	65

		precision	recall	f1-score	support
Building_windows_float	1	0.71	0.71	0.71	21
Building_windows	2	0.69	0.78	0.73	23
Vehicle_windows_float	3	0.33	0.20	0.25	5
Containers	5	1.00	0.25	0.40	4
Tableware	6	0.60	1.00	0.75	3
Headlamps	7	1.00	1.00	1.00	9
accuracy				0.72	65
macro avg		0.72	0.66	0.64	65
weighted avg		0.73	0.72	0.71	65

Parâmetros:

C=100; class\_weight= Balanced; gamma=scale, Kernel=rbf



# Hipótese 1:Dados com as 7 classes e todos os atributos

## SVM\_rbf

		precision	recall	f1-score	support
Building_windows_float	1	0.7308	0.9048	0.8085	21
Building_windows	2	0.7391	0.7391	0.7391	23
Vehicle_windows_float	3	1.0000	0.4000	0.5714	5
Containers	5	1.0000	0.5000	0.6667	4
Tableware	6	0.7500	1.0000	0.8571	3
Headlamps	7	1.0000	0.8889	0.9412	9
accuracy				0.7846	65
macro avg		0.8700	0.7388	0.7640	65
weighted avg		0.8092	0.7846	0.7776	65

### Parâmetros:

C=100; class\_weigt= Balanced; gamma=scale, Kernel=rbf

## SVM\_poly

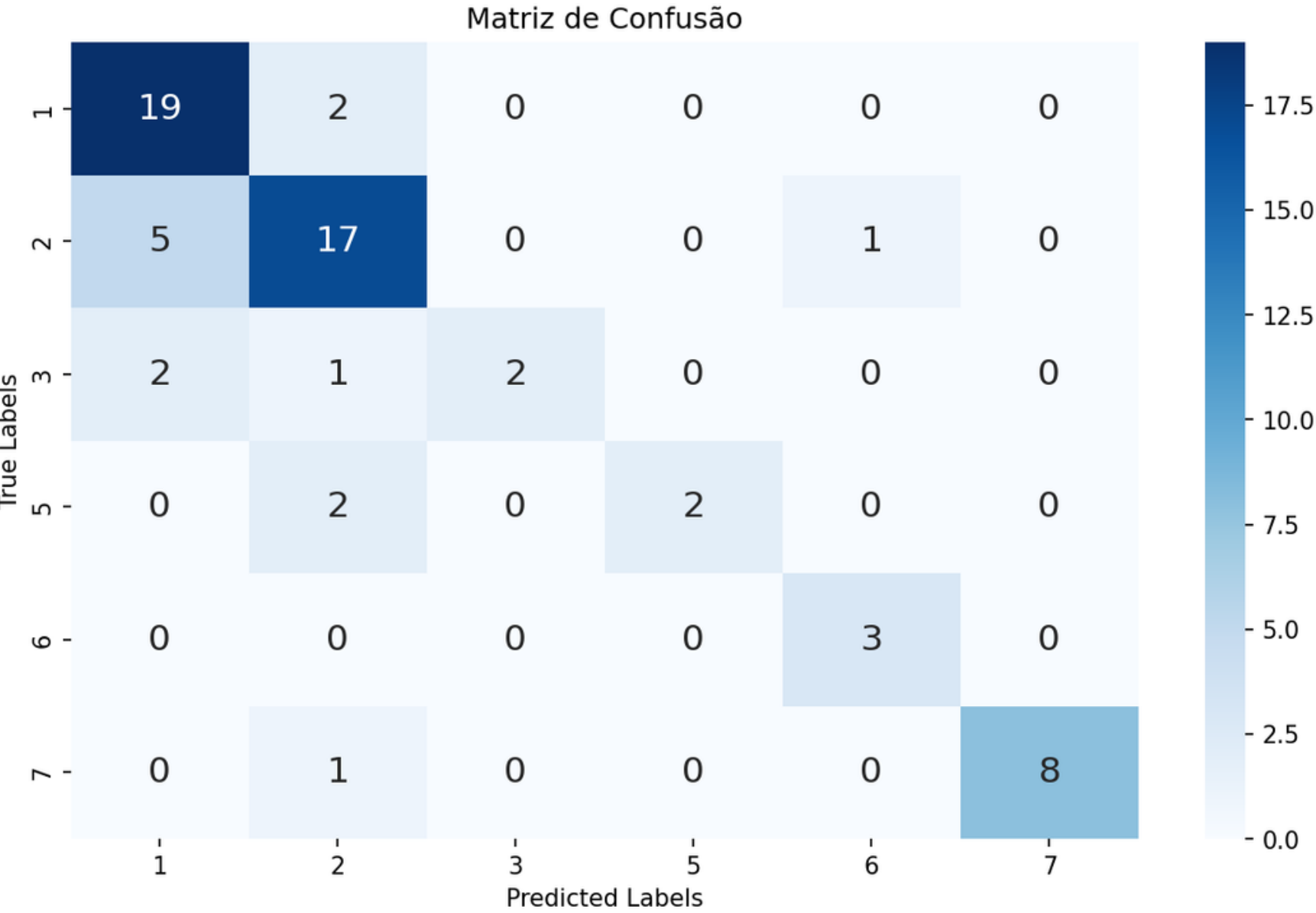
		precision	recall	f1-score	support
Building_windows_float	1	0.7619	0.7619	0.7619	21
Building_windows	2	0.7600	0.8261	0.7917	23
Vehicle_windows_float	3	0.5000	0.4000	0.4444	5
Containers	5	1.0000	0.7500	0.8571	4
Tableware	6	0.7500	1.0000	0.8571	3
Headlamps	7	1.0000	0.8889	0.9412	9
accuracy				0.7846	65
macro avg		0.7953	0.7711	0.7756	65
weighted avg		0.7882	0.7846	0.7831	65

