iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54648

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Adaboost Model-Based Approach for Effectively Detecting Spam in IoT Devices

Vulugundam Anitha[1], Donthula Aashritha[2], Gunda Rushika[3], Gaddam Meghana[4]
[1]*Assistant Professor, Dept. of ETE, G.Narayanamma Institute of Technology & Science (For Women), Hyderabad, Telangana, India*
[2, 3, 4]*B.Tech IV year students, Dept. of ETE, G.Narayanamma Institute of Technology & Science (For Women), Hyderabad, Telangana, India*

Abstract: ***The number of Internet of Things (IoT) devices in smart homes is rising swiftly, producing enormous amounts of data that are mostly transmitted over wireless communication channels. IoT devices can be at risk from a variety of threats, such as hacker attacks, cyberattacks, erratic network connectivity, data leakage, etc. By evaluating vast amounts of data using complex algorithms, machine learning may assist uncover spam in IoT data. It can also help to raise the security level of the IoT system in smart homes by utilising statistical analysis and machine learning to find anomalies in the data. In this work, two machine learning models—the Bagged model and the Adaboost model—are evaluated using a wide range of criteria employing a vast number of input feature sets. Each model generates a spam score using the improved input attributes. The suggested algorithm is used to determine the network's linked IoT devices' spamicity score. The REFIT Smart Home dataset is used to test the suggested method. The outcomes show that the suggested strategy is beneficial when compared to other current plans.***
Keywords: IoT, Cyberattack, Adaboost, Bagged model, spamicity, REFIT.

I. INTRODUCTION

Regardless of their geographical locations, The Internet of Things (IoT) enables the fusion and integration of physical objects. Implementing these network management and control mechanisms makes privacy and protection measures crucial and difficult in this setting. As shown in Fig. 1, to address security problems including eavesdropping, spam, spoofing attacks, jamming, invasions, DoS attacks, and malware, IoT applications must secure user data privacy [1-5]. The size and kind of the organization where IoT security measures are implemented determines how effective they will be. Users' actions compel the security gateways to collaborate. The smart organization's IoT security cameras, for instance, may record many parameters for analysis and wise decision-making. The greatest amount of caution should be used with web-based devices because they make up the majority of IoT devices. IoT devices placed in an organization are frequently utilized in the workplace to effectively integrate security and privacy features. For instance, wearable technology that gathers and sends user health data to a linked smartphone should guard against data leaks to preserve privacy. Both consumers and attackers are drawn to the IoT because of its increasing nature. However, IoT devices decide on a defense strategy and the important parameters in the security protocols for the trade-off between security, privacy, and computation as ML emerges in various attack scenarios. It is difficult for an IoT system with limited resources to estimate the current network and timely attack state, thus this job is complex [6-13].

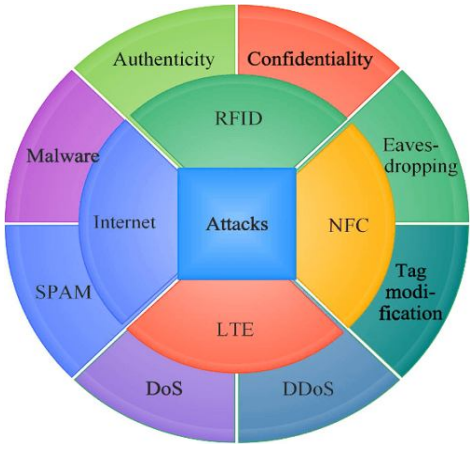


Fig. 1: Protocols with potential threats [1]

Network security has been significantly improved using a variety of machine learning techniques, including supervised learning, unsupervised learning, and reinforcement learning. In order to conserve energy and increase the lifespan of IoT systems, machine learning approaches aid in the development of protocols for light-weight access control. There are numerous tools and computational methods available for protecting IoT devices with unrestricted resources. On the other hand, there are few options for securing IoT devices with limited resources, and these options provide significant hurdles in terms of cost, performance, and service expenses. Cognitive safe shield with machine learning capabilities protects dairy IoT devices using limited resources [14]. Using Secure Shield Framework architecture, it improves the security posture of our Dairy IoT device without reducing its Usable Life (ULD). It is also possible to generate highly accurate unsupervised learning models with smaller feature set sizes, allowing for a reduction in the necessary processing resources. Another design alternative that is being considered for resource optimization is training a single common model for all IoT devices rather than a separate model for each item [15]. Supervised ML can be used to accurately identify unauthorized IoT devices. Seventeen IoT devices representing nine different device types were used to collect and manually label network traffic data for the purpose of training and evaluating a multi-class classifier [16].

