Appendices

A Combining continuous outputs

Instead of deciding the class label as output, some learning models may output the probability with which an instance belongs to a class. This probability can be thought of as the degree of support shown by the learning model towards a class.

$$\begin{aligned} s_{c,j}(x_i) &= \text{for the instance } x_i, \\ s_{c,j} &= \text{support received by the } j^{th} \text{ class from the } c^{th} \text{ classifier} \\ w_j &= \text{weight of the } j^{th} \text{ classifier} \\ \text{C} &= \text{total number of classifiers or models} \\ \mu_j(x_i) &= \text{total support for the } j^{th} \text{ class for instance } x_i \end{aligned}$$

Following are commonly used for combining the outputs of base learners.

1. *Sum rule*: According to this rule, the individual supports from all the learning models are added to obtain the final support for a particular class as shown in equation 1. The final output of the ensemble is the class with the highest support.

$$\mu_j(x_i) = \sum_{c=1}^{C} s_{c,j}(x_i) \tag{1}$$

2. *Mean rule*: According to this rule, after adding the individual supports from all the learning models, the total sum is normalized by the total number of learning models ($\frac{1}{C}$) as shown in equation 2.

$$\mu_j(x_i) = \frac{1}{C} \sum_{c=1}^{C} s_{c,j}(x_i) \tag{2}$$

3. *Weighted sum rule*: Each learning model is assigned a weight and the total support is the total sum of the product of the learning model's weights and

their supports as shown in equation 3.

$$\mu_j(x_i) = \sum_{c=1}^{C} w_t s_{c,j}(x_i) \tag{3}$$

4. *Product rule*: According to this rule, for a particular class the supports provided by the learning models are multiplied to obtain the final output.

$$\mu_j(x_i) = \prod_{c=1}^{C} s_{c,j}(x_i) \tag{4}$$

5. *Maximum rule*: According to this rule, for a particular class, the maximum support given by the participating learning models is selected as shown in equation 5.

$$\mu_j(x_i) = \max_{c=1}^{C} s_{c,j}(x_i) \tag{5}$$

6. *Minimum rule*: As the name suggests, for a particular class this rule selects the minimum support given by the participating learning models as shown in equation 6.

$$\mu_j(x_i) = \min_{c=1}^{C} s_{c,j}(x_i) \tag{6}$$

7. *Generalized mean rule*: The rules discussed above are special versions of the generalized mean rule given in equation 7.

$$\mu_{j,\infty}(x_i) = \left[\frac{1}{C} \sum_{c=1}^{C} s_{c,j}(x_i)^{\infty} \right]^{\frac{1}{\infty}} \tag{7}$$

B Hyper-parameter values

B.1 Bagging

Table 1 and Table 2 give the hyper-parameter values for Bagging ensemble method.

Table 1: Hyper-parameter Values for 2-class Security Datasets (Bagging)

		k-Nearest Neighbors	Support Vector Machine	Decision Trees	Logistic Regression
	Android Dataset 1	n_neighbors= 1	C=10, degree=1	criterion='entropy',max_depth=6, min_samples_leaf=2,min_samples_split=2	C=10,penalty="l2"
	Android Dataset 2	n_neighbors= 3	C=1, degree=1	criterion='gini',max_depth=3, min_samples_leaf=1,min_samples_split=2	C=1000,penalty="l2"
Security Datasets	SWaT	n_neighbors= 1	C=10, degree=1	criterion='entropy',max_depth=9, min_samples_leaf=1,min_samples_split=2	C=100,penalty="l2"
	Phishing	n_neighbors= 1	C=10, degree=1	criterion='entropy',max_depth=9, min_samples_leaf=1,min_samples_split=3	C=0.1,penalty="l2"
	Kitsune Network Attack	n_neighbors= 4	C=10, degree=1	criterion='gini',max_depth=8, min_samples_leaf=3,min_samples_split=2	C=0.001,penalty="l2"

Table 2: Hyper-parameter values for ransomware multiclass datasets (Bagging)

		Classifiers			
		k-Nearest Neighbors	Support Vector Machine	Decision Trees	Logistic Regression
Ransomware Multiclass	Citroni	n_neighbors= 1	C=1, degree=1	criterion='gini', max_depth=3, min_samples_leaf=1, min_samples_split=2	C=1.0, penalty="l2"
	CryptLocker	n_neighbors= 7	C=10, degree=1	criterion='gini', max_depth=7, min_samples_leaf=1, min_samples_split=3	C=10.0, penalty="l2"
	CryptoWall	n_neighbors= 3	C=10, degree=1	criterion='entropy', max_depth=6, min_samples_leaf=1, min_samples_split=2	C=10.0, penalty="l2"
	Kollah	n_neighbors= 2	C=10, degree=1	criterion='gini',max_depth=8, min_samples_leaf=1, min_samples_split=2	C=10.0, penalty="l2"
	Kovter	n_neighbors= 2	C=10, degree=1	criterion='gini',max_depth=5, min_samples_leaf=1, min_samples_split=4	C=1.0, penalty="l2"
	Locker	n_neighbors= 2	C=10, degree=1	criterion='gini',max_depth=5, min_samples_leaf=1, min_samples_split=4	C=1.0, penalty="l2"
	Matsnu	n_neighbors= 4	C=1, degree=1	criterion='entropy',max_depth=7, min_samples_leaf=1, min_samples_split=6	C=1.0, penalty="l2"
	Pgpcoder	n_neighbors= 1	C=0.1, degree=1	criterion='gini',max_depth=3, min_samples_leaf=1, min_samples_split=2	C=0.1, penalty="l2"
	Reveton	n_neighbors= 1	C=10, degree=1	criterion='gini',max_depth=8, min_samples_leaf=1, min_samples_split=7	C=1000.0, penalty="l2"
	TeslaCrypt	n_neighbors= 3	C=1, degree=1	criterion='gini',max_depth=7, min_samples_leaf=3, min_samples_split=2	C=1.0, penalty="l2"
	Trojan-Ransom	n_neighbors= 3	C=10, degree=1	criterion='entropy',max_depth=7, min_samples_leaf=3, min_samples_split=9	C=10.0, penalty="l2"

B.2 Boosting

Table 3, 4, 5, and 6 give the hyper-parameter values for the Boosting ensemble method.

Table 3: Hyper-parameter values for 2-class security datasets (Adaboost, GB and XGB)

		Classifiers		
	Dataset name	AdaBoost	Gradient Boosting	Extreme Gradient Boosting
Security Datasets	Android Dataset 1	learning_rate=0.1, n_estimators=500, algorithm='SAMME.R'	loss= 'exponential', learning_rate=0.1, n_estimators=500,max_features='log2', max_depth=5 ,criterion='mse'	colsample_bytree= 0.5, learning_rate=0.01, max_depth= 10, min_child_weight= 1, n_estimators= 500, subsample= 0.75
	Android Dataset 2	learning_rate=0.1, n_estimators=350, algorithm='SAMME.R'	loss= 'exponential', learning_rate=0.1, n_estimators=300,max_features='log2', max_depth=5 ,criterion='mse'	colsample_bytree= 1, learning_rate=0.1, max_depth= 10, min_child_weight= 1, n_estimators= 100, subsample= 0.5
	SWaT	learning_rate=1.0, n_estimators=500, algorithm='SAMME.R'	criterion='friedman_mse', learning_rate= 1.0, loss='exponential',max_depth= 8, max_features= log2, n_estimators= 500	colsample_bytree= 1, learning_rate=0.01, max_depth= 6, min_child_weight= 1, n_estimators= 300, subsample= 0.5
	Phishing	learning_rate=1.0, n_estimators=200, algorithm='SAMME'	loss= 'exponential', learning_rate=0.1, n_estimators=500, max_features='log2', max_depth=5 ,criterion='friedman_mse'	colsample_bytree= 0.5, learning_rate=0.1, max_depth= 10, min_child_weight= 1, n_estimators= 200, subsample= 1
	Kitsune Network Attack	learning_rate=1, n_estimators=500, algorithm='SAMME.R'	loss= 'exponential', learning_rate=0.1, n_estimators=500,max_features='log2', max_depth=8,criterion='mse'	colsample_bytree= 0.75, learning_rate=0.1, max_depth= 10, min_child_weight= 1, n_estimators= 500, subsample= 1

Table 4: Hyper-parameter values for 2-class security datasets (LGB and HGB)

		Classifiers	
	Dataset name	Light Gradient Boosting	Hist Gradient Boosting
Security Datasets	Android Dataset 1	boosting_type= 'dart', num_leaves= 10, learning_rate= 0.5, min_child_weight= 1, min_child_samples=100, colsample_bytree= 0.66,reg_alpha= 0.5, reg_lambda= 1, subsample= 0.5	learning_rate=0.1, max_iter=150, max_leaf_nodes=30, min_samples_leaf=20
	Android Dataset 2	boosting_type= 'gbdt',colsample_bytree= 1,learning_rate= 0.5, min_child_samples= 50,min_child_weight=1,num_leaves=10, reg_alpha=0.5,reg_lambda= 1.4,subsample= 0.5	learning_rate=0.1, max_iter=100, max_leaf_nodes=40, min_samples_leaf=5
	SWaT	boosting_type= 'dart', colsample_bytree= 0.66, learning_rate=1, min_child_samples= 50, min_child_weight=1, num_leaves=20, reg_alpha=0.5, reg_lambda= 1.2, subsample= 0.5	learning_rate=0.01, max_iter=100, max_leaf_nodes=30, min_samples_leaf=3
	Phishing	boosting_type= 'dart', num_leaves= 20, learning_rate= 1, min_child_weight= 1, min_child_samples=20, colsample_bytree= 1, reg_alpha= 0.5, reg_lambda= 1.2, subsample= 0.5	learning_rate=0.1, max_iter=200, max_leaf_nodes=40, min_samples_leaf=10
	Kitsune Network Attack	boosting_type= 'dart', num_leaves= 20, learning_rate= 0.5, min_child_weight= 1e-05, min_child_samples=20, colsample_bytree= 1, reg_alpha= 0.5, reg_lambda= 1.2, subsample= 0.5	learning_rate=0.1, max_iter=200, max_leaf_nodes=40, min_samples_leaf=10

Table 5: Hyper-parameter values for Ransomware multiclass dataset (Adaboost, GB and XGB)

