

# **Comparação Sistemática do Desempenho de Classificadores Binários Supervisionados**

Defesa de dissertação de Mestrado em  
Engenharia de Sistemas e Computação  
PESC/COPPE UFRJ



**UFRJ**  
UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



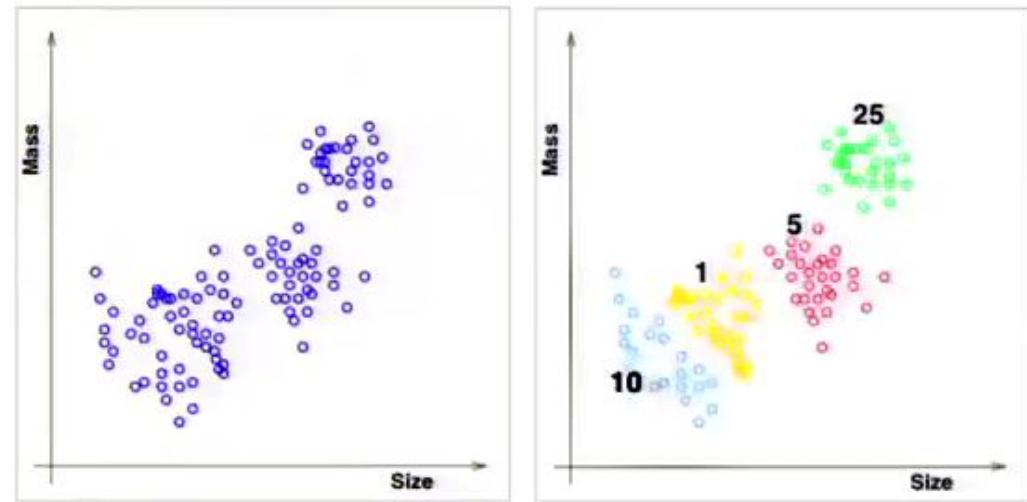
**COPPE**  
UFRJ



# 1 Introdução

## O que é um classificador?

- Realiza reconhecimento de padrões para separar um conjunto de observações em grupos – ou **classes**
- Caso só existam 2 classes, é um classificador **binário**
- Caso ele receba exemplos de treino, é um classificador **supervisionado**

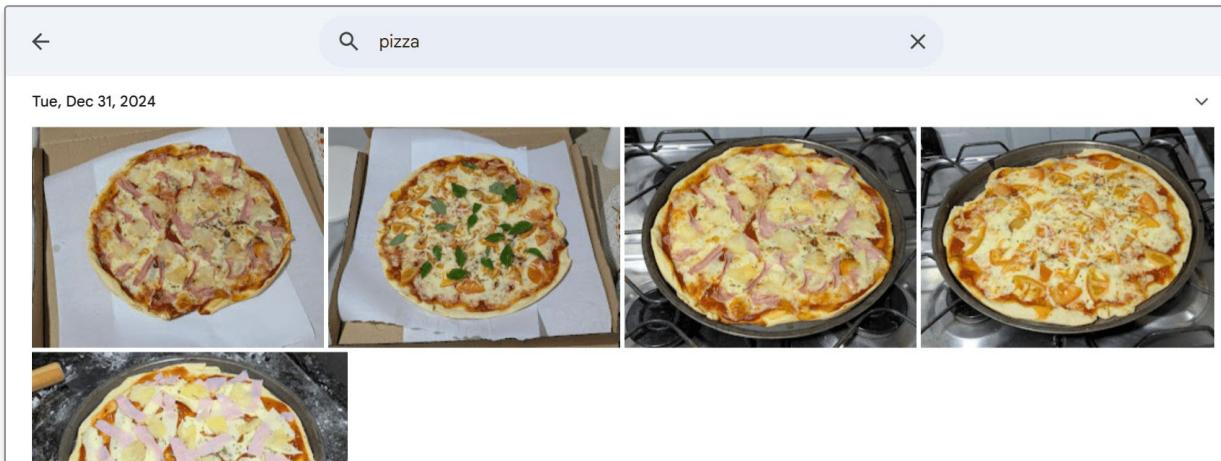


Abu-Mostafa (2012)

A screenshot of a Gmail inbox. The messages listed are:

- ADMIN HELPDESK - Mar 10: Recebeu esta mensagem porque atingiu o limite de quo... (starred)
- Bolt PureWave - Mar 10: Seu e-mail precisa de validação urgente! Ação Necessária! Olá, Detectamos que seu e-mail aind... (starred)
- Translator #1 - Mar 9: Tradutor de idiomas bidirecional instantâneo. Tradutor bidirecional instantâneo. Tradução portátil em ... (starred)
- yrdodjr@pillingers.de - Mar 6: O melhor calçado com palmilha anatômica! Conforto ortopédico para todos os dias! Experimente o ... (starred)
- Cloedar Sicurezza - Mar 5: Melhore sua postura e realce sua figura com o sutiã... Sutiã corretivo para melhorar a postura e conforto incrí... (starred)
- Insoles Magnetic - Mar 4: Palmilhas para melhorar a postura e reduzir a dor n... Sempre em pé? Akusoli alivia a dor com a terapia magn...

Interface dos websites mail.google.com e photos.google.com. Fotografias do autor.



# 2

## Trabalhos Relacionados

## 2 Trabalhos Relacionados

# Introduções de algoritmos

- A comparação de classificadores normalmente é feita na introdução de novos algoritmos
  - Japkowicz e Shah (2011) comparam com a indústria farmacêutica

*Algoritmos nos quais os autores possuem o maior interesse tendem a obter os melhores resultados*

— King *et al.* (1995)

## Comparações anteriores

- Também existem comparações imparciais, mas são imperfeitas
  - King *et al.* (1995) e Japkowicz e Shah (2011) fazem críticas:
    - única métrica de desempenho;
    - dependência de única fonte de dados;
    - baixa diversidade de algoritmos;
    - falta de transparência sobre implementação;
    - falta de transparência sobre otimização de parâmetros.



Anvisa (1 nov.2018). Disponível em:  
<https://facebook.com/photo.php?fbid=1683289781780702>.  
Acesso em 15 mar.2025.

## ② Trabalhos Relacionados

Características de trabalhos de comparação de classificadores

Artigo	Número de classificadores	Métricas	Dados
Amancio <i>et al.</i> (2014)	9	Acurácia	Artificiais
Fernández-Delgado <i>et al.</i> (2014)	179	Acurácia	96% UCI 4% mundo real
Oh (2011)	4	Acurácia	100% UCI
Lim <i>et al.</i> (2000)	33	Acurácia Tempo	100% UCI, adição de ruído
Singh <i>et al.</i> (2017)	3	Acurácia Tempo	100% UCI, adição de ruído
Zhang <i>et al.</i> (2017)	11	Acurácia AUC Tempo	UCI, KEEL, LibSVM, adição de ruído
<b>Este trabalho</b>	5	Acurácia AUC F-score LogLoss MSE RMSE Tempo	Artificiais, adição de ruído, desequilíbrio de dados + Teste com dados de mundo real

# 3

## Experimentos iniciais

### ③ Experimentos iniciais

#### Escolha de classificadores

- Escolha baseada em algoritmos com bom desempenho em trabalhos comparativos anteriores
- Ênfase em métodos com funcionamento distinto
- Implementações do scikit (Hao e Ho, 2019)

#### Otimização de parâmetros

- Treinar o modelo com diversos (hiper-)parâmetros e selecionar o conjunto com melhor desempenho
- Experimentos iniciais sem otimização

##### ① SVM

Support Vector Machine

- *Kernel* pré-computado
- *Kernel* gaussiano/RBF
- *Kernel* polinomial

##### ② GBDT

Stochastic Gradient Boosted Decision Trees

##### ③ RF

Random Forest

##### ④ kNN

k-Nearest Neighbors

##### ⑤ NB

Gaussian Naïve Bayes

### ③ Experimentos iniciais

## Escolha de métricas de desempenho

- Como medir se um classificador é melhor que outro?
- Usa-se muito **acurácia** – número de acertos sobre total

Sem câncer	Com câncer
10,923	260

He e Garcia (2009)

- Um classificador que sempre responde "**Não**" teria ~98% de acurácia
- Para dados desbalanceados, "a prática convencional de usar um único critério de avaliação não proporciona as informações adequadas" —He e Garcia (2009)
- Escolha com base em comparações de métricas: Ferri *et al.* (2009), Labatut e Cherifi (2012)

