International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Performance Prediction of Infiltration Detection system

Alakananda K[1], Dr. Shivakumar G S[2]

[1]*Computer Science and Engineering, Srinivas Institute of Technology, valachil, Mangaluru, Kranataka, India*
[2]*Head, Department of CSE, Srinivas Institute of Technology, valachil, Mangaluru, Kranataka, India*

Abstract - *This study investigates the "performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection (Iftikhar Ahmad, May 30, 2018,)", and accurately detects malicious traffic, accuracy, and recall of computer networks. Due to the support vector machine capabilities of the infiltration detection system and the non-linear classification used by a wide range of tasks, the support vector uses the machine and performs better compared to other classifiers. Many classification methods do not follow IDS to increase the efficiency of the discovery rate. Recent work has used multi-layer preceptors, support vector machines and other techniques to address performance concerns. Such techniques imply limitations and are inefficient to use on large samples. Infiltration detection performance depends on accuracy, which needs to be improved to reduce false alarms and increase detection rate. Therefore, an efficient classification method is needed to overcome the problem. This issue is considered in this thesis; providing an overview of ELM, SVM, and RF classification methods for infiltration detection. These methods are popular because of their ability in classification. NSL-KDD uses the knowledge discovery and processing data sets that are estimated due to the evaluation criteria of the infiltration detection system. The results demonstrate the feasibility and efficiency of the proposed cooperative and adaptive penetration detection method. Furthermore, the method is more specific than the methods used by a set of SVM and RF in terms of detection accuracy, precision, and recall rate. The conclusive results demonstrated that ELM overcomes different methodologies.*

Key Words: **IDS, SVM, RF, ELM, ML, NSL_KDD**

1. INTRODUCTION

Network penetration detection software protects a computer network from unauthorized users, perhaps including insides. The task of infiltration detection is to find a predictive model for example a classifier. That is able to distinguish between "unusual" connections and "normal" common connections, known as infiltration or attack. Unnecessary and irrelevant features in the data caused a long-standing problem in network traffic classification. These features not only slow down the classification process but also prevent a classifier from making accurate decisions, especially when it comes to big data. In this thesis, the optimal features for classification of specific SVM, RF, and ELM based algorithms are analyzed. This support vector machine based feature selection algorithm can handle linear, non-linear, classification, and regression dependent data features. Its effectiveness in cases of network penetration is being evaluated. An infiltration detection system named ELM, SVM, and RF is built using the selected features of the proposed feature selection algorithm. The performance of ELM, SVM, and RF is evaluated using the NSL-KDD sample. Evaluation results show that ELM achieves better accuracy and lower computational cost compared to SVM and RF. The raw preparation information is 4 gigabytes of compressed binary TCP dump information from weeks of network traffic. It can handle about five million connection records. Correspondingly, fourteen day test information gives nearly 20 million connection records.

2. METHODOLOGY

As mentioned earlier, the framework is to develop synchronous classifiers that will improve the accuracy of penetration detection. For this purpose, trained and tested information are combined into single synthesis. The ELM, SVM, and RF algorithms do not take the opinion of every expert. To straighten out, the system framework is divided into related phases:

1. NSL_KDD data pre-taking care of
2. Data classification by SVM
3. Data classification by RF
4. Data classification by ELM
5. Evaluation of results for each approach

NSL-KDD knowledge discovery and data mining (Iftikhar Ahmad, May 30, 2018), [1] analysis within dataset. The NSL-KDD contains thousands of connection data's and extracts 41 qualitative and quantitative features. Each set of 41 extracted features represents a survey during a routine or attack. The connection is a series of TCP packets that start and end at certain well-defined times, and the data from one source IP address to the target IP address between them is under some defined protocol. Each connection is labeled as normal or offensive, and exactly one specific attack type. Each connection record contains approximately 100 bytes. To properly evaluate the performance of each classifier, the data is divided into two different databases: one for preparing and other for checking and evaluating classifiers, [8] which can be divided into two stages:

1) Preparing data: (NSL-KDD 80%), this sample is used to train every expert within the fair.

2) Checking data: (NSL-KDD 20%), this sample is used to analyze the efficiency of each base classifier in the system and the efficiency of the sync classifier.

It should be noted that the test data does not come from the same probability distribution of the training sample, but also the specific attack types that are not in the training data. The samples contain a sum of 24 preparing attack types, and the test data alone contains 14 additional types.

2.1 Preprocessing data:

Due to symbolic features, the classifier is less able to process the raw sample. So, preprocessing is required, in that non-numeric or symbolic features are omitted or replaced because they are not significant partnerships to detect infiltration. Accordingly, this method creates overhead, including more preparing time; classifier's technology is complicated and memory is wasted computing resources. Therefore, numerical features are omitted from the raw sample for better efficiency of infiltration detection techniques.

2.2 Support Vector Machine Classification

SVM is a supervised ML method intended for binary classification. An example classifier SVM has been applied to different instance recognition applications that underline the objective learning strategy for re-establishment and configuration with the expansion of part works. SVM anomaly has become one of the favorite methods to detect infiltration, due to its good generalization behavior and ability to cross dimensions. This is useful for identifying specific risks around the world using structural risk minimization, because it can be well generalized using kernel strategies even in the case of a high volume of small training sample scenarios. SVM requires the handling of raw samples for classification, which increases the complexity of the architecture and reduces the exactness of penetration detection.