	Class	Adaboost	Gradient Boosting	Extreme Gradient Boosting
Ransomware Multiclass	Citroni	learning_rate=1.0, n_estimators=100, algorithm='SAMME'	loss= 'exponential', learning_rate=1.0, n_estimators=100, max_depth=8, criterion='friedman_mse'	colsample_bytree= 1, learning_rate=0.1, max_depth= 2, min_child_weight= 1, n_estimators= 100, subsample= 1
	CryptLocker	learning_rate=1.0, n_estimators=200, algorithm='SAMME.R'	loss= 'exponential', learning_rate=1.0, n_estimators=500, max_depth=5 , criterion='friedman_mse'	colsample_bytree= 0.5, learning_rate=0.01, max_depth= 2, min_child_weight= 1, n_estimators= 100, subsample= 0.5
	CryptoWall	learning_rate=0.1, n_estimators=200, algorithm='SAMME.R'	loss= 'deviance', learning_rate=0.1, n_estimators=100, max_depth=3, criterion='friedman_mse'	colsample_bytree= 0.5, learning_rate=0.01, max_depth= 2, min_child_weight= 1, n_estimators= 100, subsample= 0.5
	Kollah	learning_rate=0.1, n_estimators=50, algorithm='SAMME.R'	loss= 'exponential', learning_rate=0.1, n_estimators=200, max_depth=3, criterion='friedman_mse'	colsample_bytree= 0.5, learning_rate=0.3, max_depth= 6, min_child_weight= 1, n_estimators= 100, subsample= 1
	Kovter	learning_rate=1.0, n_estimators=500, algorithm='SAMME'	loss= 'exponential', learning_rate=0.01, n_estimators=500,max_depth=3, criterion='friedman_mse'	colsample_bytree= 0.75, learning_rate=0.1, max_depth= 2, min_child_weight= 1, n_estimators= 100, subsample= 1
	Locker	learning_rate=1.0, n_estimators=500, algorithm='SAMME'	loss= 'exponential', learning_rate=0.01, n_estimators=500, max_depth=3, criterion='friedman_mse'	colsample_bytree= 0.75, learning_rate=0.1, max_depth= 2, min_child_weight= 1, n_estimators= 100, subsample= 1
	Matsnu	learning_rate=1.0, n_estimators=200, algorithm='SAMME'	loss= 'exponential', learning_rate=1.0, n_estimators=100, max_depth=8, criterion='friedman_mse'	colsample_bytree= 0.75, learning_rate=0.1, max_depth= 10, min_child_weight= 1, n_estimators= 300, subsample= 1
	Pgpcoder	learning_rate=1.0, n_estimators=50, algorithm='SAMME'	loss= 'deviance', learning_rate=0.01, n_estimators=200, max_depth=3, criterion='friedman_mse'	colsample_bytree= 0.5, learning_rate=0.01, max_depth= 2, min_child_weight= 1, n_estimators= 300, subsample= 0.5
	Reveton	learning_rate=0.1, n_estimators=500, algorithm='SAMME.R'	loss= 'exponential', learning_rate=0.1, n_estimators=500, max_depth=8, criterion='friedman_mse'	colsample_bytree= 1, learning_rate=0.3, max_depth= 6, min_child_weight= 1, n_estimators= 100, subsample= 1
	TeslaCrypt	learning_rate=0.1, n_estimators=100, algorithm='SAMME.R'	loss= 'deviance', learning_rate=0.1, n_estimators=200, max_depth=8, criterion='friedman_mse'	colsample_bytree= 0.5, learning_rate=0.01, max_depth= 10, min_child_weight= 1, n_estimators= 200, subsample= 1
	Trojan-Ransom	learning_rate=0.1, n_estimators=350, algorithm='SAMME.R'	loss= 'exponential', learning_rate=1.0, n_estimators=300, max_depth=3, criterion='mse'	colsample_bytree= 1, learning_rate=0.01, max_depth= 10, min_child_weight= 1, n_estimators= 500, subsample= 0.75

Table 6: Hyper-parameter values for Ransomware multiclass dataset (LGB and HGB)

		Classifiers	
	Class	Light Gradient Boosting	HistGradient Boosting
Ransomware Multiclass	Citroni	boosting_type= 'gbdt', colsample_bytree= 1, learning_rate= 0.1, min_child_samples= 20, min_child_weight=1e-05, num_leaves=5, reg_alpha=0.5, reg_lambda= 1.4, subsample= 0.5	learning_rate=0.1, max_iter=150, max_leaf_nodes=10, min_samples_leaf=10
	CryptLocker	boosting_type= 'dart',colsample_bytree= 0.5, learning_rate= 1, min_child_samples= 20, min_child_weight=1, num_leaves=10, reg_alpha=1.2, reg_lambda= 1.4	learning_rate=1,max_iter=100, max_leaf_nodes=10,min_samples_leaf=3
	CryptoWall	boosting_type= 'gbdt', colsample_bytree= 1, learning_rate= 1.0, min_child_samples= 20, min_child_weight=1e-05, num_leaves=5, reg_alpha=0.5 , reg_lambda= 1.2, subsample= 0.5	learning_rate=1, max_iter=100, max_leaf_nodes=30, min_samples_leaf=5
	Kollah	boosting_type= 'dart', colsample_bytree= 0.5, learning_rate= 1, min_child_samples= 20, min_child_weight=1e-05, num_leaves=10, reg_alpha=0.5, reg_lambda= 1, subsample= 0.5	learning_rate=1, max_iter=100, max_leaf_nodes=10, min_samples_leaf=3
	Kovter	boosting_type= 'dart',colsample_bytree= 0.5, learning_rate= 0.1,min_child_samples= 20, min_child_weight=1e-05,num_leaves=5, reg_alpha=0.5,reg_lambda= 1,subsample= 0.5	learning_rate=0.01, max_iter=200, max_leaf_nodes=10, min_samples_leaf=3,
	Locker	boosting_type= 'dart',colsample_bytree= 0.5, learning_rate= 0.1,min_child_samples= 20, min_child_weight=1e-05,num_leaves=5, reg_alpha=0.5,reg_lambda= 1,subsample= 0.5	learning_rate=0.01, max_iter=200, max_leaf_nodes=10, min_samples_leaf=3
	Matsnu	boosting_type= 'gbdt',colsample_bytree= 1, learning_rate= 1.0,min_child_samples= 50, min_child_weight=1,num_leaves=5, reg_alpha=0.5,reg_lambda= 1.2,subsample= 0.5	learning_rate=0.1, max_iter=100, max_leaf_nodes=10, min_samples_leaf=3
	Pgpcoder	boosting_type= 'gbdt', colsample_bytree= 0.5, learning_rate= 1.0,min_child_samples= 20, min_child_weight=1e-05,num_leaves=5, reg_alpha=0.5,reg_lambda= 1.2,subsample= 0.5	learning_rate=0.01, max_iter=100, max_leaf_nodes=10, min_samples_leaf=3
	Reveton	boosting_type= 'gbdt',colsample_bytree= 0.5, learning_rate= 0.1,min_child_samples= 20, min_child_weight=1e-05,num_leaves=10, reg_alpha=0.5,reg_lambda= 1,subsample= 0.5	learning_rate=0.1, max_iter=100, max_leaf_nodes=30, min_samples_leaf=10
	TeslaCrypt	boosting_type= 'gbdt',colsample_bytree= 0.66, learning_rate= 1.0,min_child_samples= 20, min_child_weight=1e-05,num_leaves=10, reg_alpha=1,reg_lambda=1.4,subsample= 0.5	learning_rate=0.1, max_iter=100, max_leaf_nodes=10, min_samples_leaf=5
	Trojan-Ransom	boosting_type= 'gbdt',colsample_bytree= 0.5, learning_rate= 0.5,min_child_samples= 50, min_child_weight=1e-05,num_leaves=5, reg_alpha=1.2,reg_lambda= 1,subsample= 0.5	learning_rate=0.1, max_iter=100, max_leaf_nodes=30, min_samples_leaf=20

Bibliography

[1] Brij Gupta, Dharma P Agrawal, and Shingo Yamaguchi. *Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security.* IGI Global, 2016. ISBN 9781785399824.

[2] K Naveen Durai, R Subha, and Anandakumar Haldorai. A Novel Method to Detect and Prevent SQLIA using Ontology to Cloud Web Security. *Wireless Personal Communications*, 117(4):2995–3014, 2021.

[3] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. Fake News Detection using Machine Learning Ensemble Methods. *Complexity*, 2020:1–11, 2020.

[4] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space.* Springer Nature, 2022.

[5] Simson Garfinkel and Gene Spafford. *Web Security, privacy & Commerce.* " O'Reilly Media, Inc.", 2002.

[6] Petar Lachkov, Lo'ai Tawalbeh, and Smriti Bhatt. Vulnerability Assessment for Applications Security Through Penetration Simulation and Testing. *Journal of Web Engineering*, 21(7):2187–2208, 2022.

[7] Donald Ray and Jay Ligatti. Defining Code-Injection Attacks. *Acm Sigplan Notices*, 47(1):179–190, 2012.

[8] Abhishek Singh, Baibhav Singh, and Hirosh Joseph. Vulnerability Analysis for HTTP. In *Vulnerability Analysis and Defense for the Internet*, pages 79–110. Springer, 2008.

[9] Jai Puneet Singh. Analysis of SQL Injection Detection Techniques. *Theoretical and Applied Informatics (TAAI)*, 28(1-2):37–55, 2016.

[10] V Nithya, S Lakshmana Pandian, and C Malarvizhi. A Survey on Detection and Prevention of Cross-site Scripting Attack. *International Journal of Security and Its Applications*, 9(3):139–152, 2015.

[11] M. Backes and P. Ning. *Computer Security – ESORICS 14th European Symposium on Research in Computer Security*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009. ISBN 9783642044441.

[12] B. Hoffman and B. Sullivan. *Ajax Security*. Pearson Education, 2007. ISBN 9780132701921.

[13] V Benjamin Livshits and Weidong Cui. Spectator: Detection and Containment of JavaScript Worms. In *USENIX Annual Technical Conference*, pages 335–348, 2008.

[14] Marthony Taguinod, Adam Doupé, Ziming Zhao, and Gail-Joon Ahn. Toward a Moving Target Defense for Web Applications. In *2015 IEEE international conference on information reuse and integration*, pages 510–517. IEEE, 2015.

[15] Joe G Greener, Shaun M Kandathil, Lewis Moffat, and David T Jones. A Guide to Machine Learning for Biologists. *Nature Reviews Molecular Cell Biology*, 23(1):40–55, 2022.

[16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Machine Learning Basics. *Deep learning*, 1(7):98–164, 2016.

