Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning
Vol. 4, Issue 2 – 2020
ISSN: 2581-5156

SVM Block Based Neural Learning Technique for Identification of Fraudulent Web Pages

S. Balasubramanian[1], S.Pratheep[1], R.Rajmohan[2], T. Ananth Kumar[3], M.Pavithra[3]

[1]Scholar, Department of Computer Science and Engineering, IFET College of Engineering, Gangarampalayam, Villupuram, India.
[2]Associate Professor, Department of Computer Science and Engineering, IFET College of Engineering, Gangarampalayam, Villupuram, India.
[3]Assistant Professor, Department of Computer Science and Engineering, IFET College of Engineering, Gangarampalayam, Villupuram, India.

Abstract

Due to rapid growth in internet there are more malicious webpages are grown, So it is important to detect and control such malicious webpages to avoid problems in Social Networks, Net-Banking, Business and so on. The dangers of these websites have created a demand for safeguards that protect end-users from visiting them. In an existing technique they detect webpage using uniform resource locator (URL) by using random forest and decision trees to detect and categorize malicious websites automatically. The cookies, HTTP Headers, Lock Icon, Privacy Policy, SSL Certification, Badge Verification, URL Query Strings are utilized for malicious webpages classification using (Support Vector Machine) SVM classifier.

Keywords: Malicious webpage, URL detection, SVM Classifier, Machine Learning.

Introduction

In the today's world Internet is most common and important to everyone's life and it became more useful to do their work. It can used in different domains such as Net-Banking, E-Commerce, Business, Social Networks and Finance. Cybercrime attacks also include online frauds in transaction, malware URL attacks, data's theft and so on. Almost all cybercrimes will be in the context of websites and viruses. In terms of building an online system, it is a task of extreme intensity, as the Internet is an increasingly important factor in what people want, and how they use it, which makes it difficult to design a comprehensive security measures. To spread your Internet links, you'll need a URL. URLs are multilateral designations for resource base such as papers, folders, web pages, or graphics on the internet. is used to explore both the knowledge present and neutral web site features and harmless web

application characterizations in a model that is trained on both current and non-neutral website functionalities.

Related Works

Much current neural network [1,4] innovation is employed in detection of malicious websites, based on features like news feed. While the neural networks are known to handle both malware and benign websites, they are yet to be refined for analysis and to gather URLs, images, and details for machine learning analysis. Due to the difficulties of distinguishing all articles in whole from processed products, only a number of programmatic plays focus on understanding points focus in play. When it can identify fraudulent websites very appropriately, it can use statistical models that span various attacks

Late research assembles factual models dependent on above highlights for characterizing URLs into pernicious and amiable class [14,15]. To evaluate malicious websites, Jianhua Liu et al . introduced a Markov decision process as well as a decision tree[1]and to detect security threats and vulnerabilities for website using JavaScript. Hung Le et.al [2] presented the Convolutional neural network methods for word level, character level and special character level are used to detect url to identify unique characters in the training corpus, and represent each character as a vector and learn the semantic and sequential patterns. Dharmaraj R.Patil et.al [3] presented Content based features for threat detection using HTML and Javascript to detect malicious URLs and the type of attacks based on multi-class classification.

