RESEARCH ARTICLE OPEN ACCESS

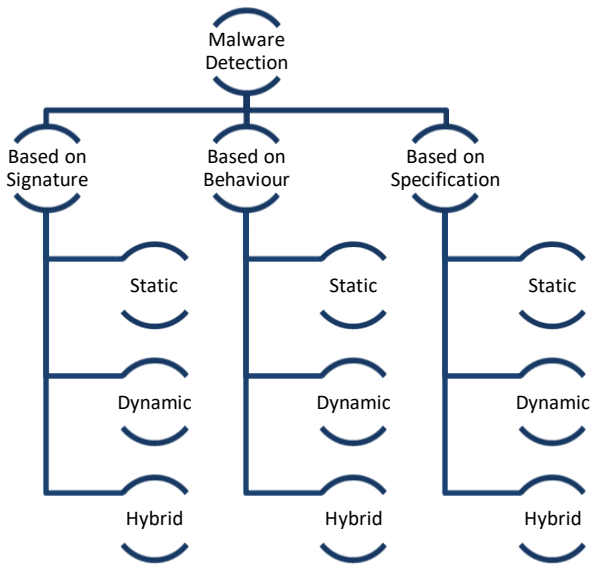
Multilayer Perception Classifier for Malware Detection using Machine Learning Algorithm

Dr D.J. Samatha Naidu, P. Balaji
Principal, Annamacharya PG College of Computer Studies,Rajampet.
Dept..,of MCA, Annamacharya PG College of Computer Studies,Rajampet.

Abstract
Mobile phones are a significant component of people's life and are progressively engaged in these technologies. Increasing customer numbers encourages the hackers to make malware. In addition, the security of sensitive data is regarded lightly on mobile devices. Based on current approaches, recent malware changes fast and thus become more difficult to detect. An alternative solution to detect malware using anomaly-based classifier is proposed.
Keywords — detection malware, cybersecurity, machine learning, Multilayer perception, classifier, analysis, memory forensics.

Introduction: The breakthrough in internet technology and computer networking have made high speed shared internet possible. The effect of this development is the daily increase in the number of computer systems that have become susceptible to malware attacks. The innovation has made the internet a huge storehouse where resources are virtualized and utilized to the need of users. Despite the immense benefits that the internet revolution has brought, there are numerous challenges that it also poses to the security of computer systems. The conventional computer system is entirely centred on a single host machine running operating system, while several machines connected to the host are running on the guest operating system. The prevalent security threat confronting the users is the attack on a computer system by malicious programs which spread to other computers that have not been infected.With an ever increasing demand of smart phones, the mobile phone market is expanding at an exponential pace. With such a boom in the smart-phone industry, there is a need to realize the holistic review of the brand and the model of phone. There are numerous brands present in the market, out of which some are dominant and occupy quite a big part of the industry. For instance, Samsung, Apple, etc. are names associated with brands which are famous throughout the world. Electronic commerce plays a vital role in increasing the sales of the mobile phones and influencing consumer buying patterns. Through both static and dynamic analysis, malware can be detected by the following:

- Behavior Analysis
- Without executing/running the code



II Proposed Work

In this project we are using various machine learning algorithms such as SVM, KNN and ANN to detect malware attack and to train machine learning algorithms we have downloaded malware dataset from link Dataset link: https://www.dropbox.com/s/ep8qjakfwh1rzk4/malimg_dataset.zip?dl=0 and this dataset contains more than 20 attacks from different malware family

III Related Work

Malware Detection using conventional methods is incompetent to detect new and generic malware. For the investigation of a variety of malware, there were no ready-made machine learning datasets available for malware detection. So we generated our dataset by downloading a variety of malware files from the world's famous malware projects. By performing unstructured data collection from the downloaded APK files and feature mining process the final dataset was generated with 16300 records and a total of 215 features. There was a need to evaluate the performance of the generated dataset with supervised machine learning classifiers. So in this paper, we propose a malware detection approach using different supervised machine learning classifiers. Here supervised algorithms, Feature Reduction Techniques, and Ensembling techniques are used to evaluate the performance of the generated dataset. Machine Learning classifiers are evaluated on the evaluation parameters like AUC, FPR, TPR, Cohen Kappa Score, Precision, and Accuracy. We also represented the results of classifiers using Bar plots of Accuracy and plotting the ROC curve. From the results of machine learning classifiers, the performance of the CatBoost Classifier is highest with Accuracy 93.15% having a value of ROC curve as 0.91 and Cohen Kappa Score as 81.56%.

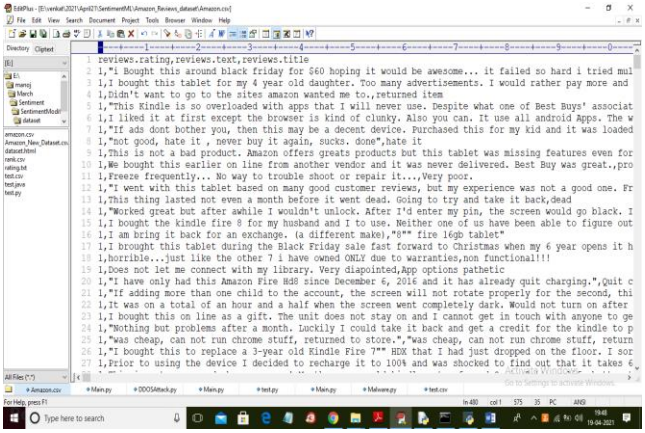
IV Future Scope

The limitations and the future scope of the Data as mentioned above Analytics technologies in light of optimal decision undertaking within organizational context while considering the various risk management approaches fostering enhanced consumer experience and improved innovation and development practices within industries have been duly considered.

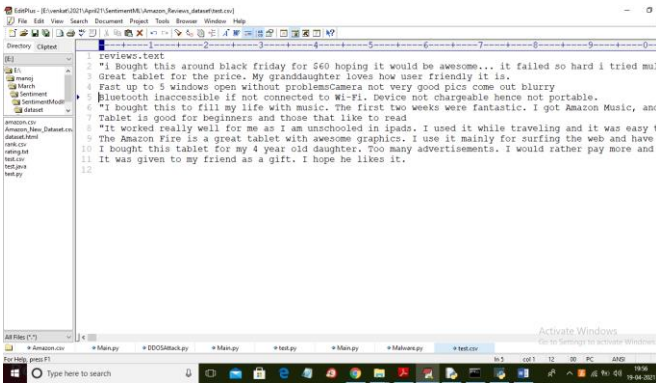
V Conclusion

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. Our research is aiming to achieve this by conducting sentiment analysis of mobile phone reviews and classifying the reviews into positive and negative sentiment. After balancing the data with almost equal ratio of positive and negative reviews, three classification models have been used to classify reviews. Out of the three classifiers, i.e., Naïve Bayes, SVM and Decision Tree, predictive accuracy of SVM is found to be the best. The accuracy results have been cross validated and the highest value of

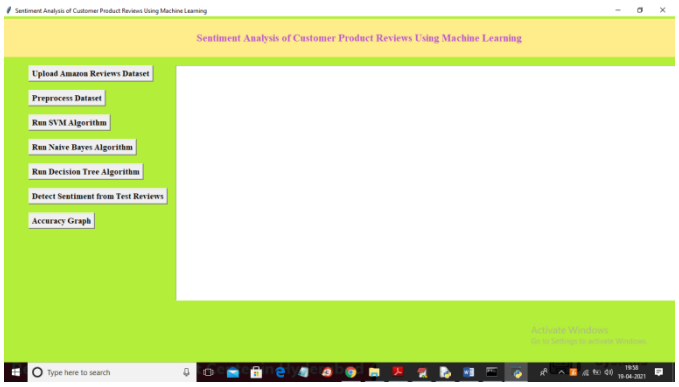
accuracy achieved was 81.75% for SVM among the three models.



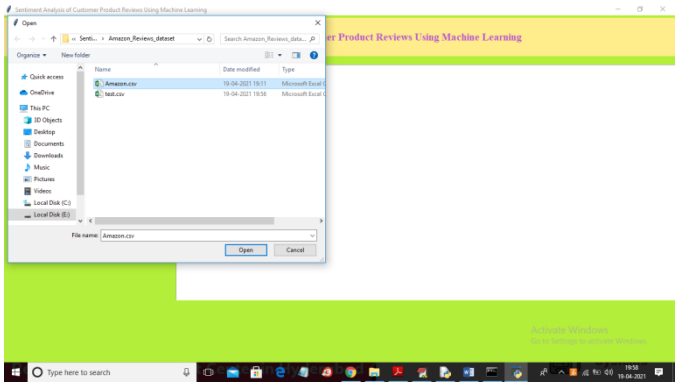
Screen 1:In above screen click on 'Upload Dataset' button to upload dataset



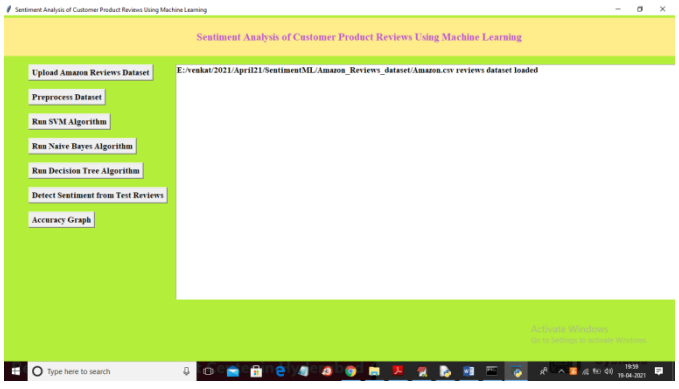
Screen 2: In above test data we have only test reviews and by applying ML trained model on above test data we can predict sentiment label



Screen 3: In above screen click on 'Upload Amazon Reviews Dataset' button to upload dataset



Screen 4: In above screen we are selecting and uploading 'Amazon.csv' file and then click on 'Open' button to load dataset and to get below screen



Screen 5 :In above screen dataset loaded and now click on 'Preprocess Dataset' button to read all reviews from dataset and then apply Preprocess steps to get below screen

References Books

[1]. Sanjay Chakrabortya and Lopamudra Dey. A rule-based probabilistic technique for malware code detection. Multiagent and Grid Systems – An International Journal, IOS Press, 12, 2016, pp. 271–286 271. DOI 10.3233/MGS-160254

[2]. Y. Zhou, Z. Wang, W. Zhou, and X. Jiang. Hey, you, get off of my market: Detecting malicious apps in official and alternative android markets. in NDSS, vol. 25, no. 4, 2012, pp. 50–52.

[3]. D. Keragala. Detecting malware and sandbox evasion techniques, SANS Institute InfoSec Reading Room, 2016. URL: https://www.sans.org/reading-room/whitepapers/ forensics/detecting-malware-sandbox-evasion-techniques-36667.

[4]. Sharif, M., Yegneswaran, V., Saidi, H., Porras, P., and Lee, W. Eureka: A framework for enabling static malware analysis. In Computer security-ESORICS 2008, pages 481- 500. Springer.

[5]. Moser, A., Kruegel, C., and Kirda, E. Limits of static analysis for malware detection. In Computer security applications conference, ACSAC 2007. Twenty-third annual, 2007, pages 421-430.