WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

IMPLEMENTATION OF MALWARE PREDICTION BASED ON INFECTION RATE

Gawade J. S*[1], Dhairyashil Jadhav*[2], Omkar Ghadge*[3], Swapnil Mhaske*[4], Vivek Kale*[5]

*[*1]Prof. HOD Dept. of Information Technology, SVPM, MALEGOAN, Maharashtra, India*
*[*2,3,4,5]Student, Dept. of Information Technology, SVPM, MALEGAON, Maharashtra, India*

ABSTRACT

When users input their passwords in a public place, they may be at risk of attackers stealing their password. In this modern, technological age, the internet has been accepted by the masses. And with it, the danger of malicious attacks by cybercriminals or hackers have increased. These attacks are done via Malware, and have resulted in billions of dollars of financial damage. This makes the prevention of malicious attacks an essential part of the battle against cybercrime. In this paper we applying machine learning algorithms to predict the malware we use supervised machine learning algorithms to predict the malware. In this we use support vector machine algorithm. SVM is basically classification algorithm. We have collected a publicly available dataset on kaggle website. The entire project is divided into two parts training and testing.

***Keywords*:** *Support vector machine, Machine learning, Malware.*

1. INTRODUCTION

Malware is software created to infect the machine. It can affect the computer without knowing it to user. The main reason of a malware could include accessing private networks, stealing sensitive information,taking over the computer system to make a use of it and disturb the whole communication things.It is a software that injected in such a way that and most of them have ability to spread itself,remain undetectable and due to this there is change in system or damage or system is infested.common types of malware includes worms,adware ,spyware, Trojan or Trojan horse prevention of malicious attacks an essential

part of the battle against cybecrime we are applying machine learning algorithms to predict the malware infection rates for these its features. For classification purposes we use Support Vector Machine algorithm(SVM). It is done by taking into account the different properties of the machines, especially ones related to the machine's malware protection status, and user habits.

2. IMPLEMENTATION

A. DFD DIAGRAMS:

Data Flow Diagram of Level 0

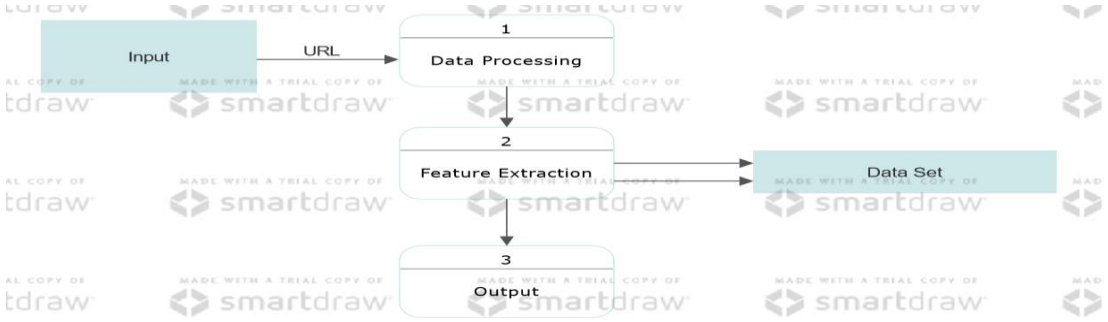


FIG. DATA FLOW DIAGRAM LEVEL 1

B. TRAINING:

In training session we take input as url's from kaggle website and upload it for data processing then on basis of some parametres we classify them and we extract the features of all of them. We get output by doing all this steps.