### Parâmetros:

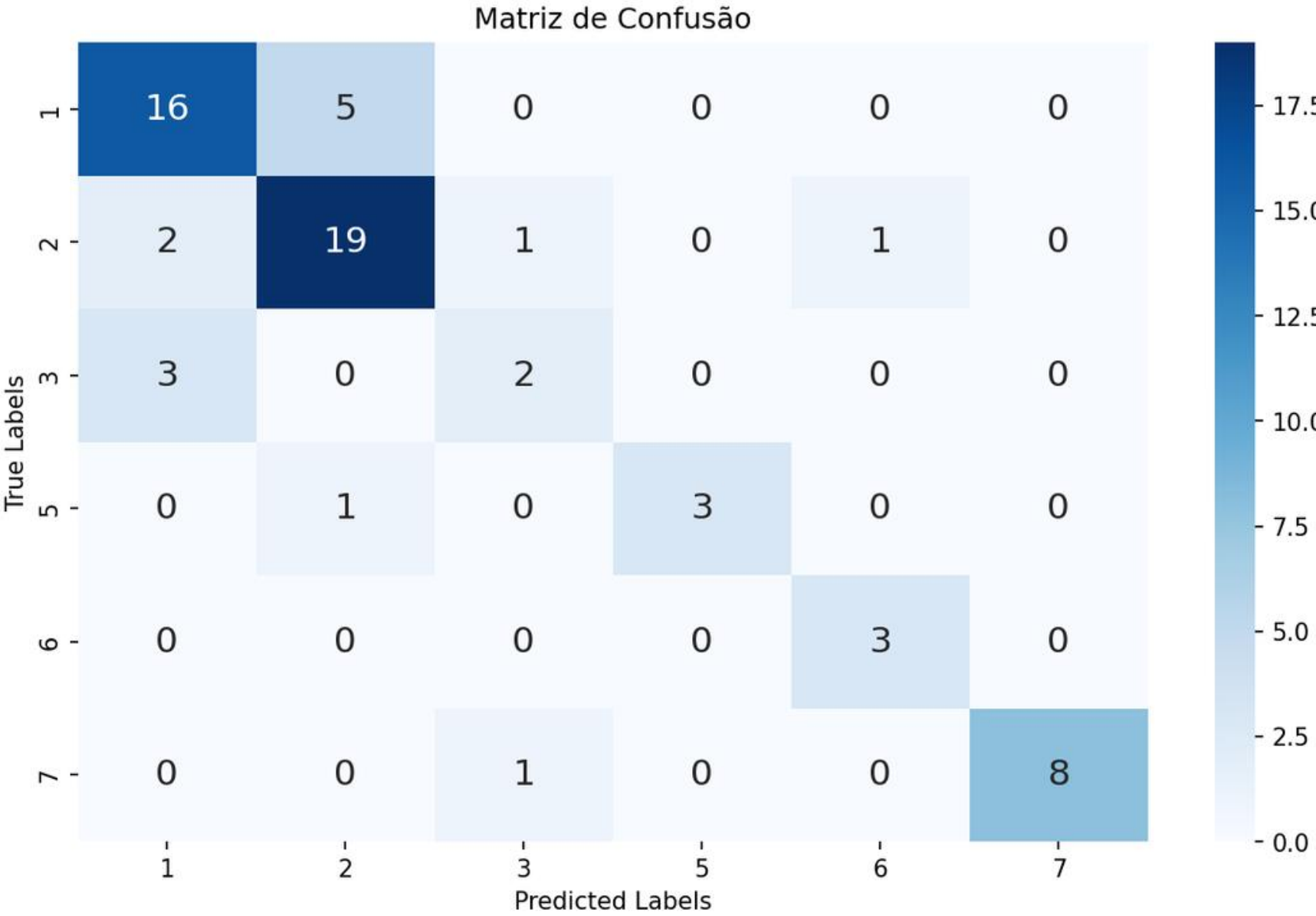
C=100; class\_weigt= Balanced; gamma=1, Kernel=poly

