Uso do TF-IDF na Comparação de Dados para Detecção de Ransomware

Augusto Parisot, Lucila M. S. Bento, Raphael C. S. Machado

Universidade Federal Fluminense
Universidade do Estado do Rio de Janeiro

Motivação

20%

Incidentes em todo o mundo causados por ransomware

92h

Duração média de um ataque de ransomware

$ 1.82 M

Custo médio para se recuperar de ataques de ransomware

Dados dos relatórios "The State of Ransomware 2023" e "X-Force Threat Intelligence Index 2024"

Problema

Métodos tradicionais têm dificuldade em acompanhar a evolução rápida dessas ameaças.

Necessidade de soluções simples e computacionalmente eficientes para grandes volumes de dados.

Relatórios de análise dinâmica, como do Cuckoo Sandbox, contêm dados textuais complexos.

Contribuições

Dataset	TF-IDF	Exploração	Scripts
Comportamento das famílias de ransomware Ryuk, Revil, NetWalker, MountLocker, LockBit, Egregor, Conti e Clop obtido com o Cuckoo SandBox.	Uso dessa técnica de processamento de linguagem natural para identificar padrões nos dados de comportamento obtidos.	Comparação dos dados de rede, assinatura, chamadas de API e strings presentes no dataset para identificar os mais eficazes para a detecção.	Criação e compartilhamento de códigos para obtenção de amostras de malwares em fontes públicas e gratuitas.

TF-IDF

$$\mathrm{TF}(t,d) = \frac{\text{número de ocorrências do termo } t \text{ no documento } d}{\text{número total de termos no documento } d}$$

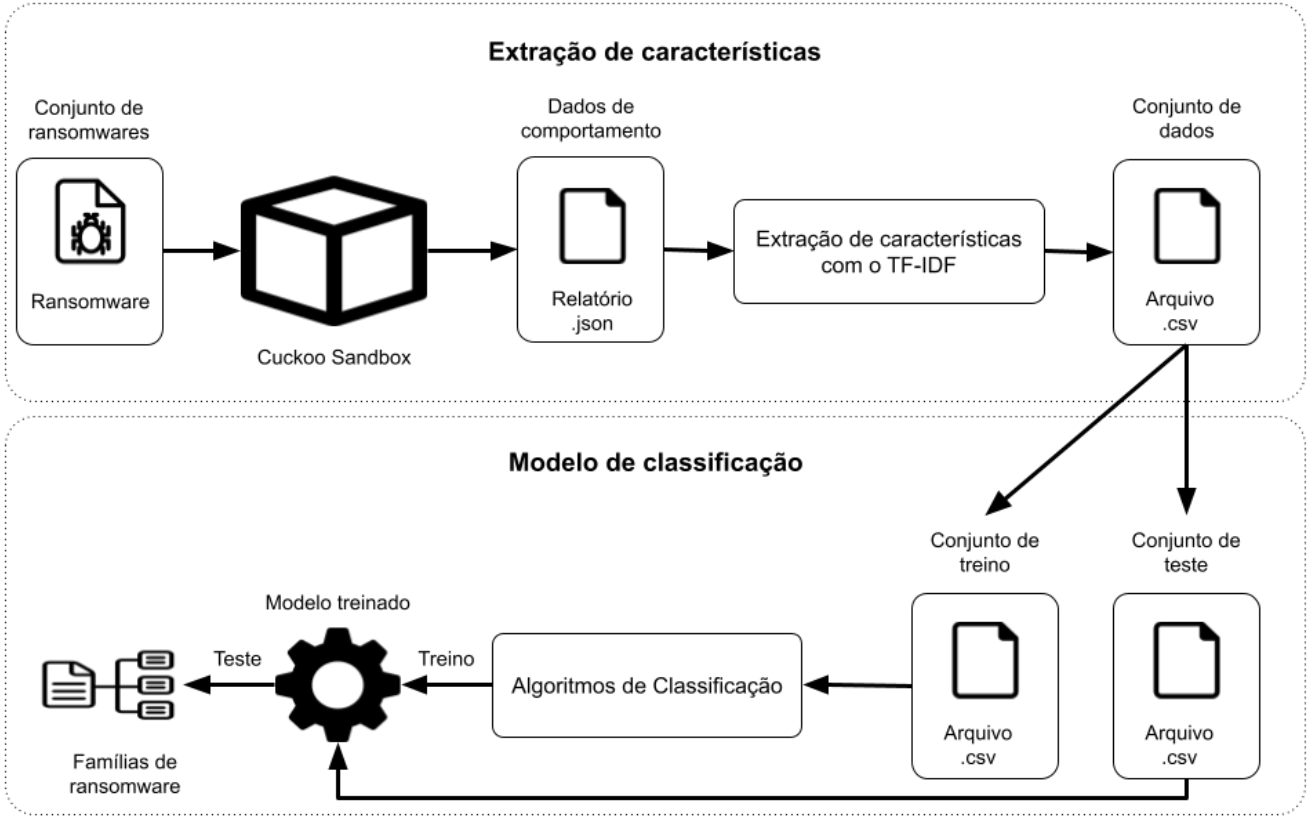
$$\mathrm{IDF}(t,D) = \log\left(\frac{\text{número total de documentos no } \textit{corpus } D}{\text{número de documentos que contêm o termo } t}\right)$$