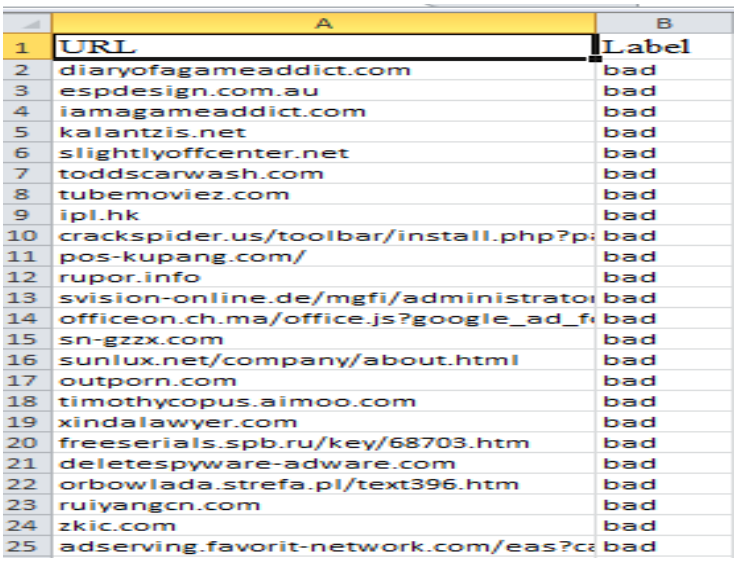
Proposed System

In Support Vector Machine it can analyze their data for classification and prediction. Contrasted and the Markov identification tree model, in the proposed framework, It has fewer hyper-parameters to tune in the conceptual methodology, and it can deal with discrete highlights of pages. [5], and accordingly may give superb execution to the incorporated discovery task. Separating variability optimises the width while resulting decisions are less sensitive to slight shifts in feature vectors. Distinct and separate information is assigned into a multidimensional space. It can generate many classification models. Researchers are looking to use URL assessment for identification rather than weight to help in the prediction of malevolent articles. If attribute points can be derived from multiple categories, a hyperplane is employed to aid in machine learning models. Creative phrase one classi URL classification technique does a mixture of blacklists and authenticates connections to determine phishing Websites.

Implementation

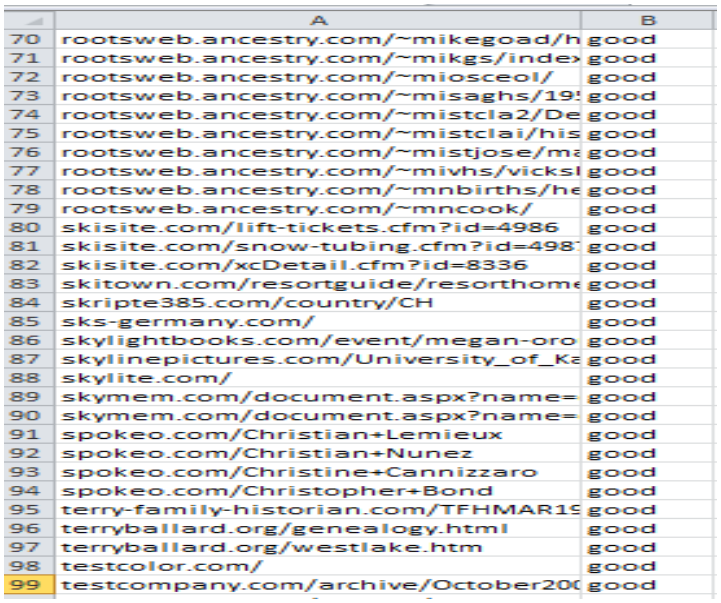
Separating variability optimizes the width while resulting decisions are less sensitive to slight shifts in feature vectors. Distinct and separate information is assigned into a multidimensional space.

Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning
Vol. 4, Issue 2 – 2020
ISSN: 2581-5156



	A	B
1	URL	Label
2	diaryofagameaddict.com	bad
3	espdesign.com.au	bad
4	iamagameaddict.com	bad
5	kalantzis.net	bad
6	slightlyoffcenter.net	bad
7	toddscarwash.com	bad
8	tubemoviez.com	bad
9	ipl.hk	bad
10	crackspider.us/toolbar/install.php?p	bad
11	pos-kupang.com/	bad
12	rupor.info	bad
13	svision-online.de/mgfi/administrato	bad
14	officeon.ch.ma/office.js?google_ad_f	bad
15	sn-gzzx.com	bad
16	sunlux.net/company/about.html	bad
17	outporn.com	bad
18	timothycopus.aimoo.com	bad
19	xindalawyer.com	bad
20	freeserials.spb.ru/key/68703.htm	bad
21	deletespyware-adware.com	bad
22	orbowlada.strefa.pl/text396.htm	bad
23	ruiyangcn.com	bad
24	zkic.com	bad
25	adserving.favorit-network.com/eas?ca	bad