① Acurácia

② F-score

③ LogLoss

④ Area Under ROC Curve (AUC)

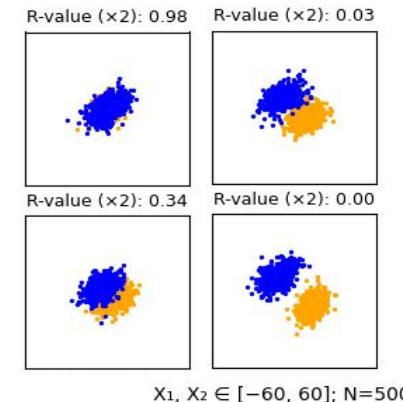
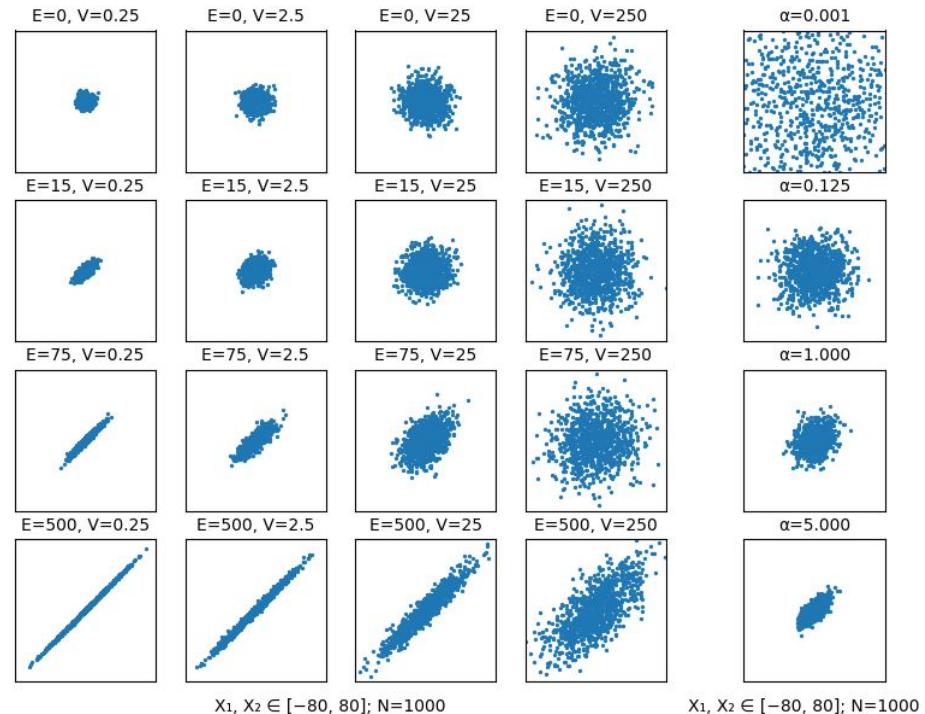
⑤ Mean Squared Error (MSE)

⑥ Tempo de treino

### ③ Experimentos iniciais

## Geração de dados artificiais

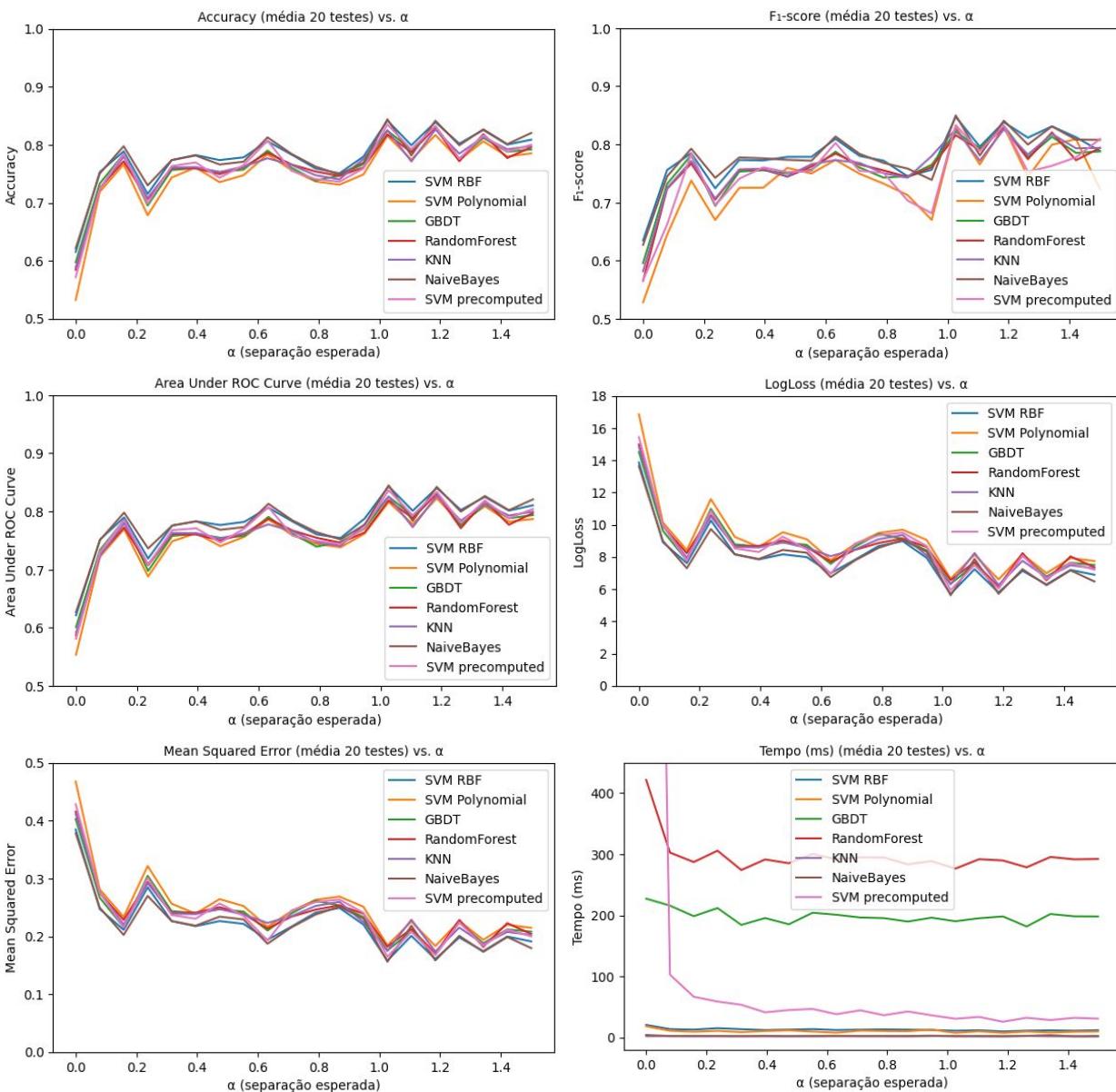
- Como evitar o viés de reuso de dados?  
Dados precisam ser inéditos
- Dados artificiais / sintéticos
- Gerador de dados adaptado de Hirschberger *et al.* (2007), inspirado por Amancio *et al.* (2014)
  - Dados a partir de matrizes de covariância aleatórias
  - Permite dados desbalanceados (e.g. 75% classe A, 25% classe B)
  - Permite adição de ruído
- Cálculo de sobreposição das classes usando a medida de R-value de Oh (2011)