$$\text{TF-IDF}(t,d,D) = TF(t,d) \times IDF(t,D)$$

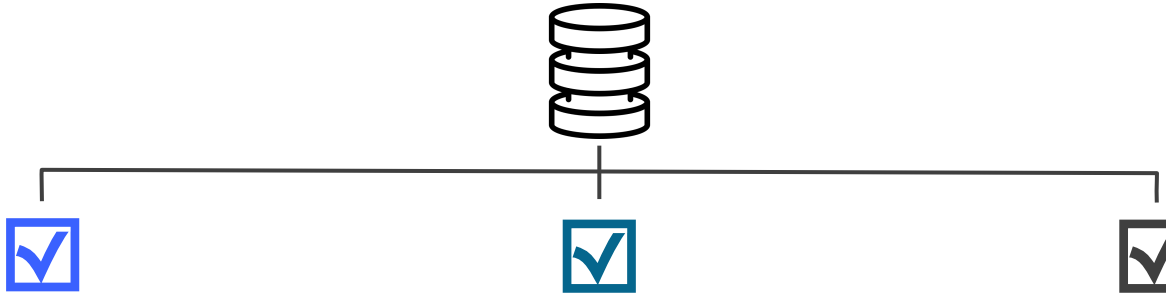
Frequência de Termo

Inverso da Frequência de documentos

Abordagem


Extração de características
Conjunto de ransomwares
Ransomware
Cuckoo Sandbox
Dados de comportamento
Relatório .json
Extração de características com o TF-IDF
Conjunto de dados
Arquivo .csv
Modelo de classificação
Conjunto de treino
Conjunto de teste
Modelo treinado
Teste
Treino
Algoritmos de Classificação
Arquivo .csv
Arquivo .csv
Famílias de ransomware

Tratamento dos dados



Normal	StandardScaler	Principal Components Analysis
Conjunto de dados original, sem qualquer tratamento.	Features ajustadas para ter média 0 e desvio padrão 1.	Foi utilizada com n=100 componentes principais.

Métricas

Precision

$$\frac{TP}{TP + FP}$$

Recall

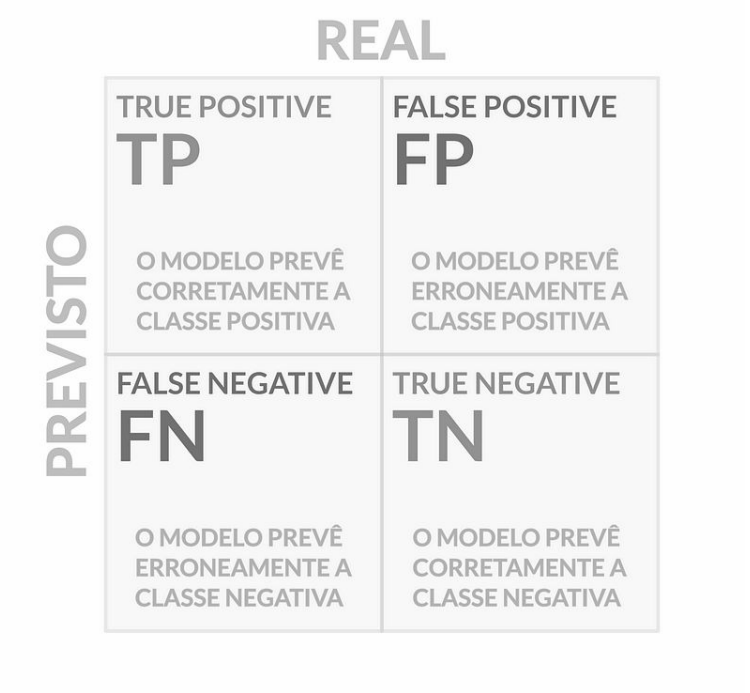
$$\frac{TP}{TP + FN}$$

F1-Score

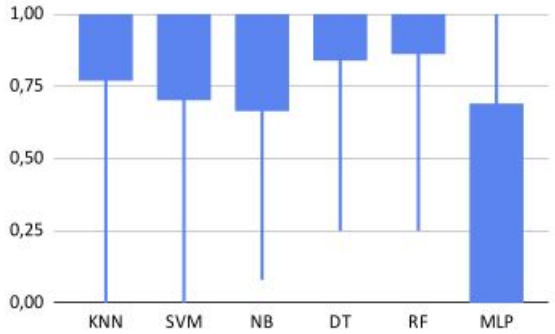
$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Acurácia

