

Métricas para Classificação

Otimização de HiperParâmetros

Seleção de Modelos

Prof. Gustavo Willam Pereira



INSTITUTO FEDERAL
Sudeste de Minas Gerais

Matriz de Confusão

- Diagonal Principal – Acertos

- Verdadeiros Negativos (VN) – Eram Negativos e foram corretamente classificados como Negativos
- Verdadeiros Positivos (VP) – Eram Positivos e foram corretamente classificados como Positivos

- Diagonal Secundária – Erros

- Falso Positivo (FP) – Eram Negativos e foram erroneamente classificados como Positivos
- Falso Negativos (FN) - Eram Positivos e foram erroneamente classificados como Negativos

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

Métricas para Classificação

Matriz de confusão

$$acurácia\ global = \frac{VN + VP}{VN + FP + FN + VP}$$

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

Métricas para Classificação

Matriz de confusão

		Predito	
		Negativo	Positivo
Observado	Negativo	9550	250
	Positivo	100	100

$$acurácia\ global = \frac{VN + VP}{VN + FP + FN + VP}$$

$$acurácia\ global = \frac{9550 + 100}{10000} = 0,965$$

Métricas para Classificação

Matriz de confusão

		Predito	
		Negativo	Positivo
Observado	Negativo	9800	0
	Positivo	200	0

- Agora considere passar todos os valores preditos “Positivos” para preditos “Negativos” ou seja, nosso preditor classifica todos como “Negativos”, então teríamos a seguinte acurácia global.

$$acurácia\ global = \frac{9800}{10000} = 0,98$$

- Logo, uma classificação ao acaso apresentou uma acurácia global maior.
- Sendo assim é necessário analisar outros tipos de métricas, dependendo do desequilíbrio que temos entre as classes.

Métricas para Classificação

Matriz de confusão $acurácia\ global = \frac{VN + VP}{VN + FP + FN + VP}$

$$precisão(exatidão) = \frac{VP}{VP + FP}$$

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

- A precisão é a avaliação dos valores classificados (preditos) como positivos.
- Percentual e 1's previstos que são realmente 1's
- Ex1: Qual a probabilidade de uma pessoa ter uma doença? Quantos foram classificados corretamente?
- Ex2: Predição para se fazer um determinado investimento -> Uma classificação FP irá te fazer perder muito dinheiro.

Métricas para Classificação

Matriz de confusão

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

$$acurácia\ global = \frac{VN + VP}{VN + FP + FN + VP}$$

$$recall(revoga\c{c}\tilde{a}o)(sensibilidade) = \frac{VP}{VP + FN}$$

- Recall -> avalia os que eram “positivos” e foram classificados como “positivos”
- Percentual de 1's corretamente classificado
- Ex1: Determinar quantos indivíduos tem uma determinada doença.
 - Uma classificação FN seria muito ruim, uma vez que a pessoa tem a doença, mas o modelo classificou como Negativo, ou seja a pessoa não iria receber o tratamento.

Métricas para Classificação

Matriz de confusão

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

$$acurácia\ global = \frac{VN + VP}{VN + FP + FN + VP}$$

$$especificidade = \frac{VN}{VN + FP}$$

- Especificidade -> avalia os que eram “Negativos” e foram classificados como “negativos”
- Percentual de 0's corretamente classificado
- Ex1: Determinar se um determinado email é spam.
 - Uma classificação FP seria muito ruim, uma vez que o email não é Spam, mas o modelo classificou como Spam, ou seja o usuário teria que checar a caixa de spam.