II. IMPLEMENTATION & RESULTS

Bagging, also known as Bootstrap aggregating, is an effective ensemble learning technique that improves the performance and accuracy of machine learning algorithms. It serves as a valuable method to handle the trade-off between bias and variance by reducing the variance of predictive models. Bagging is particularly suitable for decision tree algorithms and prevents overfitting of data. Parameters that can improve the swiftness and precision of the model in line with the given information.

Adaptive Boosting, also known as AdaBoost, is an Ensemble Method utilized in Machine Learning. The AdaBoost classifier creates a classifier on the initial dataset, then creates more copies of the classifier on the same dataset with weights that have been changed for cases that were incorrectly classified. The flowchart is displayed in Fig. 2.

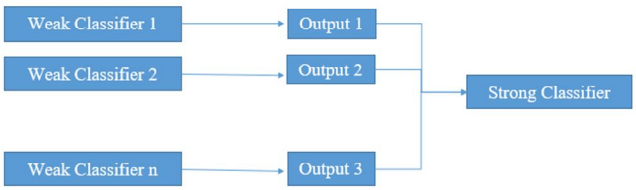


Fig. 2: Flow chart of Adaboost model

The main parameters that have significant impact are base_estimator, n_estimators, and learning_curve. Adaboost model is shown in Fig. 3.

1) base_estimator: The learner used to train the model is a weak one. The weak learner used by default for training is the Decision Tree Classifier. You can, however, specify other machine learning techniques.
2) n_estimators: This parameter indicates the number of weak learners that are trained iteratively.
3) predict: The data to be tested is typically the only argument required by the predict() function. It returns the labels of the data supplied as an argument based on the trained data retrieved from the model.
4) learning_curve: To fit the model with a reasonable bias-variance trade-off, the right amount of training data should be determined with the use of a learning curve.

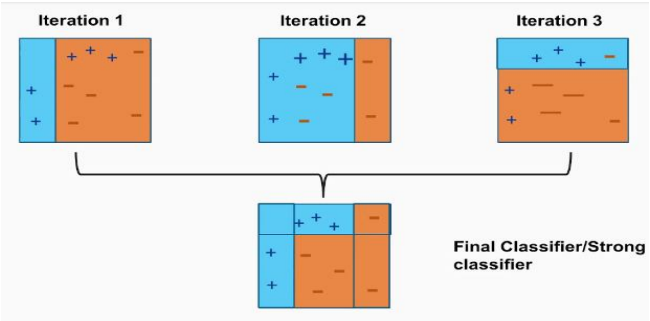


Fig. 3: Adaboost model

To Implement Bagged model and Adaboost Model, Python is used. To run Python code, Visual Studio Code (IDE) is used. The suggested method finds the spam parameters that are affecting IoT devices. To get the best results, the suggested approach is validated using the IoT dataset. In the context of spam identification in IoT devices, Principal Component Analysis (PCA) can be helpful. Due to its capacity to reduce dimensionality, extract pertinent features, identify abnormalities and facilitate data visualization, PCA is crucial for spam detection in IoT devices. By using PCA, spam detection algorithms may be made more reliable, efficient, and capable of handling the enormous volumes of data that IoT devices produce. This will increase the security and dependability of these devices by allowing them to detect and stop spam or other unwanted behavior.

