

Programa de Pós-graduação em Sistemas de Informação

SIN5007 - Reconhecimento de Padrões (2023)

Censo da Educação Superior - Cursos 2021

MSc. Leonardo Cunha dos Santos
Gabriel Francisco dos Santos Silva

São Paulo / 2023



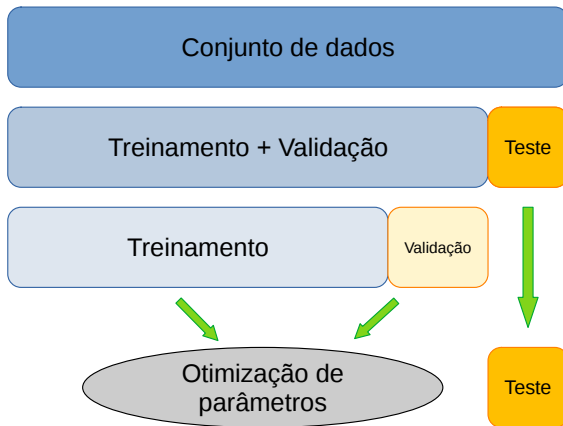
Agenda

① Estimação de desempenho

Estimação de desempenho

Separação de conjuntos de dados

Treinamento, validação e teste



Planejamento do experimento

Validação cruzada estratificada

- Estratégia: grid-search
- Maximização: F1-score
- Valor de k: 5
- Modelo: Naive Bayes
- Valores de hiperparâmetros testados:
'var_ssmoothing' : *np.logspace(0, -15, num = 100)*

* *var_ssmoothing*float, default=1e-9

Portion of the largest variance of all features that is added to variances for calculation stability.

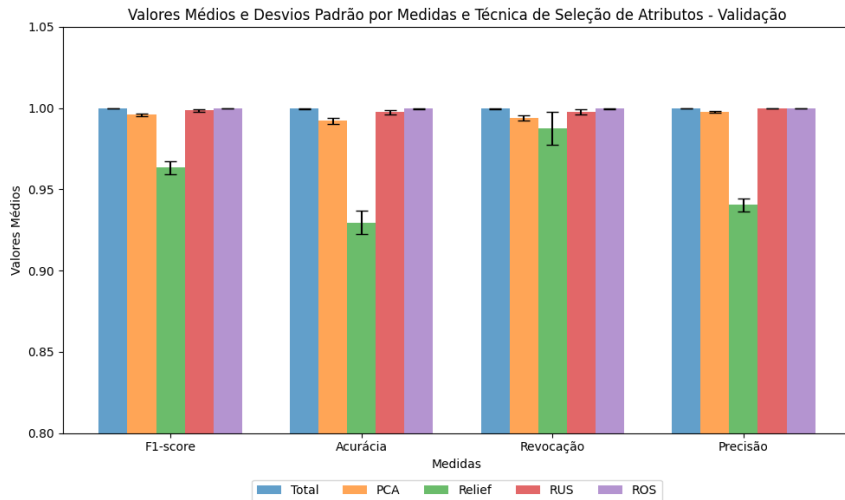
Planejamento do experimento

Validação cruzada estratificada

Médias	F1-score	Acurácia	Revocação	Precisão	var_smoothing(avg)
Total	0.9998	0.9997	0.9997	0.9999	5.3e-04
PCA	0.9958	0.9921	0.9939	0.9977	5.1e-01
Relief	0.9634	0.9296	0.9876	0.9404	8.0e-01
RUS	0.9986	0.9974	0.9975	0.9998	1.0e-01
ROS	0.9998	0.9997	0.9997	0.9999	7.4e-05

Planejamento do experimento

Resultados



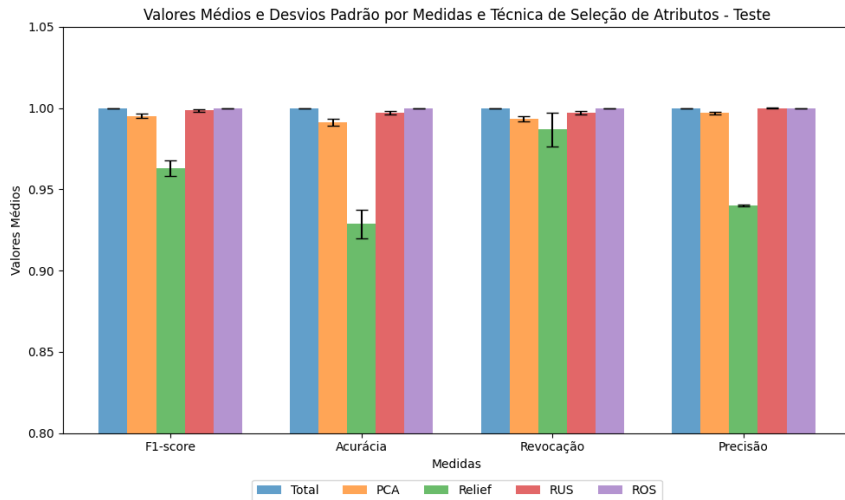
Planejamento do experimento

Valores obtidos com os conjuntos de teste

Médias	F1-score	Acurácia	Revocação	Precisão
Total	0.9999	0.9998	0.9999	0.9999
PCA	0.9952	0.9911	0.9934	0.9971
Relief	0.9629	0.9287	0.9868	0.9402
RUS	0.9985	0.9972	0.9972	0.9999
ROS	0.9999	0.9998	0.9999	0.9999

Planejamento do experimento

Resultados utilizando o conjunto de teste



Obrigado!

Thanks! / ¡Gracias!

Leonardo Cunha dos Santos

`lattes.cnpq.br/5620610314140397`

`leonardo.cunha.santos@usp.br`

Gabriel Francisco dos Santos Silva

`gabfssilva@gmail.com`