[17] Nazrul Hoque, Mihir Singh, and Dhruba K Bhattacharyya. EFS-MI: An Ensemble Feature Selection Method for Classification. *Complex & Intelligent Systems*, 4(2):105–118, 2018.

[18] Andrew S Tanenbaum. *Computer Networks*. Pearson Education India, 2003.

[19] Larry L Peterson and Bruce S Davie. *Computer Networks: A Systems Approach*. Elsevier, 2007.

[20] Joseph Migga Kizza, Wheeler Kizza, and Wheeler. *Guide to Computer Network Security*, volume 8. Springer, 2013.

[21] Joseph Migga Kizza. *Computer Network Security*. Springer Science & Business Media, 2005.

[22] Linda S Rutledge and Lance J Hoffman. A Survey of Issues in Computer Network Security. *Computers & Security*, 5(4):296–308, 1986.

[23] Jie Wang. *Computer Network Security*. Springer, 2009.

[24] Bert-Jaap Koops. The Internet and its Opportunities for Cybercrime. *Transnational criminology manual, M. Herzog-Evans, ed*, 1:735–754, 2010.

[25] Nazrul Hoque, Monowar H Bhuyan, Ram Charan Baishya, Dhruba K Bhattacharyya, and Jugal K Kalita. Network Attacks: Taxonomy, Tools and Systems. *Journal of Network and Computer Applications*, 40:307–324, 2014.

[26] Owasp. Xss Filter Evasion Cheat Sheet. `https://www.owasp.org/index.php/XSS_Filter_Evasion_Cheat_Sheet`, 2018. [Accessed: 10-8-2018].

[27] Techopedia. Raw Data. `https://www.techopedia.com/definition/1230/raw-data`.

[28] Alexa Internet. Alexa Top 500 Websites. `http://www.alexa.com/topsites`.

[29] DP and KF. XSSed Repository. `http://www.xssed.com/`.

[30] Brian Chess and Gary McGraw. Static Analysis for Security. *IEEE Security & Privacy*, 2(6):76–79, 2004.

[31] Daniel Bates, Adam Barth, and Collin Jackson. Regular Expressions Considered Harmful in Client-Side XSS Filters. In *Proceedings of the 19th International Conference on World Wide Web*, pages 91–100. ACM, 2010.

[32] Riccardo Pelizzi and R Sekar. Protection, Usability and Improvements in Reflected XSS Filters. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, page 5, 2012.

[33] G Maone. NoScript. `https://noscript.net`. [Accessed: 12-11-2016].

[34] DavidRoss. IEBlog, IE8 Security Part IV: The XSS Filter. `https://blogs.msdn.microsoft.com/ie/2008/07/02/ie8-security-part-iv-the-xss-filter/`, 2008. [Accessed: 14-11-2016].

[35] Shashank Gupta and Brij Bhooshan Gupta. XSS-immune: A Google Chrome Extension-based XSS Defensive Framework for Contemporary Platforms of Web Applications. *Security and Communication Networks*, 9(17):3966–3986, 2016.

[36] Kanpata Sudhakara Rao, Naman Jain, Nikhil Limaje, Abhilash Gupta, Mridul Jain, and Bernard Menezes. Two for the price of one: A Combined Browser Defense Against XSS and Clickjacking. In *2016 International Conference on Computing Networking and Communications (ICNC)*, pages 1–6. IEEE, 2016.

[37] Chih-Hung Wang and Yi-Shauin Zhou. A New Cross-Site Scripting Detection Mechanism Integrated with HTML5 and CORS Properties by Using Browser Extensions. In *2016 International Computer Symposium (ICS)*, pages 264–269. IEEE, 2016.

[38] V Benjamin Livshits and Monica S Lam. Finding Security Vulnerabilities in Java Applications with Static Analysis. In *USENIX Security Symposium*, volume 14, pages 18–18, 2005.

[39] Omer Tripp, Marco Pistoia, Stephen J Fink, Manu Sridharan, and Omri Weisman. TAJ: Effective Taint Analysis of Web Applications. In *ACM Special Interest Group on Programming Languages (SIGPLAN) Notices*, volume 44, pages 87–97. ACM, 2009.

[40] Nenad Jovanovic, Christopher Kruegel, and Engin Kirda. Pixy: A Static Analysis Tool for Detecting Web Application Vulnerabilities. In *2006 IEEE Symposium on Security and Privacy*, page 6. IEEE, 2006.

[41] Philipp Vogt, Florian Nentwich, Nenad Jovanovic, Engin Kirda, Christopher Kruegel, and Giovanni Vigna. Cross-Site Scripting Prevention with Dynamic Data Tainting and Static Analysis. In *14th Annual Network and Distributed System Security Symposium, NDSS*, volume 2007, page 12, 2007.

[42] Yichen Xie and Alex Aiken. Static Detection of Security Vulnerabilities in Scripting Languages. In *USENIX Security Symposium*, volume 15, pages 179–192, 2006.

[43] Michael D Ernst. Static and Dynamic Analysis: Synergy and Duality. In *In proceedings of the ICSE Workshop on Dynamic Analysis*, pages 24–27, 2003.

[44] Philip Jia Guo. *A Scalable Mixed-level Approach to Dynamic Analysis of C and C++ Programs.* PhD thesis, Massachusetts Institute of Technology, 2006.

[45] Hossain Shahriar and Mohammad Zulkernine. Mitigating Program Security Vulnerabilities: Approaches and Challenges. *ACM Computing Surveys (CSUR)*, 44(3):11, 2012.

[46] Davide Balzarotti, Marco Cova, Vika Felmetsger, Nenad Jovanovic, Engin Kirda, Christopher Kruegel, and Giovanni Vigna. Saner: Composing Static and Dynamic Analysis to Validate Sanitization in Web Applications. In *IEEE Symposium on Security and Privacy*, pages 387–401. IEEE, 2008.

[47] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly Detection: A Survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

[48] Nana K Ampah, Cajetan M Akujuobi, Mathew NO Sadiku, and Shumon Alam. An Intrusion Detection Technique based on Continuous Binary Communication Channels. *International Journal of Security and Networks*, 6(2-3): 174–180, 2011.

[49] Dhruba Kumar Bhattacharyya and Jugal Kumar Kalita. *Network Anomaly Detection: A Machine Learning Perspective*. CRC Press, 2013.

[50] Prasanta Gogoi, Bhogeswar Borah, and Dhruba K Bhattacharyya. Anomaly Detection Analysis of Intrusion Data using Supervised & Unsupervised Approach. *Journal of Convergence Information Technology*, 5(1):95–110, 2010.

[51] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. Network Anomaly Detection: Methods, Systems and Tools. *IEEE Communications Surveys & Tutorials*, 16(1):303–336, 2014.

[52] Charles Reis, John Dunagan, Helen J Wang, Opher Dubrovsky, and Saher Esmeir. BrowserShield: Vulnerability-Driven Filtering of Dynamic HTML. *ACM Transactions on the Web (TWEB)*, 1(3):11, 2007.

[53] Charlie Curtsinger, Benjamin Livshits, Benjamin G Zorn, and Christian Seifert. ZOZZLE: Fast and Precise In-Browser JavaScript Malware Detection. In *USENIX Security Symposium*, pages 33–48, 2011.

[54] Shashank Gupta and BB Gupta. Xss-Secure as a Service for the Platforms of Online Social Network-based Multimedia Web Applications in Cloud. *Multimedia Tools and Applications*, 77(4):4829–4861, 2018.

[55] Xiaobing Guo, Shuyuan Jin, and Yaxing Zhang. XSS Vulnerability Detection using Optimized Attack Vector Repertory. In *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pages 29–36. IEEE, 2015.

[56] Wafa Ben Jaballah and Nizar Kheir. A Grey-Box Approach for Detecting Malicious User Interactions in Web Applications. In *Proceedings of the 2016 International Workshop on Managing Insider Security Threats*, pages 1–12. ACM, 2016.

[57] Gary Wassermann and Zhendong Su. Static Detection of Cross-site Scripting Vulnerabilities. In *Proceedings of the 30th International Conference on Software Engineering*, pages 171–180. ACM, 2008.

[58] Swaswati Goswami, Nazrul Hoque, Dhruba K Bhattacharyya, and Jugal Kalita. An Unsupervised Method for Detection of XSS Attack. *International Journal of Network Security*, 19(5):761–775, 2017.

[59] Peter Likarish, Eunjin Jung, and Insoon Jo. Obfuscated Malicious JavaScript Detection using Classification Techniques. In *4th International Conference on Malicious and Unwanted Software (MALWARE)*, pages 47–54. IEEE, 2009.

[60] Angelo Eduardo Nunan, Eduardo Souto, Eulanda M dos Santos, and Eduardo Feitosa. Automatic Classification of Cross-Site Scripting in Web Pages using Document-based and URL-based Features. In *2012 IEEE Symposium on Computers and Communications (ISCC)*, pages 000702–000707. IEEE, 2012.

[61] Christopher Kruegel and Giovanni Vigna. Anomaly Detection of Web-based Attacks. In *Proceedings of the 10th ACM conference on Computer and Communications Security*, pages 251–261. ACM, 2003.

[62] William Robertson, Giovanni Vigna, Christopher Kruegel, Richard A Kemmerer, et al. Using Generalization and Characterization Techniques in the Anomaly-based Detection of Web Attacks. In *13th Annual Symposium on Network and Distributed System Security*, 2006.

[63] Yingbo Song, Angelos D Keromytis, and Salvatore J Stolfo. Spectrogram: A Mixture-of-Markov-Chains Model for Anomaly Detection in Web Traffic. In *16th Annual Network and Distributed System Security Symposium, NDSS*, volume 9, pages 1–15, 2009.

[64] Rasim Alguliyev, Yadigar Imamverdiyev, and Lyudmila Sukhostat. Cyber-Physical Systems and their Security Issues. *Computers in Industry*, 100: 212–223, 2018.

[65] Abdulmalik Humayed, Jingqiang Lin, Fengjun Li, and Bo Luo. Cyber-physical Systems Security—a Survey. *IEEE Internet of Things Journal*, 4 (6):1802–1831, 2017.

[66] Claudia Aradau. Security that matters: Critical infrastructure and objects of protection. *Security dialogue*, 41(5):491–514, 2010.

[67] Tianbo Lu, Jiaxi Lin, Lingling Zhao, Yang Li, and Yong Peng. A Security Architecture in Cyber-Physical Systems: Security Theories, Analysis, Simulation and Application Fields. *International journal of security and its applications*, 9(7):1–16, 2015.