Métricas para Classificação

Matriz de confusão

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

$$acurácia\ global = \frac{VN + VP}{VN + FP + FN + VP}$$

$$precisão(exatidão) = \frac{VP}{VP + FP}$$

$$recall(revogação)(sensibilidade) = \frac{VP}{VP + FN}$$

$$especificidade = \frac{VN}{VN + FP}$$

$$F1 = 2 * \frac{precisão * recall}{precisão + recall}$$

Métricas para Classificação

Matriz de confusão

		Predito	
		Negativo	Positivo
Observado	Negativo	9550	250
	Positivo	100	100

$$acurácia\ global = \frac{9550 + 100}{10000} = 0,965$$

$$precisão = \frac{100}{350} = 0,285$$

$$recall = \frac{100}{200} = 0,50$$

$$specificity = \frac{9550}{9800} = 0,97$$

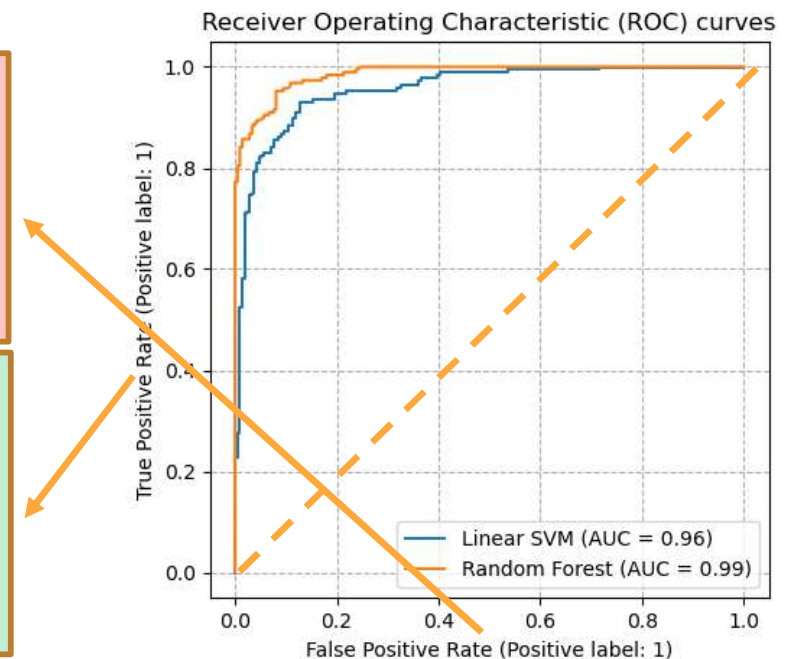
$$F1 = 2 \frac{0,285 \times 0,50}{0,285 + 0,50} = 0,36$$

Métricas para Classificação

ROC curve

Area Under Curve (AUC)

		Predito		
		Negativo	Positivo	
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)	$FPR = \frac{FP}{VN + FP}$
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)	$TPR = \frac{VP}{VP + FN}$



Otimização de HiperParâmetros

- No aprendizado de máquina, a otimização ou ajuste de hiperparâmetros é o problema de escolher um conjunto de hiperparâmetros ideais para um algoritmo de aprendizado.
- Um hiperparâmetro é um parâmetro cujo valor é usado para controlar o processo de aprendizado.
- Por outro lado, os valores de outros parâmetros são aprendidos.
- O mesmo tipo de modelo de aprendizado de máquina pode exigir diferentes restrições, para generalizar diferentes padrões de dados.
- Essas medidas são chamadas de hiperparâmetros e precisam ser ajustadas para que o modelo possa resolver o problema de aprendizado de máquina de maneira otimizada.
- A validação cruzada é frequentemente usada para realizar essa otimização.

Otimização de HiperParâmetros - Grid Search

- A maneira tradicional de realizar a otimização de hiperparâmetros tem sido a pesquisa em grade, ou uma varredura de parâmetros, que é simplesmente uma pesquisa exaustiva através de um subconjunto especificado manualmente do espaço de hiperparâmetros de um algoritmo de aprendizado.
- Um algoritmo de pesquisa de grade deve ser guiado por alguma métrica de desempenho, normalmente medida por validação cruzada no conjunto de treinamento.