# Hipótese 1:Dados com as 7 classes e todos os atributos

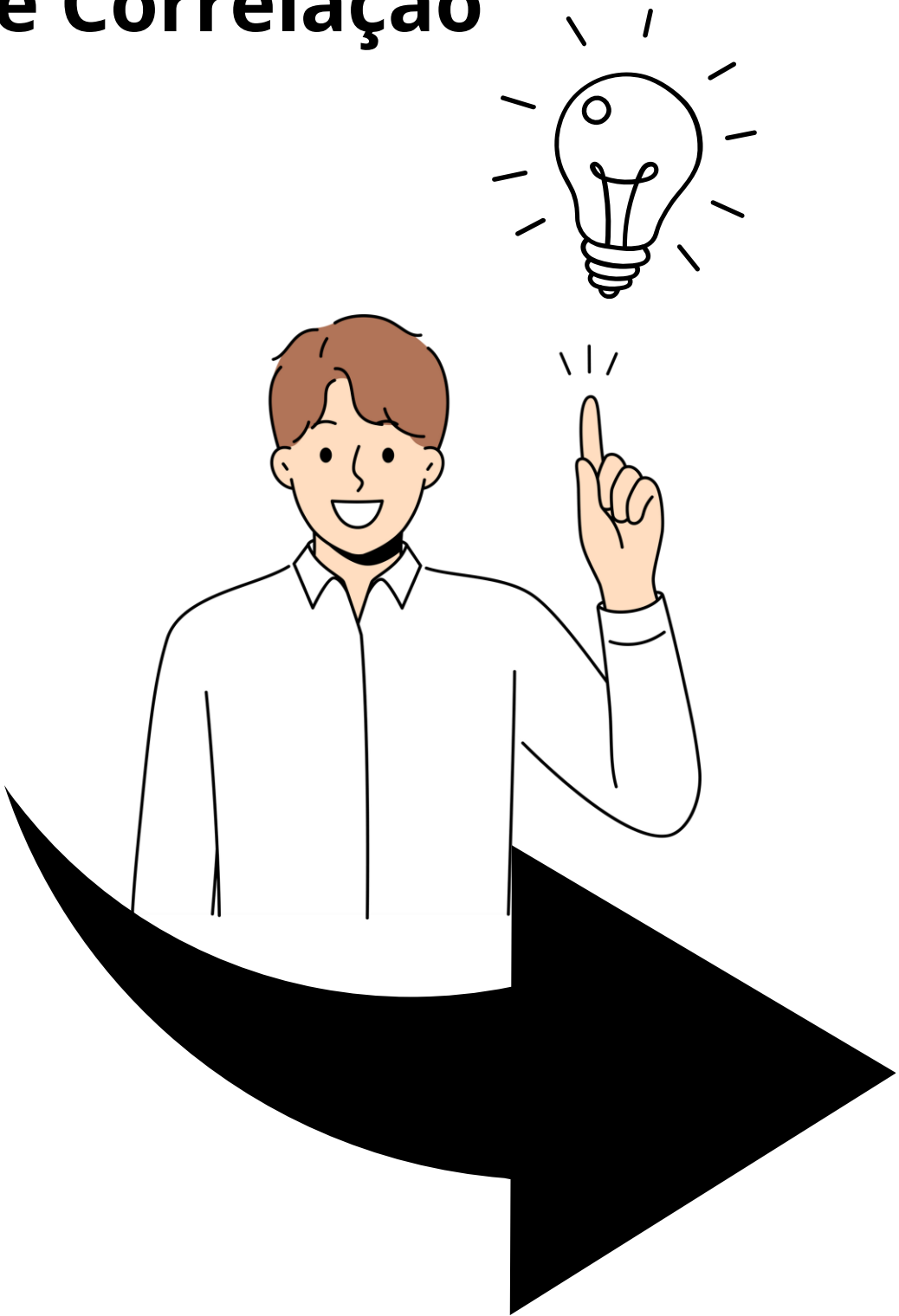
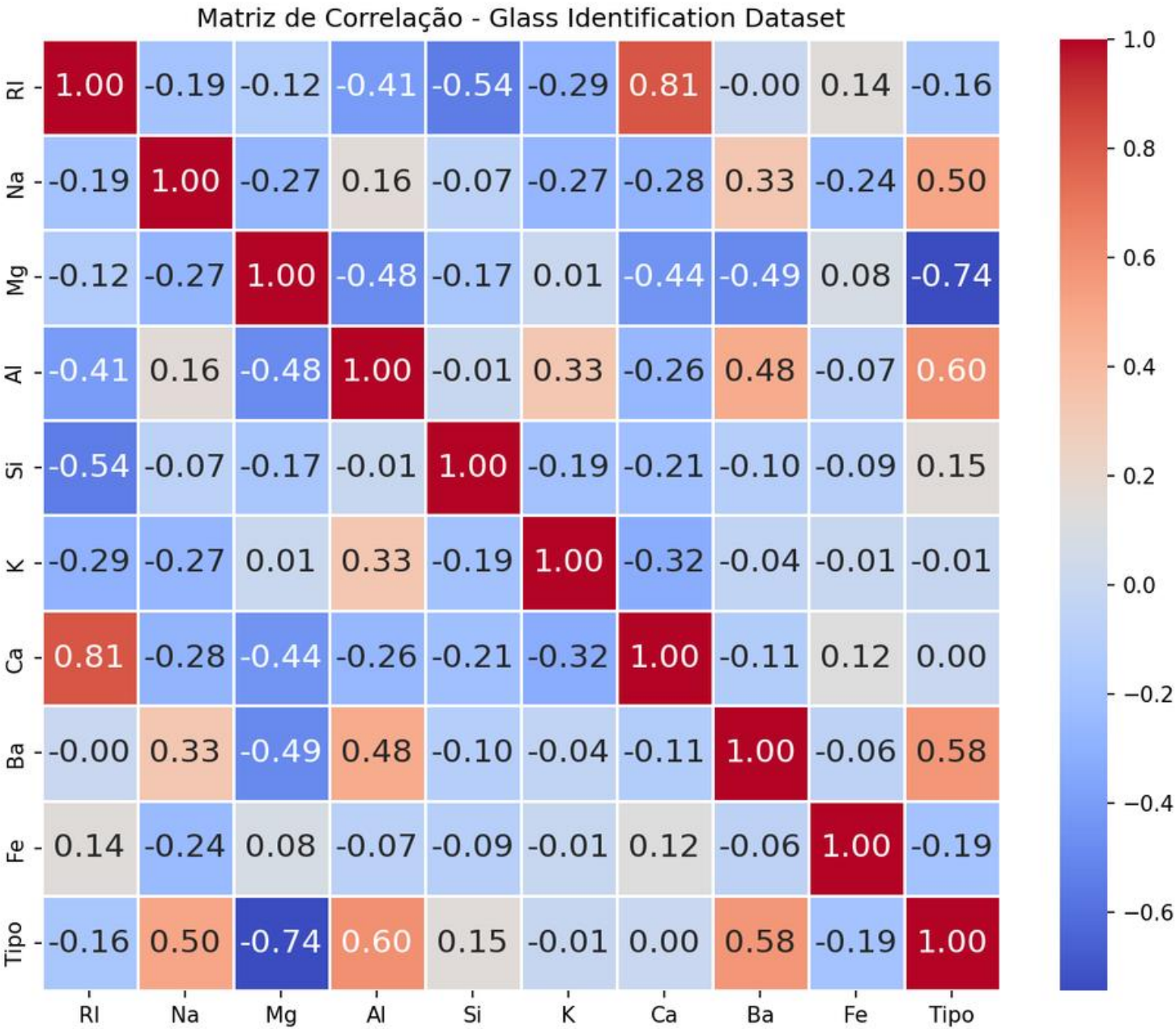
SVM\_rbf