[68] Cyber Physical Systems Security: Analysis, Challenges and Solutions, author=Ashibani, Yosef and Mahmoud, Qusay H, journal=Computers & Security, volume=68, pages=81–97, year=2017, publisher=Elsevier.

[69] Bing Zhang, Xin-Xin Ma, and Zhi-Guang Qin. Security Architecture on the Trusting Internet of Things. *Journal of Electronic Science and Technology*, 9 (4):364–367, 2011.

[70] Rafiullah Khan, Sarmad Ullah Khan, Rifaqat Zaheer, and Shahid Khan. Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges. In *2012 10th international conference on frontiers of information technology*, pages 257–260. IEEE, 2012.

[71] Rwan Mahmoud, Tasneem Yousuf, Fadi Aloul, and Imran Zualkernan. Internet of Things (IoT) Security: Current Status, Challenges and Prospective Measures. In *2015 10th international conference for internet technology and secured transactions (ICITST)*, pages 336–341. IEEE, 2015.

[72] Yong Peng, Tianbo Lu, Jingli Liu, Yang Gao, Xiaobo Guo, and Feng Xie. Cyber-Physical System Risk Assessment. In *2013 ninth international conference on intelligent information hiding and multimedia signal processing*, pages 442–447. IEEE, 2013.

[73] Ragunathan Rajkumar, Insup Lee, Lui Sha, and John Stankovic. Cyber-Physical Systems: The Next Computing Revolution. In *Proceedings of the 47th Design Automation Conference*, DAC '10, pages 731–736, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0002-5. doi: 10.1145/1837274.1837461. URL http://doi.acm.org/10.1145/1837274.1837461.

[74] Sunil Singh, Neha Yadav, and Pawan Kumar Chuarasia. A Review on Cyber Physical System Attacks: Issues and Challenges. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, pages 1133–1138. IEEE, 2020.

[75] Bill Miller and Dale Rowe. A Survey of SCADA and Critical Infrastructure Incidents. In *Proceedings of the 1st Annual conference on Research in information technology*, pages 51–56, 2012.

[76] Robert Mitchell and Ing-Ray Chen. A Survey of Intrusion Detection Techniques for Cyber-Physical Systems. *ACM Computing Surveys (CSUR)*, 46 (4):1–29, 2014.

[77] Kai Zhao and Lina Ge. A Survey on the Internet of Things Security. In *2013 Ninth international conference on computational intelligence and security*, pages 663–667. IEEE, 2013.

[78] Alvaro Cardenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, Shankar Sastry, et al. Challenges for Securing Cyber Physical Systems. In *Workshop on future directions in cyber-physical systems security*, volume 5, 2009.

[79] Yilin Mo, Tiffany Hyun-Jin Kim, Kenneth Brancik, Dona Dickinson, Heejo Lee, Adrian Perrig, and Bruno Sinopoli. Cyber-Physical Security of a Smart Grid Infrastructure. *Proceedings of the IEEE*, 100(1):195–209, 2012.

[80] Levente Buttyán, Dennis Gessner, Alban Hessler, and Peter Langendoerfer. Application of Wireless Sensor Networks in Critical Infrastructure Protection: Challenges and Design Options [Security and Privacy in Emerging Wireless Networks]. *IEEE Wireless Communications*, 17(5), 2010.

[81] Alecsandru Patrascu and Victor-Valeriu Patriciu. Cyber Protection of Critical Infrastructures using Supervised Learning. In *2015 20th International*

Conference on Control Systems and Computer Science (CSCS), pages 461–468. IEEE, 2015.

[82] Shengyi Pan, Thomas Morris, and Uttam Adhikari. Classification of Disturbances and Cyber-attacks in Power Systems using Heterogeneous Time-Synchronized Data. *IEEE Transactions on Industrial Informatics*, 11(3): 650–662, 2015.

[83] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack Detection and Identification in Cyber-Physical Systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[84] Steven M Rinaldi, James P Peerenboom, and Terrence K Kelly. Identifying Understanding and Analyzing Critical Infrastructure Interdependencies. *IEEE Control Systems*, 21(6):11–25, 2001.

[85] Gaoqi Liang, Steven R Weller, Junhua Zhao, Fengji Luo, and Zhao Yang Dong. The 2015 Ukraine Blackout: Implications for False Data Injection Attacks. *IEEE Transactions on Power Systems*, 32(4):3317–3318, 2017.

[86] Gaoqi Liang, Junhua Zhao, Fengji Luo, Steven R Weller, and Zhao Yang Dong. A Review of False Data Injection Attacks against Modern Power Systems. *IEEE Transactions on Smart Grid*, 8(4):1630–1638, 2017.

[87] Beibei Li, Rongxing Lu, Wei Wang, and Kim-Kwang Raymond Choo. Distributed Host-based Collaborative Detection for False Data Injection Attacks in Smart Grid Cyber-Physical System. *Journal of Parallel and Distributed Computing*, 103:32–41, 2017.

[88] Liang Hu, Zidong Wang, Qing-Long Han, and Xiaohui Liu. State Estimation under False Data Injection Attacks: Security Analysis and System Protection. *Automatica*, 87:176–183, 2018.

[89] Yuan Chen, Soummya Kar, and José MF Moura. Dynamic Attack Detection in Cyber-Physical Systems with Side Initial State Information. *IEEE Transactions on Automatic Control*, 62(9):4618–4624, 2017.

[90] Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. Cyber-Physical System Security for the Electric Power Grid. *Proceedings of the IEEE*, 100(1):210–224, 2012.

[91] Chuadhry Mujeeb Ahmed, Carlos Murguia, and Justin Ruths. Model-based Attack Detection Scheme for Smart Water Distribution Networks. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 101–113. ACM, 2017.

[92] Ahmed A Abokifa, Kelsey Haddad, Cynthia S Lo, and Pratim Biswas. Detection of Cyber Physical Attacks on Water Distribution Systems via Principal Component Analysis and Artificial Neural Networks. In *World Environmental and Water Resources Congress 2017*, pages 676–691, 2017.

[93] Riccardo Taormina, Stefano Galelli, Nils Ole Tippenhauer, Elad Salomons, and Avi Ostfeld. Characterizing Cyber-Physical Attacks on Water Distribution Systems. *Journal of Water Resources Planning and Management*, 143 (5):04017009, 2017.

[94] Ahmad Ali AlZubi, Mohammed Al-Maitah, and Abdulaziz Alarifi. Cyber Attack Detection in Healthcare using Cyber Physical System and Machine Learning Techniques. *Soft Computing*, 25(18):12319–12332, 2021.

[95] William Schneble and Geethapriya Thamilarasu. Optimal Feature Selection for Intrusion Detection in Medical Cyber-Physical Systems. In *2019 11th International Conference on Advanced Computing (ICoAC)*, pages 238–243. IEEE, 2019.

[96] Dayu Yang, Alexander Usynin, and J Wesley Hines. Anomaly-based Intrusion Detection for SCADA Systems. In *5th International Topical Meeting on Nuclear Plant Instrumentation Control and Human Machine Interface Technologies*, pages 12–16, 2006.

[97] Robert Mitchell and Ray Chen. Behavior Rule Specification-based Intrusion Detection for Safety Critical Medical Cyber Physical Systems. *IEEE Transactions on Dependable and Secure Computing*, 12(1):16–30, 2015.

[98] Christopher Zimmer, Balasubramanya Bhat, Frank Mueller, and Sibin Mohan. Time-based Intrusion Detection in Cyber-Physical Systems. In *Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems*, pages 109–118. ACM, 2010.

[99] Paul Oman and Matthew Phillips. Intrusion Detection and Event Monitoring in SCADA Networks. In *International Conference on Critical Infrastructure Protection*, pages 161–173. Springer, 2007.

[100] Steven Cheung, Bruno Dutertre, Martin Fong, Ulf Lindqvist, Keith Skinner, and Alfonso Valdes. Using Model-based Intrusion Detection for SCADA Networks. In *Proceedings of the SCADA security scientific symposium*, volume 46, pages 1–12. Citeseer, 2007.

[101] Jared Verba and Michael Milvich. Idaho National Laboratory Supervisory Control and Data Acquisition Intrusion Detection System (SCADA IDS). In *Technologies for Homeland Security, 2008 IEEE Conference on*, pages 469–473. IEEE, 2008.

[102] Nazrul Hoque, Dhruba K Bhattacharyya, and Jugal K Kalita. Botnet in DDoS Attacks: Trends and Challenges. *IEEE Communications Surveys & Tutorials*, 17(4):2242–2270, 2015.

[103] Esraa Alomari, Selvakumar Manickam, BB Gupta, Shankar Karuppayah, and Rafeef Alfaris. Botnet-based Distributed Denial of Service (ddos) Attacks on Web Servers: Classification and Art. *arXiv preprint arXiv:1208.0403*, 2012.

[104] Jose Nazario. Ddos Attack Evolution. *Network Security*, 2008(7):7–10, 2008.

[105] Saman Taghavi Zargar, James Joshi, and David Tipper. A Survey of Defense Mechanisms against Distributed Denial of Service (DDoS) Flooding Attacks. *IEEE Communications Surveys & Tutorials*, 15(4):2046–2069, 2013.

[106] Amit Praseed and P Santhi Thilagam. DDoS Attacks at the Application Layer: Challenges and Research Perspectives for Safeguarding Web Applications. *IEEE Communications Surveys & Tutorials*, 2018.

[107] Khundrakpam Johnson Singh and Tanmay De. MLP-GA based algorithm to detect Application layer DDoS Attack. *Journal of Information Security and Applications*, 36:145–153, 2017.

[108] Yuntao Zhao, Wenbo Zhang, Yongxin Feng, and Bo Yu. A Classification Detection Algorithm Based on Joint Entropy Vector against Application-layer DDoS Attack. *Security and Communication Networks*, 2018, 2018.

[109] Huey-Ing Liu and Kuo-Chao Chang. Defending Systems Against Tilt DDoS Attacks. *2011 6th International Conference on Telecommunication Systems Services and Applications (TSSA)*, pages 22–27, 2011.

[110] Indraneel Sreeram and Venkata Praveen Kumar Vuppala. HTTP Flood Attack Detection in Application Layer using Machine Learning Metrics and Bio Inspired Bat Algorithm. *Applied Computing and Informatics*, 2017. doi: https://doi.org/10.1016/j.aci.2017.10.003.

[111] Sheng Wen, Weijia Jia, Wei Zhou, Wanlei Zhou, and Chuan Xu. CALD: Surviving various Application-layer DDoS Attacks that mimic Flash Crowd. In *4th international conference on Network and System Security (NSS)*, pages 247–254. IEEE, 2010.

