BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 206-213
DOI: 10.2478/bjlp-2022-004021

Analysis of Malware Detection Using Naive Bayes Algorithm Comparing Support Vector Machine Algorithm.

Borra Madhan Mohan Reddy
Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India: 602105.

P.Sriramya
Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering,Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India: 602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: To enhance the accuracy in Detection of Malware in Analysis of Novel Malware Detection Using Naive Bayes Algorithm comparing Support Vector Machine Algorithm. **Materials and Methods:** This study contains two groups one is the Novel Naive Bayes Algorithm comparing Support Vector Machine Algorithm. Each group consists of a sample size of 30. Their accuracies are compared with each other using different sample sizes also. **Results:** SPSS was used to calculate the sample size. The pre-test analysis was maintained at 80%. G-power is used to calculate sample size. The Support Vector Machine Algorithm is 64.1% more accurate than the Naive Bayes Algorithm of 62.8% in classifying the malware Detection. There is a statistically insignificant distinction in accuracy for 2-Algorithms is $p>0.05$ by performing independent samples t-tests which is 0.206. **Conclusion:** Through this, prediction is done for The Naive Bayes Algorithm is significantly better than the Support Vector Machine (SVM) in identifying Malware detection. It can be also considered as a better option for the classification of malware detection.

Keywords

Malware Detection, Machine Learning, Cloud Storage, Novel Malware Analysis, Support Vector Machine, Naive Bayes Algorithm.

INTRODUCTION

Network safety is an essential factor of computing that protects the safety of data kept on PCs connected by a single network. Information on the internet has recently become commonplace in our lives (El-Khouly and El-Seoud 2017). Every day, the number of network attackers grows, and the risks they pose evolve as well. For foundations such as universities, unusual ventures, and businesses, network security is a critical memory. Many critical capacities for the nation's security can be delivered through these institutions (Le and Markopoulou 2010). Clients these days are particularly interested in web-based services. Clients may now speak with one another and share data and information with one another. By employing Information Technology (IT) affiliations and Internet-Service-Providers, these services are now less costly and more useful (ISPs). Malware has the

potential to put networks at danger. Malware is an application software that may bring digital devices into a network, such as computers, laptops, advanced mobile phones, and tablets (Dorf 2018). It affects these devices by misdirecting them and erasing their own knowledge and data; for example: Adware may wish to do the nefarious task. Malware is the most dangerous threat to networks. To carry out its attack, the virus may take on a variety of shapes. It returns as a bundle deal and checks out to access the network on a regular basis. New types and forms of malware are discovered on a daily basis. Virus programmers are always making decisions on how to shield their malware against anti-malware applications such as Kaspersky, McAfee, and NOD (Termanini 2018).

According to an increasing number of computer virus malware within networks these days, it has evolved into a big menace to our machines. Network attackers created the worms specifically to carry out these assaults (Chen 2007). A well-designed system model is essential to fight these assaults and prevent them from proliferating and spreading throughout the network, inflicting harm to our computers. In this work, we designed a detection system model for this topic (Termanini 2018). The planned framework recognises the worm malware using data from a dataset obtained from the Kaspersky organization website; the framework will obtain the data bundle and then dissect it; the Nave Bayesian arrangement procedure will then begin to work and begin to characterize the bundle; by utilizing the data mining Naive Bayesian order method, the framework worked quickly and produced excellent results in recognising the worm. The Naive Bayesian classification approach, which employs probability mathematical equations for both danger and benign data, identifies malware and classifies data as threat or benign (Xi-cheng et al. 2018). The studies found that using the proposed dataset enhances the effectiveness and effectiveness of detecting worm malware by 95% worm detection accuracy and 98 percent detection rate with 21% false positives, indicating that using the proposed dataset enhances the effectiveness and effectiveness of detecting worm malware (Pooryousef and Amini 2017).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The literature review revealed a research gap: classification algorithms based on Naive Bayes are not adequate for handling huge datasets. When the dataset has more target classes with greater noise and overlaps, it performs poorly. Naive Bayes will underperform and have low accuracy in these circumstances wherein the volume of highlights for each record factor surpasses the vloume of getting ready statistics tests. The goal of this study is to use the Naive Bayes Algorithm to identify cyberbullying and enhance classification accuracy by comparing it to the Support Vector Machine Algorithm to decrease false detection.

MATERIALS AND METHODS

The research work was performed in the Data analytics Lab in the Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences. The sample size taken for conducting the experiment was 10. Two groups are considered as classifiers algorithms in order to classify prediction of fare amount, machine learning classification algorithms are used. The work was carried out on 100000 records from a data-master dataset (Joslin 2010). The accuracy in classifying the malware was performed by evaluating two groups. A total of 10 iterations were performed on each group to achieve better accuracy. The Study uses a dataset-master image dataset downloaded from kaggle. Fig. 1 shows the architecture of attack classification.