SVM\_poly

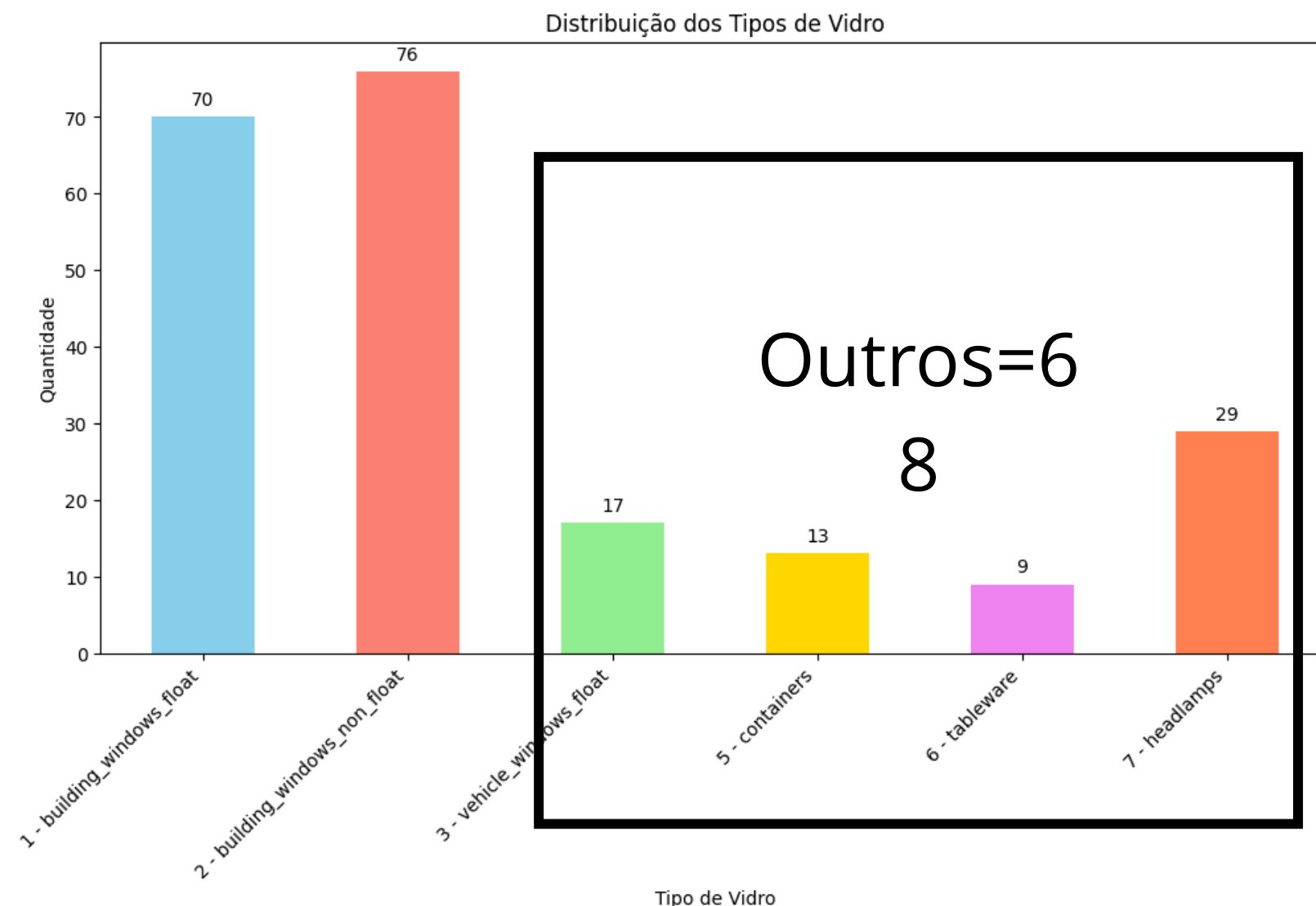


# Estudo da Matriz de Correlação



-Ca  
-RI





Passamos a resolver um problema com 3 classes...

Devido a quantidade de amostras do tipo 1 e 2:

Permanecemos com essas e somamos todos os restantes na classe **"outros"** e ficamos com um total de 68 amostras



# Hipótese 2.1: Problema reduzido a 3 classes com todos os atributos

SVM\_rbf

SVM\_poly

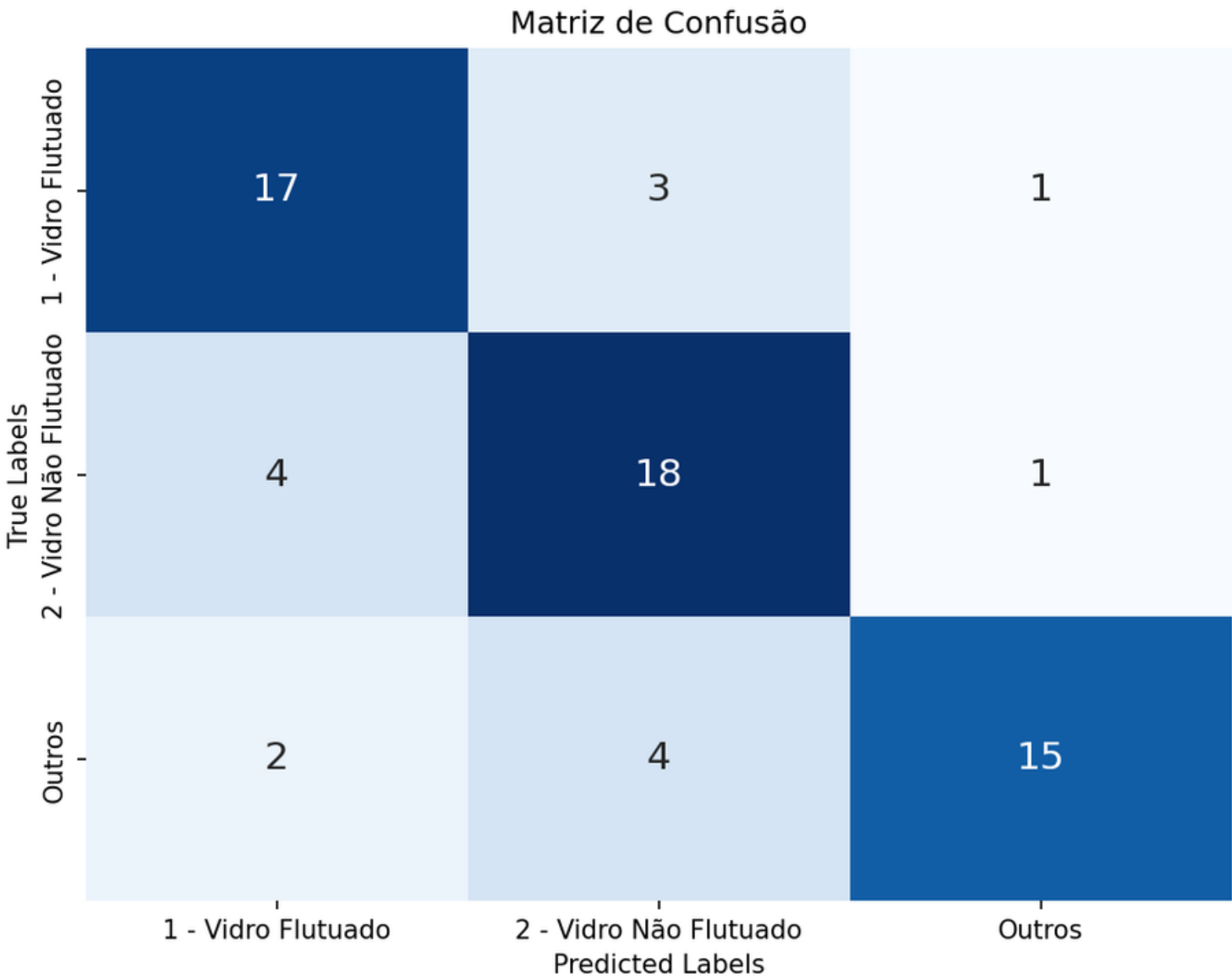
Classification Report:					Classification Report:						
		precision	recall	f1-score	support			precision	recall	f1-score	support
Building_windows_float	1	0.7391	0.8095	0.7727	21	Building_windows_float	1	0.8421	0.7619	0.8000	21
Building_windows	2	0.7200	0.7826	0.7500	23	Building_windows	2	0.7200	0.7826	0.7500	23
Outros	3	0.8824	0.7143	0.7895	21	Outros	3	0.7619	0.7619	0.7619	21
accuracy				0.7692	65	accuracy				0.7692	65
macro avg				0.7707	65	macro avg				0.7706	65
weighted avg				0.7701	65	weighted avg				0.7700	65