# ③ Experimentos iniciais

## Resultados

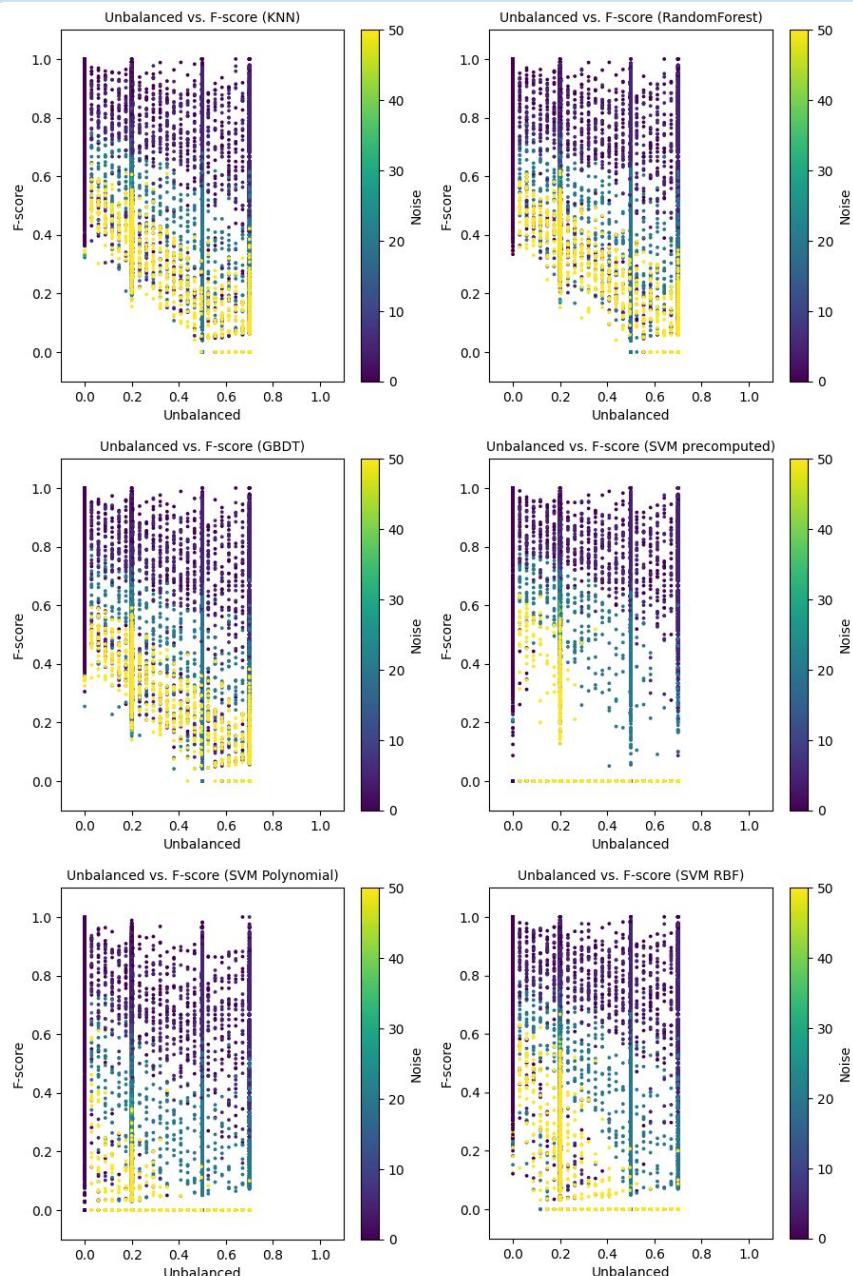
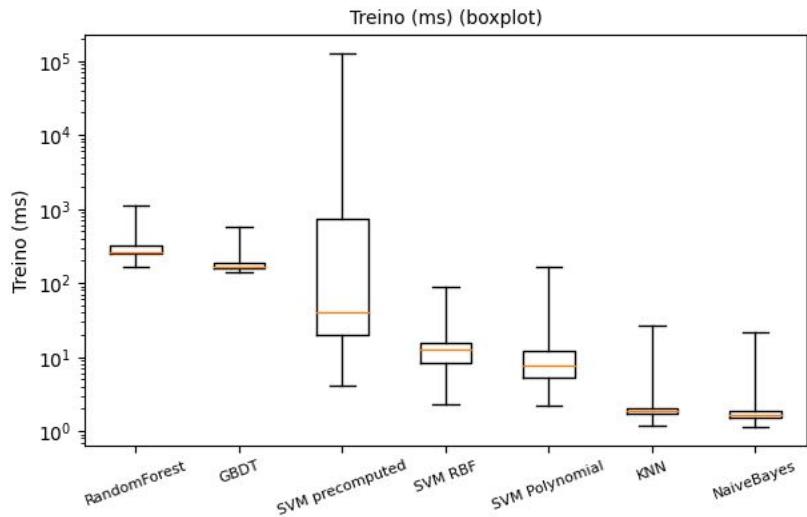
- Primeiros resultados: desempenho vs.  $\alpha$
- Desempenho muito similar entre classificadores
- Exceto em tempo de treino



# ③ Experimentos iniciais

## Resultados

- Tempo de treino
- Resistência a ruído



# 4

# Experimentos revisados

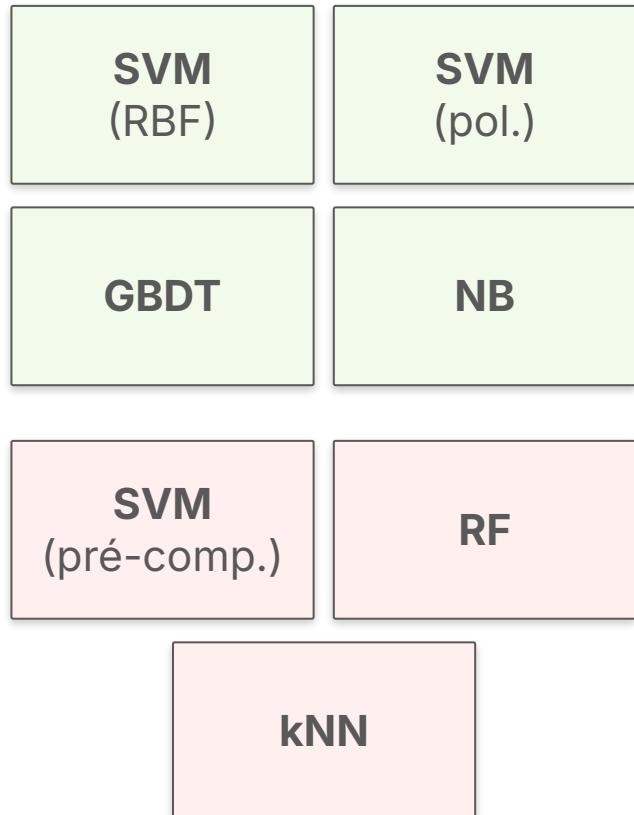
## ④ Experimentos revisados

### Modificações

- Tempo de treino é injusto
- Desempenho muito similar
- Otimização de parâmetros ajudaria?

### Otimização de parâmetros

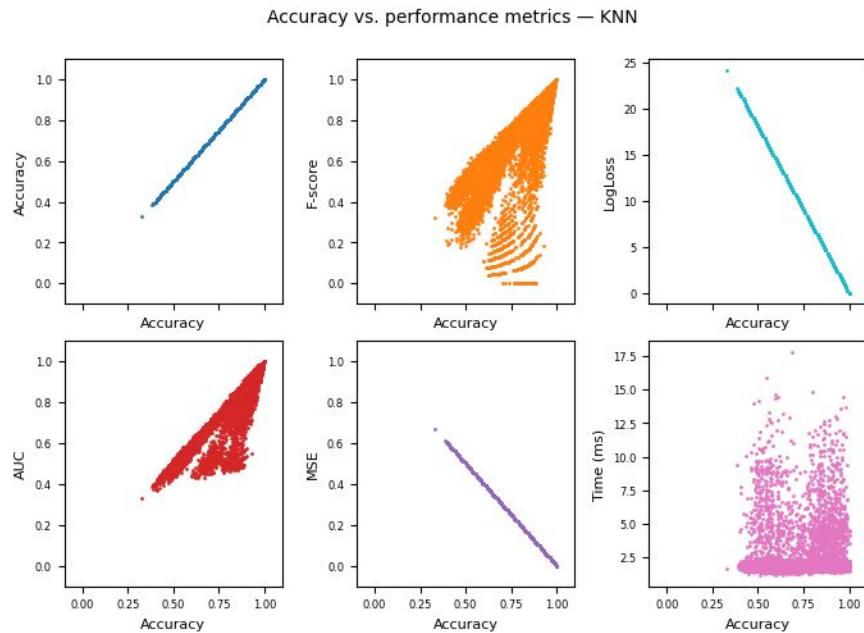
- Otimização é custosa
- Quais classificadores otimizar?
- Quais parâmetros?
- Random Search, validação cruzada 5 folds com 25 valores para cada parâmetro otimizado