Naive Bayes (NB) Algorithm

NB is a probabilistic machine learning algorithm that can be utilized in a wide assortment of grouping tasks. The name naive is utilized on the grounds that it accepts the provisions that go into the model are free of one another. Equation (1) gives the numerically calculation formula of Bayesian calculation is addressing a class variable and the arrangement of qualities are, Conditional probability of A given B can be registered as:

$$P(A \mid B) = P(A \cap B) / P(B) \quad (1)$$

Support Vector Machine (SVM) Algorithm

SVM is one of the maximum widely known Supervised Learning calculations, that's applied for Classification as appropriately as Regression issues. Be that as it may, primarily, used for Classification problems in ML. We offer a novel method for detecting malware. An approach proposes an SVM based malware detection system with the capacity to categorize and prevent harmful programmes. The suggested method contains steps for learning and detection.

The steps in the learning stage are as follows:

1. Knowledge development based on factors that might indicate the existence of mobile malware.
2. Knowledge about mobile malware activity is presented as a number of feature vectors.
3. Using SVM, create a collection of mobile malware classes.

The following steps make up the monitoring stage:

1. Collecting characteristics on the mobile device that may indicate the presence of malware attacks.
2. Using the information acquired, construct the feature vectors.
3. In the detecting step, the collected feature vectors are classified using SVM so it is possible to allocate them to one of the malicious programs classes.
4. Stopping the harmful application from running.

Statistical Analysis

The SPSS (Statistical Package for the Social Sciences) statistical software was used in the research for statistical analysis. Group statistics and independent sample t-tests were performed on the experimental results and the graph was built for two groups with two parameters under study. Support vectors are the data points that lie closest to the decision surface (or hyperplane) They are the data points most difficult to classify. They have direct bearing on the optimum location of the decision surface (Bhosale, Ade, and Deshmukh 2014).

RESULTS

The proposed algorithm Naive Bayes and existing algorithm Support Vector Machine algorithm were run at a time in an Anaconda-Jupyter. As the sample sets are executed for a number of iterations the accuracy values of Naive Bayes and Support Vector Machine Algorithm classifiers vary for the classification of accuracy as shown in Table 1. Analysis of the overall classification of Detection of Malware in Analysis of Malware Detection Using Naive Bayes Algorithm comparing Support Vector Machine Algorithm models shows the classification of the detecting malware. Naive Bayes (62.7%) shows better accuracy than Support Vector Machine (64.1%). Statistical Analysis of Mean, Standard deviation and Standard Error and Accuracy of Naive Bayes and Support Vector Machine Algorithm is done and shown in table 2. There is a statistically significant difference in Accuracy values between the algorithms. Support Vector Machine had obtained higher accuracy compared to Naive Bayes as shown in Fig. 2. There is a statistical significant difference in accuracy for two algorithms is $p>0.05$ by performing independent

samples t-tests which is 0.206 and hence it is insignificant. Hence the statistical analysis is done and tabulated in table 3.

DISCUSSION

The Naive Bayes and Support Vector Machine Algorithm classifiers on a dataset acquired from diverse sources like Kaggle, Github, et al. are compared during this section. After completing preprocessing and extraction on the dataset, the dataset was separated into portions for training and testing. The accuracy is calculated using both NB and SVM Algorithm. Surprisingly, the Naive Bayes outperformed the SVM in every way. The accuracy of a classifier is critical in determining the efficacy of Detection of Novel Cloud Malware in Cloud storage to reduce false detection. There is a statistical significant difference in accuracy for two algorithms is $p>0.05$ by performing independent samples t-tests which is 0.206 and hence it is insignificant. Hence the statistical analysis is done.

The shares of correct predictions divided by the whole number of guesses is known as accuracy. We evaluated the accuracy of each machine learning technique to figure out which was the foremost effective. We used sci-kits sklearn Metrics accuracy score to calculate the classifier accuracy for NB and SVM algorithm (Jin and Zhan 2008). From the database, the algorithm will get matched Analysis of Novel Cloud Malware Detection, also as basic profile information about. These findings are being provided to an interface that will display and populate a machine learning algorithm that finds and formalizes the ideas that underlie the information it sees (Simangunsong, Zarlis, and Tulus 2019). Despite the actual fact that the presented methodology yielded good results, the approach's shortcoming is that it needs to be enhanced to reduce false detection of malware. This may be avoided in the future by combining Naive Bayes with other approaches (Mishra and Maheshwary 2017).