Otimização de HiperParâmetros - Grid Search

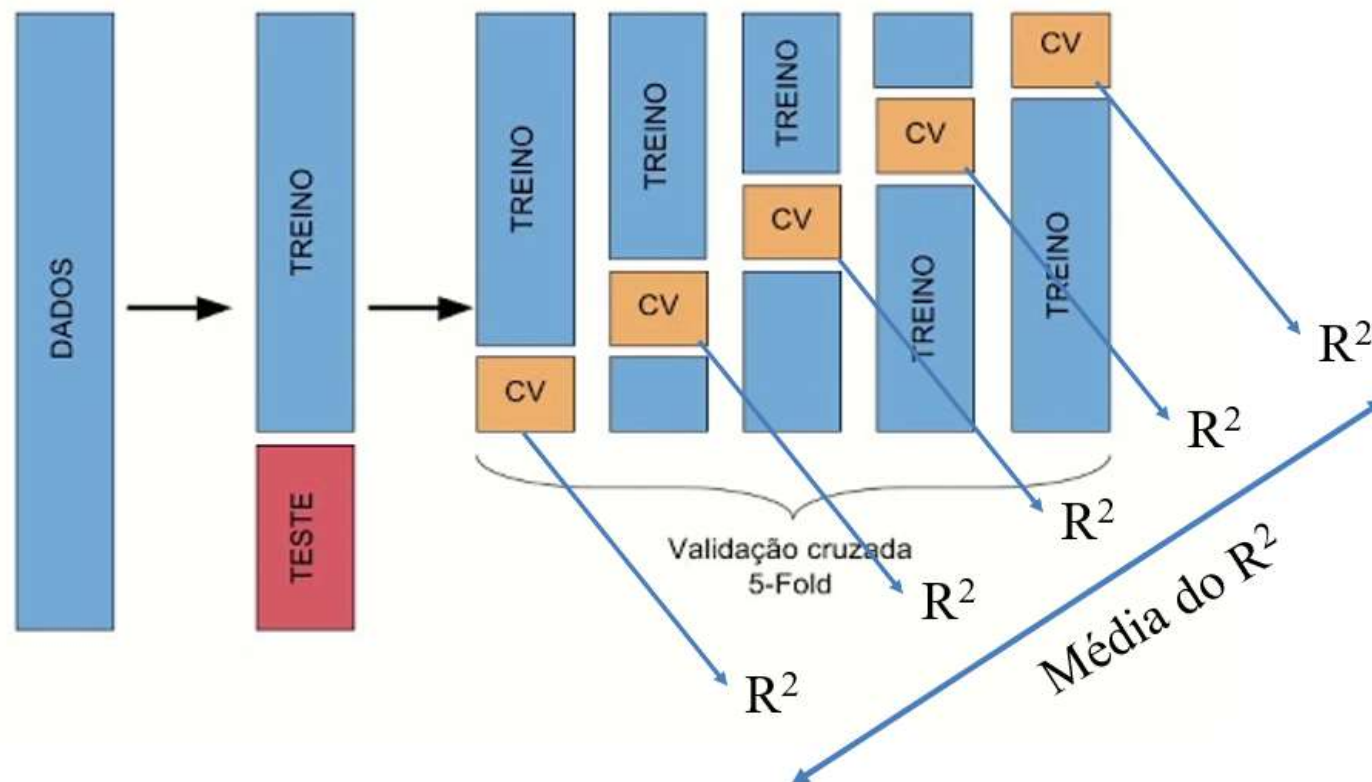
- Como o espaço de parâmetro de um aprendizado de máquina pode incluir espaços de valor com valor real ou ilimitado para determinados parâmetros, pode ser necessário definir manualmente os limites e a discretização antes de aplicar a pesquisa de grade.
- Por exemplo, um algoritmo SVM equipado com um kernel RBF tem pelo menos dois hiperparâmetros que precisam ser ajustados para um bom desempenho em dados não vistos: uma constante de regularização C e um hiperparâmetro de kernel γ (Gama).
- Ambos os parâmetros são contínuos, portanto, para realizar a pesquisa de grade, seleciona-se um conjunto finito de valores "razoáveis" para cada um, digamos:

$$C \in \{10, 100, 1000\}$$

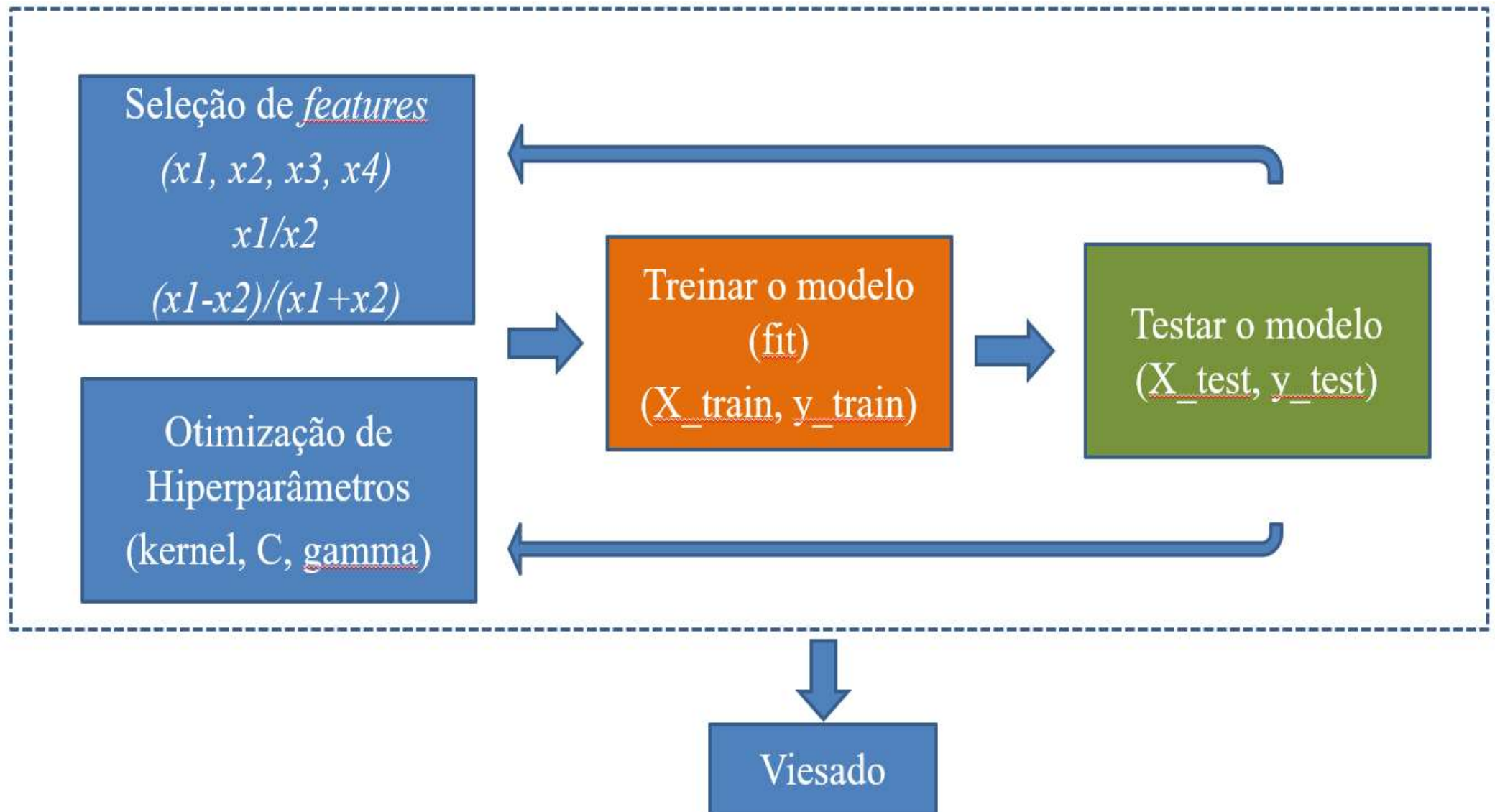
$$\gamma \in \{0.1, 0.2, 0.5, 1.0\}$$

Otimização de HiperParâmetros e Avaliação de Modelos

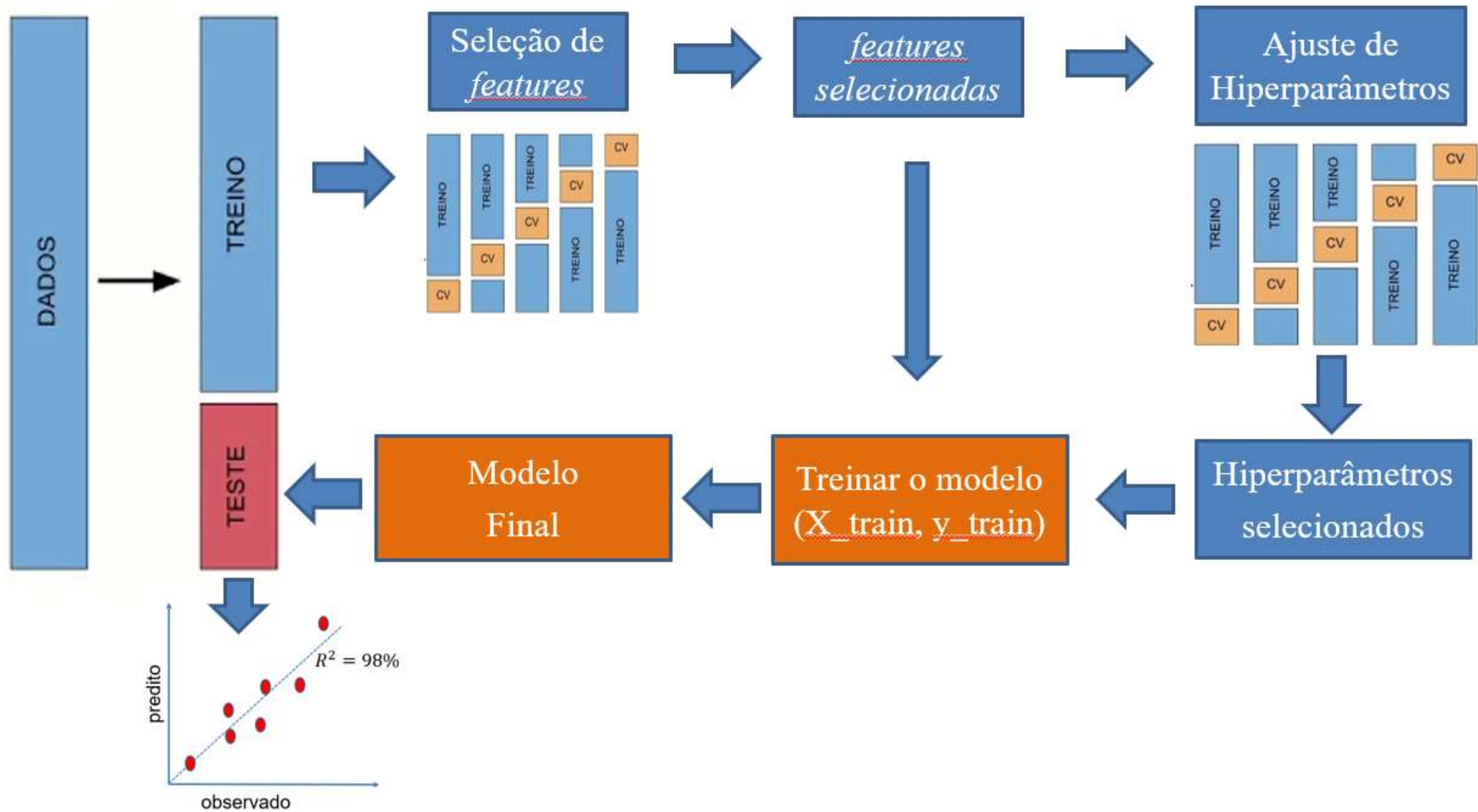
- A melhor forma de se fazer a seleção do modelo e a otimização do Hiperparâmetros é utilizando a validação cruzada.



Avaliação de Modelos



Avaliação de Modelos





INSTITUTO FEDERAL
Sudeste de Minas Gerais