DETECTION OF MALWARE USING MACHINE LEARNING ALGORITHMS

Tahir Naquash1, Yusuf Ahmad2, Satyam Singh3, Shubhanshu Kumar4, Yash Anjana5
1Professor, 2,3,4,5Student
Department of CSE, HKBKCE, Bangalore, India
*Corresponding Author
E-mail Id :- shubhanshubb@gmail.com

INTRODUCTION

Malware is a type of cyberattack that can take many forms and is designed to harm computers, users, businesses, or computer systems. Malicious software is typically installed and run without the user's knowledge or consent, and can include viruses, Trojan horses, ransomware, spyware, adware, rogue software, wipers, scareware, and other types of threats.

This study highlights the potential of machine learning algorithms to detect harmful traffic on computer systems and improve the security of computer networks. The proposed approach involved using malware analysis and detection with machine learning algorithms to compute the difference in correlation symmetry integrals. Several classification algorithms, including Random Forest, AdaBoost, GNB, Gradient Boosting, DT, and the proposed approach, were evaluated in terms of their detection accuracy rates.

PROPOSED SYSTEM:

Malicious components of malware can be detected through static analysis, which involves analyzing the malware binaries to identify harmful strings, or dynamic analysis, which involves monitoring the software as it operates in a controlled environment. While both methods have their pros and cons, it's best to use both when analyzing malware. To improve the accuracy of malware detection, it may be helpful to reduce the number of dangerous features and focus on more robust characteristics. This process begins with identifying potential methods or algorithms for feature selection. Ideally, solutions should be able to detect previously unseen malware while reducing the number of required characteristics.

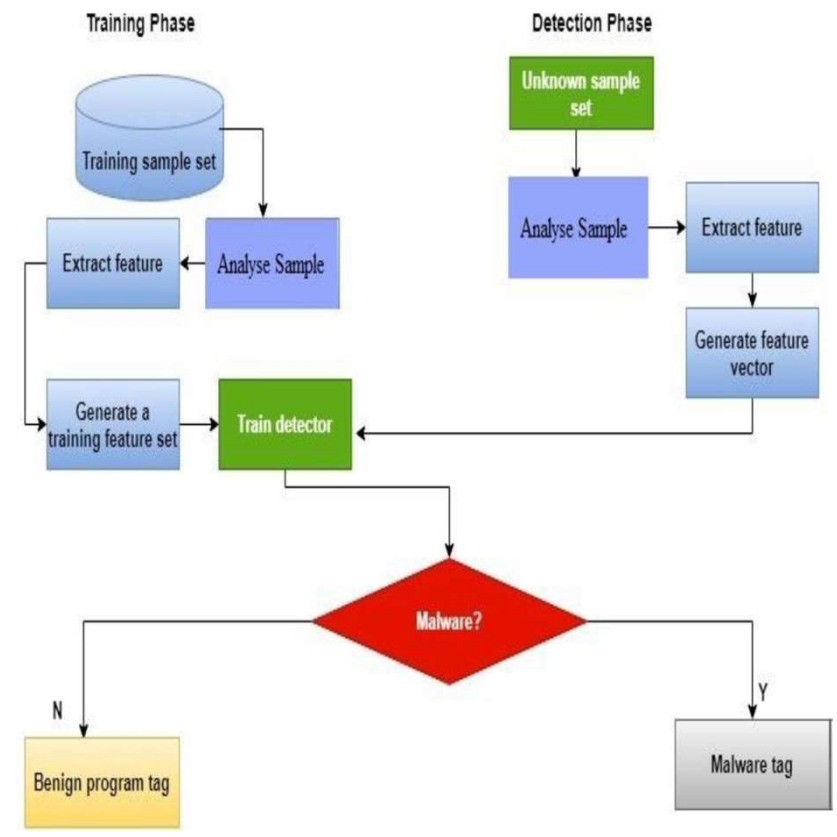


Figure Proposed ML malware detection method.

METHODOLOGY

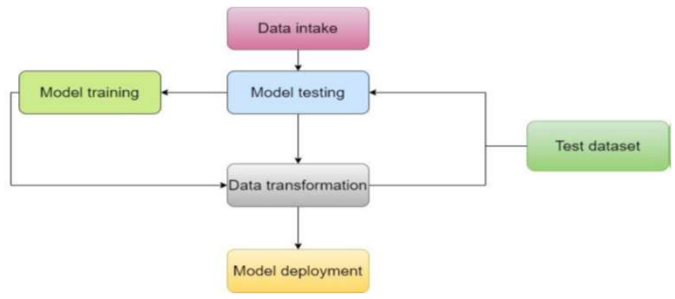
This introduces the different steps and elements of a typical machine learning workflow for malware detection and classification, investigates the difficulties and constraints of such a workflow, and evaluates the most recent advancements and trends in the field, with a focus on deep learning methods. Below is a description of the planned research approach for this study.

DATASET

PRE-PROCESSING

FEATURES EXTRACTION

FEATURES SELECTION



RESULTS:

Training and testing were the two key stages of the classification procedure. Both dangerous and safe files were supplied to a system to teach it. With each batch of data it annotated, each classifier (RF, AdaBoost, GNB, GB, or DT) got more intelligent. A classifier was provided a set of fresh files during testing, some of which were hazardous and some of which were not. The classifier evaluated whether the files were malicious or not. The second-best model for malware detection is the convolutional neural network (RF) with an accuracy of 98.76%, followed by support vector machine (GB) with an accuracy of 98.41%. In terms of true positive rate (TPR), RANDOM FOREST had the highest rate of 99.22%, followed by DT with a rate of 99.07%, and GB with a rate of 98%.

CONCLUSIONS

The study presented a protective mechanism that evaluated four ML algorithm approaches to malware detection and selected the most appropriate one. The RF ML method had the highest accuracy (99.42%) among all the classifiers evaluated. In addition to potentially providing the highest detection accuracy and accurately characterizing malware, static analysis based on PE information and carefully selected data showed promise in experimental findings. The significant benefit of this approach is that it does not require executing anything to determine if data are malicious. Overall, the study highlights the importance of using machine learning techniques in malware detection and characterisation. The results demonstrate that DT, RF, and GB algorithms can effectively identify dangerous versus benign data, and static analysis based on PE information and carefully selected data can improve malware detection accuracy.

REFERENCES

1 Deshmukh V.M. Performance evaluation of machine learning classifiers in malware detection.
2. Akhtar, M.S.; Feng, T. IOTA based anomaly detection machine learning in mobile sensing
3. Bera, P.; Patra, P.K. A novel machine learning based malware detection and classification framework.