# 4 Experimentos revisados

## Revisão de métricas

- Relações lineares entre métricas
- Tempo de treino não é justo – kNN e NB não "treinam", por exemplo
- Remoção de LogLoss, MSE
- Introdução de RMSE (Caruana e Niculescu-Mizil, 2004), tempo de teste e total



① Acurácia

② F-score

③ Root Mean Squared Error (RMSE)

④ Area Under ROC Curve (AUC)

⑤ Tempo (treino)

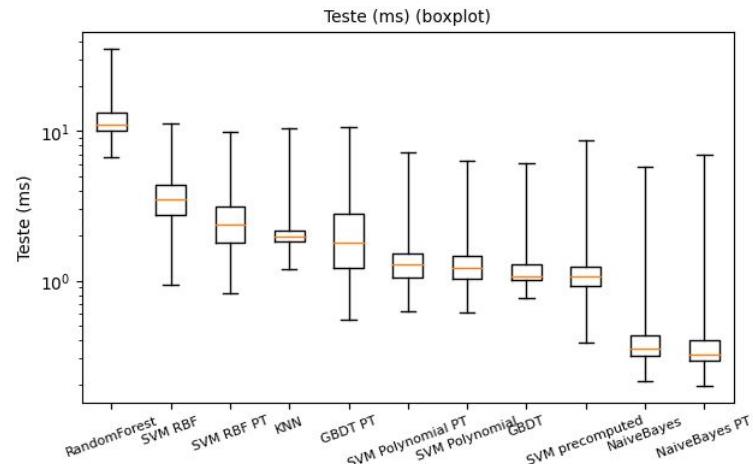
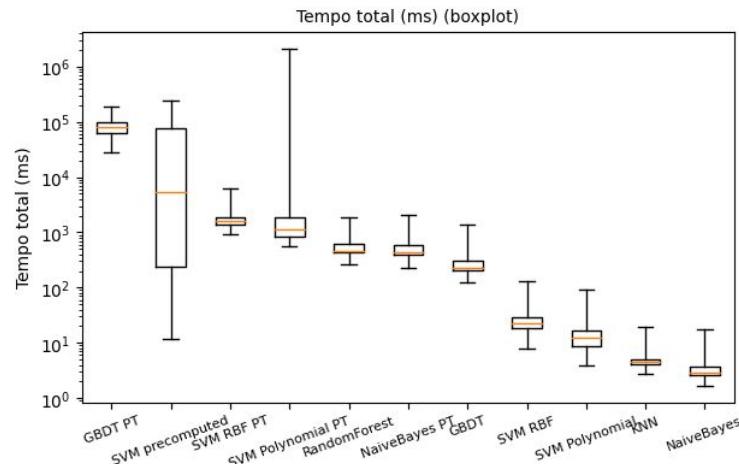
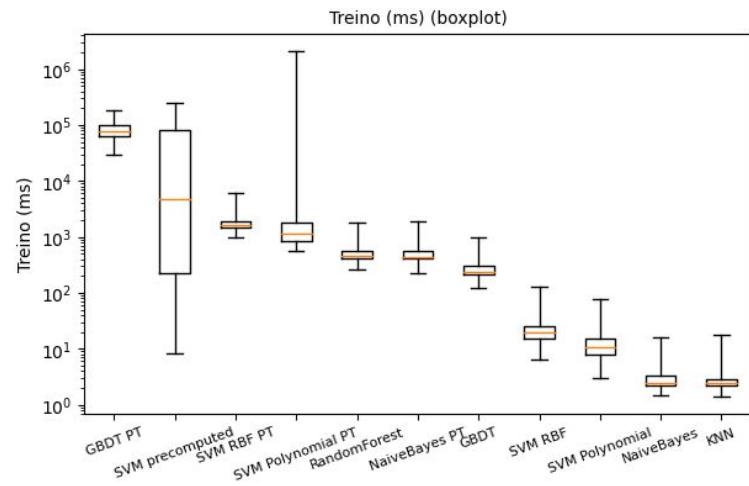
⑥ Tempo (teste)

⑦ Tempo (total)

# 4) Experimentos revisados

## Resultados – Tempo

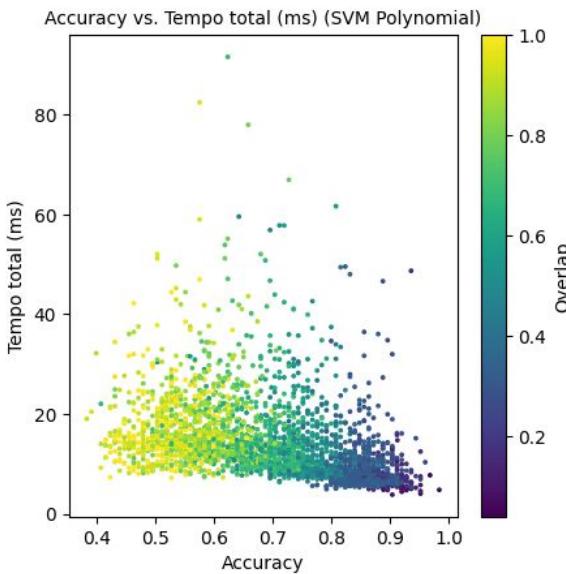
- Tempo de teste não afeta tanto assim
- No geral a ordem se mantém, mas agora modelos otimizados entram na conta
- kNN e NB se mantém como mais rápidos
- GBDT otimizado é o mais lento



# 4 Experimentos revisados

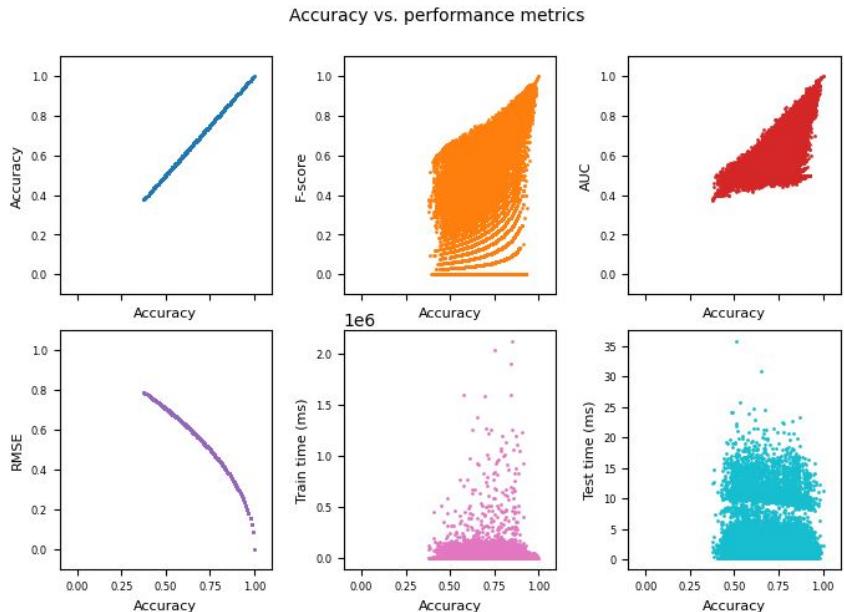
## Tempo vs. desempenho

- Correlação entre desempenho e tempo em alguns casos
- Não significa que modelos mais rápidos são melhores
- Mas sim que problemas mais "fáceis" são mais rápidos de resolver
- "Facilidade" tem correlação com sobreposição



