Data mining is readily growing and accepted technology in recent years. It is utilized for finding instant decisions by analyzing the historical records. A formal decision making technique can also be helpful for information security. In this presented work the demonstration of a data mining application is provided. The proposed data mining application contributes on the information security. Therefore URL classification problem is taken in consideration. In this context we can apply here the any supervised learning algorithm but in this work the association rule mining based technique is proposed for solving the URL classification. That technique is used for analyzing the URL patterns of two kinds of class labels i.e. phishing and legitimate. In this context a rule based classification technique is proposed. That technique is computing the association rules and we can use these patterns to classify the URL data. The idea is taken from [1] where apriori algorithm is implemented for generation and classification of phishing URLs. Apriori algorithm is computationally complex and requires significant amount of time and memory for generating candidate sets. Therefore we usages the FP-Tree algorithm which efficient develops the association rules with less resource requirements. The system can be
used for designing the phishing tool bars. This technique is used with the phish tank dataset with different set of data for experimentations. The obtained results shows the proposed technique requires less amount of time and memory. In near future it is tried to reduce time and improve the accuracy of the proposed phishing URL classification system.

References

1. S. Carolin Jeeva and Elijah Blessing Rajsingh, “Intelligent phishing url detection using association rule mining”, Hum. Cent. Comput. Inf. Sci. (2016) 6:10, DOI 10.1186/s13673-016-0064-3
2. Chapter 3: Data Mining: an Overview, available online at: http://shodhganga.inflibnet.ac.in/bitstream/10603/11075/7/07_chapter3.pdf
3. Mohammed J. Zaki and Wagner MeiraJr, “Data Mining and Analysis Fundamental Concepts and Algorithms”, Cambridge University Press Hardback, 2014 [Book]
4. Michael Goebel and Le Gruenwald – A Survey of Data Mining and Knowledge Discovery Software Tools, ACM, 1999
5. Neelam adhabPadhy, Dr. Pragnyaban Mishra, “The Survey of Data Mining Applications and Feature Scope”, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), PP. 43-58 Vol.2, No.3, June 2012.
6. Sundaravaradan, Naren, Manish Marwah, Amip Shah, and Naren Ramakrishnan. "Data mining approaches for life cycle assessment." In Sustainable Systems and Technology (ISSST), 2011 IEEE International Symposium on, pp. 1-6. IEEE, 2011.
7. Manoj and Jatinder Singh, “Applications of Data Mining for Intrusion Detection”, International Journal of Educational Planning & Administration. Volume 1, Number 1 (2011), pp. 37-42
8. M. Rajalakshmi, M. Sakti, “Max-Miner Algorithm Using Knowledge Discovery Process in Data Mining”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 11, November 2015
9. Smriti Srivastava & Anchal Garg, “Data Mining For Credit Card Risk Analysis: A Review”, International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), Vol. 3, Issue 2, Jun 2013, 193-200
10. Dipti Verma and Rakesh Nashine, “Data Mining: Next Generation Challenges and Future Directions”, International Journal of Modeling and Optimization, Vol. 2, No. 5, October 2012
11. Gaurav Varshney, Manoj Misra and Pradeep K. Atrey, “A survey and classification of web phishing detection schemes”, SECURITY AND COMMUNICATION NETWORKS, Security Comm. Networks 2016; 9:6266–6284, Copyright © 2016 John Wiley & Sons, Ltd
12. Hassan Y. A. Abutaira, Abdeljelil Belghitha, “Using Case-Based Reasoning for Phishing Detection”, Procedia Computer Science 109C (2017) 281–288, 2017 The Authors Published by Elsevier B.V.
13. Rakesh Verma, Avisha Das, “What’s in a URL: Fast Feature Extraction and Malicious URL Detection”, IWSPA ’17, March 24-24 2017, Scottsdale, AZ, USA, c 2017 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-4909-3/17/03.

Index Terms
Keywords

classification, rule based classification, association rule mining, phishing URLs, FP-Tree;