There are restrictions with different cloud environments in identifying malware and furthermore proposes a cloud-based malware recognition structure, which utilizes a hybrid way to deal with recognizing malware. Cloud malware investigation apparatuses are growing better than ever features, which could possibly address such vulnerabilities. These findings are being provided to an interface that will display and populate a machine learning algorithm that finds and formalizes the ideas that underlie the information it sees. Despite the actual fact that the presented methodology yielded good results, the approach's shortcoming is that it needs to be enhanced to reduce false detection of malware. This may be avoided in the future by combining Naive Bayes with other approaches.

CONCLUSION

The studies on prediction are completed using the device getting to machine learning algorithms. Naive Bayes Algorithms comparing Support Vector Machine are giving the accuracy of 62.8%, 64.1% accuracy separately. The studies can be in addition prolonged with diverse datasets and diverse attributes for the ensemble of the device getting to know algorithms.

DECLARATION

Conflicts of Interest

The author declares no conflict of Interest.

Authors Contributions

Author BMM was involved in data collection, data analysis, manuscript writing. Author PSR was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. Qbec Infosol Pvt. Ltd., Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Bhosale, Dipali, Roshani Ade, and P. R. Deshmukh. 2014. "Feature Selection Based Classification Using Naive Bayes, J48 and Support Vector Machine." *International Journal of Computer Applications*. https://doi.org/10.5120/17456-8202.

Chen, Zesheng. 2007. *Modeling and Defending Against Internet Worm Attacks*.

Dorf, Richard C. 2018. "Computers, Software Engineering, and Digital Devices." https://doi.org/10.1201/9781315220659.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Jin, Zhan, and J. I. N. Zhan. 2008. "Spam Message Self-Adaptive Filtering System Based on Naive Bayes and Support Vector Machine." *Journal of Computer Applications*. https://doi.org/10.3724/sp.j.1087.2008.00714.

Joslin, Ann. 2010. "Regional Fare Policy and Fare Allocation, Innovations in Fare Equipment and Data Collection." https://doi.org/10.5038/cutr-nctr-rr-2006-04.

Le, Anh, and Athina Markopoulou. 2010. "Locating Byzantine Attackers in Intra-Session Network Coding Using SpaceMac." *2010 IEEE International Symposium on Network Coding (NetCod)*. https://doi.org/10.1109/netcod.2010.5487673.

Mishra, Ashish, and Preeti Maheshwary. 2017. "A Novel Technique for Fingerprint Classification Based on Naive Bayes Classifier and Support Vector Machine." *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2017914806.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Pooryousef, Shahrooz, and Morteza Amini. 2017. "Enhancing Accuracy of Android Malware Detection Using Intent Instrumentation." *Proceedings of the 3rd International Conference on Information Systems Security and Privacy*. https://doi.org/10.5220/0006195803800388.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Simangunsong, Juanto, Muhammad Zarlis, and Tulus. 2019. "Analysis of Algorithms Support Vector Machine with Naive Bayes Kernel in Data Classification." *Proceedings of the International Conference on Natural Resources and Technology*. https://doi.org/10.5220/0008553302870291.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

Termanini, Rocky Dr. 2018. "DDoS Malware: The Curse of Virus Rain™." *The Nano Age of Digital Immunity Infrastructure Fundamentals and Applications*. https://doi.org/10.1201/9781315167404-5.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Xi-cheng, Chen, Bing-Bing Fan, Wen-Jia Yang, Gui-Long Tian, and Kai Wu. 2018. "Air-Attack Weapon Identification Model of Weighted Navie Bayes Based on SOA." *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1060/1/012054.

TABLES AND FIGURES

Table 1. Comparison of Accuracy and Sensitivity achieved during the evaluation of Naive Bayes and Support Vector Machine (SVM) models for classification with different iterations.

Iterations	Naive Bayes	Support Vector Machine
1	62%	64.1%
2	62.5%	64.0%
3	61.5%	64.5%
4	62.3%	63.7%
5	61.9%	64.8%

Table 2. Statistical Analysis of Mean, Standard deviation and Standard Error of and Sensitivity of Naive Bayes and Support Vector Machine(SVM). There is a Statistically significant difference in Accuracy values in the algorithms. Support Vector Machine had the highest Accuracy (64%) and Sensitivity (62%) compared with Naive Bayes. The Standard error is also less in Naive Bayes in comparison to Support Vector Machine (SVM).

GROUP	N	Mean	Std. Deviation	Std.Error Mean
Accuracy Naive Bayes	5	62.000	.79057	.35355
SVM	5	64.200	.40000	.17889

Table 3. Comparison of the significance level for Naive Bayes and Support Vector Machine algorithms with value p = 0.001. Both Naive Bayes and Support Vector Machine have a significance level less than 0.05 in terms of accuracy with a 95% confidence interval.