## Correlação entre métricas

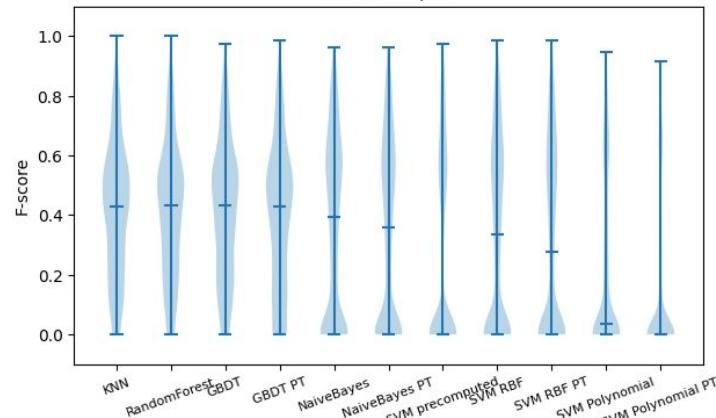
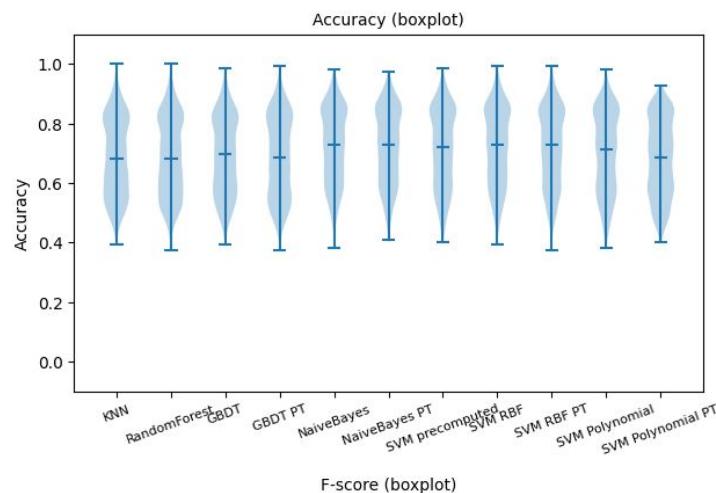
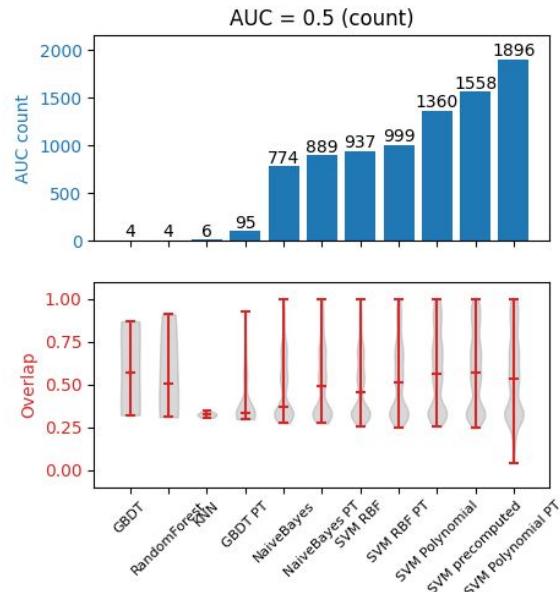
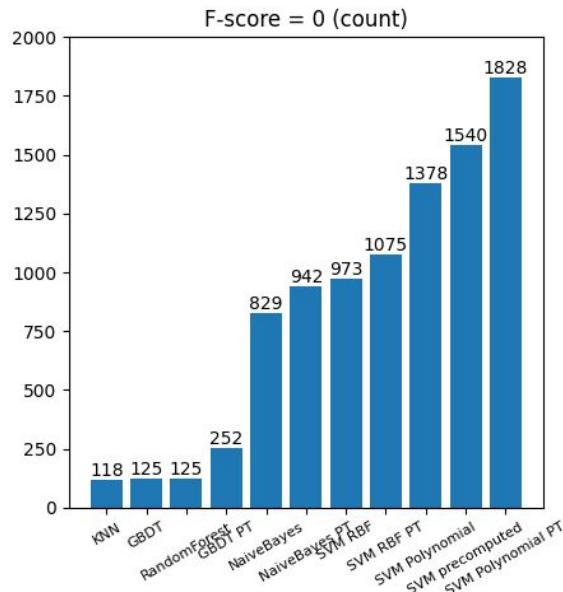
- As métricas revisadas apresentam, claro, alguma relação, mas não mais linear
- A troca foi positiva



# 4 Experimentos revisados

## Resultados

- Desempenho ainda muito similar
- Mesmo com otimização de parâmetros!
- **Erros catastróficos:** F-score = 0 e AUC = 0.5 indicam modelos que não conseguiram aprender os dados
- SVMs e NB possuem ocorrências mais altas de F-score = 0 e AUC = 0.5

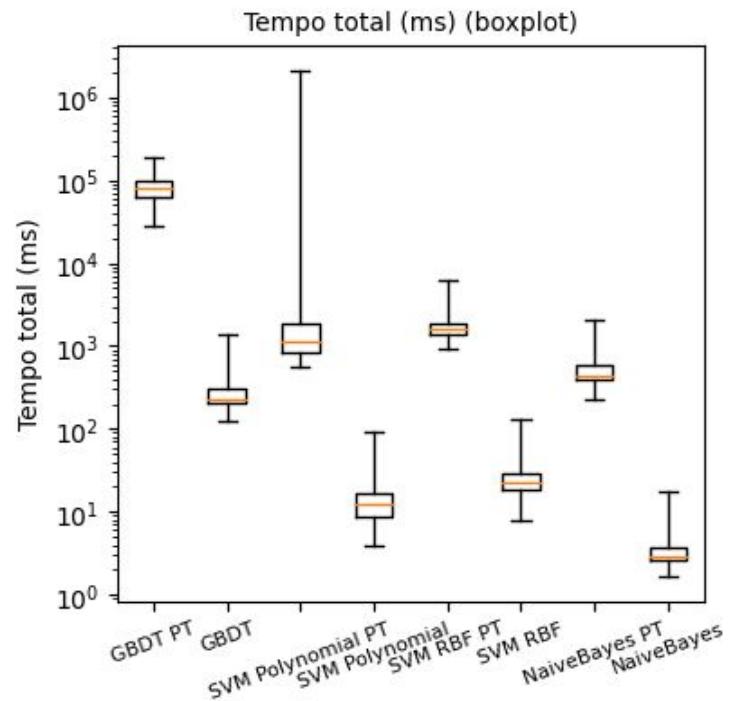


# 4) Experimentos revisados

## Resultados

- Otimização aumenta significativamente o tempo de execução do modelo
- Vale a pena?

Modelo	Métrica	Diferença de média entre modelos Teste-t pareado (p-value)			
		Todos	<i>noise, unbal = 0</i>	<i>noise, unbal &gt; 0</i>	<i>overlap &gt; 0.85</i>
SVM RBF	Acurácia	-0.00324 0.000	+0.00128 0.545	-0.00301 0.000	-0.00470 0.001
	F-score	-0.01836 0.000	-0.00187 0.572	-0.02150 0.000	-0.04569 0.000
	AUC	-0.00449 0.000	+0.00043 0.844	-0.00505 0.000	-0.00688 0.000
	RMSE	+0.00310 0.000	+0.00006 0.980	+0.00296 0.000	+0.00355 0.001
SVM Pol.	Acurácia	-0.02119 0.000	-0.11568 0.000	-0.01315 0.000	+0.00081 0.601
	F-score	-0.07792 0.000	-0.19312 0.000	-0.06517 0.000	+0.00869 0.186
	AUC	-0.03258 0.000	-0.10944 0.000	-0.02312 0.000	+0.00054 0.667
	RMSE	+0.02145 0.000	+0.10896 0.000	+0.01333 0.000	-0.00058 0.612
NaiveBayes	Acurácia	-0.00019 0.385	-0.00176 0.188	+0.00022 0.373	-0.00074 0.235
	F-score	-0.01679 0.000	-0.00643 0.039	-0.02052 0.000	-0.02393 0.000
	AUC	-0.00395 0.000	-0.00129 0.313	-0.00443 0.000	-0.00197 0.001
	RMSE	+0.00019 0.348	+0.00168 0.119	-0.00022 0.347	+0.00057 0.222
GBDT	Acurácia	-0.00209 0.005	+0.00080 0.893	-0.00131 0.144	-0.00875 0.000
	F-score	-0.01503 0.000	-0.00289 0.752	-0.01830 0.000	-0.00195 0.534
	AUC	-0.00645 0.000	-0.00019 0.975	-0.00669 0.000	-0.00747 0.000
	RMSE	+0.00084 0.195	-0.00179 0.724	+0.00004 0.957	+0.00649 0.000



# 5

## Dados reais

## 5 Dados reais

### Conjuntos de dados

- Experimentação limitada com dados reais
- Duas ocorrências de SVM (pol.) otimizado com erros catastróficos

Higher Education Students Performance Evaluation

McDonald's Stock Price

NBA Rookies' Career

Problema 1: Matriz de confusão SVM (pol.) otim. [F-score=0]

	Predição Positiva	Predição Negativa	Total:
Positivo	0 (True Positive)	11 (False Negative)	11 29.73%
Negativo	0 (False Positive)	26 (True Negative)	26 70.27%
Total:	0 0%	37 100%	37 100%