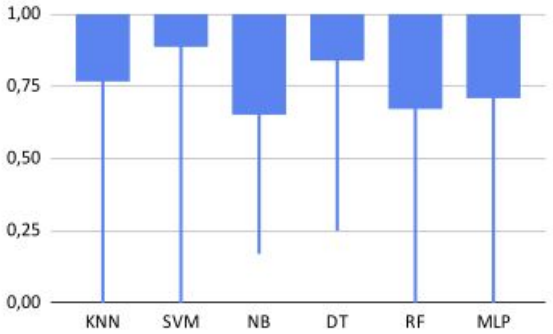
$$\frac{TP + TN}{TP + TN + FP + FN}$$



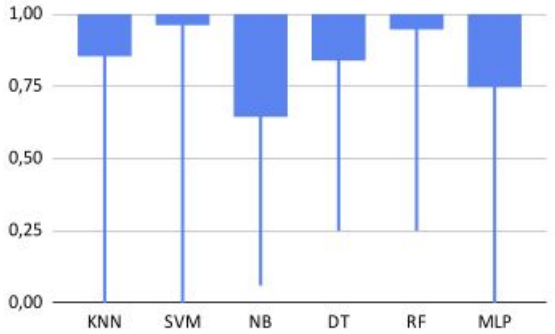
Avaliação - Precision (70:30)



Normal

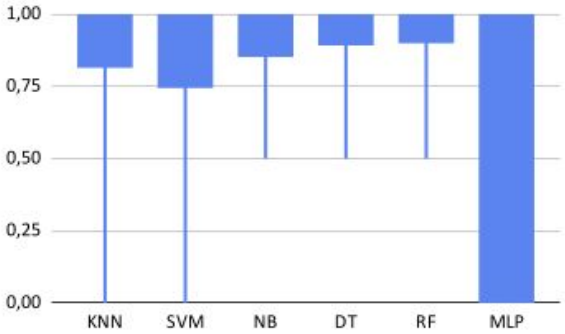


PCA

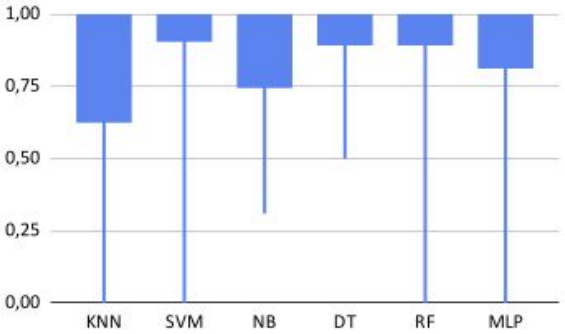


StandardScaler

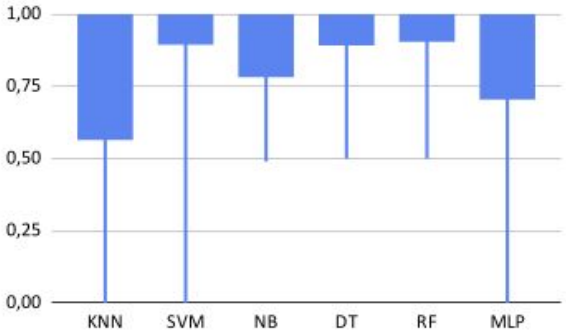
Avaliação - Recall (70:30)