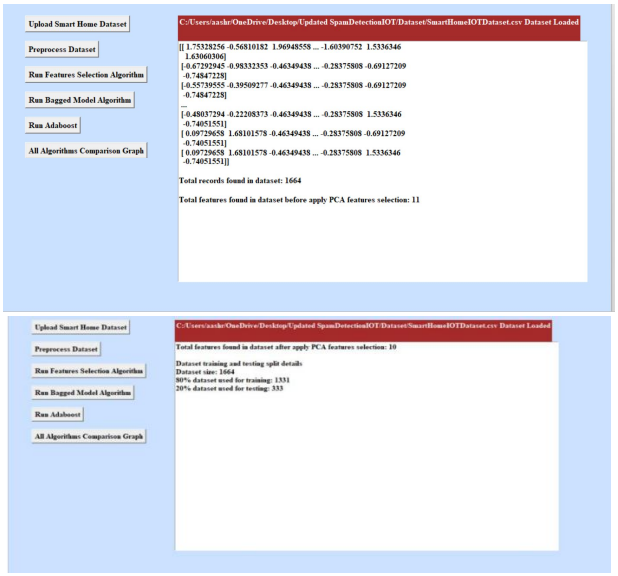


Fig. 4: PCA feature selection

Spamicity Score defines the amount of non-spam percentage present in the data. As shown in Fig. 4, after applying PCA, the number of features was reduced to 11, and there were 1664 records in the dataset. Of these, 1331 records were used for training and 29 records for testing. The train and test data were now ready, and you could select "Run Bagged Model" and “Adaboost Model” to proceed.

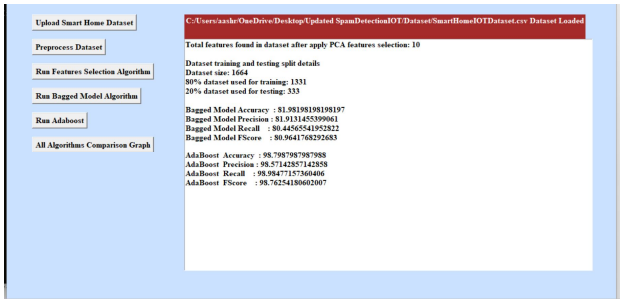


Fig. 5: Parameters of bagged model and adaboost model

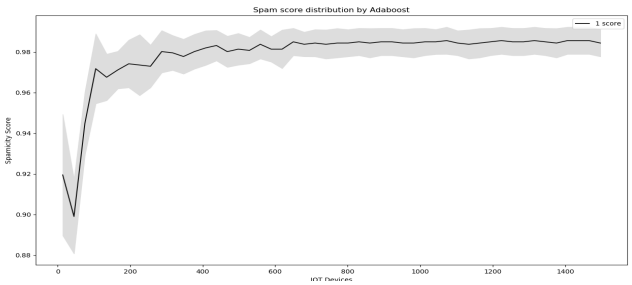


Fig. 6: Spam score distribution by Adaboost Model

From Figs. 5, 6 & 7, several metrics are used to evaluate and compare different models:

a) *Accuracy:* Accuracy is a measure used to assess the performance of a model in identifying relationships and patterns between variables based on the input or training data. A higher accuracy indicates better predictive capability.

b) *Precision:* Precision is a metric that measures the ability of a model to produce true positive predictions while minimizing false positives. A model with a precision of 1.0 indicates that it does not produce any false positive predictions.

c) *Recall:* A model's capacity to accurately identify positive samples, independent of the quantity of negative samples, is measured by recall, also known as sensitivity or true positive rate. It measures the completeness of the model's predictions for positive instances.

d) *F1-Score:* The F1-score is a machine learning evaluation metric that combines the precision and recall scores of a model. By taking into account both false positives and false negatives, it offers a fair assessment of a model's accuracy. In cases where the dataset is unbalanced, the F1-score is especially helpful.

The comparison graph uses these metrics to assess and compare the performance of different models. Additionally, the accuracy metric evaluates how many correct predictions a model made across the entire dataset, providing an overall measure of its predictive accuracy. Adaboost model is performing better in spam detection with respect to above said metrics.

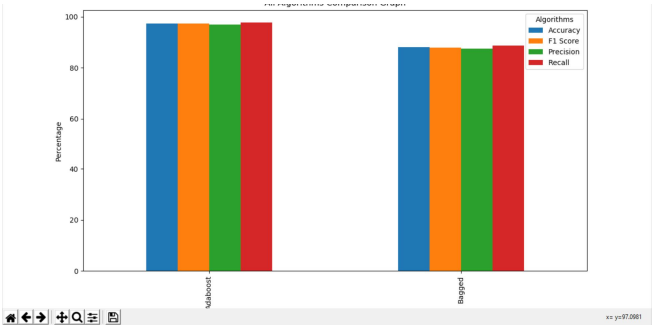


Fig. 7: Comparison graph

III. CONCLUSION

This paper focuses on utilizing the spamicity score to assess the reliability of IoT devices within a smart home network. The proposed algorithm is employed to calculate the spam score of each IoT device, and a framework is presented to detect spam parameters using machine learning models. Prior to conducting experiments, the IoT dataset undergoes pre-processing through feature engineering procedures. Different machine learning algorithms are used to analyse time-series data produced by smart metres after thorough testing. In the context of a smart home, ensemble methods are utilised to determine the contribution levels of IoT devices by giving them a spam score.