Figure 1.URL list



	A	B
70	rootsweb.ancestry.com/~mikegoad/h	good
71	rootsweb.ancestry.com/~mikgs/inde	good
72	rootsweb.ancestry.com/~miosceol/	good
73	rootsweb.ancestry.com/~misaghs/19	good
74	rootsweb.ancestry.com/~mistcla2/De	good
75	rootsweb.ancestry.com/~mistclai/his	good
76	rootsweb.ancestry.com/~mistjose/ma	good
77	rootsweb.ancestry.com/~mivhs/vicksl	good
78	rootsweb.ancestry.com/~mnbirths/he	good
79	rootsweb.ancestry.com/~mncook/	good
80	skisite.com/lift-tickets.cfm?id=4986	good
81	skisite.com/snow-tubing.cfm?id=498	good
82	skisite.com/xcDetail.cfm?id=8336	good
83	skitown.com/resortguide/resorthome	good
84	skripte385.com/country/CH	good
85	sks-germany.com/	good
86	skylightbooks.com/event/megan-oro	good
87	skylinepictures.com/University_of_Ka	good
88	skylite.com/	good
89	skymem.com/document.aspx?name=	good
90	skymem.com/document.aspx?name=	good
91	spokeo.com/Christian+Lemieux	good
92	spokeo.com/Christian+Nunez	good
93	spokeo.com/Christine+Cannizzaro	good
94	spokeo.com/Christopher+Bond	good
95	terry-family-historian.com/TFHMAR19	good
96	terryballard.org/genealogy.html	good
97	terryballard.org/westlake.htm	good
98	testcolor.com/	good
99	testcompany.com/archive/October20(	good

Figure 2.Classifier list

In this above table, it is demonstrated the great URLS used in the dataset. It can generate many classification models. Researchers are looking to use URL assessment for identification rather than weight to help in the prediction of malevolent articles. If attribute points can be derived from multiple categories, a hyperplane is employed to aid in machine learning models. Creative phrase One classi URL classification technique does a mixture of blacklists and authenticate connections to determine phishing Websites.

Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning
Vol. 4, Issue 2 – 2020
ISSN: 2581-5156

SVM Algorithm

In Support Vector Machine [23] it can analyze their data for classification and prediction. Contrasted and the Markov location tree model, in the proposed framework it has less hyper-parameters to tune and can deal with discrete highlights of site pages [15, 25], and subsequently may give superb execution to the unified discovery task. Since the hyper plane significantly increases the spacing and the reports highlight boundaries remain the same, SVM is a more appropriate choice for short extracted features. When the data does not fit into a linear model, they partition the dataset with an activation functions into a Multi-dimensional space and extract the relevant features.

Application of URL Detection

The URL discovery assists with distinguishing phish ID, phish URL, phish detail URL, accommodation and furthermore it decreases remote correspondence cost [9,10] and transmitter pre-handling cost[11,13]. URL phishing also useful in load balancing HTTP requests across several content servers using Application Request Routing [19].

Results & Discussion

Accuracy of URL

Based on the URLS they can be checked with datasets and shown their accuracy in the window.

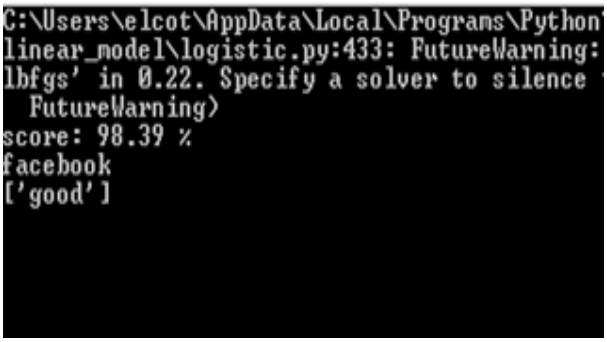


Figure 3.URL status

Here the URL status can be shown with their status in which it can predict their webpages are good or bad. In this they can insist the users to avoid such maliciousssplit data into train and test data and they prevent from their attacking websites.

Webpage Status

The webpage can show their status about their current webpage.

Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning
Vol. 4, Issue 2 – 2020
ISSN: 2581-5156

Figure 4.Webpage model

In order to maintain quality performance, this specific prolonged learning will be necessary.