Problema 2: Matriz de conf. SVM (pol.) otim. [F-score=0.73]

	Predição Positiva	Predição Negativa	Total:
Positivo	182 (True Positive)	0 (False Negative)	182 57.78%
Negativo	133 (False Positive)	0 (True Negative)	133 42.22%
Total:	315 100%	0 0%	315 100%

# 6

# Discussão

## ⑥ Discussão

### Resultados agregados

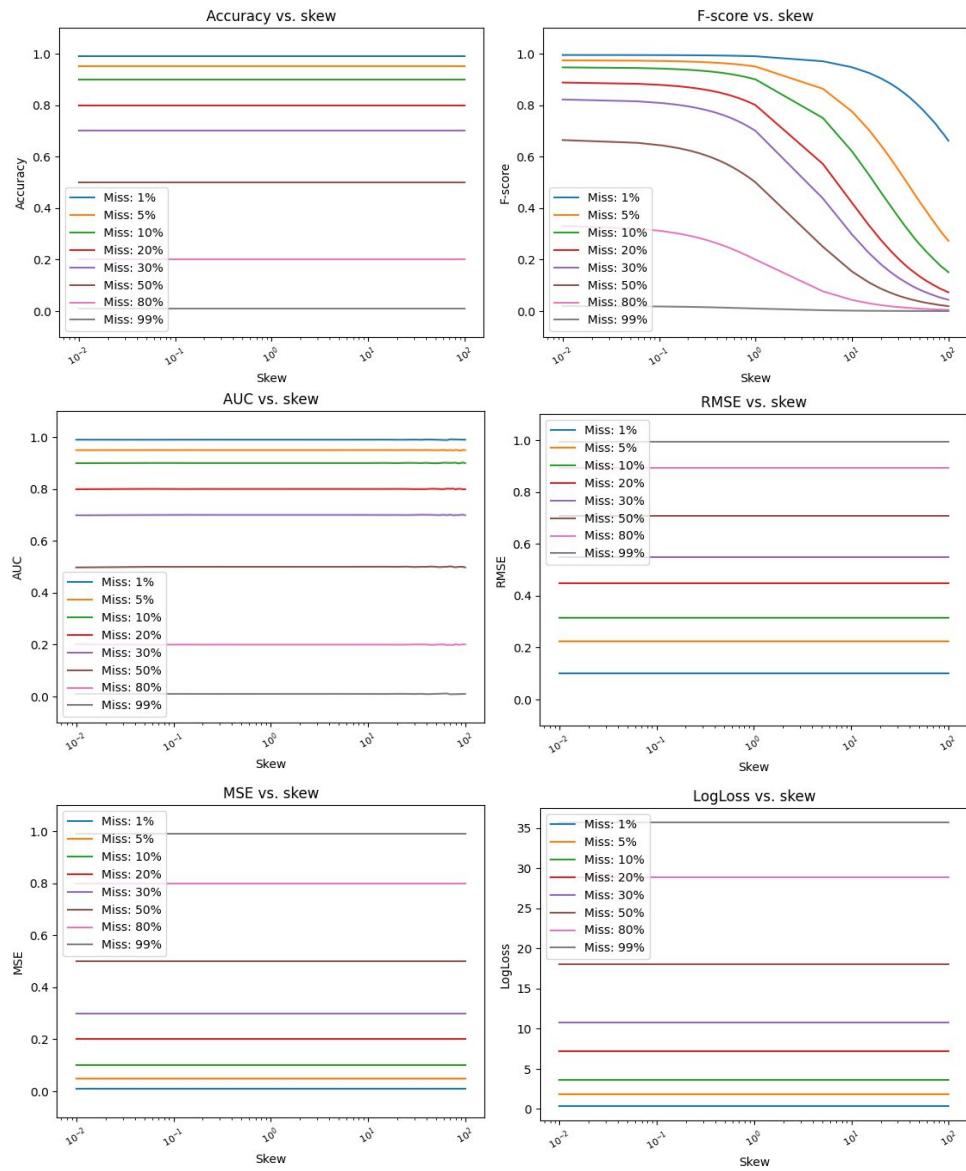
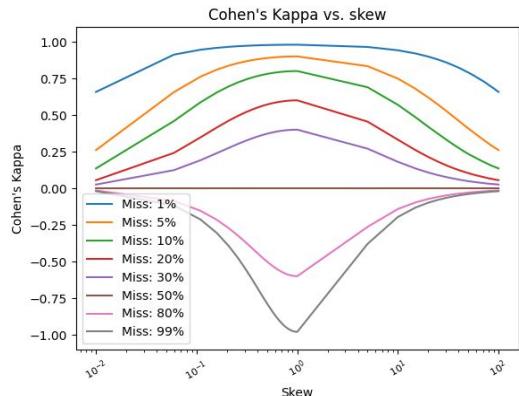
- Todos os classificadores tiveram desempenhos muito similares
- Contudo, é possível montar uma tabela de desempenho relativo entre modelos

Modelo	Resistência a deseq., ruído	Tempo de execução	Propensão a erros catastróficos
kNN	Maior	Rápido	Menor
RF	Maior	Lento	Menor
GBDT	Maior	Lento	Menor
SVMs	Menor	Meio-termo	Maior
NB	Menor	Rápido	Maior

# 6 Discussão

## Skew

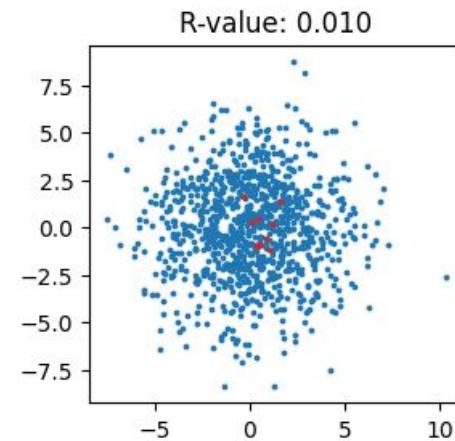
- Jeni et al. (2013) apresenta uma forma de visualizar a resistência a desequilíbrio de dados
- $Skew = \frac{\text{Qtd. pontos neg.}}{\text{Qtd. pontos pos.}}$
- Gera-se uma pseudo-classificação com percentual de acertos simulados
- Mede-se o valor de cada métrica de desempenho
- F-score só é resistente a skew para um lado
- Outras métricas como Kappa de Cohen (1960) resistem melhor



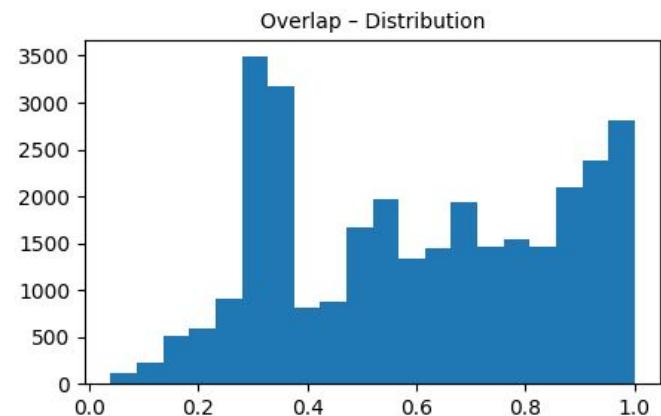
## ⑥ Discussão

### Considerações finais

- Métrica de sobreposição não é resistente a dados desequilibrados
- Pico inexplicado em ~30% de sobreposição
- Otimização em muitas vezes piorou o desempenho
  - Random Search não cobre 100% do espaço
  - Implementação otimiza por acurácia!
  - Propagação de falhas: otimiza-se para uma métrica errônea, e obtém-se um bom desempenho, medido pela mesma métrica



Duas classes com parâmetros idênticos com exceção de N, com valores de 1000 e 10.



**7**

# **Conclusão**

# 7 Conclusão

## Resumo do trabalho

- Testamos 5 classificadores distintos, mas 7 modelos
- Fazendo uso de 9 métricas de desempenho no total
- Otimização de parâmetros
  - Em nosso experimento, não compensou o tempo de treino adicional
- Dados artificiais
- Implementações da biblioteca scikit
- Desempenhos muito similares
- Em geral, kNN, RF e GBDT mostraram maior resiliência em comparação com SVMs e NB
- Grande diferença em termos de tempo: RF e GBDT são os mais lentos, e kNN e NB são os mais velozes

## Trabalhos futuros

- Muito ainda a ser explorado
- Não experimentamos com variações de qtd. de pontos, de dimensões ou de classes
- Outras distribuições, e.g. multimodal
- Melhores métricas de desempenho, levando em consideração Jeni *et al.* (2013)
- Melhor cálculo de sobreposição de classes
- Outras implementações? Linguagens?