The outcomes show that the spamicity score helps to improve the prerequisites for IoT devices in a smart home to operate successfully. Each IoT equipment is given a spam score by applying the framework with machine learning models, improving the circumstances for their proper operation. The framework incorporates machine learning models to identify spam attributes of IoT devices, while the feature engineering approach is utilized for preliminary processing of the IoT dataset. This process results in a more specific set of requirements that must be met for IoT devices to function effectively in a smart home. Furthermore, future plans involve considering environmental and contextual factors to enhance the reliability and safety of IoT devices. Additionally, the extension of this work has yielded favorable outcomes compared to existing methods. The Bagged model achieved an accuracy of 85%, while Adaboost achieved 99% accuracy. Moving forward, there are intentions to incorporate climatic and surrounding features of IoT devices to further enhance their security and trustworthiness.

REFERENCES

[1] Dr. Aaisha Makkar, Dr. Neeraj Kumar, Prof. Ahmed Ghoneim, “An Efficient Spam Detection Technique for IoT Devices using Machine Learning”, IEEE Transactions on Industrial Informatics, 2021.

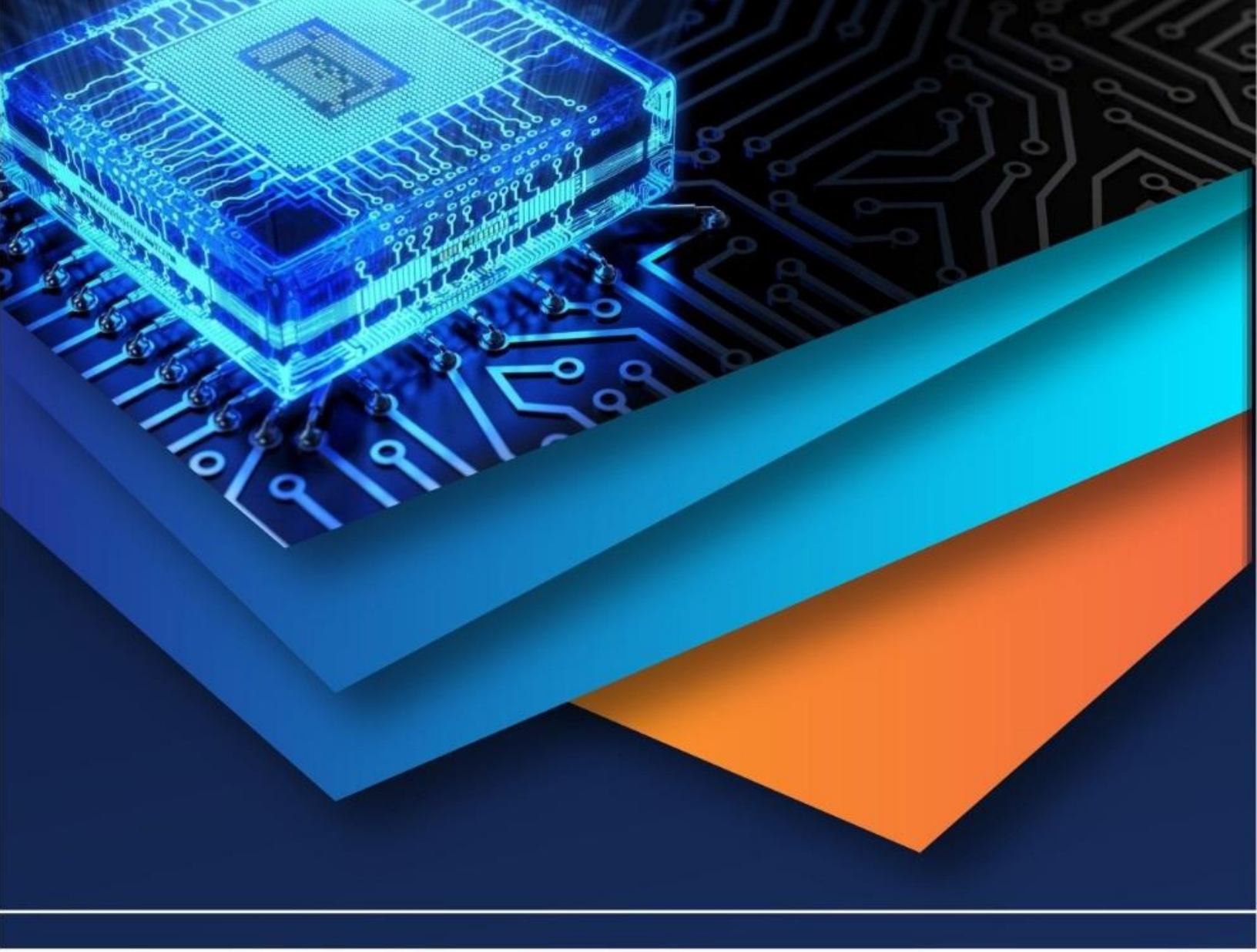
[2] W. Kim, O.-R. Jeong, C. Kim, and J. So, “The dark side of the internet: Attacks, costs and responses,” Information systems, vol. 36, no. 3, pp. 675–705, 2011.