Conclusion

This paper has proposed an Support vector machine methodology to detect malicious webpages using cookies, URL query strings, HTTP headers etc. By transforming the data to a high-dimensional space, SVM classifiers can utilize a kernel function to separate the dataset into subsets and assign attributes to each subset. Also, malware-hunting webpages may enhance your overall site's performance, as well as giving good results. The paper outlines some substantive issues for use in the object model and then raises some other questions that require further study.

References

1. Liu, Jianhua, Mengda Xu, Xin Wang, Shigen Shen, and Minglu Li. "A Markov detection tree-based centralized scheme to automatically identify malicious webpages on cloud platforms." IEEE Access 6 (2018): 74025-74038.
2. Le, Hung, Quang Pham, Doyen Sahoo, and Steven CH Hoi. "URLNet: Learning a URL representation with deep learning for malicious URL detection." arXiv preprint arXiv: 1802.03162 (2018).
3. “Feature-based Malicious URL and Attack Type Detection Using Multi-class Classification” Dharmaraj R. Patil, and Jayantrao B. Patil 1 1Department of Computer Engineering, R. C. Patel Institute of Technology, Shirpur-425405, India.
4. “Malicious URL Detection using Machine Learning” Doyen Sahoo, Chenghao Liu and Steven C.H. Hoi, India.
5. Big Data: Deep Learning for detecting Malware” Emmanuel Masabo and Kyanda Swaib Kaawaase, Makerere University Kampala, Uganda.
6. Ahmed, A. Ali, and Nik QuosthoniSunaidi. "Malicious Website Detection: A Review." J. Forensic Sci. Crim. Investig 7, no. 3 (2018): 1-4.

7. Hou, Yung-Tsung, Yimeng Chang, Tsuhan Chen, Chi-Sung Laih, and Chia-Mei Chen. "Malicious web content detection by machine learning." expert systems with applications 37, no. 1 (2010): 55-60.
8. Jayasri, K., R. Rajmohan, and D. Dinagaran. "Analyzing the query performances of description logic based service matching using Hadoop." In 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 1-7. IEEE, 2015.
9. Vishvaksenan K.S. and Rajmohan R 2017 Performance analysis of multi-carrier IDMA system for co-operative networks Cluster Computing https://doi.org/10.1007/s10586-017-1186-8.
10. Vishvaksenan, K. S., R. Rajmohan, and R. Kalaiarasan. "Multi-carrier IDMA system for relay aided cooperative downlink communication with transmitter preprocessing." In 2017 International Conference on Communication and Signal Processing (ICCSP), pp. 2206-2210. IEEE, 2017.
11. Rajmohan, R., K. S. Vishvaksenan, M. Mira, and Sudharssun Subramanian. "Performance of a turbo-coded downlink IDMA system using transmitter pre-processing." Computers & Electrical Engineering 53 (2016): 385-393.
12. D. Jayakumar, R.Rajmohan, D.Saravanan and MO. Ramkumar, 2019, Detection of Bacterial Contamination and Ph Quantity Using Digitalization Strategy, Journal of Physics: Conference Series, Volume 1362. https://iopscience.iop.org/article/10.1088/1742-6596/1362/1/012081/meta
13. Rajmohan, R., K. S. Vishvaksenan, and Anjana Krishnan. "Cooperative downlink Multi-Carrier IDMA system using Transmitter Preprocessing." In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1800-1803. IEEE, 2016.
14. Nadanam, Padmapriya, and R. Rajmohan. "QoS evaluation for web services in cloud computing." In 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), pp. 1-8. IEEE, 2012.
15. Rajmohan, R., Padmapriya, N, A domain ontology based service matching for CHORD based super peer network, Proceedings- International Conference on Data Science and Engineering, ICDSE 2012.
16. Padmapriya, N., Rajmohan, R, Reliability evaluation suite for cloud services, 3rd International Conference on Computing, Communication and Networking Technologies, ICCCNT 2012.
17. D Jayakumar, R Rajmohan, D Saravanan, Detection of Bacterial Contamination and Ph Quantity Using Digitalization Strategy, Journal of Physics: Conference Series, 2019.
18. Anuprabhavathi, G., and R. Rajmohan. "Energy-efficient and cost-effective resource provisioning framework for map reduce workloads using dcc algorithm." International Journal of Engineering Science Invention Research & Development 2, no. 9 (2016): 623-628.