Normal



PCA

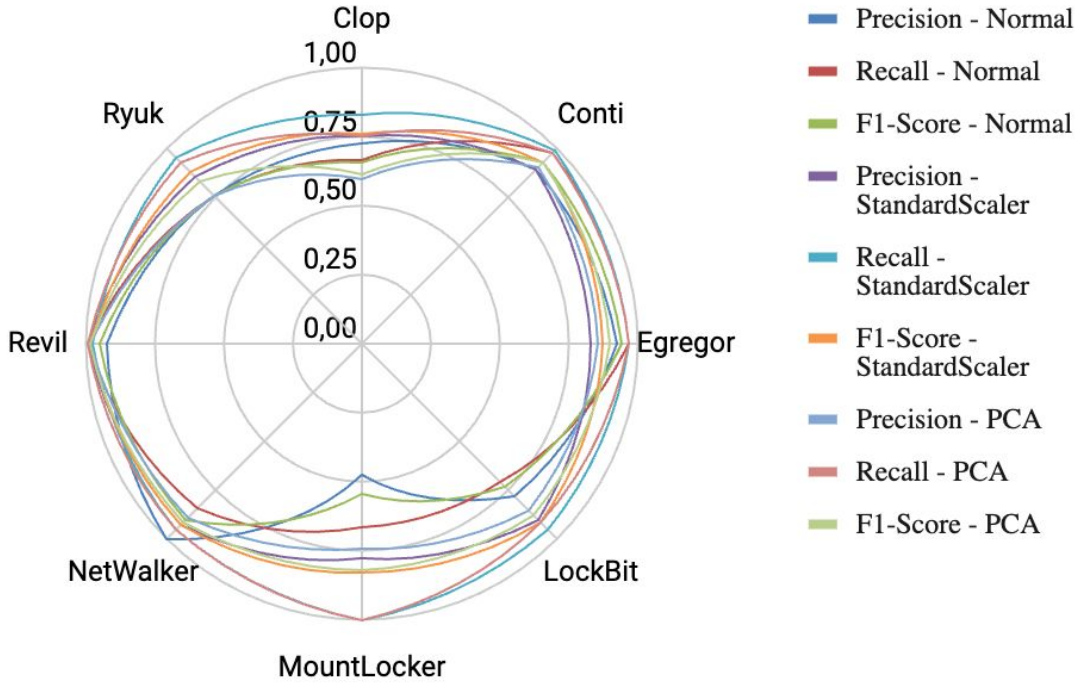


StandardScaler

Avaliação (70:30)

Algoritmo	Normal			StandardScaler			PCA			Normal			StandardScaler			PCA		
	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1
	Assinatura									Rede								
KNN	0,85	0,83	0,84	0,95	0,90	0,91	0,95	0,90	0,92	0,85	0,90	0,85	0,91	0,81	0,83	0,86	0,83	0,83
SVM	0,88	0,88	0,82	0,86	0,79	0,81	0,94	0,85	0,88	0,91	0,91	0,89	0,87	0,92	0,87	0,91	0,93	0,89
NB	0,76	0,93	0,80	0,69	0,91	0,72	0,77	0,81	0,74	0,67	0,85	0,69	0,66	0,85	0,68	0,70	0,70	0,64
DT	0,93	0,85	0,88	0,93	0,85	0,88	0,93	0,85	0,88	0,86	0,88	0,85	0,86	0,88	0,85	0,86	0,88	0,85
RF	0,94	0,92	0,91	0,94	0,93	0,91	0,87	0,93	0,89	0,87	0,90	0,84	0,88	0,92	0,88	0,68	0,83	0,73
MLP	0,43	0,49	0,49	0,74	0,80	0,77	0,92	0,93	0,91	0,39	0,39	0,32	0,89	0,90	0,88	0,86	0,88	0,86
	String									API								
KNN	0,94	0,85	0,88	0,78	0,91	0,81	0,84	0,92	0,85	0,69	0,75	0,71	0,73	0,43	0,50	0,69	0,48	0,55
SVM	0,74	0,69	0,71	0,98	0,97	0,97	0,94	0,97	0,95	0,65	0,76	0,68	0,87	0,71	0,77	0,84	0,77	0,79
NB	0,94	0,97	0,95	0,94	0,91	0,92	0,76	0,97	0,81	0,82	0,81	0,80	0,75	0,75	0,71	0,85	0,84	0,83
DT	0,83	0,97	0,88	0,83	0,97	0,88	0,83	0,97	0,88	0,88	0,85	0,85	0,88	0,85	0,85	0,88	0,85	0,85
RF	0,89	0,96	0,91	0,91	0,98	0,93	0,89	0,98	0,93	0,90	0,84	0,86	0,90	0,84	0,86	0,80	0,77	0,77
MLP	0,53	0,51	0,49	0,76	0,99	0,82	0,67	0,81	0,73	0,25	0,37	0,30	0,82	0,67	0,67	0,71	0,76	0,73

Avaliação (70:30)



String - Melhor desempenho médio com PCA e StandardScaler

Avaliação (50:50)

- A distribuição das métricas Precision e Recall foram semelhantes àquelas da divisão 70:30.
- O desempenho médio geral o RF se destaque em todas as métricas e tipos de dados
  - Destaque para os dados de Assinatura e String
- Os dados de String continuam sendo os mais eficazes
- As famílias NetWalker e Revil apresentaram resultados robustos e consistentemente altos

Considerações finais

- Os dados de String tratados com StandardScaler e analisados com RF e SVM formaram as combinações mais eficazes.
- O dataset possui mais amostras maliciosas do que benignas, potencialmente contribuindo para os resultados obtidos.
- No futuro, pretendemos usar arquiteturas de aprendizagem profunda para comparar com os indutores rasos usados.
- Também pretendemos realizar novos experimentos em ambientes Windows 10 e 11.

Obrigado!

Augusto Parisot
aparisot@id.uff.br

Lucila M. S. Bento
lucila.bento@ime.uerj.br

Raphael C. S. Machado
raphaelmachado@ic.uff.br