Parâmetros:  
C=100; class\_weigt= Balanced; gamma=3, Kernel=rbf,

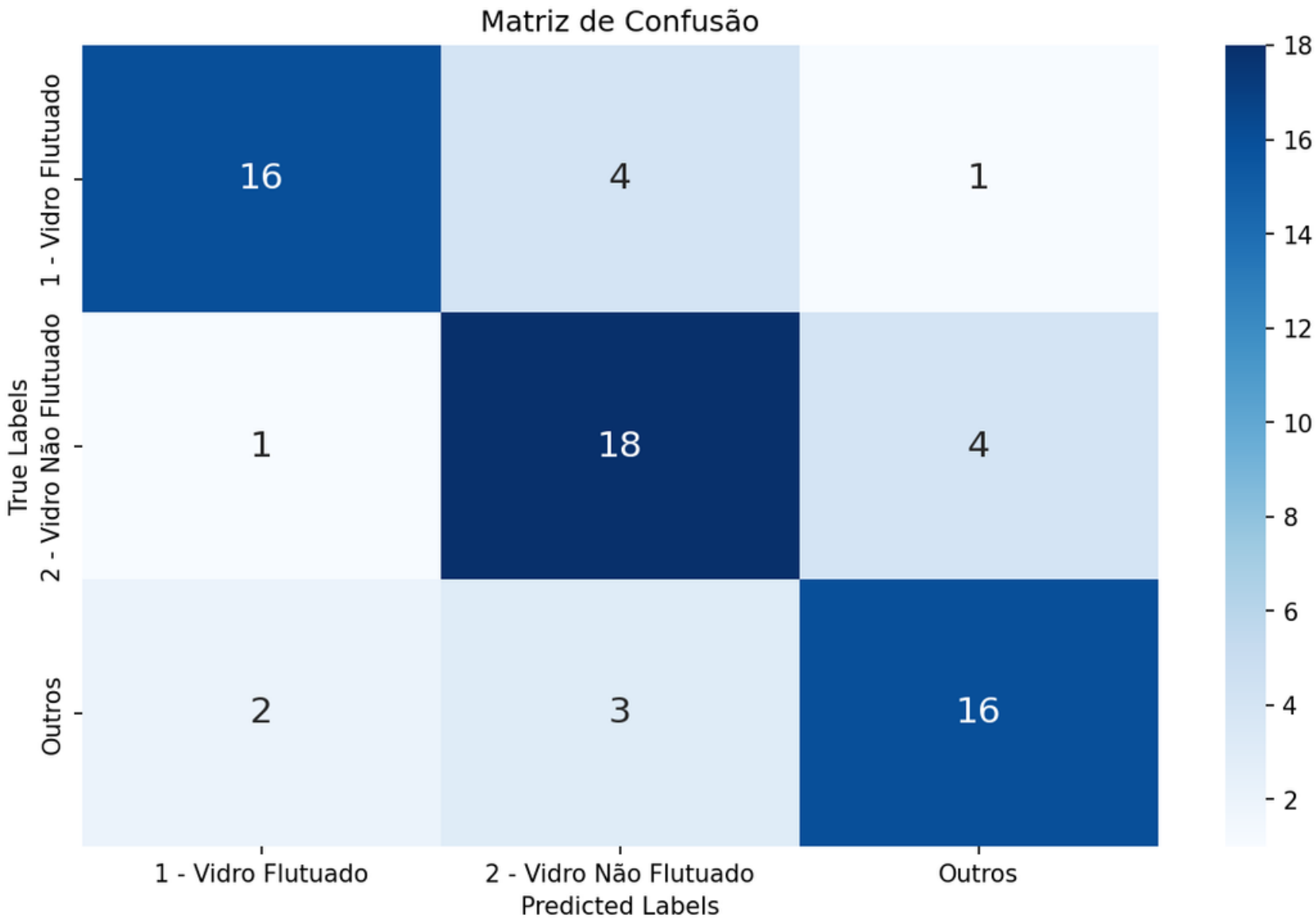
Parâmetros:  
C=100; class\_weigt= Balanced; gamma=1, Kernel=poly, degree=5

# Hipótese 2.1: Problema reduzido a 3 classes com todos os atributos

SVM\_rbf



SVM\_poly





## Hipótese 2.2: Problema reduzido a 3 classes e retirando a coluna **Ca**

### SVM\_poly

		precision	recall	f1-score	support
Building_windows_float	1	0.7500	0.8571	0.8000	21
Building_windows	2	0.7083	0.7391	0.7234	23
Outros	3	0.7647	0.6190	0.6842	21
accuracy				0.7385	65
macro avg		0.7410	0.7384	0.7359	65
weighted avg		0.7400	0.7385	0.7355	65

