IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

Spam Detection Technique Using Machine Learning With Principle Component Analysis

Mrs. P. Immaculate Rexi Jenifer[1], Abinaya S[2], Banu Rithika.R[3], Madhu Bala. R[4]
Assistant Professor, Department of Science and Computer Engineering[1]
Students, Department of Science and Computer Engineering[2,3,4]
Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tiruvarur, Tamil Nadu, India

Abstract: *A collection of millions of devices with sensors and actuators that are linked via wired or wireless channels for data transmission. Over the last decade, it has grown rapidly, with more than 25 billion devices expected to be connected by 2020. The amount of data released by these devices will multiply many times over in the coming years. In addition to increased volume, the device generates a large amount of data in a variety of modalities with varying data quality defined by its speed in terms of time and position dependency. In such an environment, machine learning algorithms can play an important role in ensuring biotechnology-based security and authorization, as well as anomalous detection to improve usability and security.On the other hand, attackers frequently use learning algorithms to exploit system vulnerabilities. As a result of these considerations, we propose that the security of devices be improved by employing machine learning to detect spam. Spam Detection Using Machine Learning Framework is proposed to attain this goal. Four machine learning models are assessed using multiple metrics and a vast collection of input feature sets in this framework. Each model calculates a spam score based on the input attributes that have been adjusted. This score represents the device's trustworthiness based on a variety of factors. In comparison to other current systems, the findings collected demonstrate the effectiveness of the proposed method.*

Keywords: Collection of data, Authorization, Anomalous detection, Support Vector Machine, K-nearest neighbour, Spam.

REFERENCES

[1]. Aaisha Makkar, Sahil (GE) Garg, Neeraj Kumar, M. Shamim Hossain, Ahmed Ghoneim, Mubarak Alrashoud,"An Efficient Spam Detection Technique for IoT Devices using Machine Learning" ,IEEE Transactions on Industrial Informatics (Volume: 17, Issue: 2, Feb. 2021)

[2]. Z.-K. Zhang, M. C. Y. Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, andS. Shieh, "Iot security: ongoing challenges and research opportunities,"in 2014 IEEE 7th international conference on service-oriented computing and applications. IEEE, 2014, pp. 230–234.

[3]. A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for iot security and privacy: The case study of a smarthome," in 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops). IEEE, 2017, pp. 618–623.

[4]. E. Bertino and N. Islam, "Botnets and internet of things security," Computer, no. 2, pp. 76–79, 2017.

[5]. C. Zhang and R. Green, "Communication security in internet of thing: preventive measure and avoid ddos attack over iot network," Proceedings of the 18th Symposium on Communications & Networking. Society for Computer Simulation International, 2015, pp. 8–15.

[6]. W. Kim, O.-R. Jeong, C. Kim, and J. So, "The dark side of the internet: Attacks, costs and responses," Information systems, vol. 36, no. 3, pp.675–705, 2011.

[7]. H. Eun, H. Lee, and H. Oh, "Conditional privacy preserving security protocol for nfc applications," IEEE Transactions on Consumer Electronics, vol. 59, no. 1, pp. 153–160, 2013.

[8]. R. V. Kulkarni and G. K. Venayagamoorthy, "Neural network based secure media access control protocol for wireless sensor networks," in 2009 International Joint Conference on Neural Networks. IEEE, 2009, pp. 1680–1687.

[9]. M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

strategies, and applications,"IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 1996–2018, 2014.

[10]. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2015.

[11]. F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," Soft Computing, vol. 20, no. 1, pp. 343–357, 2016.

[12]. N. Sutta, Z. Liu, and X. Zhang, "A study of machine learningalgorithms on email spam classification," in Proceedings of the 35th International Conference, ISC High Performance 2020, vol. 69, pp. 170–179,Frankfurt,Germa.

[13]. L. Xiao, Y. Li, X. Huang, and X. Du, "Cloud-based malware detection game for mobile devices with offloading," IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2742–2750, 2017.

[14]. J. W. Branch, C. Giannella, B. Szymanski, R. Wolff, and H. Kargupta, "In-network outlier detection in wireless sensor networks," Knowledge and information systems, vol. 34, no. 1, pp. 23–54, 2013.

[15]. I. Jolliffe, Principal component analysis. Springer, 2011.

[16]. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of machine learning research, vol. 3, no. Mar, pp. 1157–1182, 2003.