[112] Jema David Ndibwile, A Govardhan, Kazuya Okada, and Youki Kadobayashi. Web Server Protection against Application Layer DDoS Attacks using Machine Learning and Traffic Authentication. In *IEEE 39th Annual Computer Software and Applications Conference (COMPSAC)*, volume 3, pages 261–267. IEEE, 2015.

[113] Hakem Beitollahi, Dyari Mohammed Sharif, and Mahdi Fazeli. Application Layer DDoS Attack Detection using Cuckoo Search Algorithm-trained Radial Basis Function. *IEEE Access*, 10:63844–63854, 2022.

[114] Junho Choi, Chang Choi, Byeongkyu Ko, and Pankoo Kim. A Method of DDoS Attack Detection using HTTP Packet Pattern and Rule Engine in Cloud Computing Environment. *Soft Computing*, 18:1697–1703, 2014.

[115] Furqan Rustam, Muhammad Faheem Mushtaq, Ameer Hamza, Muhammad Shoaib Farooq, Anca Delia Jurcut, and Imran Ashraf. Denial of Service Attack Classification using Machine Learning with Multi-features. *Electronics*, 11(22):3817, 2022.

[116] Samuel Black and Yoohwan Kim. An Overview on Detection and Prevention of Application Layer DDoS Attacks. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0791–0800. IEEE, 2022.

[117] Tom M Mitchell. Machine Learning. 1(9), 1997.

[118] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT press, 2018.

[119] Rich Caruana and Alexandru Niculescu-Mizil. An Empirical Comparison of Supervised Learning Algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, 2006.

[120] Thomas G Dietterich et al. Ensemble Learning. *The handbook of brain theory and neural networks*, 2(1):110–125, 2002.

[121] Robi Polikar. Ensemble Learning. In *Ensemble machine learning*, pages 1–34. Springer, 2012.

[122] Omer Sagi and Lior Rokach. Ensemble Learning: A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1249, 2018.

[123] Sajid Nagi and Dhruba Kr Bhattacharyya. Classification of Microarray Cancer Data using Ensemble Approach. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2(3):159–173, 2013.

[124] Sotiris B Kotsiantis. Bagging and Boosting Variants for Handling Classification Problems: A Survey. *The Knowledge Engineering Review*, 29(1):78–100, 2014.

[125] J. Van Hulse C. Seiffert, T. M. Khoshgoftaar and A. Napolitano. Resampling or Reweighting: A Comparison of Boosting Implementations. In *Proceedings of the 20th IEEE International Conference on Tools Artificial Intelligence*, pages 445–451, 2008.

[126] Kenji Kira and Larry A Rendell. A Practical Approach to Feature Selection. In *Machine learning proceedings 1992*, pages 249–256. Elsevier, 1992.

[127] Isabelle Guyon and André Elisseeff. An Introduction to Variable and Feature Selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[128] Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, and María Tombilla-Sanromán. Filter Methods for Feature Selection–A Comparative Study. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 178–187. Springer, 2007.

[129] Naoual El Aboudi and Laila Benhlima. Review on Wrapper Feature Selection Approaches. In *2016 International Conference on Engineering & MIS (ICEMIS)*, pages 1–5. IEEE, 2016.

[130] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014.

[131] Hercules Dalianis and Hercules Dalianis. Evaluation Metrics and Evaluation. *Clinical Text Mining: secondary use of electronic patient records*, pages 45–53, 2018.

[132] Redleg. Malicious JavaScript Examples. `https://aw-snap.info/articles/js-examples.php`.

[133] Justin Ma, Lawrence K Saul, Stefan Savage, and Geoffrey M Voelker. Identifying Suspicious URLs: An Application of Large-scale Online Learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 681–688. ACM, 2009.

[134] Ma. Dataset Reputation. `http://www.sysnet.ucsd.edu/projects/url/`.

[135] Justin Ma, Lawrence K Saul, Stefan Savage, and Geoffrey M Voelker. Beyond Blacklists: Learning to Detect Malicious Web sites from Suspicious URLs. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1245–1254. ACM, 2009.

[136] Kurt Thomas, Chris Grier, Justin Ma, Vern Paxson, and Dawn Song. Design and Evaluation of a Real-time URL spam filtering service. In *IEEE Symposium on Security and Privacy (SP)*, pages 447–462. IEEE, 2011.

[137] Nour Moustafa and Jill Slay. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set). In *2015 military communications and information systems conference (MilCIS)*, pages 1–6. IEEE, 2015.

[138] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A Ghorbani. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, 1:108–116, 2018.

[139] Jonathan Goh, Sridhar Adepu, Khurum Nazir Junejo, and Aditya Mathur. A Dataset to Support Research in the Design of Secure Water Treatment

Systems. In *International conference on critical information infrastructures security*, pages 88–99. Springer, 2016.

[140] Thomas Morris and Wei Gao. Industrial Control System Traffic Data Sets for Intrusion Detection Research. In *International Conference on Critical Infrastructure Protection*, pages 65–78. Springer, 2014.

[141] Thomas H Morris, Zach Thornton, and Ian Turnipseed. Industrial Control System Simulation and Data Logging for Intrusion Detection System Research. *7th annual southeastern cyber security summit*, pages 3–4, 2015.

[142] Parthajit Borah, DK Bhattacharyya, and JK Kalita. Malware Dataset Generation and Evaluation. In *2020 IEEE 4th Conference on Information & Communication Technology (CICT)*, pages 1–6. IEEE, 2020.

[143] Daniele Sgandurra, Luis Muñoz-González, Rabih Mohsen, and Emil C Lupu. Automated Dynamic Analysis of Ransomware: Benefits, Limitations and use for Detection. *arXiv preprint arXiv:1609.03020*, 2016.

[144] Rami M Mohammad, Fadi Thabtah, and Lee McCluskey. Phishing Websites Features. *School of Computing and Engineering, University of Huddersfield*, 2015.

[145] Yisroel Mirsky, Tomer Doitshman, Yuval Elovici, and Asaf Shabtai. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. *arXiv preprint arXiv:1802.09089*, 2018.

[146] Upasana Sarmah, DK Bhattacharyya, and JK Kalita. A Survey of Detection Methods for XSS Attacks. *Journal of Network and Computer Applications*, 118:113–143, 2018.

[147] Internet Archive. Heritrix Homepage. `https://webarchive.jira.com/wiki/display/Heritrix`.

[148] Stephen Merity. Navigating the WARC file format. `http://commoncrawl.org/2014/04/navigating-the-warc-file-format/`, 2014. [Accessed: 7-8-2018].

[149] Zhi-Hua Zhou. Ensemble Learning. In *Machine Learning*, pages 181–210. Springer, 2021.

[150] J Elder and Daryl Pregibon. A Statistical Perspective on KDD. *Advances in knowledge discovery and data mining*, pages 83–116, 1996.

[151] Thomas G Dietterich. Machine-learning Research. *AI magazine*, 18(4):97–97, 1997.

[152] Dan Steinberg and P Colla. Cart. *"Classification and Regression Trees, Tree Structured Nonparametric Data Analysis", Interface Documentation, Salford Systems*, 1995.

[153] Leo Breiman. Arcing Classifiers. Technical report, Technical Report 486, Statistics Department, University of California, Berkeley, CA 94720, 1996.

[154] Sandro Vega-Pons and José Ruiz-Shulcloper. A Survey of Clustering Ensemble Algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03):337–372, 2011.

[155] Reza Ghaemi, Md Nasir Sulaiman, Hamidah Ibrahim, Norwati Mustapha, et al. A Survey: Clustering Ensembles Techniques. *World Academy of Science, Engineering and Technology*, 50:636–645, 2009.

[156] Fan Yang, Xuan Li, Qianmu Li, and Tao Li. Exploring the Diversity in Cluster Ensemble Generation: Random Sampling and Random Projection. *Expert Systems with Applications*, 41(10):4844–4866, 2014.

[157] Xiaoli Z Fern and Carla E Brodley. Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 186–193, 2003.

[158] Ishtiaq Ahmed, Rahman Ali, Donghai Guan, Young-Koo Lee, Sungyoung Lee, and TaeChoong Chung. Semi-supervised Learning using Frequent Itemset and Ensemble Learning for SMS Classification. *Expert Systems with Applications*, 42(3):1065–1073, 2015.

[159] Guoxian Yu, Guoji Zhang, Zhiwen Yu, Carlotta Domeniconi, Jane You, and Guoqiang Han. Semi-supervised Ensemble Classification in Subspaces. *Applied Soft Computing*, 12(5):1511–1522, 2012.

[160] Fazia Bellal, Haytham Elghazel, and Alex Aussem. A semi-supervised feature ranking method with ensemble learning. *Pattern Recognition Letters*, 33(10): 1426–1433, 2012.

[161] Lei Shi, Xinming Ma, Lei Xi, Qiguo Duan, and Jingying Zhao. Rough set and Ensemble Learning based Semi-supervised Algorithm for Text Classification. *Expert Systems with Applications*, 38(5):6300–6306, 2011.

[162] Kun-Hong Liu and Chun-Gui Xu. A Genetic Programming-based Approach to the Classification of Multiclass Microarray Datasets. *Bioinformatics*, 25 (3):331–337, 2009.

[163] Hong-Bin Shen and Kuo-Chen Chou. Ensemble Classifier for Protein fold Pattern Recognition. *Bioinformatics*, 22(14):1717–1722, 2006.

[164] Hyun-Joo Oh, Mutiara Syifa, Chang-Wook Lee, and Saro Lee. Land Subsidence Susceptibility Mapping using Bayesian, Functional, and Meta-ensemble Machine Learning Models. *Applied Sciences*, 9(6):1248, 2019.

[165] Matthew Middlehurst, James Large, Michael Flynn, Jason Lines, Aaron Bostrom, and Anthony Bagnall. HIVE-COTE 2.0: A New Meta Ensemble for Time Series Classification. *Machine Learning*, 110(11):3211–3243, 2021.

[166] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive Logistic Regression: A Statistical View of Boosting. *Annals of statistics*, pages 337–374, 2000.

[167] Marcel Dettling. BagBoosting for Tumor Classification with Gene Expression Data. *Bioinformatics*, 20(18):3583–3593, 2004.

[168] Yvan Saeys, Thomas Abeel, and Yves Van de Peer. Robust Feature Selection using Ensemble Feature Selection Techniques. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 313–325. Springer, 2008.