Sloane MS 2435 (circa séc. XIII). British Library.

# Obrigado!



**Matheus Avellar**

**Contato:**

avellar@cos.ufrj.br

**Slides:**

[s.avl.la/dissertacao-ppt](https://s.avl.la/dissertacao-ppt)



**UFRJ**

UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



**COPPE**  
UFRJ



Prof. Orientador: Carlos Eduardo Pedreira  
Prof.<sup>a</sup> Orientadora: Laura de Oliveira Fernandes Moraes

# Referências Bibliográficas

- BLANK, G., REISDORF, B. C. "The Participatory Web", *Information, Communication & Society*, v. 15, n. 4, pp. 537–554, Mar. 2012. doi: 10.1080/1369118X.2012.665935. Disponível em: <<https://doi.org/10.1080/1369118X.2012.665935>>.
- MCGRADY, R., ZHENG, K., CURRAN, R., et al. "Dialing for Videos: A Random Sample of YouTube", *Journal of Quantitative Description: Digital Media*, v. 3, Dez. 2023. ISSN: 2673-8813. doi: 10.51685/jqd.2023.022. Disponível em: <<https://journalqd.org/article/view/4066>>.
- SAIDANI, N., ADI, K., ALLILI, M. S. "A semantic-based classification approach for an enhanced spam detection", *Computers & Security*, v. 94, pp. 101716, 2020. ISSN: 0167-4048. doi: <https://doi.org/10.1016/j.cose.2020.101716>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167404820300043>>.
- KARIM, A., AZAM, S., SHANMUGAM, B., et al. "A Comprehensive Survey for Intelligent Spam Email Detection", *IEEE Access*, v. 7, pp. 168261–168295, 2019. doi: 10.1109/ACCESS.2019.2954791.
- ABDALLAH, A., MAAROF, M. A., ZAINAL, A. "Fraud detection system: A survey", *Journal of Network and Computer Applications*, v. 68, pp. 90–113, 2016. ISSN: 1084-8045. doi: <https://doi.org/10.1016/j.jnca.2016.04.007>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1084804516300571>>.
- HORIGUCHI, S., AMANO, S., OGAWA, M., et al. "Personalized Classifier for Food Image Recognition", *IEEE Transactions on Multimedia*, v. 20, n. 10, pp. 2836–2848, 2018. doi: 10.1109/TMM.2018.2814339.
- GUTTA, S., WECHSLER, H. "Face recognition using hybrid classifier systems". In: *Proceedings of International Conference on Neural Networks (ICNN'96)*, v. 2, pp. 1017–1022 vol.2, 1996. doi: 10.1109/ICNN.1996.549037.
- HOSSAIN, M. E., KABIR, M. A., ZHENG, L., et al. "A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions", *Artificial Intelligence in Agriculture*, v. 6, pp. 138–155, 2022. ISSN: 2589-7217. doi: 10.1016/j.aiia.2022.09.002. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2589721722000125>>.
- ANGHEL, A., PAPANDREOU, N., PARNELL, T., et al. "Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms". 2019. Disponível em: <<https://arxiv.org/abs/1809.04559>>.
- YILMAZ, N., SEKEROGLU, B. "Student Performance Classification Using Artificial Intelligence Techniques". In: Aliev, R. A., Kacprzyk, J., Pedrycz, W., et al. (Eds.), *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019*, pp. 596–603, Cham, CHE, 2020. Springer International Publishing. ISBN: 978-3-030-35249-3. doi: 10.1007/978-3-030-35249-3\_76.
- JENI, L. A., COHN, J. F., DE LA TORRE, F. "Facing Imbalanced Data-Recommendations for the Use of Performance Metrics". In: *2013 Human-Computer Interaction Conference on Affective Computing and Intelligent Interaction*, pp. 245–251, Geneva, CHE, Set. 2013. doi: 10.1109/ACII.2013.47. Disponível em: <<https://sites.pitt.edu/~jeffcohn/skew/PID2829477.pdf>>.
- ELGELDAWI, E., SAYED, A., GALAL, A. R., et al. "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis", *Informatics*, v. 8, n. 4, 2021. ISSN: 2227-9709. doi: 10.3390/informatics8040079. Disponível em: <<https://www.mdpi.com/2227-9709/8/4/79>>.
- REJANI, Y. I. A., SELVI, S. T. "Early Detection of Breast Cancer using SVM Classifier Technique". 2009. Disponível em: <<https://arxiv.org/abs/0912.2314>>.
- KARABATAK, M. "A new classifier for breast cancer detection based on Naïve Bayesian", *Measurement*, v. 72, pp. 32–36, 2015. ISSN: 0263-2241. doi: <https://doi.org/10.1016/j.measurement.2015.04.028>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263224115002419>>.
- ABU-MOSTAFA, Y. S., MAGDON-ISMAIL, M., LIN, H.-T. *Learning From Data: A Short Course*. AMLBook, 2012. ISBN: 9781600490064.
- ABU-MOSTAFA, Y. S. "The Learning Problem". In: *Machine Learning Course (CS 156)*, Pasadena, USA, Abr. 2012. California Institute of Technology. Disponível em: <<https://youtu.be/mbyG85GZ0PI>>.
- DUDA, R. O., HART, P. E., STORK, D. G. *Pattern Classification*. 2nd. ed. California, USA, Wiley, Nov. 2000. ISBN: 978-0-471-05669-0. Disponível em: <<https://www.wiley.com/en-us/Pattern+Classification%2C2nd+Edition-p-9780471056690>>.
- FERNÁNDEZ-DELGADO, M., CERNADAS, E., BARRO, S., et al. "Do we need hundreds of classifiers to solve real world classification problems?", *Journal of Machine Learning Research*, v. 15, n. 1, pp. 3133–3181, Jan. 2014. ISSN: 1532-4435. doi: 10.5555/2627435.2697065. Disponível em: <<https://dl.acm.org/doi/10.5555/2627435.2697065>>.
- AMANCIO, D. R., COMIN, C. H., CASANOVA, D., et al. "A Systematic Comparison of Supervised Classifiers", *PLOS ONE*, v. 9, n. 4, pp. 1–14, Abr. 2014. doi: 10.1371/journal.pone.0094137.
- LIM, T.-S., LOH, W.-Y., SHIH, Y.-S. "A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms", *Machine Learning*, v. 40, n. 3, pp. 203–228, Set. 2000. ISSN: 1573-0565. doi: 10.1023/A:1007608224229.
- COHEN, J. "A Coefficient of Agreement for Nominal Scales", *Educational and Psychological Measurement*, v. 20, n. 1, pp. 37–46, 1960. doi: 10.1177/001316446002000104.
- SINGH, A., HALGAMUGE, M. N., LAKSHMIGANTHAN, R. "Impact of Different Data Types on Classifier Performance of Random Forest, Naïve Bayes, and K-Nearest Neighbors Algorithms", *International Journal of Advanced Computer Science and Applications*, v. 8, n. 12, 2017. doi: 10.14569/IJACSA.2017.081201.
- ZHANG, C., LIU, C., ZHANG, X., et al. "An up-to-date comparison of state-of-the-art classification algorithms", *Expert Systems with Applications*, v. 82, pp. 128–150, 2017. ISSN: 0957-4174. doi: 10.1016/j.eswa.2017.04.003. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417417302397>>.
- KING, R. D., FENG, C., SUTHERLAND, A. "STATLOG: Comparison of Classification Algorithms on Large Real-World", *Applied Artificial Intelligence*, v. 9, n. 3, pp. 289–333, 1995. doi: 10.1080/08839519508945477.
- JAPKOWICZ, N., SHAH, M. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge, GBR, Cambridge University Press, Ago. 2011. ISBN: 9780511921803. doi: 10.1017/CBO9780511921803. Disponível em: <<https://www.cambridge.org/core/books/evaluating-learning-algorithms/3CB22D16AB609D1770C24CA2CB5A11BF>>.
- OH, S. "A new dataset evaluation method based on category overlap", *Computers in Biology and Medicine*, v. 41, n. 2, pp. 115–122, 2011. ISSN: 0010-4825. doi: 10.1016/j.combiomed.2010.12.006. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0010482510001757>>.
- HAO, J., HO, T. K. "Machine Learning Made Easy: A Review of 'Scikit-learn' Package in Python Programming Language", *Journal of Educational and Behavioral Statistics*, v. 44, n. 3, pp. 348–361, 2019. ISSN: 1076-9986, 1935-1054. Disponível em: <<http://www.jstor.org/stable/45278427>>.
- FRIEDMAN, J. H. "Stochastic gradient boosting", *Computational Statistics & Data Analysis: Nonlinear Methods and Data Mining*, v. 38, n. 4, pp. 367–378, 2002. ISSN: 0167-9473. doi: 10.1016/S0167-9473(01)00065-2. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167947301000652>>.
- FAWAGREH, K., GABER, M. M., ELYAN, E. "Random forests: from early developments to recent advancements", *Systems Science & Control Engineering*, v. 2, n. 1, pp. 602–609, 2014. doi: 10.1080/21642583.2014.956265. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/21642583.2014.956265>>.

