INTRUSION DETECTION SYSTEM USING OPTIMUM-PATH FOREST

MATHEUS ROCHA E ALEXANDRE SELANI

TRABALHO ORIGINAL

- SVM-RBF (Support Vector Machine com kernel Radial Basis Function)
- SOM (Self Organized Maps)
- Classificador Bayesiano

TRABALHO REALIZADO

- SVM-RBF (Support Vector Machine com kernel Radial Basis Function)
- KNN (K-Nearest Neighbors)

OPF (Optimum Path Forest)

MATERIAIS E MÉTODOS

Computador: Intel Core i5-10300H, 8GB RAM e Windows 11

- **Base de dados: KDD-CUP**
  - 41 features
  - 23 classes
  - 10% → 494021 amostras
  - Três variantes
  - One-Hot Encoder e remoção de duplicatas
- **Treinamento**
  - 20 folds com 10 iterações
  - 1 iteração → 1 par de folds com treino e teste
  - 50% treino e 50% teste
  - 5-fold cross validation
- **Implementação**
  - Python
  - Sickit-Learn
  - OPFython
- **Métricas**
  - tempo de treinamento e de teste
  - F1-score macro
  - Acurácia
  - Accuracy OPF-style

MATERIAIS E MÉTODOS

MATERIAIS E MÉTODOS

KDD-CUP 5

labels	
b'normal.'	87832
b'back.'	968
b'guess_passwd.'	53
b'buffer_overflow.'	30
b'imap.'	12
b'ftp_write.'	8

KDD-CUP 10

labels	
b'normal.'	87832
b'neptune.'	51820
b'back.'	968
b'ipsweep.'	651
b'guess_passwd.'	53
b'buffer_overflow.'	30
b'land.'	19
b'imap.'	12
b'loadmodule.'	9
b'ftp_write.'	8
b'multihop.'	7