[169] David Jonathan Miller and Siddharth Pal. An Extension of Iterative Scaling for Joint Decision-level and Feature-level Fusion in Ensemble Classification. In *2005 IEEE Workshop on Machine Learning for Signal Processing*, pages 61–66. IEEE, 2005.

[170] Florian Lingenfelser, Johannes Wagner, Thurid Vogt, Jonghwa Kim, and Elisabeth André. Age and Gender Classification from Speech using Decision Level Fusion and Ensemble based Techniques. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[171] Kyle T Peterson, Vasit Sagan, Paheding Sidike, Elizabeth A Hasenmueller, John J Sloan, and Jason H Knouft. Machine Learning-based Ensemble Prediction of Water-quality Variables using Feature-level and Decision-level Fusion with Proximal Remote Sensing. *Photogrammetric Engineering & Remote Sensing*, 85(4):269–280, 2019.

[172] Anders Krogh, Jesper Vedelsby, et al. Neural Network Esembles, Cross Validation, and Active Learning. *Advances in neural information processing systems*, 7:231–238, 1995.

[173] Leo Breiman. Bias, variance, and arcing classifiers. Technical report, Tech. Rep. 460, Statistics Department, University of California, Berkeley, 1996.

[174] Yoav Freund, Robert E Schapire, et al. Experiments with a new Boosting Algorithm. In *icml*, volume 96, pages 148–156. Citeseer, 1996.

[175] Yoav Freund, Robert Schapire, and Naoki Abe. A Short Introduction to Boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780): 1612, 1999.

[176] David H Wolpert. Stacked Generalization. *Neural networks*, 5(2):241–259, 1992.

[177] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing Value Estimation Methods for DNA Microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[178] Shichao Zhang. Nearest neighbor selection for iteratively knn imputation. *Journal of Systems and Software*, 85(11):2541–2552, 2012.

[179] Nazanin Vafaei, Rita A Ribeiro, and Luis M Camarinha-Matos. Data Normalisation Techniques in Decision Making: Case Study with TOPSIS Method. *International journal of information and decision sciences*, 10(1):19–38, 2018.

[180] SGOPAL Patro and Kishore Kumar Sahu. Normalization: A Preprocessing Stage. *arXiv preprint arXiv:1503.06462*, 2015.

[181] Manoranjan Dash and Huan Liu. Feature Selection for Classification. *Intelligent Data Analysis*, 1(1-4):131–156, 1997.

[182] Mario Di Mauro, Giovanni Galatro, Giancarlo Fortino, and Antonio Liotta. Supervised Feature Selection Techniques in Network Intrusion Detection: A Critical Review. *Engineering Applications of Artificial Intelligence*, 101: 104216, 2021.

[183] Huawen Liu, Xindong Wu, and Shichao Zhang. A New Supervised Feature Selection Method for Pattern Classification. *Computational Intelligence*, 30 (2):342–361, 2014.

[184] Jun Chin Ang, Andri Mirzal, Habibollah Haron, and Haza Nuzly Abdull Hamed. Supervised, Unsupervised, and Semi-supervised Feature Selection: A Review on Gene Selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(5):971–989, 2015.

[185] Pabitra Mitra, CA Murthy, and Sankar K. Pal. Unsupervised Feature Selection using Feature Similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):301–312, 2002.

[186] Mingjie Qian and Chengxiang Zhai. Robust Unsupervised Feature Selection. In *Twenty-third International Joint Conference on Artificial Intelligence*, 2013.

[187] Cosmin Lazar, Jonatan Taminau, Stijn Meganck, David Steenhoff, Alain Coletta, Colin Molter, Virginie de Schaetzen, Robin Duque, Hugues Bersini, and Ann Nowe. A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(4):1106–1119, 2012.

[188] Alan Jović, Karla Brkić, and Nikola Bogunović. A Review of Feature Selection Methods with Applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1200–1205. Ieee, 2015.

[189] Ali Muhammad Usman, Umi Kalsom Yusof, and Syibrah Naim. Filter-based Multi-objective Feature Selection using NSGA III and Cuckoo Optimization Algorithm. *IEEE Access*, 8:76333–76356, 2020.

[190] Yingying Zhu, Junwei Liang, Jianyong Chen, and Zhong Ming. An Improved NSGA-III Algorithm for Feature Selection used in Intrusion Detection. *Knowledge-Based Systems*, 116:74–85, 2017.

[191] Nazrul Hoque, Dhruba K Bhattacharyya, and Jugal K Kalita. MIFS-ND: A Mutual Information-based Feature Selection Method. *Expert Systems with Applications*, 41(14):6371–6385, 2014.

[192] Martin Binder, Julia Moosbauer, Janek Thomas, and Bernd Bischl. Multi-objective Hyperparameter Tuning and Feature Selection using Filter Ensembles. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pages 471–479, 2020.

[193] Roberto Battiti. Using Mutual Information for Selecting Features in Supervised Neural Net Learning. *IEEE Transactions on neural networks*, 5(4): 537–550, 1994.

[194] Kenji Kira, Larry A Rendell, et al. The Feature Selection Problem: Traditional Methods and A New Algorithm. In *Association for the Advancement of Artificial Intelligence (AAAI)*, volume 2, pages 129–134, 1992.

[195] Rajen B Bhatt and M Gopal. On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters*, 26(7):965–975, 2005.

[196] Rasim Cekik and Alper Kursat Uysal. A Novel Filter Feature Selection Method using Rough Set for Short Text Data. *Expert Systems with Applications*, 160:113691, 2020.

[197] Patrick E Meyer and Gianluca Bontempi. On the use of Variable Complementarity for Feature Selection in Cancer Classification. In *Workshops on applications of evolutionary computation*, pages 91–102. Springer, 2006.

[198] Qinbao Song, Jingjie Ni, and Guangtao Wang. A Fast Clustering-based Feature Subset Selection Algorithm for High-dimensional Data. *IEEE transactions on knowledge and data engineering*, 25(1):1–14, 2011.

[199] William H Press, Brian P Flannery, Saul A Teukolsky, and William T Vetterling. Numerical Recipes, 1989.

[200] François Fleuret. Fast Binary Feature Selection with Conditional Mutual Information. *Journal of Machine learning research*, 5(9), 2004.

[201] Chris Ding and Hanchuan Peng. Minimum Redundancy Feature Selection from Microarray Gene Expression Data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.

[202] Jianyu Miao and Lingfeng Niu. A Survey on Feature Selection. *Procedia Computer Science*, 91:919–926, 2016.

[203] Ron Kohavi and George H John. Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[204] Sebastián Maldonado and Richard Weber. A Wrapper Method for Feature Selection using Support Vector Machines. *Information Sciences*, 179(13): 2208–2217, 2009.

[205] Gang Chen and Jin Chen. A Novel Wrapper Method for Feature Selection and its Applications. *Neurocomputing*, 159:219–226, 2015.

[206] Hui-Huang Hsu, Cheng-Wei Hsieh, and Ming-Da Lu. Hybrid Feature Selection by Combining Filters and Wrappers. *Expert Systems with Applications*, 38(7):8144–8150, 2011.

[207] Upasana Sarmah and DK Bhattacharyya. Cost-Effective Detection of Cyber Physical System Attacks. In *Advances in Machine Learning for Big Data Analysis*, pages 33–69. Springer, 2022.

[208] Alexey Tsymbal, Mykola Pechenizkiy, and Pádraig Cunningham. Diversity in Search Strategies for Ensemble Feature Selection. *Information Fusion*, 6 (1):83–98, 2005.

[209] Weizeng Ni. A Review and Comparative Study on Univariate Feature Selection Techniques. 2012.

[210] Jianping Hua, Waibhav D Tembe, and Edward R Dougherty. Performance of Feature-Selection Methods in the Classification of High-dimension Data. *Pattern Recognition*, 42(3):409–424, 2009.

[211] Chap T Le. *Introductory Biostatistics*. John Wiley & Sons, 2003.

[212] Patrick Schober, Christa Boer, and Lothar A Schwarte. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & analgesia*, 126(5): 1763–1768, 2018.

[213] Agustin Garcia Asuero, Ana Sayago, and AG González. The Correlation Coefficient: An Overview. *Critical reviews in analytical chemistry*, 36(1): 41–59, 2006.

[214] Reginald Smith. A Mutual Information Approach to Calculating Nonlinearity. *Stat*, 4(1):291–303, 2015.

[215] Alexander L Young, Willem van den Boom, Rebecca A Schroeder, Vijay Krishnamoorthy, Karthik Raghunathan, Hau-Tieng Wu, and David B Dunson. Mutual Information: Measuring Nonlinear Dependence in Longitudinal Epidemiological Data. *Plos one*, 18(4):e0284904, 2023.

[216] Carlos D Correa and Peter Lindstrom. The Mutual Information Diagram for Uncertainty Visualization. *International Journal for Uncertainty Quantification*, 3(3), 2013.

[217] Joost CF De Winter, Samuel D Gosling, and Jeff Potter. Comparing the Pearson and Spearman Correlation Coefficients across Distributions and Sample Sizes: A Tutorial using Simulations and Empirical Data. *Psychological methods*, 21(3):273, 2016.

[218] Fatemeh Amiri, MohammadMahdi Rezaei Yousefi, Caro Lucas, Azadeh Shakery, and Nasser Yazdani. Mutual Information-based Feature Selection for Intrusion Detection systems. *Journal of Network and Computer Applications*, 34(4):1184–1199, 2011.

[219] Jorge R Vergara and Pablo A Estévez. A Review of Feature Selection Methods based on Mutual Information. *Neural computing and applications*, 24:175–186, 2014.

[220] H Yang and John Moody. Feature Selection based on Joint Mutual Information. In *Proceedings of international ICSC symposium on advances in intelligent data analysis*, volume 23. Citeseer, 1999.

[221] Carolin A Rickert, Manuel Henkel, and Oliver Lieleg. An Efficiency-driven, Correlation-based Feature Elimination Strategy for Small Datasets. *APL Machine Learning*, 1(1), 2023.

[222] Hae-Young Kim. Statistical Notes for Clinical Researchers: Covariance and Correlation. *Restorative dentistry & endodontics*, 43(1), 2018.

[223] Pablo A Jaskowiak, Ricardo JGB Campello, Thiago F Covoes, and Eduardo R Hruschka. A Comparative Study on the use of Correlation Coefficients for Redundant Feature Elimination. In *2010 Eleventh Brazilian Symposium on Neural Networks*, pages 13–18. IEEE, 2010.