# Referências Bibliográficas

- WICKRAMASINGHE, I., KALUTARAGE, H. "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation", *Soft Computing*, v. 25, n. 3, pp. 2277–2293, Fev. 2021. ISSN: 1433-7479. doi: 10.1007/s00500-020-05297-6. Disponível em: <<https://link.springer.com/article/10.1007/s00500-020-05297-6>>.
- SYARIF, I., PRUGEL-BENNETT, A., WILLS, G. "SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance", *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, v. 14, n. 4, pp. 1502–1509, Dez. 2016. ISSN: 1693-6930. doi: 10.12928/telkomnika.v14i4.3956. Disponível em: <<https://telkomnika.uan.ac.id/index.php/TELKOMNIKA/article/view/3956>>.
- XIA, Y., LIU, C., LI, Y., et al. "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring", *Expert Systems with Applications*, v. 78, pp. 225–241, 2017. ISSN: 0957-4174. doi: 10.1016/j.eswa.2017.02.017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417417301008>>.
- HE, H., GARCIA, E. A. "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, pp. 1263–1284, 2009. doi: 10.1109/TKDE.2008.239. Disponível em: <<https://ieeexplore.ieee.org/document/5128907>>.
- FERRI, C., HERNÁNDEZ-ORALLO, J., MODROIU, R. "An experimental comparison of performance measures for classification", *Pattern Recognition Letters*, v. 30, n. 1, pp. 27–38, 2009. ISSN: 0167-8655. doi: 10.1016/j.patrec.2008.08.010. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865508002687>>.
- LABATUT, V., CHERIFI, H. "Accuracy Measures for the Comparison of Classifiers". 2012. Disponível em: <<https://arxiv.org/abs/1207.3790>>.
- HIRSCHBERGER, M., QI, Y., STEUER, R. E. "Randomly generating portfolio-selection covariance matrices with specified distributional characteristics", *European Journal of Operational Research*, v. 177, n. 3, pp. 1610–1625, 2007. ISSN: 0377-2217. doi: 10.1016/j.ejor.2005.10.014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0377221705006600>>.
- PROBST, P., BOULESTEIX, A.-L. "To Tune or Not to Tune the Number of Trees in Random Forest", *Journal of Machine Learning Research*, v. 18, n. 181, pp. 1–18, 2018. Disponível em: <<http://jmlr.org/papers/v18/17-269.html>>.
- OSHIRO, T. M., PEREZ, P. S., BARANAUSKAS, J. A. "How Many Trees in a Random Forest?" In: Perner, P. (Ed.), *Machine Learning and Data Mining in Pattern Recognition*, pp. 154–168, Berlin, DEU, 2012. Springer Science+Business Media. ISBN: 978-3-642-31537-4. doi: 10.1007/978-3-642-31537-4\_13.
- LOUPPE, G. "Understanding Random Forests: From Theory to Practice". 2015. Disponível em: <<https://arxiv.org/abs/1407.7502>>.
- INYANG, U. G., IJEBU, F. F., OSANG, F. B., et al. "A Dataset-Driven Parameter Tuning Approach for Enhanced K-Nearest Neighbour Algorithm Performance", *International Journal on Advanced Science, Engineering and Information Technology*, v. 13, n. 1, pp. 380–391, Jan. 2023. doi: 10.18517/ijaseit.13.1.16706. Disponível em: <<https://ijaseit.insightsociety.org/index.php/ijaseit/article/view/16706>>.
- BATISTA, G. E. D. A. P. A., SILVA, D. F. "How k-Nearest Neighbor Parameters Affect its Performance". In: *X Argentine Symposium on Artificial Intelligence*, pp. 1–12, Mar del Plata, ARG, Ago. 2009. Sociedad Argentina de Informática, 38th Jornadas Argentinas de Informática e Investigación Operativa. Disponível em: <<https://s.avl.la/6ucxg>>.
- HANCOCK, J., KHOSHGOFTAAR, T. M. "Impact of Hyperparameter Tuning in Classifying Highly Imbalanced Big Data". In: *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 348–354, 2021. doi: 10.1109/IRI51335.2021.00054.
- CARUANA, R., NICULESCU-MIZIL, A. "Data mining in metric space: an empirical analysis of supervised learning performance criteria". In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, p. 69–78, New York, USA, 2004. Association for Computing Machinery. ISBN: 1581138881. doi: 10.1145/1014052.1014063.
- KOWALSKI, C. J. "On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient", *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 21, n. 1, pp. 1–12, 1972. ISSN: 0035-9254, 1467-9876. doi: 10.2307/2346598. Disponível em: <<http://www.jstor.org/stable/2346598>>.
- CHAPELLE, O., VAPNIK, V., BOUSQUET, O., et al. "Choosing Multiple Parameters for Support Vector Machines", *Machine Learning*, v. 46, n. 1, pp. 131–159, Jan. 2002. ISSN: 1573-0565. doi: 10.1023/A:1012450327387. Disponível em: <<https://link.springer.com/article/10.1023/A:1012450327387>>.
- LAVESSON, N., DAVIDSSON, P. "Quantifying the impact of learning algorithm parameter tuning". In: *Proceedings of the 21st National Conference on Artificial Intelligence*, v. 1, AAAI'06, p. 395–400. AAAI Press, 2006. ISBN: 9781577352815. doi: 10.5555/1597538.1597602. Disponível em: <<https://dl.acm.org/doi/10.5555/1597538.1597602>>.
- ABU ALFEILAT, H. A., HASSANAT, A. B., LASASSMEH, O., et al. "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review", *Big Data*, v. 7, n. 4, pp. 221–248, 2019. doi: 10.1089/big.2018.0175. Disponível em: <<https://www.liebertpub.com/doi/10.1089/big.2018.0175>>. PMID: 31411491.
- EL HINDI, K. M., ALJULAI DAN, R. R., ALSALMAN, H. "Lazy fine-tuning algorithms for naïve Bayesian text classification", *Applied Soft Computing*, v. 96, pp. 106652, 2020. ISSN: 1568-4946. doi: 10.1016/j.asoc.2020.106652. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494620305901>>.
- PROBST, P., WRIGHT, M. N., BOULESTEIX, A.-L. "Hyperparameters and tuning strategies for random forest", *WIREs Data Mining and Knowledge Discovery*, v. 9, n. 3, pp. e1301, 2019. doi: 10.1002/widm.1301. Disponível em: <<https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1301>>.
- EL HINDI, K., ALSALMAN, H., QASEM, S., et al. "Building an Ensemble of Fine-Tuned Naïve Bayesian Classifiers for Text Classification", *Entropy*, v. 20, n. 11, 2018. ISSN: 1099-4300. doi: 10.3390/e20110857. Disponível em: <<https://www.mdpi.com/1099-4300/20/11/857>>.
- ALONSO, S. G., DE LA TORRE DÍEZ, I., RODRIGUES, J. J. P. C., et al. "A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector", *Journal of Medical Systems*, v. 41, n. 11, pp. 183, Out. 2017. ISSN: 1573-689X. doi: 10.1007/s10916-017-0832-2. Disponível em: <<https://link.springer.com/article/10.1007/s10916-017-0832-2>>.