	Levene's Test for Equality of Variances		T-test for Equality of means						
Accuracy	F	Sig.	t	df	Sig(2-tailed)	Mean Difference	Std. Error Difference	95% confidence interval of the Difference	
								Lower	Upper
Equal variances assumed	1.894	.206	-5.552	8	.001	-2.20000	.39623	-3.11371	-1.28629
Equal variances not assumed			-5.552	5.922	.002	-2.20000	.39623	-3.17265	-1.22735

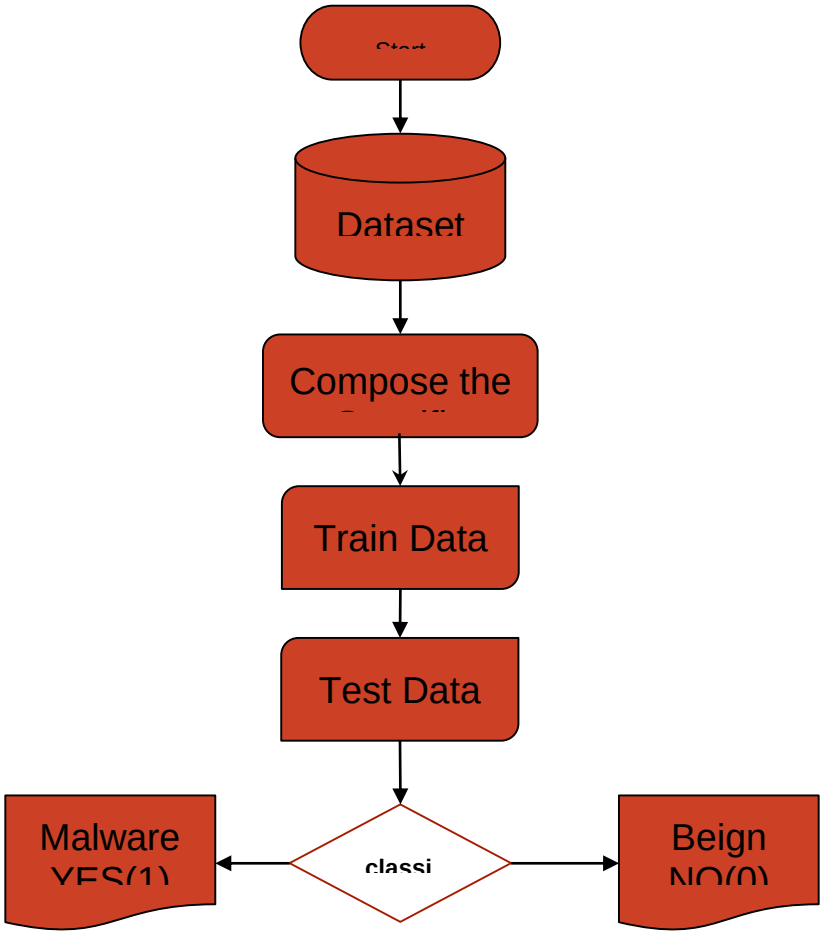

Fig. 1. Architecture Diagram Malware Analysis

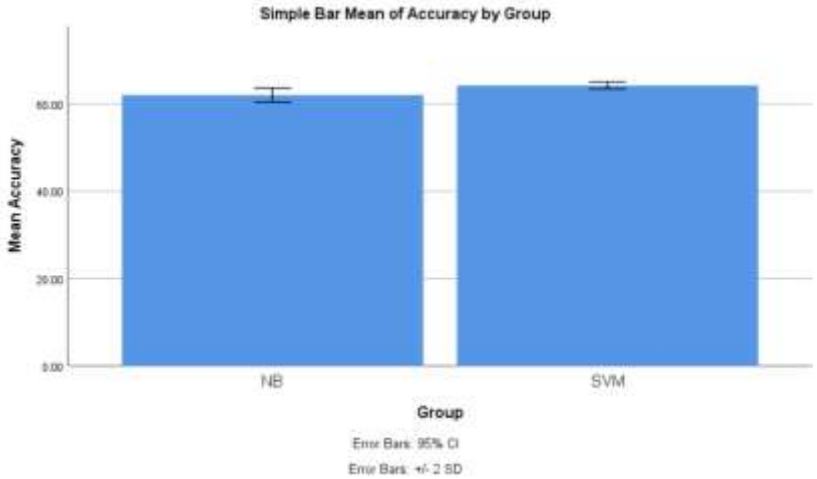

Fig. 2. Comparison of mean sensitivity of NB and SVM algorithm. The standard errors appear to be less in Support Vector Machine compared to Naive Bayes. Support Vector Machine Algorithm appears to produce more consistent results with higher sensitivity. X-Axis: Support Vector Machine vs Naive Bayes Algorithm. Y-Axis: Mean sensitivity of detection +/- 2 SD,Error Bars 95% CI.