[224] Jerome H Friedman. Stochastic Gradient Boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

[225] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, Rory Mitchell, Ignacio Cano, Tianyi Zhou, et al. Xgboost: Extreme Gradient Boosting. *R package version 0.4-2*, 1(4):1–4, 2015.

[226] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[227] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely Randomized Trees. *Machine learning*, 63(1):3–42, 2006.

[228] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature Selection based on Mutual Information Criteria of Max-dependency, Max-relevance, and Min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

[229] Maria E Orlowska and Marian W Orlowski. Maintenance of Knowledge in Dynamic Information Systems. In *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, pages 315–329. Springer, 1992.

[230] Takeshi Yatagai, Takamasa Isohara, and Iwao Sasase. Detection of HTTP-GET Flood Attack based on Analysis of Page Access Behavior. In *2007 IEEE Pacific rim conference on communications, computers and signal processing*, pages 232–235. IEEE, 2007.

[231] A Dhanapal and P Nithyanandam. The Slow HTTP DDOS Attacks: Detection, Mitigation and Prevention in the Cloud Environment. *Scalable Comput. Pract. Exp.*, 20(4):669–685, 2019.

[232] A Dhanapal and P Nithyanandam. An OpenStack based Cloud Testbed Framework for Evaluating HTTP Flooding Attacks. *Wireless Networks*, 27 (8):5491–5501, 2021.

[233] Hala Albaroodi, Selvakumar Manickam, and Mohammed Anbar. A Proposed Framework for Outsourcing and Secure Encrypted Data on OpenStack Object Storage (Swift). *Journal of Computer Science*, 11(3):590, 2015.

[234] Reza Mohammadi, Chhagan Lal, and Mauro Conti. HTTPScout: A Machine Learning based Countermeasure for HTTP Flood Attacks in SDN. *International Journal of Information Security*, 22(2):367–379, 2023.

[235] Noe Marcelo Yungaicela-Naula, Cesar Vargas-Rosales, and Jesus Arturo Perez-Diaz. SDN-based Architecture for Transport and Application Layer DDoS Attack Detection by using Machine and Deep Learning. *IEEE Access*, 9:108495–108512, 2021.

[236] Ahmad Zainudin, Love Allen Chijioke Ahakonye, Rubina Akter, Dong-Seong Kim, and Jae-Min Lee. An Efficient Hybrid DNN for DDoS Detection and Classification in Software Defined IIoT Networks. *IEEE Internet of Things Journal*, 2022.

[237] Noe M Yungaicela-Naula, Cesar Vargas-Rosales, Jesús Arturo Pérez-Díaz, and Diego Fernando Carrera. A Flexible SDN-based Framework for Slow-rate DDoS Attack Mitigation by using Deep Reinforcement Learning. *Journal of network and computer applications*, 205:103444, 2022.

[238] Omerah Yousuf and Roohie Naaz Mir. DDoS Attack Detection in Internet of Things using Recurrent Neural Network. *Computers and Electrical Engineering*, 101:108034, 2022.

[239] George H John, Ron Kohavi, and Karl Pfleger. Irrelevant Features and the Subset Selection Problem. In *Machine learning proceedings 1994*, pages 121–129. Elsevier, 1994.

[240] Pat Langley. Selection of Relevant Features in Machine learning. In *Proceedings of the AAAI Fall symposium on relevance*, volume 184, pages 245–271. California, 1994.

[241] Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova, and Ali A Ghorbani. Detecting HTTP-based Application Layer DoS Attacks on Web Servers in the presence of Sampling. *Computer Networks*, 121:25–36, 2017.

[242] Gavin Brown, Adam Pocock, Ming-Jie Zhao, and Mikel Luján. Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. *The journal of machine learning research*, 13(1):27–66, 2012.

[243] Nam Yong Kim, Shailendra Rathore, Jung Hyun Ryu, Jin Ho Park, and Jong Hyuk Park. A Survey on Cyber Physical System Security for IoT: Issues, Challenges, Threats, Solutions. *Journal of Information Processing Systems*, 14(6):1361–1384, 2018.

[244] Ajeet Singh and Anurag Jain. Study of Cyber Attacks on Cyber-Physical System. In *Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, pages 26–27, 2018.

[245] Gerald Brown, Matthew Carlyle, Javier Salmerón, and Kevin Wood. Defending Critical Infrastructure. *Interfaces*, 36(6):530–544, 2006.

[246] Silvio Ereno Quincozes, Daniel Mossé, Diego Passos, Célio Albuquerque, Luiz Satoru Ochi, and Vinícius Figueiredo dos Santos. On the Performance of GRASP-based Feature Selection for CPS Intrusion Detection. *IEEE Transactions on Network and Service Management*, 19(1):614–626, 2021.

[247] Mauricio GC Resende and Celso C Ribeiro. Greedy Randomized Adaptive Search Procedures: Advances, Hybridizations, and Applications. *Handbook of metaheuristics*, pages 283–319, 2010.

[248] Silvio E Quincozes, Diego Passos, Célio Albuquerque, Daniel Mossé, and Luiz Satoru Ochi. An Extended Assessment of Metaheuristics-based Feature Selection for Intrusion Detection in CPS Perception Layer. *Annals of Telecommunications*, 77(7):457–471, 2022.

[249] Saoreen Rahman, Muhammad Ahmed, and M Shamim Kaiser. ANFIS based Cyber Physical Attack Detection System. In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pages 944–948. IEEE, 2016.

[250] Alshaibi Ahmed Jamal, Al-Ani Mustafa Majid, Anton Konev, Tatiana Kosachenko, and Alexander Shelupanov. A Review on Security Analysis of Cyber Physical Systems using Machine Learning. *Materials today: proceedings*, 80:2302–2306, 2023.

[251] Yuriy Zacchia Lun, Alessandro D'Innocenzo, Francesco Smarra, Ivano Malavolta, and Maria Domenica Di Benedetto. State of the art of Cyber-Physical Systems Security: An Automatic Control Perspective. *Journal of Systems and Software*, 149:174–216, 2019.

[252] Weizhong Yan, Lalit K Mestha, and Masoud Abbaszadeh. Attack Detection for Securing Cyber Physical Systems. *IEEE Internet of Things Journal*, 6 (5):8471–8481, 2019.

[253] Firooz B Saghezchi, Georgios Mantas, Manuel A Violas, A Manuel de Oliveira Duarte, and Jonathan Rodriguez. Machine Learning for DDoS Attack Detection in Industry 4.0 CPPSs. *Electronics*, 11(4):602, 2022.

[254] Ranjit Panigrahi, Samarjeet Borah, Moumita Pramanik, Akash Kumar Bhoi, Paolo Barsocchi, Soumya Ranjan Nayak, and Waleed Alnumay. Intrusion Detection in Cyber Physical Environment using Hybrid Naïve Bayes—Decision Table and Multi-objective Evolutionary Feature Selection. *Computer Communications*, 188:133–144, 2022.

[255] Fernando Jiménez, Gracia Sánchez, José M García, Guido Sciavicco, and Luis Miralles. Multi-objective Evolutionary Feature Selection for Online Sales Forecasting. *Neurocomputing*, 234:75–92, 2017.

[256] Chih-Wen Chen, Yi-Hong Tsai, Fang-Rong Chang, and Wei-Chao Lin. Ensemble Feature Selection in Medical Datasets: Combining Filter, Wrapper, and Embedded Feature Selection Results. *Expert Systems*, 37(5):e12553, 2020.

[257] Amin Hashemi, Mohammad Bagher Dowlatshahi, and Hossein Nezamabadi-pour. Ensemble of Feature Selection Algorithms: A Multi-Criteria Decision-

Making Approach. *International Journal of Machine Learning and Cybernetics*, 13(1):49–69, 2022.

[258] Jong Hyen Kim and Byeong Seok Ahn. Extended VIKOR Method using Incomplete Criteria Weights. *Expert Systems with Applications*, 126:124–132, 2019.

[259] Serafim Opricovic. Multicriteria Optimization of Civil Engineering Systems. *Faculty of civil engineering, Belgrade*, 2(1):5–21, 1998.

[260] Mohammad Aizat Basir, Mohamed Saifullah Hussin, and Yuhanis Yusof. Ensemble Feature Selection Method based on Bio-Inspired Algorithms for Multi-Objective Classification Problem. In *Advances on Smart and Soft Computing: Proceedings of ICACIn 2020*, pages 167–176. Springer, 2021.

[261] Haleh Vafaie and Kenneth A De Jong. Genetic Algorithms as a Tool for Feature Selection in Machine Learning. In *ICTAI*, pages 200–203. Citeseer, 1992.

[262] Gypsy Nandi. An Enhanced Approach to Las Vegas Filter (LVF) Feature Selection Algorithm. In *2011 2nd National Conference on Emerging Trends and Applications in Computer Science*, pages 1–3. IEEE, 2011.

[263] Asif Ekbal, Sriparna Saha, and Christoph S Garbe. Feature Selection using Multiobjective Optimization for Named Entity Recognition. In *2010 20th International Conference on Pattern Recognition*, pages 1937–1940. IEEE, 2010.

[264] Wing WY Ng, Yuxi Tuo, Jianjun Zhang, and Sam Kwong. Training Error and Sensitivity-based Ensemble Feature Selection. *International Journal of Machine Learning and Cybernetics*, 11:2313–2326, 2020.

[265] Julian Blank, Kalyanmoy Deb, and Proteek Chandan Roy. Investigating the Normalization Procedure of NSGA-III. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 229–240. Springer, 2019.

[266] Yousef Sanjalawe and Turke Althobaiti. DDoS Attack Detection in Cloud Computing Based on Ensemble Feature Selection and Deep Learning. *Computers, Materials & Continua*, 75(2), 2023.

[267] James Kennedy and Russell Eberhart. Particle Swarm Optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. ieee, 1995.

[268] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey Wolf Optimizer. *Advances in engineering software*, 69:46–61, 2014.

[269] Amir Hossein Gandomi and Amir Hossein Alavi. Krill Herd: A New Bio-inspired Optimization Algorithm. *Communications in nonlinear science and numerical simulation*, 17(12):4831–4845, 2012.

[270] Seyedali Mirjalili and Andrew Lewis. The Whale Optimization Algorithm. *Advances in engineering software*, 95:51–67, 2016.

[271] Arun Kumar Dey, Govind P Gupta, and Satya Prakash Sahu. A Metaheuristic-based Ensemble Feature Selection Framework for Cyber Threat Detection in IoT-enabled Networks. *Decision Analytics Journal*, 7: 100206, 2023.