C. PARAMETERS OF CLASSIFICATION:

1) Ranking

2) islp

3) valid

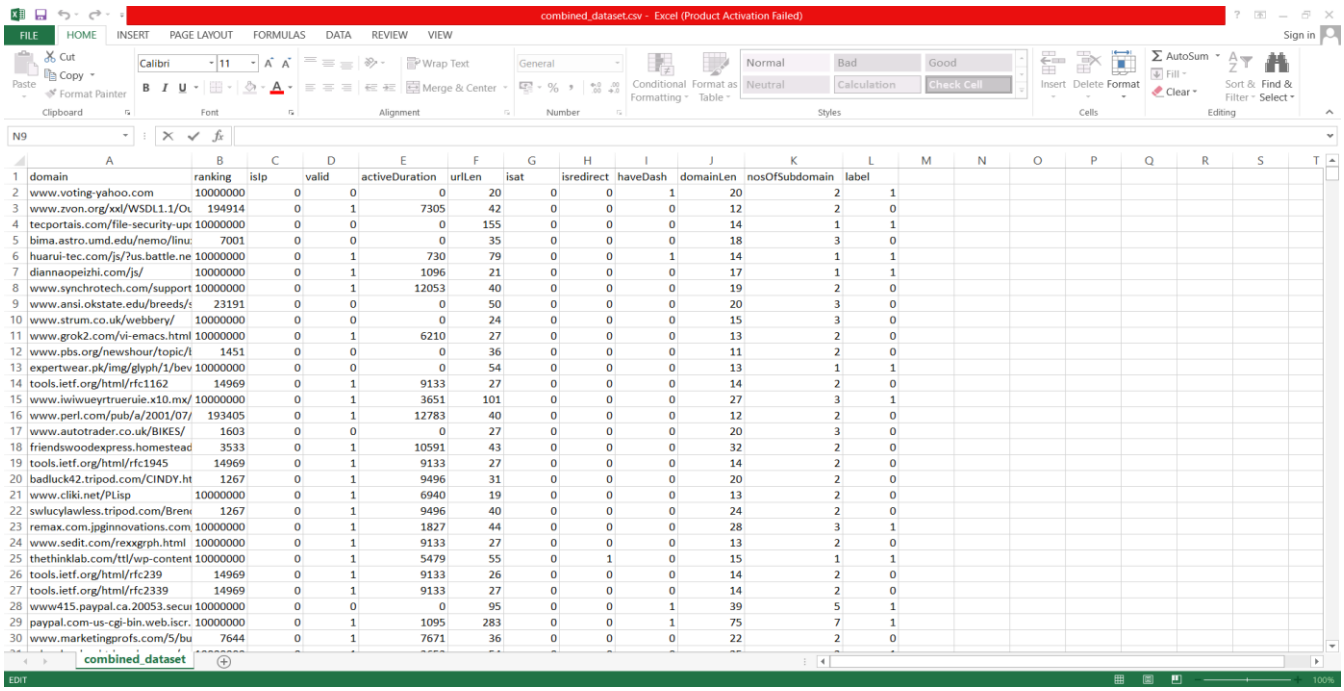
4) Active duration

5) Urllen

6) Havedash

7) Domain legnth

D. INPUT DATASET:



combined_dataset.csv - Excel (Product Activation Failed)

	domain	ranking	islp	valid	activeDuration	urlLen	isat	isredirect	haveDash	domainLen	nosOfSubdomain	label
2	www.voting-yahoo.com	10000000	0	0	0	20	0	0	1	20	2	1
3	www.zvon.org/xxl/WSDL1.1/Ou	194914	0	1	7305	42	0	0	0	12	2	0
4	tecportais.com/file-security-upd	10000000	0	0	0	155	0	0	0	14	1	1
5	bima.astro.umd.edu/nemo/linu	7001	0	0	0	35	0	0	0	18	3	0
6	huarui-tec.com/js/?us.battle.ne	10000000	0	1	730	79	0	0	1	14	1	1
7	diannaopeizhi.com/js/	10000000	0	1	1096	21	0	0	0	17	1	1
8	www.synchrotech.com/support	10000000	0	1	12053	40	0	0	0	19	2	0
9	www.ansi.okstate.edu/breeds/s	23191	0	0	0	50	0	0	0	20	3	0
10	www.strum.co.uk/webbery/	10000000	0	0	0	24	0	0	0	15	3	0
11	www.grok2.com/vi-emacs.html	10000000	0	1	6210	27	0	0	0	13	2	0
12	www.pbs.org/newshour/topic/l	1451	0	0	0	36	0	0	0	11	2	0
13	expertwear.pk/img/glyph/1/bev	10000000	0	0	0	54	0	0	0	13	1	1
14	tools.ietf.org/html/rfc1162	14969	0	1	9133	27	0	0	0	14	2	0
15	www.iwiwueyrtrueruie.x10.mx/	10000000	0	1	3651	101	0	0	0	27	3	1
16	www.perl.com/pub/a/2001/07/	193405	0	1	12783	40	0	0	0	12	2	0
17	www.autotrader.co.uk/BIKES/	1603	0	0	0	27	0	0	0	20	3	0
18	friendswoodexpress.homestead	3533	0	1	10591	43	0	0	0	32	2	0
19	tools.ietf.org/html/rfc1945	14969	0	1	9133	27	0	0	0	14	2	0
20	badluck42.tripod.com/CINDY.ht	1267	0	1	9496	31	0	0	0	20	2	0
21	www.cliki.net/PLisp	10000000	0	1	6940	19	0	0	0	13	2	0
22	swlucylawless.tripod.com/Bren	1267	0	1	9496	40	0	0	0	24	2	0
23	remax.com.jpginnovations.com	10000000	0	1	1827	44	0	0	0	28	3	1
24	www.sedit.com/rexxgrph.html	10000000	0	1	9133	27	0	0	0	13	2	0
25	thethinklab.com/ttl/wp-content	10000000	0	1	5479	55	0	1	0	15	1	1
26	tools.ietf.org/html/rfc239	14969	0	1	9133	26	0	0	0	14	2	0
27	tools.ietf.org/html/rfc2339	14969	0	1	9133	27	0	0	0	14	2	0
28	www415.paypal.ca.20053.secu	10000000	0	0	0	95	0	0	1	39	5	1
29	paypal.com-us-cgi-bin.web.iscr	10000000	0	1	1095	283	0	0	1	75	7	1
30	www.marketingprofs.com/5/bu	7644	0	1	7671	36	0	0	0	22	2	0

FIG. DATASET

3. TESTING

For testing purposes we created a GUI where we can test the all acitive url's.

OUTPUT 1: MALWARE DETECTED IN GIVEN URL:

FIG. MALWARE DETECTION SYSTEM

OUTPUT 2: MALWARE NOT DETECTED IN GIVEN URL:

FIG. MALWARE DETECTION SYSTEM

4. CONCLUSION

Using this malware detection system we definitely know that which website is safe and which is not safe to open. Using this system we avoid millions of financial damage also avoid taking over computer system also stealing sensitive information. It has a great potential to be implemented in future.

REFERENCES

[1] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., "Lightgbm: A highly efficient gradient boosting decision tree." Advances in Neural Information Processing Systems, 2017. pp 3146-3154 .

[2] M. Heiderich, M. Niemietz, F. Schuster, T. Holz, and J. Schwenk, "Scriptless attacks: stealing the pie without touching the sill," in Proc. 2019.