[17]. L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in Proceedings of the 20th international conference on machine learning (ICML-03), 2003, pp. 856–863.

[18]. A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque, "Artificial intelligence driven mechanism for edge computing based industrial applications," IEEE Transactions on Industrial Informatics, 2019.

[19]. A. H. Sodhro, Z. Luo, G. H. Sodhro, M. Muzamal, J. J. Rodrigues, and V. H. C. de Albuquerque, "Artificial intelligence based qos optimization for multimedia communication in iov systems," Future Generation Computer Systems, vol. 95, pp. 667–680, 2019.

[20]. L. University, "Refit smart home dataset," https://repository.lboro.ac.uk/ articles/REFIT Smart Home dataset/2070091, 2019 (accessed April 26, 2019).

[21]. R, "Rstudio," 2019 (accessed October 23, 2019

[22]. T. Vyas, P. Prajapati, and S. Gadhwal, "A survey and evaluation of supervised machine learning techniques for spam e-mail filtering," in Proceedings of the 2015 IEEE international conference on electrical, computer and communication technologies (ICECCT), IEEE, Tamil Nadu, India, March 2015.

[23]. L. N. Petersen, "(e ageing body in monty Python live (mostly)," European Journal of Cultural Studies, vol. 21, no. 3, pp. 382–394, 2018.

[24]. L. Zhuang, J. Dunagan, D. R. Simon, H. J. Wang, and J. D. Tygar, "Characterizing botnets from email spam records," LEET, vol. 8, pp. 1–9, 2008.

[25]. W. N. Gansterer, A. G. K. Janecek, and R. Neumayer, "Spam filtering based on latent semantic indexing," in Survey of Text Mining II, pp. 165–183, Springer, New York, NY, USA, 2008.

[26]. D. Lee, M. J. Lee, and B. J. Kim, "Deviation-based spamfiltering method via stochastic approach," EPL (Europhysics Letters), vol. 121, no. 6, Article ID 68004, 2018.

[27]. A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," Telecommunication Systems, vol. 68, no. 4, pp. 687–700, 2018.

[28]. M. F. N. K. Pathan and V. Kamble, "A review various techniques for content based spam filtering," Engineering and Technology, vol. 4, 2018.

[29]. A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list," EURASIP Journal on Information Security, vol. 2016, no. 1, p. 9, 2016.

[30]. A. Bhowmick and S. M. Hazarika, "Machine learning for E-mail spam filtering: review, techniques and trends," 2016, https://www.researchgate.net/publication/303812063_Machine_Learning_for_ E-mail_Spam_Filtering_ReviewTechniques_and_Trends

[31]. M. Bassiouni, M. Ali, and E. A. El-Dahshan, "Ham and spam e-mails classification using machine learning techniques," Journal of Applied Security Research, vol. 13, no. 3, pp. 315– 331, 2018.

[32]. J. R. M´endez, T. R. Cotos-Yañez, and D. Ruano-Ord´as, "A new semantic-based feature selection method for

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

spam filtering," Applied Soft Computing, vol. 76, pp. 89–104, 2019.

[33]. R. Alguliyev and S. Nazirova, "Two approaches on implementation of CBR and CRM technologies to the spam filtering problem," Journal of Information Security, vol. 3, no. 1, Article ID 16724, 2012.

[34]. E. Alpaydin, Introduction to Machine Learning, MIT Press, Cambridge, UK, 2020.

[35]. E. P. Sanz, J. M. Gomez Hidalgo, and J. C. Cortizo P ´ erez, ´ "Chapter 3 email spam filtering," Advances in Computers, vol. 74, pp. 45–114, 2008.

[36]. S. Pitchaimani, V. P. Kodaganallur, and C. Newell, "Systems and methods for controlling email access," Google Patents, 2020.

[37]. A. d. A. Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, and S. N. Tran, "Neural-symbolic computing: an effective methodology for principled integration of machine learning and reasoning," Journal of Applied Logic, vol. 6, 2019.

[38]. A. Singh, N. (akur, and A. Sharma, "A review of supervised machine learning algorithms," in Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, New Delhi, India, March 2016.

[39]. J. Tanha, M. van Someren, and H. Afsarmanesh, "Semi-supervised self-training for decision tree classifiers," International Journal of Machine Learning and Cybernetics, vol. 8, no. 1, pp. 355–370, 2017