[272] Yuyang Zhou, Guang Cheng, Shanqing Jiang, and Mian Dai. Building an Efficient Intrusion Detection System based on Feature Selection and Ensemble Classifier. *Computer networks*, 174:107247, 2020.

[273] Burak Kolukisa and Burcu Bakir-Gungor. Ensemble Feature Selection and Classification Methods for Machine Learning-based Coronary Artery Disease Diagnosis. *Computer Standards & Interfaces*, 84:103706, 2023.

[274] Borja Seijo-Pardo, Iago Porto-Díaz, Verónica Bolón-Canedo, and Amparo Alonso-Betanzos. Ensemble Feature Selection: Homogeneous and Heterogeneous Approaches. *Knowledge-Based Systems*, 118:124–139, 2017.

[275] Thorsten Joachims. Optimizing Search Engines using Clickthrough Data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, 2002.

[276] Amin Hashemi, Mohammad Bagher Dowlatshahi, and Hossein Nazamabadi-pour. Minimum Redundancy Maximum Relevance Ensemble Feature Selection: A Bi-objective Pareto-based Approach. *Journal of Soft Computing and Information Technology (JSCIT)*, 12(1), 2023.

[277] Alexey Tsymbal, Seppo Puuronen, and David W Patterson. Ensemble Feature Selection with the Simple Bayesian Classification. *Information fusion*, 4 (2):87–100, 2003.

[278] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[279] Sridhar Adepu and Aditya Mathur. An Investigation into the Response of a Water Treatment System to Cyber Attacks. In *2016 IEEE 17th International Symposium on High Assurance Systems Engineering (HASE)*, pages 141–148. IEEE, 2016.

[280] Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M Poskitt, and Jun Sun. Anomaly Detection for a Water Treatment System using Unsupervised Machine Learning. In *2017 IEEE international conference on data mining workshops (ICDMW)*, pages 1058–1065. IEEE, 2017.

[281] Marta Catillo, Antonio Pecchia, and Umberto Villano. CPS-GUARD: Intrusion Detection for Cyber-Physical Systems and IoT Devices using Outlier-aware Deep Autoencoders. *Computers & Security*, 129:103210, 2023.

[282] Konstantinos Demertzis, Lazaros Iliadis, and Ilias Bougoudis. Gryphon: A Semi-supervised Anomaly Detection System based on One-Class Evolving Spiking Neural Network. *Neural Computing and Applications*, 32(9):4303–4314, 2020.

[283] Emmanuel Aboah Boateng, JW Bruce, and Douglas A Talbert. Anomaly Detection for a Water Treatment System based on One-Class Neural Network. *IEEE access*, 10:115179–115191, 2022.

[284] Justin M Beaver, Raymond C Borges-Hink, and Mark A Buckner. An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications. In *2013 12th international conference on machine learning and applications*, volume 2, pages 54–59. IEEE, 2013.

[285] Konstantinos Demertzis, Lazaros Iliadis, and Stefanos Spartalis. A Spiking One-class Anomaly Detection Framework for Cyber-security on Industrial

Control Systems. In *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings*, pages 122–134. Springer, 2017.

[286] Mariam Elnour, Nader Meskin, Khaled Khan, and Raj Jain. A Dual-Isolation-Forests-based Attack Detection Framework for Industrial Control Systems. *IEEE Access*, 8:36639–36651, 2020.

[287] Dan Li, Dacheng Chen, Jonathan Goh, and See-kiong Ng. Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series. *arXiv preprint arXiv:1809.04758*, 2018.

Glossary

Crawl The act of searching and indexing of Web content by a software program.. 4

DDoS A form of an attack where an army of bots simultaneously send HTTP traffic to an application to disrupt its services. 7

Dynamic Analysis Analysis done by executing the software.. 5

IP Address A unique address with which every device connected to the Internet is identified.. 4

Static Analysis Analysis done by examining the source code and not by way of execution.. 4

Vulnerabilities Weaknesses or flaws existing in the system.. 2

Web Application Software which runs on an application server and are accessed with the help of browsers,. 1

Worms Malicious software which tries to self-replicate and spread to other devices. . 7

List of Publications

Journal Publications

1. Upasana Sarmah, Dhruba Kr Bhattacharyya, Jugal K. Kalita, “A survey of Detection Methods for XSS Attacks”, Journal of Network and Computer Applications, 118, 113-143, 2018. (Scopus Indexed)

2. Upasana Sarmah, Parthajit Borah, Dhruba Kr Bhattacharyya, “Supervised Ensemble Learning Approaches and Methods: An experimental Investigation”, Springer Nature Computer Science, 5(7), 924, 2024. (Scopus Indexed)

3. Parthajit Borah, Upasana Sarmah, Dhruba Kr Bhattacharyya, Jugal K. Kalita, “Unmasking the common traits: an ensemble approach for effective malware detection", International Journal of Information Security, 1-11, 2024. (Scopus Indexed)

4. Upasana Sarmah, Dhruba Kr Bhattacharyya, Jugal K. Kalita, “MICC-UD: A Mutual Information and Correlation based Feature Selection Algorithm”. (Under Review)

5. Upasana Sarmah, Dhruba K. Bhattacharyya, “INFS-MICC: An Incremental Feature Selection Algorithm based on Mutual Information and Correlation”. (Under Review)

Conference Publications

1. Upasana Sarmah, Dhruba Kr Bhattacharyya, Jugal K. Kalita,“XSSD: A Cross-site Scripting Attack Dataset and its Evaluation”, 3rd ISEA Conference on Security and Privacy (ISEA-ISAP), 21-30, 2020.

Book Chapter Publications

1. Upasana Sarmah, Dhruba Kr. Bhattacharyya, "Cost-Effective Detection of Cyber Physical System Attacks", Advances in Machine Learning for Big Data Analysis, Springer Nature Singapore, 33-69.
   DOI: https://doi.org/10.1007/978-981-16-8930-7_2

The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

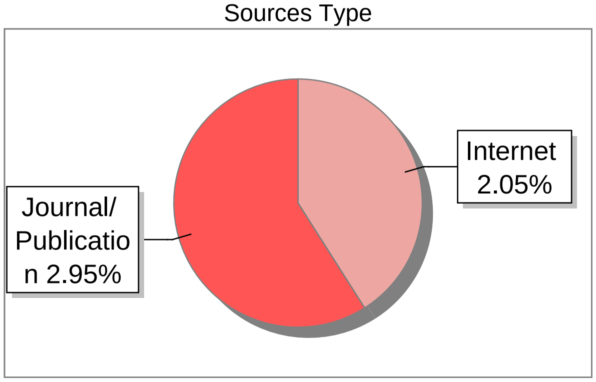
Author Name	Upasana Sarmah
Title	Detection of Web-based Attacks using Machine Learning Techniques
Paper/Submission ID	1902035
Submitted by	nabin@tezu.ernet.in
Submission Date	2024-05-30 11:25:29
Total Pages, Total Words	238, 69214
Document type	Thesis

Result Information

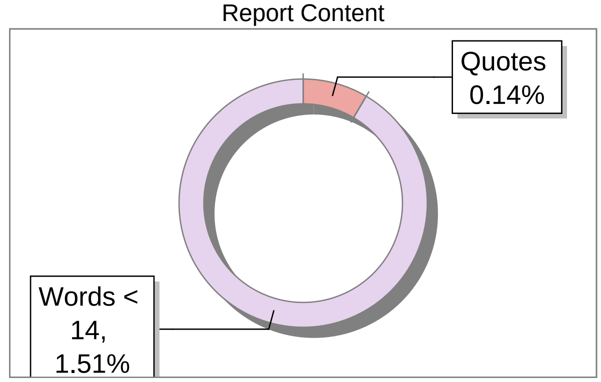
Similarity **5 %**

1 10 20 30 40 50 60 70 80 90

Sources Type



Report Content



Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	**8 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

DrillBit Similarity Report

5	79	A	A-Satisfactory (0-10%) B-Upgrade (11-40%) C-Poor (41-60%) D-Unacceptable (61-100%)
SIMILARITY %	MATCHED SOURCES	GRADE	

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
7	moam.info	<1	Internet Data
8	link.springer.com	<1	Internet Data
9	www.mdpi.com	<1	Publication
10	gredos.usal.es	<1	Publication
11	link.springer.com	<1	Internet Data
12	www.researchgate.net	<1	Internet Data
13	link.springer.com	<1	Internet Data
14	A Comparative Study of Feature Selection Methods for Stress Hotspot C, by Mangal, Ankita Hol- 2018	<1	Publication
15	www.mdpi.com	<1	Internet Data
16	cra.org	<1	Publication
17	Thesis Submitted to Shodhganga, shodhganga.inflibnet.ac.in	<1	Publication
18	www.gandhinagaruni.ac.in	<1	Publication
19	Thesis Submitted to Shodhganga Repository	<1	Publication

73	IEEE 2015 Ninth International Conference on Frontier of Computer Sci by	<1	Publication
74	An Object-Based Method for Urban Land Cover Classification Using Airbo by Chen-2014	<1	Publication
75	businessdocbox.com	<1	Internet Data
76	econjournals.sgh.waw.pl	<1	Publication
77	CHASE, a charge-assisted sequencing algorithm for automated homology-based prote by Giusepp-2005	<1	Publication
78	climateerinvest.blogspot.com	<1	Internet Data
79	Morphological statistical features for automatic segmentation of blood vessel st by Bharkad-2017	<1	Publication
80	redcol.minciencias.gov.co	<1	Publication
81	www.dx.doi.org	<1	Publication
82	biofarmaka.ipb.ac.id	<1	Publication
83	docplayer.net	<1	Internet Data
84	Geospatial-Temporal and Demand Models for Opioid Admissions, Implications for Po by Fulton-2019	<1	Publication
85	The trial of Jesus in the Gospel of Mark by Bammel-1996	<1	Publication

EXCLUDED SOURCES

1	A survey of detection methods for XSS attacks, by Sarmah, Upasana Bh-2018	3	Publication
2	A survey of detection methods for XSS attacks, by Sarmah, Upasana Bh-2018	3	Publication

3	A survey of detection methods for XSS attacks, by Sarmah, Upasana Bh- 2018	1	Publication
4	A survey of detection methods for XSS attacks, by Sarmah, Upasana Bh- 2018	1	Publication
5	A survey of detection methods for XSS attacks, by Sarmah, Upasana Bh- 2018	1	Publication
6	A survey of detection methods for XSS attacks, by Sarmah, Upasana Bh- 2018	<1	Publication