#### Parâmetros:

C=100; class\_weight= Balanced; gamma=4,  
Kernel=poly, degree=4

### RNA

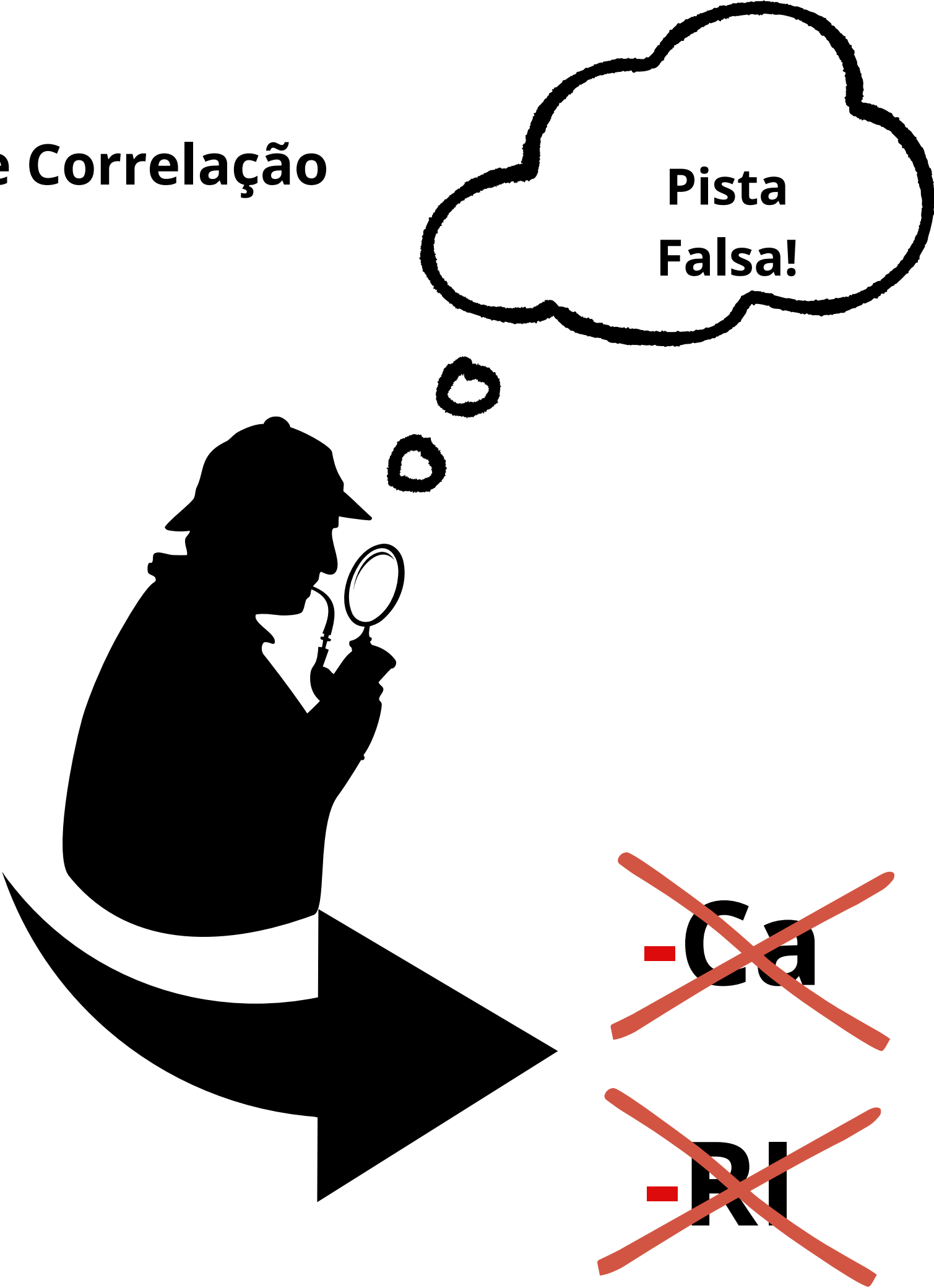
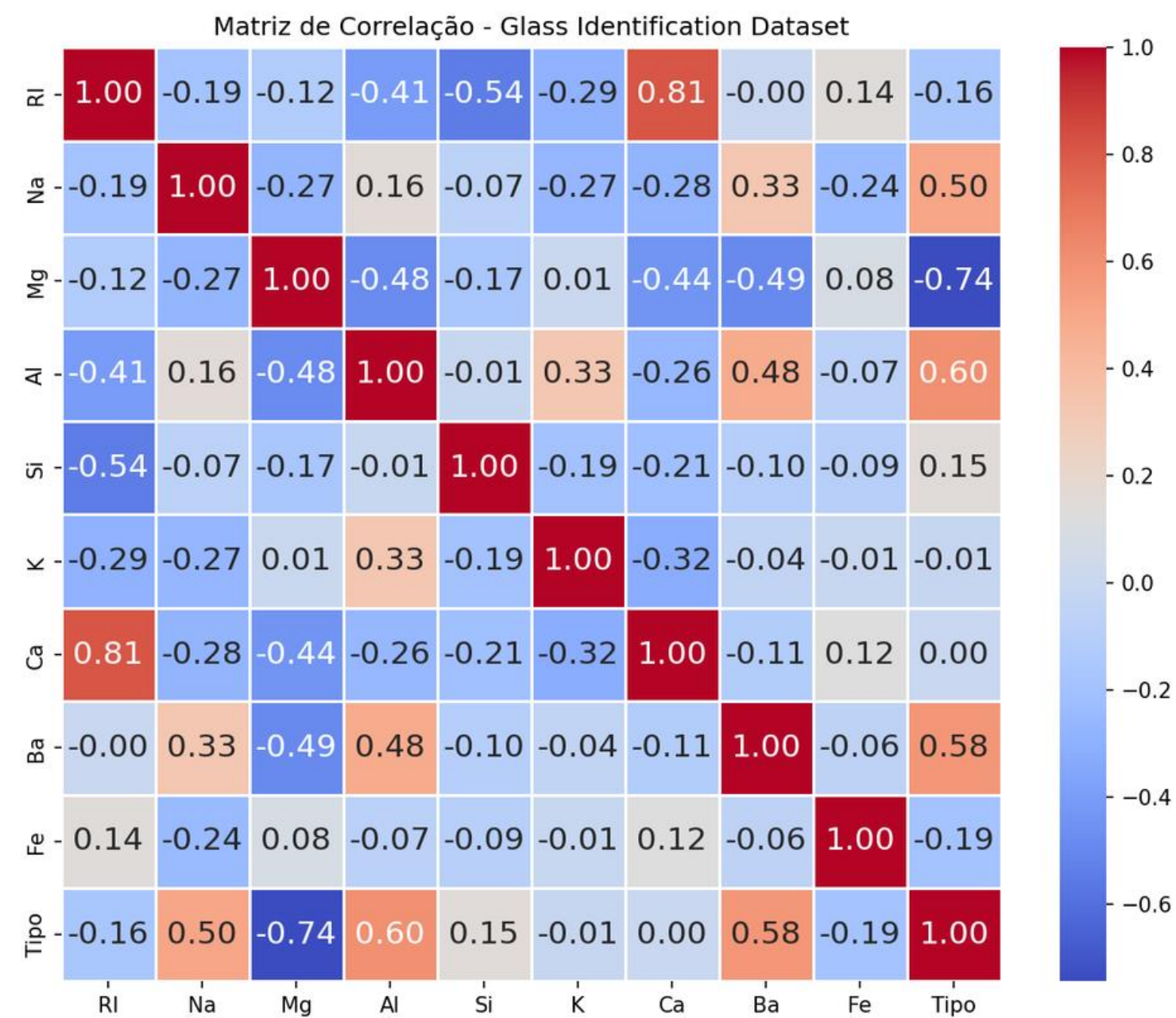
		precision	recall	f1-score	support
Building_windows_float	1	0.80	0.76	0.78	21
Building_windows	2	0.73	0.70	0.71	23
Outros	3	0.74	0.81	0.77	21
accuracy				0.75	65
macro avg		0.76	0.76	0.75	65
weighted avg		0.75	0.75	0.75	65

#### Parâmetros:

Hidden=32, Epochs=458, lr=0.1, batch=10

Obtivemos mesma acurácia para o teste com a **RGB**, mostrando que a retirada do **Ca** não levou a bons resultados aqui.

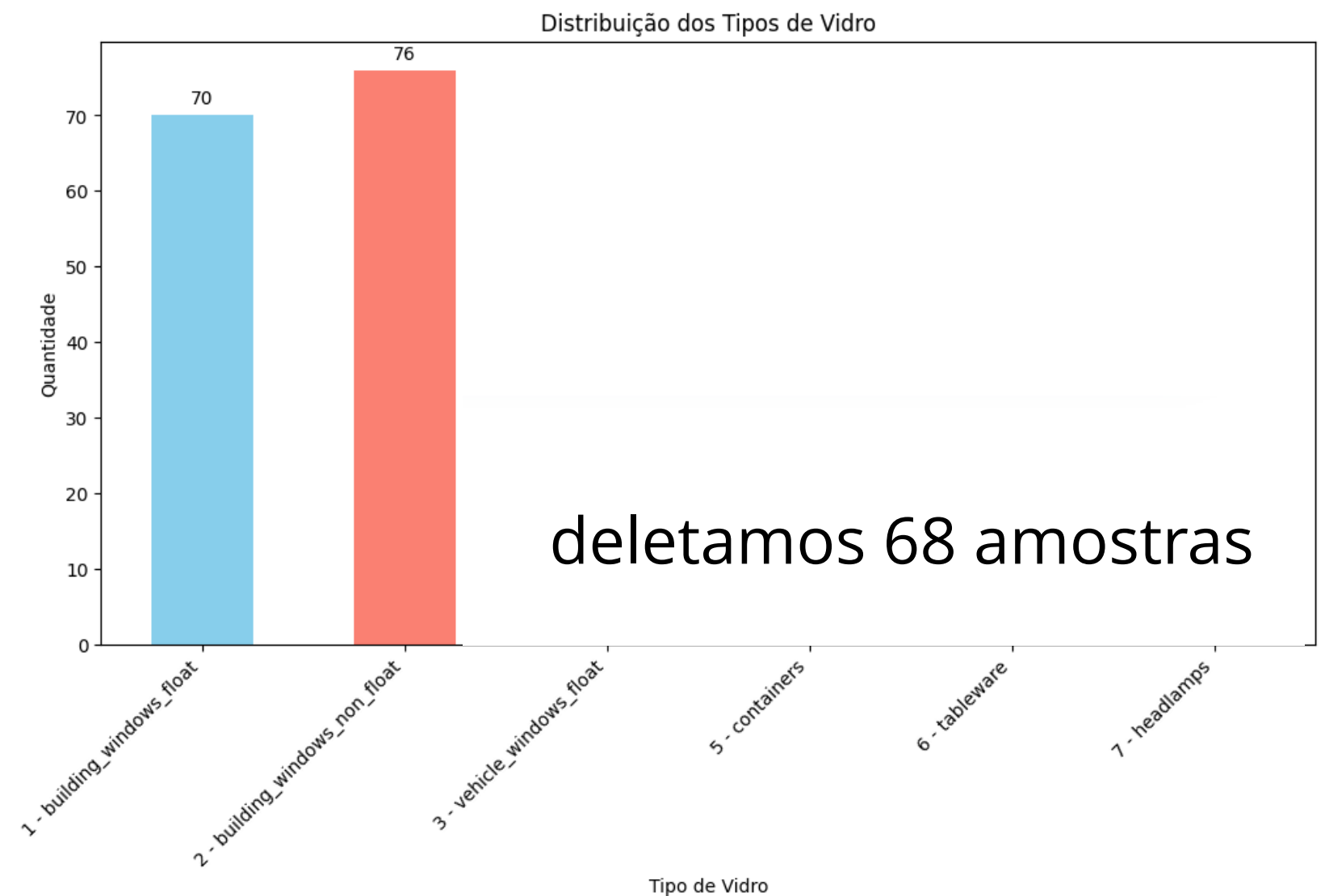
# Estudo da Matriz de Correlação



A ideia para a hipótese a seguir surgiu da análise dos dados. Observamos que tanto a classe 1 quanto a classe 2 desempenham a mesma função, diferenciando-se apenas pelo fato de uma ser Flutuada e a outra não.

Observamos que essas duas classes estão mais presentes em cenas de crime, provavelmente devido à sua utilidade. Por essa razão, as consideramos como as mais importantes para a análise

**Hipótese 4: Divisão em Flutuado e não Flutuado **excluindo** a classe “outros”**





# Hipótese 4: Divisão em Flutuado e não Flutuado **excluindo** a classe “outros”

## SVM com todos os atributos

## RNA sem o **Ca**

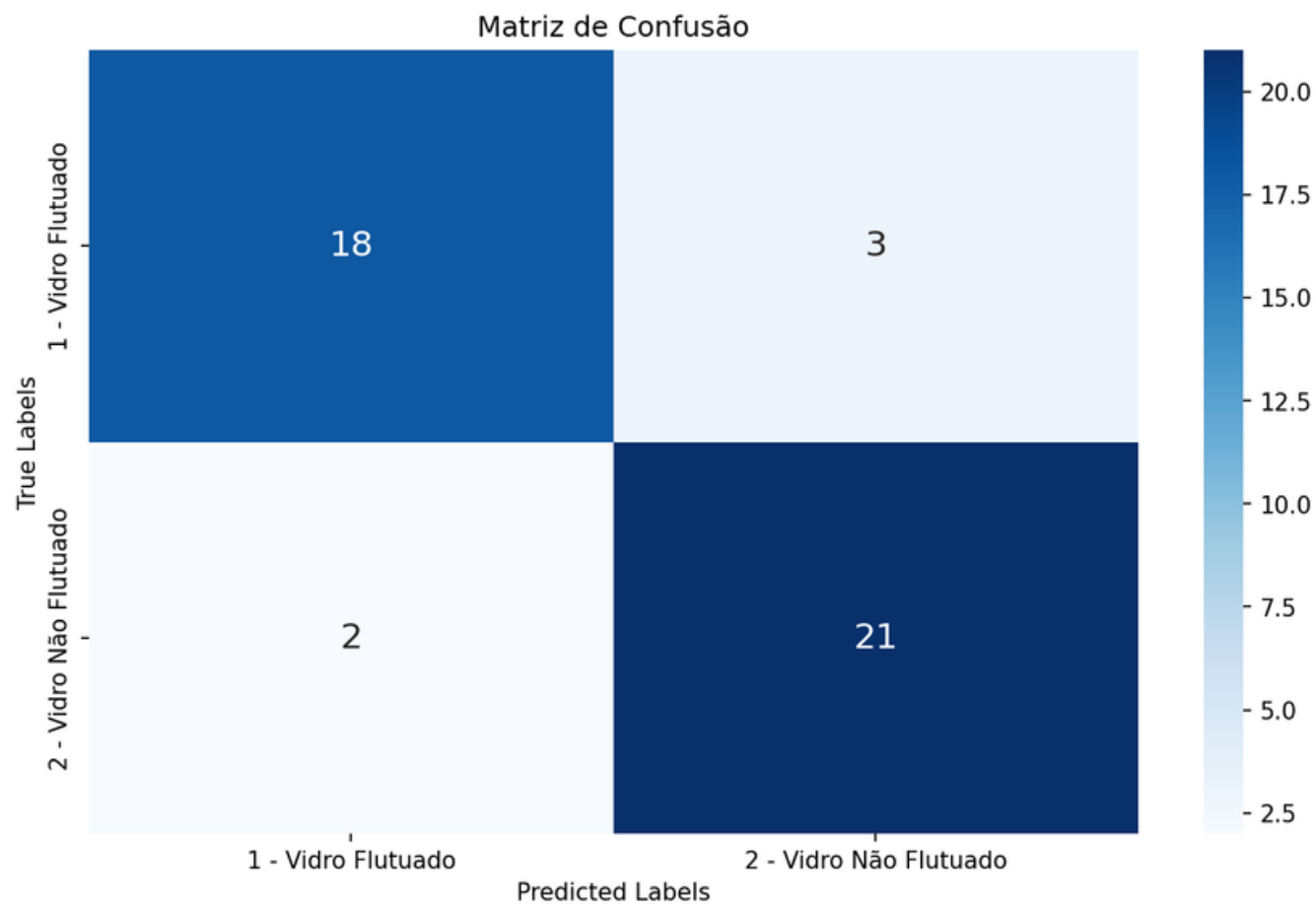
		precision	recall	f1-score	support			precision	recall	f1-score	support
Flutuado	1	0.9000	0.8571	0.8780	21	Flutuado	1	0.89	0.81	0.85	21
Não Flutuado	2	0.8750	0.9130	0.8936	23	Não Flutuado	2	0.84	0.91	0.88	23
accuracy				0.8864	44	accuracy				0.86	44
macro avg		0.8875	0.8851	0.8858	44	macro avg		0.87	0.86	0.86	44
weighted avg		0.8869	0.8864	0.8862	44	weighted avg		0.87	0.86	0.86	44

**Parâmetros:**  
C=40; class\_weight= Balanced; gamma=scale,  
Kernel=poly, degree=3

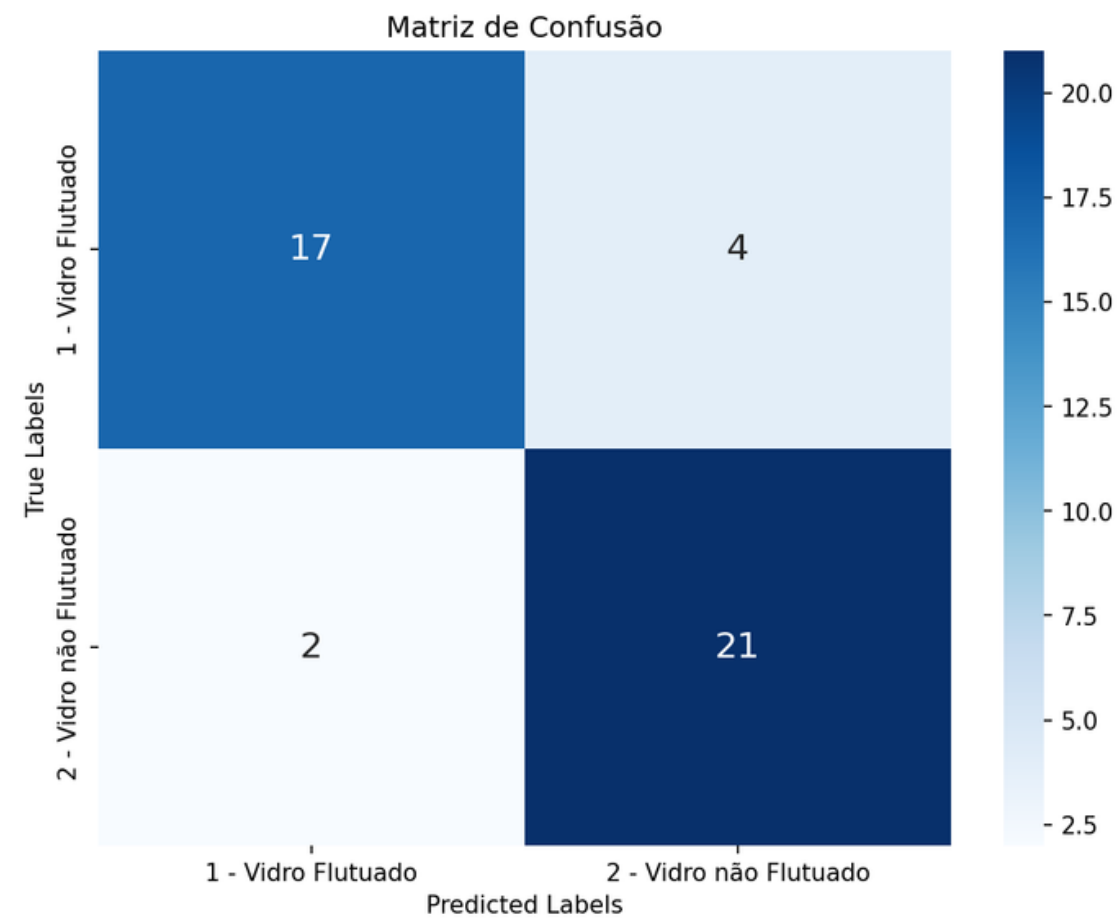
Quando o crime ocorre em um ambiente físico imóvel, como prédios, casas ou outras estruturas, temos uma acurácia de **88,64%** de sucesso na coleta de evidências

# Hipótese 4: Divisão em Flutuado e não Flutuado **excluindo** a classe “outros”

**SVM com todos os atributos**



**RNA sem o **Ca****



**Será que conseguiremos incriminar o Malfeitor?!**

Depende, se na cena do crime encontrarmos estilhaços de vidro e quisermos tirar a dúvida se o vidro pertencia a alguma janela ou vidraça de algum local temos melhores resultados.





**Muito Obrigado!**