DOI: 10.56892/bima.v8i2B.716

An Advanced DDOS Attack Detection Model with an Ensembled SVM and Baruta Selection Technique

Auwal Adamu Ajiya[1], Fatima Umar Zambuk[2]*, Badamasi Imam Ya'u, Mukhtar Abdullahi and Hussaini Dan-azumi

[1]M.I.S Unit, Computer Science Department, Abubakar Tatari Ali Polytechnic, Bauchi, Bauchi State Nigeria.

[2]Department of Mathematical Science, Abubakar Tafawa Balewa University, Bauchi, Bauchi State Nigeria.

Corresponding Author: fuzambuk@atbu.edu.ng

ABSTRACT

The paper proposed the use of an ensembled SVM model with the Boruta selection technique to improve cloud DDoS attack detection. DDoS attacks are the most common cloud security attacks, with a 16% level of use. They can render the entire system useless, with resources offline for 24 hours, multiple days, or even a week depending on the severity of the attack. In the event of successful attacks, about $ 20,000 can be lost by a company. DDoS attacks can also make the cloud environment vulnerable to hacking, due to bad hosting or shared hosting, failure to prepare against the attack, outdated codes, and other issues. This study aims to improve the performance of Support Vector Machine (SVM) to better detect Cloud DDoS attacks by eliminating key problems and improving memory efficiency, effectiveness, and high dimensional space. Several Machine learning techniques like Decision Tree, Random Forest, KNN, and SVM were used to detect DDoS attacks in a cloud environment. In terms of detection accuracy SVM is the best among the used techniques with 84.94%. A proposed ensembled SVM with the Boruta selection technique was modeled to improve the performance of DDoS attack detection techniques in the cloud. Five different models were designed using distinct machine-learning techniques and compared to the proposed model for better performance. Logistic regression, Random Forest Classification, Support Vector Machine, K-Nearest Neighbor, and Linear Discriminant Analysis. All five Classifiers were used independently and with the Bagging technique, giving different results in all aspects. From their performance found that after the boruta selection extract 51 features out of the 79 original features of the and the data that was summed up to 1048575 was reduced to 1025 for optimal performance, Random Forest Classifier and K-Nearest Neighbor was said to perform better than the proposed SVM classifier in both Individual modeling and with Bagging Ensembled learning. A great improvement was achieved by the model performance with a detection accuracy of 95.7%, 10.8% more than the traditional SVM, an improvement the accuracy. The implementation of KNN, Random Forest, and Linear Discriminant analysis in ensembled learning shows that their performance is better than the proposed system.

Keywords: DDOS Attack, SVM

INTRODUCTION

Cloud computing has made significant progress in the computing world, providing simple and affordable infrastructure to many global economy organizations and contributing to economic development. However, security threats, particularly DDoS attacks, pose a significant challenge to this progress. To ensure the continuation of this development, an efficient mechanism is required to mitigate these threats, and machine learning and artificial intelligence are increasingly being used to achieve this. While different approaches have been discussed, there is still room for improvement in the performance of machine learning techniques. This study aims to modify, improve, and hybridize machine learning techniques to tackle the problem of DDoS attacks in the cloud environment and reduce their effects on users, industries, and economic growth. Cloud applications have experienced tremendous growth, with many companies investing in the cloud for affordable service delivery. However, DDoS attacks remain the most common security threat, causing significant downtime and financial losses. Machine learning approaches, particularly SVM, have been used to detect and prevent DDoS attacks in the cloud environment. Still, there are limitations to its effectiveness, including difficulties in kernel selection, poor performance in noisy datasets, and underperformance in datasets with more features than training samples. This study aims to address these limitations to improve the performance of SVM and enhance its advantages, such as memory efficiency and effectiveness in high-dimensional space, for better detection of DDoS attacks in the cloud environment(Kundu, 2022).

RELATED WORK

Several studies have been conducted and various models have been developed to detect DDoS attacks on cloud computing environments. Jyoti & Behal (2021) compared machine learning techniques such as BayesNet, NaiveBayes, J48, and Random Forest, and found that Random Forest had the best performance. Zekri et al. (2017) compared selection algorithms like Fuzzy logic, Ensemble-based multiplier, and SVM, and discovered that the SVM-based model performed better for classification. Alarqan et al. (2019) surveyed various techniques including machine learning, data mining, artificial intelligence, classifiers, and statistical based techniques, and found that the statistical-based technique performed better. Dong et al. (2019) used ensemble unsupervised machine learning techniques and concluded that One-class SVM was the best-performing technique. Sharma et al. (2019) compared various techniques including the Time series Model, Genetic Algorithm model, Finite State Machine, and Outlier Detection Model, and found that Isolation Forests Technique gave the best performance. Aldhyani & Alkahtani (2022) designed an artificial intelligence algorithm-based system for detecting Economic Denial of Sustainability Attacks in cloud computing environments and found that Random Forest achieved 98% accuracy for binary classification and SVM achieved 97.54% for multi-classification. Nassif et al. (2021) conducted a systematic review of several machine learning techniques such as SVM, RF, KNN, DT, BP NN, ANN, LR, LMC, Neural Network, One Class SVM, Naïve Bayes, K-Means, Linear Kernel, FLD classifier, and Bayes Net, and found that SVM performed better than all the surveyed algorithms. Ensemble learning is a technique that combines the key characteristics of two or more models to arrive at predictions that are more accurate and durable than the individual models that make up the ensemble (Kundu, 2022).

DDOS

The denial of service (DoS) attack is a major threat that disrupts the availability of cloud services and prevents authorized users from accessing them. This attack is particularly problematic because it aims to impede the functionality of cloud services precisely when they are needed. Distributed denial of service (DDoS) attacks are even more severe as multiple attackers work together to overwhelm a single target, thereby preventing legitimate users from accessing its services (Alsaleem, et al., 2019).

An Ensembled Learning Technique

Ensemble learning is a technique in machine learning that combines the strengths of multiple models to improve the accuracy of predictions. This meta-learning strategy involves three main categories, namely bagging, stacking, and boosting. Familiarity with each of these categories is necessary for effectively implementing predictive models in any project (Jason, 2021).

Ensemble SVM

Ensemble learning that employs SVM as a base model is referred to as ensemble SVM. An open-source software tool known as EnsembleSVM is available, which incorporates efficient procedures for ensemble learning using SVM-based models. Currently, it offers ensemble techniques based on binary SVM models. By merging multiple SVM models trained on small subsamples of the training dataset, EnsembleSVM utilizes a divide-and-conquer approach. Although more models need to be trained, dividing the data drastically reduces the overall training time (Claesen, *et, al.*, 2014).

Boruta Selection Technique

The Boruta approach is an all-relevant method of feature selection that aims to identify all features in a dataset that are relevant to a given task. In our study, we utilized the Boruta selection technique through its scikit-learn interface and found it to have a faster runtime compared to other algorithms (Singh, 2021).

Major Words Discussed

Ensembled Learning, DDoS Attack, SVM, Bagging Classifier, Boruta selection technique

MATERIALS AND METHODS

3System Requirement

The model was designed and implemented using Python 3.9.16 in the Notebook environment.

Dataset

The most up-to-date Intrusion Detection Modelling dataset used in this study is CSE-CICIDS2018. The dataset, which was created through a systematic approach that utilizes profiles to develop cybersecurity datasets, includes 79 columns and 331125 rows. It was developed collaboratively by the Canadian Institute for Cybersecurity (CIC) and the Communications Security Establishment (CSE) and includes detailed intrusion definitions, as well as abstract distribution models for applications, protocols, or lower-level network elements. The dataset contains seven different attack scenarios, such as brute force, heartbleed, botnet, denial-of-service (DoS), distributed denial-of-service (DDoS), web attacks, and network infiltration. The attacking infrastructure comprises 50 machines, while the target organization consists of 420 PCs and 30 servers across five departments. The dataset features 80 network traffic features obtained from traffic collected using CICFlowMeter-V3 in addition to the network traffic and logs files of each victim-side workstation (Sharafaldin, 2018).

The model consists of three (3) major steps, which are the Boruta Feature technique, the Ensemble learning classifier, and the performance evaluation.

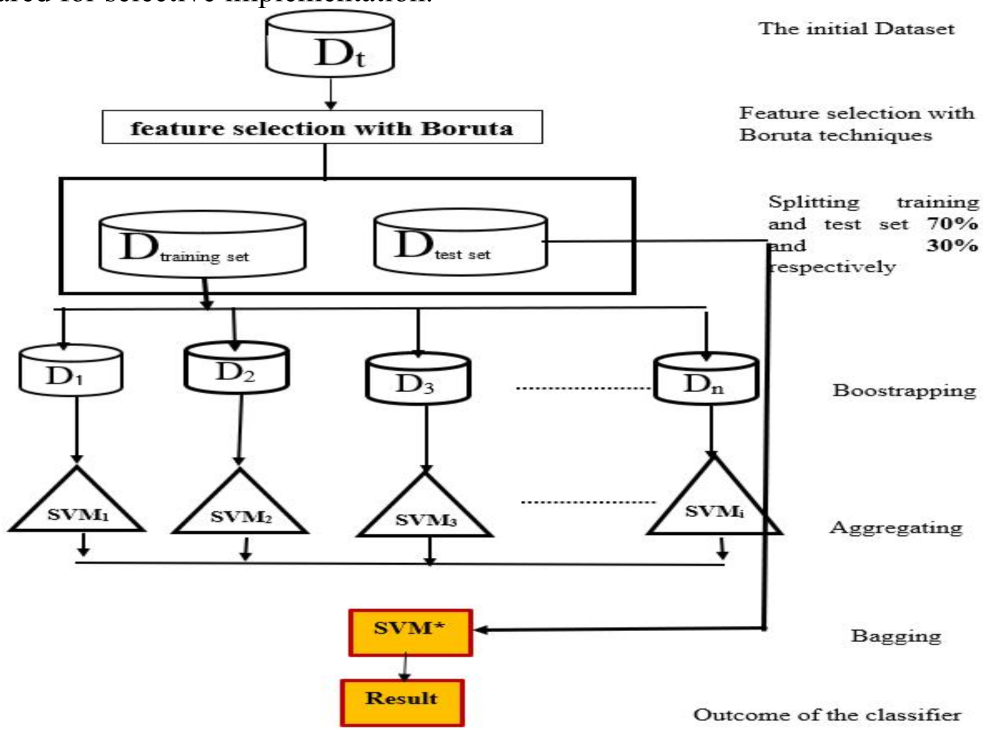
The Boruta feature selection processes were carried out to extract the best-performing features of CSE-CICID2018. The dataset is made of 79 features which were drastically reduced to 51 features after the application of the Boruta technique on the dataset.

After identifying 51 significant features from the dataset, the selected features were scaled to standardize the dataset to the same scale. This will boost the performance of the model built with the data. The scaled dataset was fit into a test and train split as demonstrated in Figure 3.

Both the train samples and the test samples were considered in Figure 3, with the training sample of 922 and the test sample of 103. Those samples were used in the modeling of all five proposed machine-learning techniques. A cross-validation technique was applied to a model to improve its performance. After applying the cross-validation technique, the bagging model with all five classifiers was built distinctly. Each

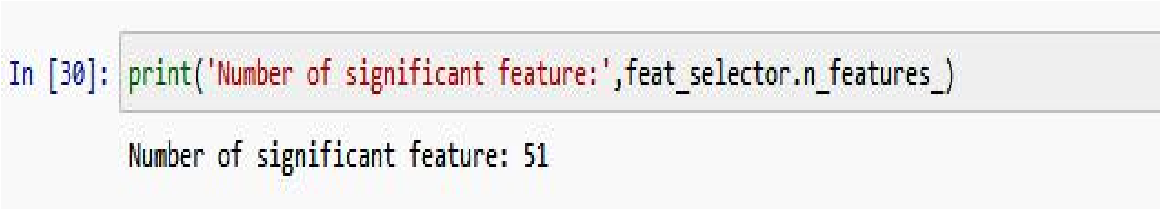
DOI: 10.56892/bima.v8i2B.716

of the classifiers presents define results that were compared for selective implementation.



Dt
feature selection with Boruta
D training set
D test set
D1
D2
D3
Dn
SVM1
SVM2
SVM3
SVMi
SVM*
Result
The initial Dataset
Feature selection with Boruta techniques
Splitting training and test set 70% and 30% respectively
Boostrapping
Aggregating
Bagging
Outcome of the classifier

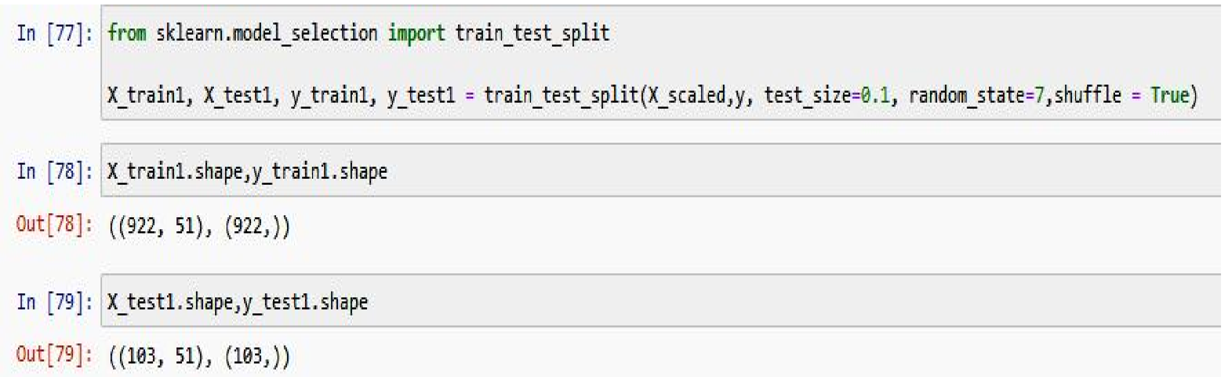
Figure 1: The Proposed model architecture of the SVM Ensemble classifier



```
In [30]: print('Number of significant feature:',feat_selector.n_features_)

         Number of significant feature: 51
```

Figure 2: The number of significant features after the Boruta feature selection



```
In [77]: from sklearn.model_selection import train_test_split

         X_train1, X_test1, y_train1, y_test1 = train_test_split(X_scaled,y, test_size=0.1, random_state=7,shuffle = True)

In [78]: X_train1.shape,y_train1.shape

Out[78]: ((922, 51), (922,))

In [79]: X_test1.shape,y_test1.shape

Out[79]: ((103, 51), (103,))
```

Figure 3: The scaled data fit into the train test split model

DOI: 10.56892/bima.v8i2B.716

RESULTS

Bagging Classification with the five different models; LR, RFC, KNN, SVC, and LDA were developed and the accuracy score of each model was collected and put in Table 1 for analysis and comparative evaluation.

Table 1: the Bagging Classifier accuracy score of five model

SN	Model	Accuracy score (%)
1	Logistic Regression (LR)	95
2	Random Forest Classifier (RFC)	99
3	K Nearest Neighbor (KNN)	99
4	Support Vector Classifier (SVC)	95.7
5	Linear Discriminant Analysis (LDA)	93.8

This information can be vividly described in Figure 1. This graphically captured the accuracy of each of the used classifiers, when in terms of the Bagging technique.

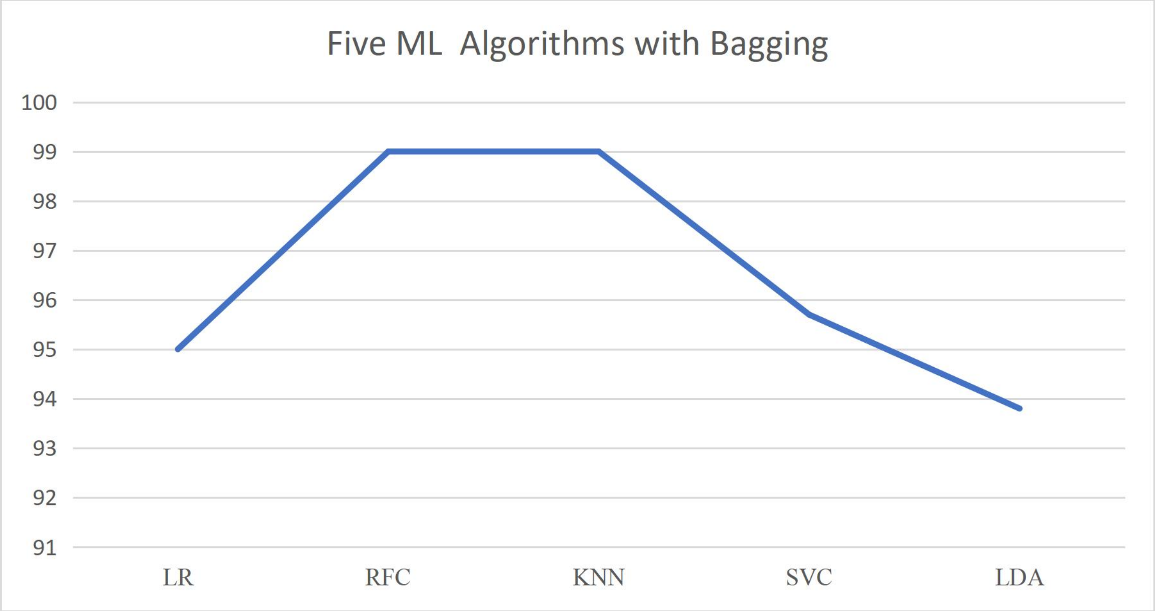


Figure 4: The Five Models in Bagging Technique

The score shows that the Random Forest algorithm and the K-nearest Neighbor were the highest-performing classifiers, followed by the proposed Support Vector Machine (SVM). Then Logistic Regression while Linear Discriminant is the lowest among the five (5) models. Figure 1 shows the five ensemble learning classifiers compared. As stated already in the review (Nassif, *et, al*., 2021), the best-performing classifier was SVM with 85.94% accuracy when compared with the proposed ensembled SVM model with a performance of 95.7, which records a great improvement. Hence our proposed system can be demonstrated in Figure 2 as having better performance than the single SVM which is the best when compared with the other single classifiers (Nassif, *et, al*., 2021) and (Sen, *et, al.,* 2020).

As quoted and analyzed in (Nassif, *et, al*., 2021) and (Sen, *et, al.,* 2020), the best-performing technique among the classifiers in cloud security detection is the SVM classifier. When aggregated in an ensembled learning technique using bagging methods, the performance of the SVM improves, which is a clear indication of better security detection.

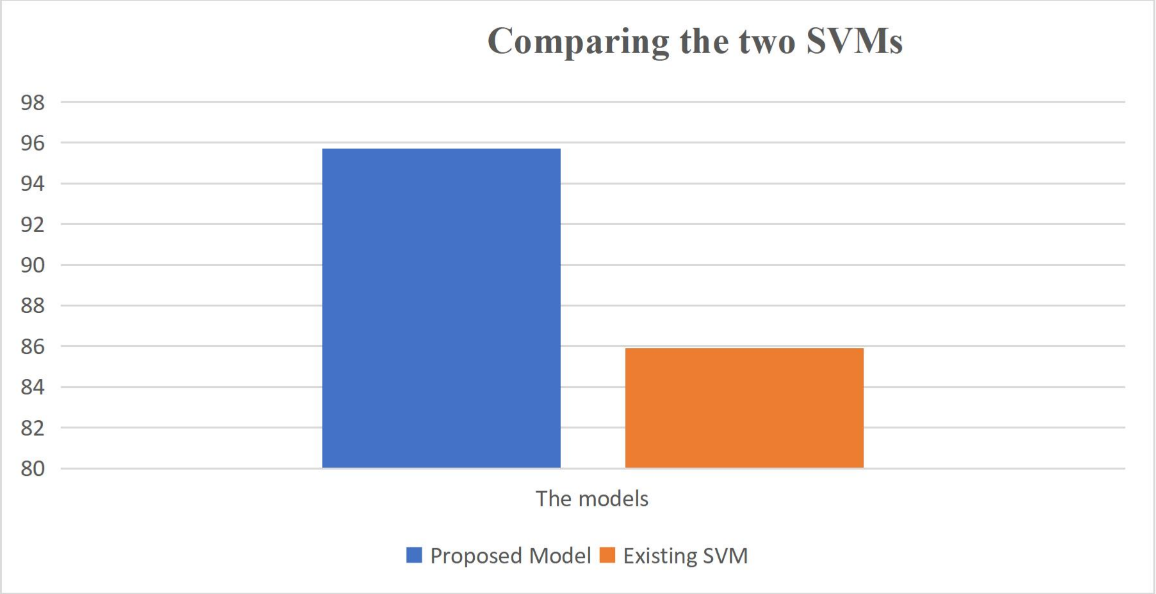


Figure 5: the Proposed Ensemble SVM and the Single SVM techniques

CONCLUSION

The Research was carried out to enhance the detection of Cloud DDoS attacks by improving the Support Vector Machine (SVM) model using the Boruta selection technique,SVM was found to be the best performing technique among Decision Tree, Random Forest, KNN, and SVM, with an accuracy of 84.94%.An ensembled SVM model with Boruta selection significantly improved the detection accuracy by 10.8% compared to traditional SVM, achieving a detection accuracy of 95.7% Random Forest Classifier and K-Nearest Neighbor outperformed the proposed SVM model, showing better performance in both individual modeling and ensemble learning setups. The further shows the importance of feature selection and ensemble techniques in enhancing DDoS attack detection in the cloud environment. The implementation of KNN, Random Forest, and Linear Discriminant analysis in ensembled learning shows that their performance is better than the proposed system.

Further Work

The number of the proposed CSE-CICID2018 dataset was drastically reduced to fit in the model efficiency and the classification used in the modeling was only five (5) out of the many used in the state-of-the-art techniques. Hence a need to use the whole dataset is the call for better performance of the real model. More classification models can also be considered to explore their performance for the appropriate selection of the model in DDoS attack detection

REFERENCES

McCollin, R. (2020, January 31). DDoS Attacks Explained: Causes, Effects, and How to Protect Your Site. kinsta.com/blog/what-is-a-DDoS-attack/

Zekri, M., El Kafhali, S., Aboutabit, N., & Saadi, Y. (2017, October). DDoS attack detection using machine learning techniques in cloud computing environments. In 2017 3rd international conference of cloud computing technologies and Applications (CloudTech) (pp. 1-7). IEEE.

El Kafhali, S., El Mir, I., & Hanini, M. (2022). Security threats, defense mechanisms, challenges, and future directions in cloud computing. Archives

DOI: 10.56892/bima.v8i2B.716

of Computational Methods in Engineering, 29(1), 223-246.

Alarqan, M. A., Zaaba, Z. F., & Almomani, A. (2019, July). Detection mechanisms of DDoS attack in cloud computing environment: A survey. In International Conference on Advances in Cyber Security (pp. 138-152). Springer, Singapore.

Dong, S., Abbas, K., & Jain, R. (2019). A survey on distributed denial of service (DDoS) attacks in SDN and cloud computing environments. IEEE Access, 7, 80813-80828.

Sharma, V., Verma, V., & Sharma, A. (2019, June). Detection of DDoS attacks using machine learning in cloud computing. In International Conference on Advanced Informatics for Computing Research (pp. 260-273). Springer, Singapore.

Aldhyani, T. H., & Alkahtani, H. (2022). Artificial Intelligence Algorithm-Based Economic Denial of Sustainability Attack Detection Systems: Cloud Computing Environments. Sensors, 22(13), 4685.

Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine learning for cloud security: a systematic review. IEEE Access, 9, 20717-20735.

Kundu, R. (2022, July 19). *Ensemble Learning: Methods, Techniques & Examples*. www.v7labs.com/blog/ensemble-learning

ALSALEEM, L. S., ALQAHTANI, S. A., ALHARBI, S. F., & AGROUBA, R. (2019). CLOUD COMPUTING BASED ATTACKS AND COUNTERMEASURES: A SURVEY. Journal of Theoretical and Applied Information Technology, 97(19), 5185-5203.

Jason, B. (2021, April). *Machinelearningmastery.Com*. machinelearningmastery.com/tour-of-ensemble-learning-algorithms/

Claesen, M., De Smet, F., Suykens, J., & De Moor, B. (2014). EnsembleSVM: A library for ensemble learning using support vector machines. *arXiv preprint arXiv:1403.0745*.

Singh, A. (2021, March 7). *Hands-On Guide To Automated Feature Selection Using Boruta*. https://analyticsindiamag.com/hands-on-guide-to-automated-feature-selection-using-boruta/

Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 1, 108-116.