KDD-CUP 15

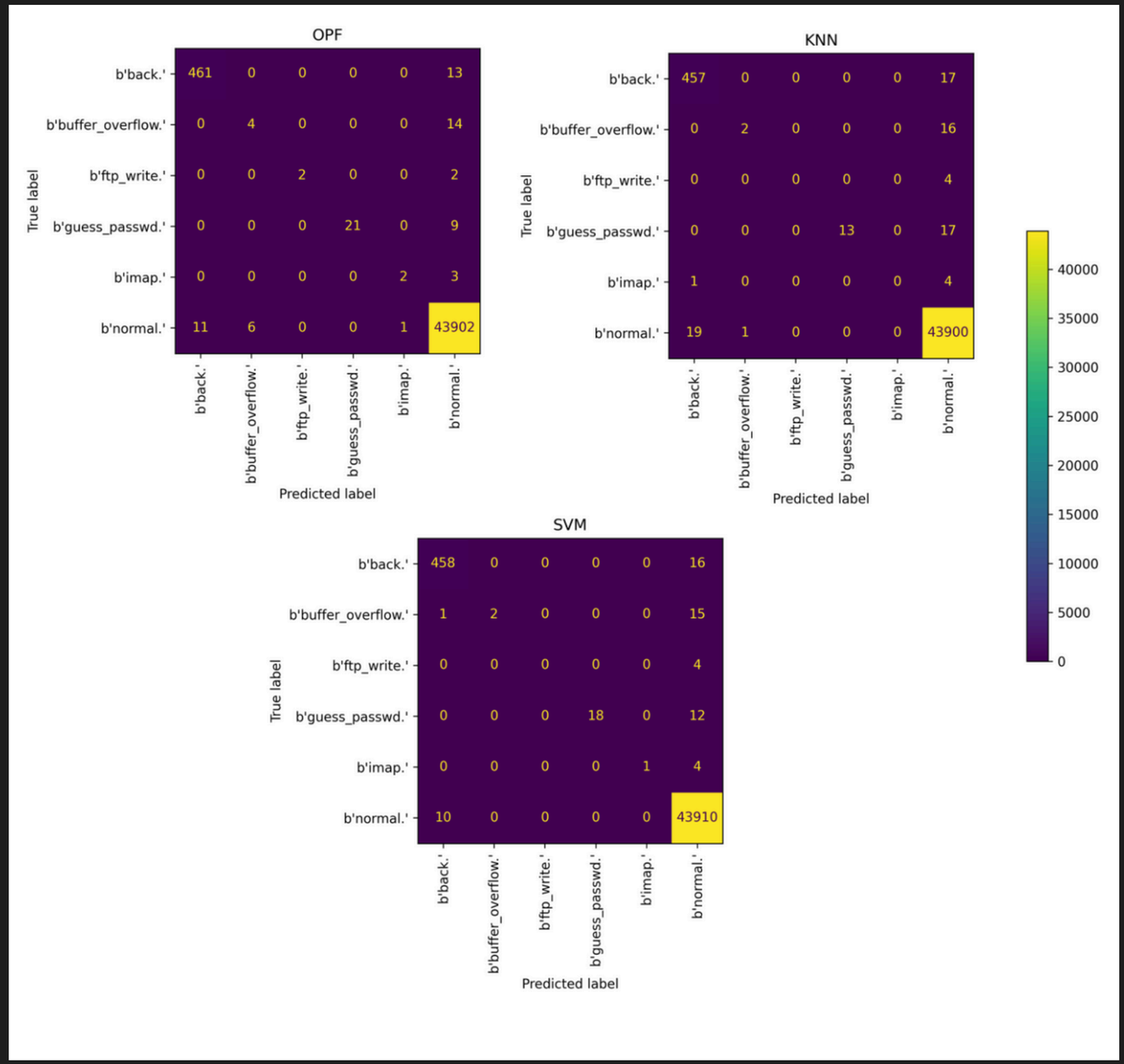
labels	
b'normal.'	87832
b'neptune.'	51820
b'back.'	968
b'ipsweep.'	651
b'portsweep.'	416
b'pod.'	206
b'nmap.'	158
b'guess_passwd.'	53
b'buffer_overflow.'	30
b'land.'	19
b'imap.'	12
b'loadmodule.'	9
b'ftp_write.'	8
b'multihop.'	7
b'phf.'	4
b'perl.'	3

RESULTADOS

Dataset	Modelos	Acc OPF-Style	Acurácia	F1-Score	Tempo Treino (s)	Tempo Teste (s)
KDD-CUP 5	OPF	0.888 ± 0.078	0.999 ± 0.000	0.699 ± 0.162	33.450 ± 5.898	21.102 ± 4.665
	SVM	0.850 ± 0.086	0.999 ± 0.001	0.640 ± 0.164	0.083 ± 0.069	0.067 ± 0.046
	KNN	0.829 ± 0.084	0.998 ± 0.001	0.583 ± 0.135	0.002 ± 0.000	0.055 ± 0.006
KDD-CUP 10	OPF	0.903 ± 0.026	0.998 ± 0.000	0.684 ± 0.073	92.525 ± 4.083	69.758 ± 6.502
	SVM	0.879 ± 0.036	0.998 ± 0.001	0.652 ± 0.074	0.262 ± 0.093	0.246 ± 0.051
	KNN	0.863 ± 0.030	0.998 ± 0.001	0.601 ± 0.090	0.004 ± 0.000	0.131 ± 0.021
KDD-CUP 15	OPF	0.903 ± 0.030	0.997 ± 0.001	0.697 ± 0.055	94.915 ± 3.932	73.122 ± 6.557
	SVM	0.879 ± 0.043	0.996 ± 0.002	0.667 ± 0.089	0.553 ± 0.295	0.354 ± 0.080
	KNN	0.875 ± 0.036	0.996 ± 0.001	0.633 ± 0.064	0.004 ± 0.000	0.163 ± 0.024