[3] H. Eun, H. Lee, and H. Oh, “Conditional privacy preserving security protocol for nfc applications,” IEEE Transactions on Consumer Electronics, vol. 59, no. 1, pp. 153–160, 2013.

[4] R. V. Kulkarni and G. K. Venayagamoorthy, “Neural network based secure media access control protocol for wireless sensor networks,” in 2009 International Joint Conference on Neural Networks. IEEE, 2009, pp. 1680–1687.

[5] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, “A system for denial-of-service attack detection based on multivariate correlation analysis,” IEEE transactions on parallel and distributed systems, vol. 25, no. 2, pp. 447–456, 2013.

[6] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "Sinr-based dos attack on remote state estimation: A game-theoretic approach," IEEE Transactions on Control of Network Systems, vol. 4, no. 3, pp. 632–642, 2016.
[7] L. Xiao, Y. Li, X. Huang, and X. Du, "Cloud-based malware detection game for mobile devices with offloading," IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2742–2750, 2017.
[8] J. W. Branch, C. Giannella, B. Szymanski, R. Wolff, and H. Kargupta, "In-network outlier detection in wireless sensor networks," Knowledge and information systems, vol. 34, no. 1, pp. 23–54, 2013.
[9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of machine learning research, vol. 3, no. Mar, pp. 1157–1182, 2003.
[10] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in Proceedings of the 20th international conference on machine learning (ICML-03), 2003, pp. 856–863.
[11] A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque, "Artificial intelligence driven mechanism for edge computing based industrial applications," IEEE Transactions on Industrial Informatics, 2019.
[12] A. H. Sodhro, Z. Luo, G. H. Sodhro, M. Muzamal, J. J. Rodrigues, and V. H. C. de Albuquerque, "Artificial intelligence based qos optimization for multimedia communication in iov systems," Future Generation Computer Systems, vol. 95, pp. 667–680, 2019.
[13] E. Bertino and N. Islam, "Botnets and internet of things security,"Computer, no. 2, pp. 76–79, 2017.
[14] Jaya Shankar Vuppalapati, Santhosh Kedari, Anitha Ilapakurti, Chandrasekar Vuppalapati, "Cognitive Secure Shield – A Machine Learning enabled threat shield for resource constrained IoT Device", 17th IEEE International Conference on Machine Learning and Applications (ICMLA), USA, 2018.
[15] Sven Nomm, Hayretdin Bahsi "Unsupervised Anomaly Based Botnet Detection in IoT Networks", 17th IEEE International Conference on Machine Learning and Applications (ICMLA), USA, 2018.
[16] Yair Meidan, Michael Bohadana, Asaf Shabtai, Martin Ochoa, Nils Ole Tippenhauer, Juan Davis Guarnizo, Yuval Elovici, "Detection of Unauthorized IoT Devices Using Machine Learning Techniques", Computer Science, Cryptography and security, 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY