Performance Analysis of Bayesian-based Algorithms for Network Attack Detection

A. Amir[1,3*], M. Nawir[1,3], S.N. Azemi[2,3], F. H. Soon[1,3], C. B. M. Rashidi[2,3]

[1]Advanced Computing Centre of Excellence, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia
[2]Advanced Communication Engineering Centre of Excellence, Universiti Malaysia Perlis 01000 Kangar, Perlis, Malaysia
[3]Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia

[*]Corresponding author's email: amizaamir@unimap.edu.my

ABSTRACT: Network traffic data is usually dynamic and massive. A low complexity ML algorithm that can build an accurate classifier is necessary to fast-process the data and ensure that the classifier is continuously updated. The multi-classification task to classifier the network attacks into several categories adds more complexity in model building. Hence, in this paper, we investigate the performance of three Bayesian-based algorithms for network attack classification by using the public UNSW-NB15 dataset. The hold-out evaluation method is applied since the dataset involves a large amount of data. Our experiment results show that AODE performs the best with an accuracy 84% with low training tim

Keywords: *Network attack detection; Naïve Bayes; Bayesian Network; multi-classification*

1. INTRODUCTION

Network security is in a critical stage because most IoT systems developed without concerning security matters. Therefore, network attack classification by using machine learning is desirable to build a better defense system for network security [1]. There are two types of classification (binary or multi-class classification). Binary classification in a network system is to classify two classes, either the data in a computer network is normal or attack data. In contrast, multi-classification is a process to recognize and overcome classifying the network traffic into more than two classes. The classes might be the sub-category of attack data such as Reconnaissance, Worms, DoS Exploits, etc. A high complexity algorithm like deep learning has become popular for classification tasks. It has also been used for developing a network attack classifier model [2][3]. Classifying the attacks into many classes may significantly increase the complexity of model development since network traffic data is usually streaming, large, and frequently updated. Bayesian-based approaches generally require less complex training processes than deep learning [4]. Hence, they are well-suited to handle the fast and streaming nature of network traffic [6]. Additionally, the Bayesian-based algorithm can be designed incrementally or online for fast learning the pattern of network data in a system.

Hence, in this paper, we investigate the performance of three Bayesian-based algorithms for network attack detection. In addition to the basic Bayesian Network algorithm, two simplified Bayesian-based algorithms: Naive Bayes [4] and Averaged-One Dependence Estimator (AODE)[5]; were evaluated in this paper.

2. METHODOLOGY

This section explains the dataset and algorithms used in this paper. The procedure used in the development of network attack detection is also presented in this section.

2.1 Dataset

UNSW-NB15 dataset is used in this current work because it represents the real normal and the synthetic network traffic [7]. Moreover, the UNSW-NB15 dataset consists a large number of data instances and features (257 673 instances with 44 features).

This dataset is available in public and labeled. Therefore, the actual network traffic scenario can be investigated by implementing the multi-class classification consisting of ten classes, including normal and attacks data. In this paper, no pre-processing and feature selection was performed to fairly evaluate the performance of the algorithms.

2.2 Bayes Family Algorithms

A Bayesian Network is a model that conceals probabilistic connections among variably of concern. One of the methods in the Bayesian family is called Naive Bayes (a strong independent feature model). Attack detection in data streams of network traffic required algorithms that can perform online learning. Hence, simple Bayes algorithms solve the issues since the models develop well in updated the dynamic data. Specifically, the development of Naive Bayes is simple [4], and classification is accomplished in a direct time.

Hence, naive Bayes is suitable for anomaly detection due to large-scale data in a network.

Another algorithm in the Bayes family is the Averaged-One Dependence Estimator (AODE) algorithm [5]. AODE algorithm performs averaging over all of a small space of alternative Naive-Bayes-like models that have weaker (and hence less detrimental) independence assumptions than Naive Bayes. The resulting algorithm is computationally efficient while delivering highly accurate classification on many learning tasks. The AODE algorithm is suitable for network attack detection since it can handle a large dataset with high accuracy. It also can classify the data instances consisting of more than two classes and predict the class probabilities for each category present in a given dataset. Additionally, it has low variance.

2.2 Bayesian-based Network Attack Detection Development

Figure 1 shows the block diagram of Bayesian-based network attack detection system development.

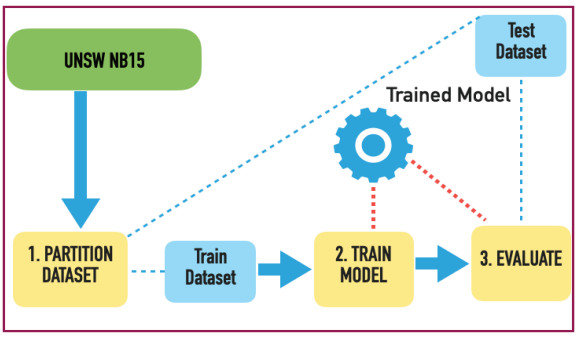

Figure 1 Bayesian-based Network Attack Detection Development Flow

In our experiment, 257,673 samples in the dataset were used. We partitioned the dataset into train and test dataset by using hold-out approach where 66% of dataset was used for training and the rest for testing. The training data was used to develop the trained model and then the model was evaluated.

3. RESULT AND DISCUSSION

The algorithms were implemented by using Weka -3.9.3 on the laptop with 3.3 GHz Dual-Core Intel Core i5 and 16 GB RAM. Table 1 shows the experimental results where we evaluated the accuracy and recorded the training time taken by each algorithm. The accuracy represents the number of correctly predicted attack class out of the total number of tests.

Table 1 Experimental results comparing Naïve Bayes, AODE and Bayesian Net by using Tabu Search

Algorithm	**Training Time (secs)**	**Accuracy**
Naïve Bayes	**1.92**	0.5167
AODE	22.12	**0.8459**
Bayesian Net – Tabu Search	75.5	0.7099

4. CONCLUSION

We have presented an experimental analysis of three Bayesian-based approaches: Naïve Bayes, AODE and Bayesian Network for network attack classification on UNSW-NB 15 dataset. The result shows that AODE algorithm performs the best among the evaluated algorithms with reasonable training time.

ACKNOWLEDGEMENT

The Fundamental Research Grant Scheme (FRGS/1/2018/ICT02/UNIMAP/02/6) has supported this work.

REFERENCES

[1] N. Elmrabit, F. Zhou , F. Li, and H.Zhou, "Evaluation of Machine Learning Algorithms for Anomaly Detection," 2020 Int. Conf. Cyber Secur. Prot. Digit. Serv. (Cyber Secur.), pp. 1--8, 2020.

[2] T. S. Pooja and P. Shrinivasacharya, "Evaluating Neural Networks using Bi-Directional LSTM for Network IDS (Intrusion Detection Systems) in Cyber Security," Glob. Transitions Proc., pp. 0–13, 2021.

[3] P. Singh, J. J. P, A. Pankaj, and R. Mitra, "Edge-Detect : Edge-centric Network Intrusion Detection using Deep Neural Network," 2021 IEEE 18th Annu. Consum. Commun. Netw. Conf., pp. 1–6, 2021.

[4] N. Hashim, S. O. Al-mamory, and A. H. Al-shakarchi, "Constructing decision rules from naive bayes model for robust and low complexity classification," Int. J. Adv. Intell. Informatics, vol. 7, no. No.1, pp. 76–88, 2021.

[5] R. Godahewa, T. Yann, C. Bergmeir, and F. Petitjean, "Seasonal Averaged One-Dependence Estimators : A Novel Algorithm to Address Seasonal Concept Drift in High-Dimensional Stream Classification," 2020 Int. Jt. Conf. Neural Networks, pp. 19–24, 2020.

[6] J. A. Perusquia, J. E. Griffin and C. Villa, "Bayesian Models Applied to Cyber Security Anomaly Detection Problems," pp. 1–40, 2021.

[7] A. Thakkar and R. Lohiya, "A Review of the Advancement in Intrusion Detection Datasets," Procedia Comput. Sci., vol. 167, no. 2019, pp. 636–645, 2020