19. Rajmohan, R., M. Pajany, R. Rajesh, D. Raghu Raman, and U. Prabu. "Smart paddy crop disease identification and management using deep convolution neural network and SVM classifier." International journal of pure and applied mathematics 118, no. 15 (2018): 255-264.
20. T. Ananth Kumar and R. S. Rajesh, "Towards power efficient wireless NoC router for SOC," 2014 International Conference on Communication and Network Technologies, Sivakasi, India, 2014, pp. 254-259, doi: 10.1109/CNT.2014.7062765.
21. Design and Development of an Efficient Branch Predictor for an In-order RISC-V Processor [Текст] / C. Arul Rathi, G. Rajakumar, T. Ananth Kumar, T.S. Arun Samuel // Журналнано- таелектронноїфізики.–2020.–Т. 12, № 5.–05021.–DOI: 10.21272/jnep. 12(5).05021.
22. S. A. Selvi, T. A. kumar, R. S. Rajesh and M. A. T. Ajisha, "An Efficient Communication Scheme for Wi-Li-Fi Network Framework," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019, pp. 697-701, doi: 10.1109/I-SMAC47947.2019.9032650.
23. Ananth kumar, T., Arun Samuel, T. S., Praveen kumar, P., Pavithra, M., & Raj Mohan, R. (2021). LIFI-Based Radiation-Free Monitoring and Transmission Device for Hospitals/ Public Places. In Tyagi, A. K. (Ed.), Multimedia and Sensory Input for Augmented, Mixed, and Virtual Reality (pp. 195-205). IGI Global. http://doi: 10.4018/978-1-7998-4703-8.ch010
24. Adimoolam M., John A., Balamurugan N.M., Ananth Kumar T. (2021) Green ICT Communication, Networking and Data Processing. In: Balusamy B., Chilamkurti N., Kadry S. (eds) Green Computing in Smart Cities: Simulation and Techniques. Green Energy and Technology. Springer, Cham. https://doi.org/10.1007/978-3-030-48141-4_6
25. John A., Ananth Kumar T., Adimoolam M., Blessy A. (2021) Energy Management and Monitoring Using IoT with CupCarbon Platform. In: Balusamy B., Chilamkurti N., Kadry S. (eds) Green Computing in Smart Cities: Simulation and Techniques. Green Energy and Technology. Springer, Cham. https://doi.org/10.1007/978-3-030-48141-4_10.
26. S. Devadharshini, R. Kalaipriya, R. Rajmohan, M. Pavithra and T. Ananthkumar, "Performance Investigation of Hybrid YOLO-VGG16 Based Ship Detection Framework Using SAR Images," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN49426.2020.9262440.
27. S. Narmadha, S. Gokulan, M. Pavithra, R. Rajmohan and T. Ananthkumar, "Determination of various Deep Learning Parameters to Predict Heart Disease for Diabetes Patients," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN 49426.2020.9262317.
28. Gopalakrishnan, P. Manju Bala and T. Ananth Kumar, "An Advanced Bio-Inspired Shortest Path Routing Algorithm for SDN Controller over VANET," 2020 International

Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262276.

29. R. Kalaipriya, S. Devadharshini, R. Rajmohan, M. Pavithra and T. Ananthkumar, "Certain Investigations on Leveraging Blockchain Technology for Developing Electronic Health Records," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-5, doi: 10.1109/ ICSCAN49426.2020.9262391.
30. S. Gokulan, S. Narmadha, M. Pavithra, R. Rajmohan and T. Ananthkumar, "Determination of Various Deep Learning Parameter for Sleep Disorder," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN49426. 2020.926 2331.
31. R. K. Gajalakshmi, T. Ananthkumar, P. Manjubala and R. Rajmohan, "An Optimized ASM based Routing Algorithm for Cognitive Radio Networks," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN49426.2020.9262397.