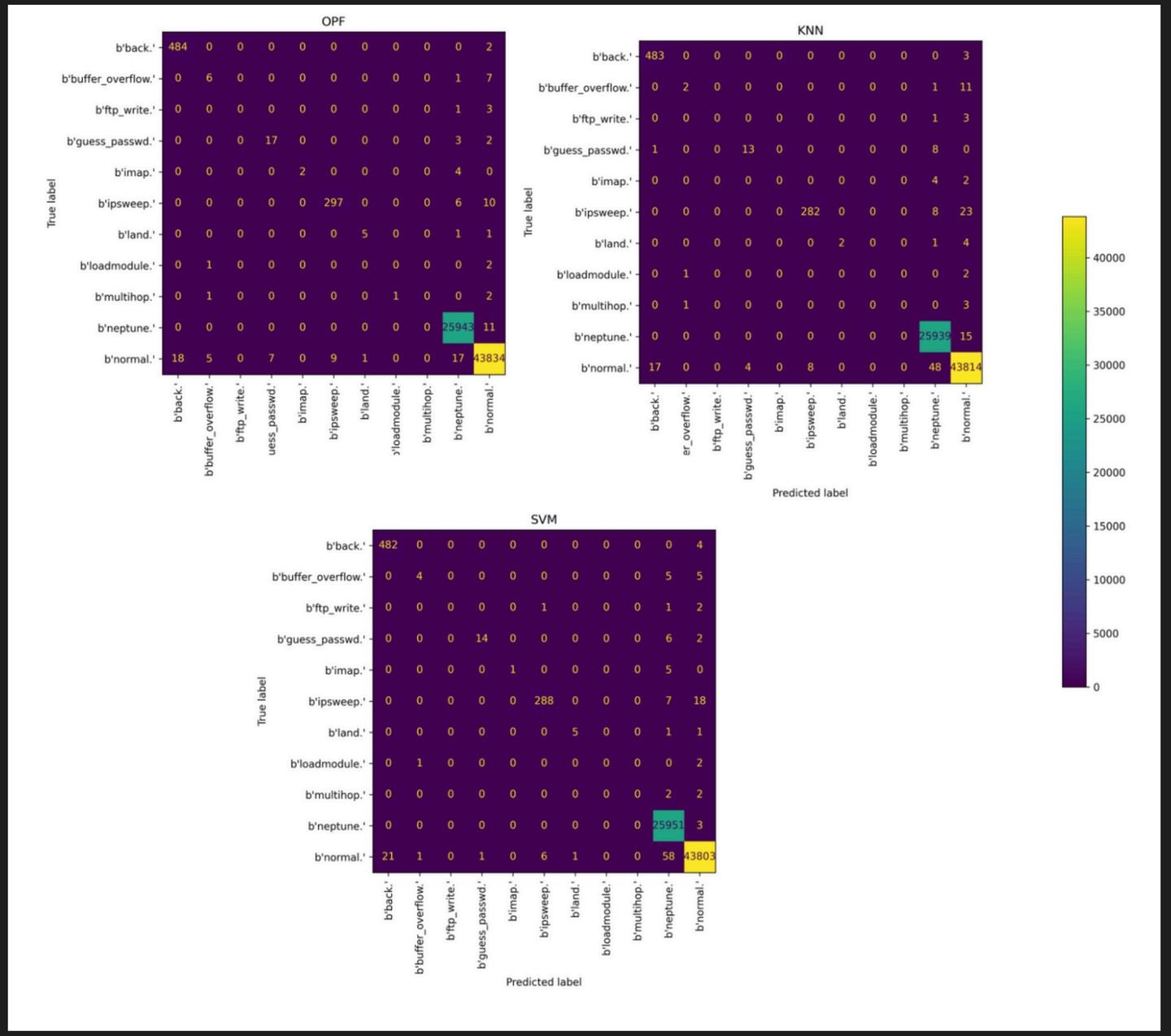
RESULTADOS

KDD-CUP 5



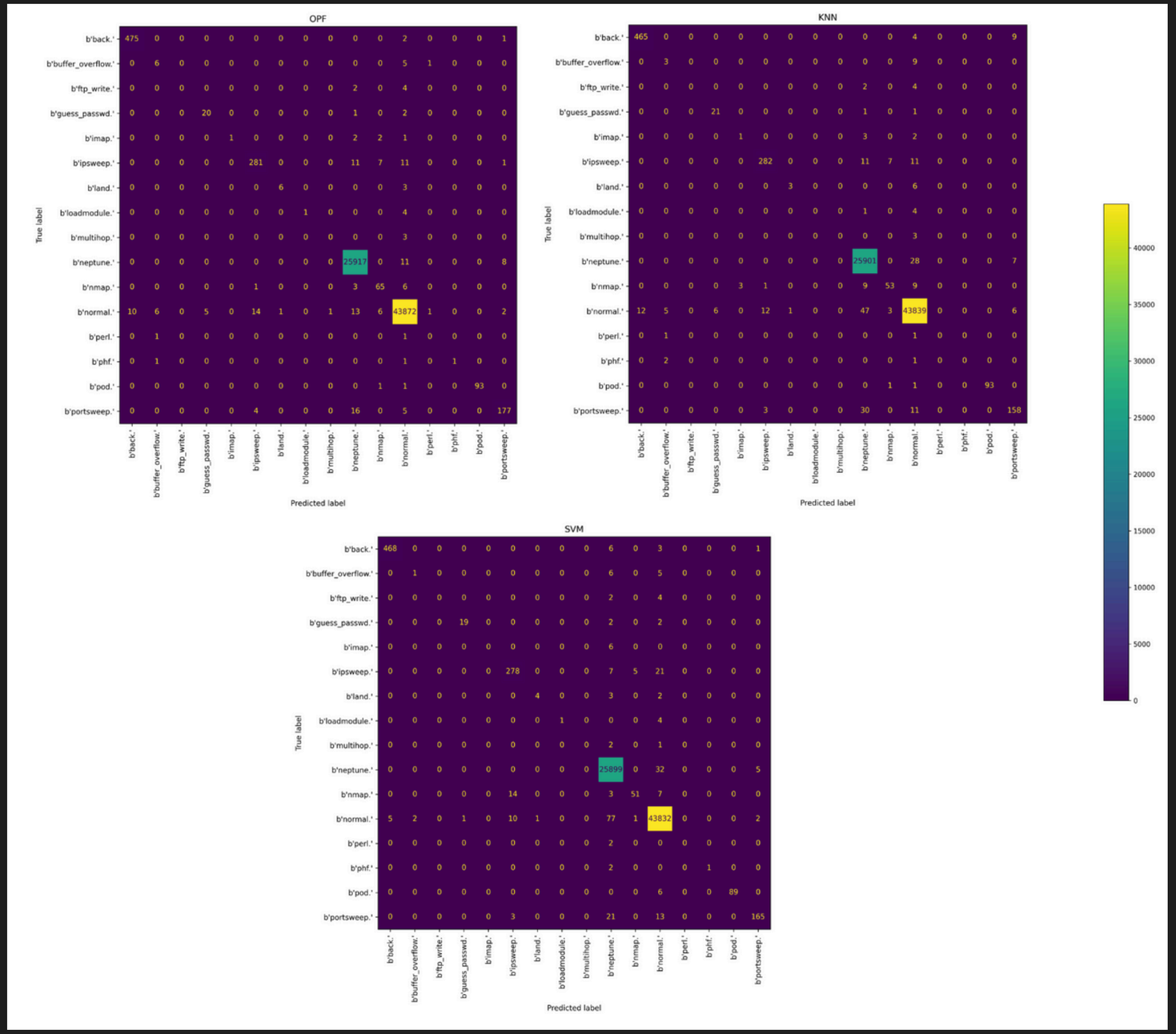
RESULTADOS

KDD-CUP 10



RESULTADOS

KDD-CUP 15



COMPARAÇÃO DE RESULTADOS

TRABALHO REALIZADO

Dataset	Modelos	Acc OPF-Style	Acurácia	F1-Score	Tempo Treino (s)	Tempo Teste (s)
KDD-CUP 5	OPF	0.888 ± 0.078	0.999 ± 0.000	0.699 ± 0.162	33.450 ± 5.898	21.102 ± 4.665
	SVM	0.850 ± 0.086	0.999 ± 0.001	0.640 ± 0.164	0.083 ± 0.069	0.067 ± 0.046
	KNN	0.829 ± 0.084	0.998 ± 0.001	0.583 ± 0.135	0.002 ± 0.000	0.055 ± 0.006
KDD-CUP 10	OPF	0.903 ± 0.026	0.998 ± 0.000	0.684 ± 0.073	92.525 ± 4.083	69.758 ± 6.502
	SVM	0.879 ± 0.036	0.998 ± 0.001	0.652 ± 0.074	0.262 ± 0.093	0.246 ± 0.051
	KNN	0.863 ± 0.030	0.998 ± 0.001	0.601 ± 0.090	0.004 ± 0.000	0.131 ± 0.021
KDD-CUP 15	OPF	0.903 ± 0.030	0.997 ± 0.001	0.697 ± 0.055	94.915 ± 3.932	73.122 ± 6.557
	SVM	0.879 ± 0.043	0.996 ± 0.002	0.667 ± 0.089	0.553 ± 0.295	0.354 ± 0.080
	KNN	0.875 ± 0.036	0.996 ± 0.001	0.633 ± 0.064	0.004 ± 0.000	0.163 ± 0.024

TRABALHO ORIGINAL

Classifier	Dataset	Acc	Training [s]	Testing [s]
OPF	**KddCup-5**	**80.54±1.66**	**5357.29**	**5083.1900**
Bayes	**KddCup-5**	**80.54±1.66**	**2092.87**	**21964.7300**
SVM-RBF	**KddCup-5**	**81.75±0.86**	**6412.76**	**5014.6600**
SOM	KddCup-5	70.34±4.55	8027.69	24578.00
OPF	**KddCup-10**	**89.31±0.39**	**5568.9798**	**5531.8046**
Bayes	**KddCup-10**	**89.31±0.39**	**2132.3876**	**27407.0000**
SVM-RBF	KddCup-10	85.26±0.12	7035.4401	5348.6321
SOM	KddCup-10	80.73±0.62	9045.0997	27196.865
OPF	**KddCup-15**	**91.49±1.74**	**5970.7832**	**5772.4331**
Bayes	**KddCup-15**	**90.49±1.72**	**2704.8383**	**36475.0000**
SVM-RBF	**KddCup-15**	**92.43±0.64**	**8884.3632**	**6895.0371**
SOM	KddCup-15	85.44±1.81	10513.5331	34614.928

CONCLUSÃO

- Existe viabilidade, mas há falhas ao classificar classes com poucas amostras
- Comparação: diferenças provavelmente relativas as condições do experimento, porém com similaridade em ambos os trabalhos.

REFERÊNCIAS

- de Rosa, Gustavo H., and João P. Papa. 2021. “OPFython: A Python implementation for Optimum-Path Forest.” Software Impacts, 100113. ISSN: 2665-9638. https://doi.org/https://doi.org/10.1016/j.simpa.2021.100113.
- Pereira, Clayton, Rodrigo Nakamura, João Paulo Papa, and Kelton Costa. 2011. “Intrusion detec-tion system using optimum-path forest.” In 2011 IEEE 36th Conference on Local Computer Networks, 183–186. IEEE.s

OBRIGADO