2.3 Random Forest Classification

Random Forest is a synthesizer classifier used to increase exactness. The RF encompasses several decision perspectives. RF has a lower classification error compared to other conventional classification algorithms. The number of trees used to separate each hub, the minimum hub size, and some highlights. The benefits of RF are listed below:

1) Generated forests can be preserved for future reference
2) Random forest defeat the issue of over-fitting
3) The RF is naturally produced in terms of accuracy and variable importance

When developing individual trees in a random forest, randomization is applied to select the best hub to be part of.

2.4 Extreme Learning Machine Classification

Extreme Learning Machine is a special type of machine learning system that applies to single layer or multiple layers. The ELM contains the number of hidden neurons that are randomly allocated to the input weight. ELM uses random projection and early preceptor models for critical thinking. The ELM algorithm is the simplest and best implementation algorithm. The ELM algorithm produces better results with less computation time. ELM performance is comparable to SVM or other AI classifiers. ELM has the potential to perform well on extraordinarily complex databases.

2.5 Evaluation

Assessments are based on the quality of the system being designed data NSL-KDD, which is randomly divided, three parts, namely, an iteration1 dataset, an iteration 2 dataset, and so an iteration 3 dataset. The entire dataset contains 125973 samples, the iteration 2 dataset contains 22576 samples and the iteration 3 of the dataset contains 25764 samples. Accuracy, precision, and recall are used as assessment dimensions.

3. PROPOSED APPROACH

1) Algorithms: SVM, RF, and ELM model for IDS
2) Input: NSL_KDD informational samples
3) Output: Classification of various types of algorithms and evaluate the accuracy, precision, and recall
4) Step 1: Upload the dataset.
5) Step2: Process the preprocessing techniques.
6) Step3: Partition of the dataset into training and testing
7) Step4: Classification.
8) Step5: The data set is uploaded to each classification algorithm, namely, SVM, RF, and ELM for training
9) Step6: The test dataset is then fed to each classification algorithm, namely, SVM, RF, and ELM.
10) Step7: Evaluate the accuracy, precision, and recall

4. RESULT

The results are investigated, contemplated, and acknowledged using the NSL-KDD datasets, as indicated in the performance testing of various ML strategies. The detailed accuracy of the ELM, SVM and RF algorithms is shown in Figure 1 with a 20% verification view and an 80% prepared data example. Here, the ELM provides better

accuracy compared to the RF and SVM in iteration 1 sample. Specifically, RF shows greater accuracy than SVM and ELM in the iteration 2 data sample. As seen in Figure 1, SVM surpasses different strategies in periodic information analysis of iteration 3.

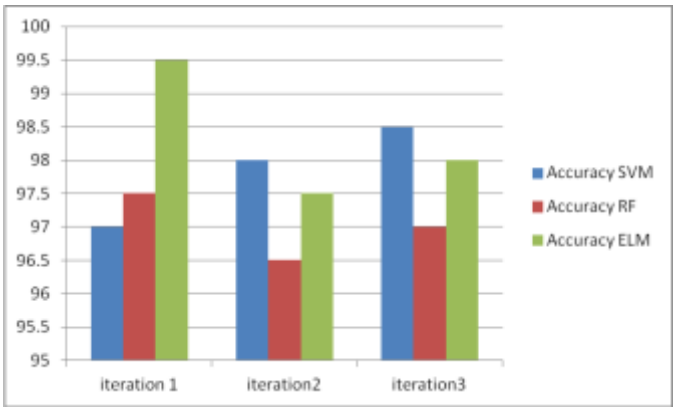

Chart 1: The detailed view of Accuracy data.

Detailed visualization precision of ELM, SVM and RF algorithms is shown in 80% prepared example and the 20% testing example in Chart 2. Here show that the accuracy of the ELM is better compared to the SVM and RF in the periodic information tests, and so on in iteration 1 samples override the RF. SVM precision is high when compared to iteration 2 dataset, RF and ELM. Furthermore, SVM provides better accuracy compared to ELM and RF in the iteration3 dataset.

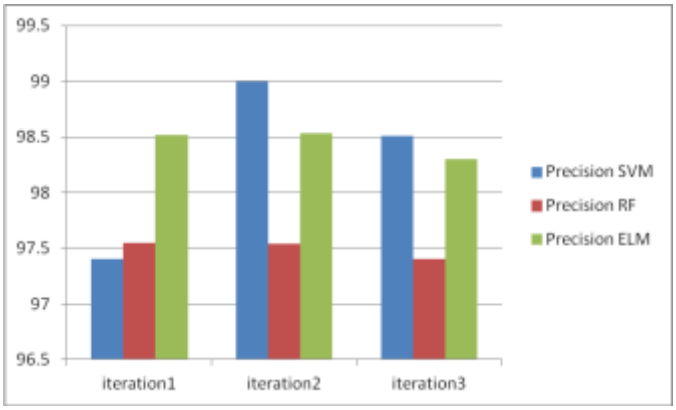

Chart 2: The detailed view of Precision data.

A detailed recall view of the ELM, SVM and RF algorithms is shown in Chart 3 for 20% of the test and 80% of the prepared data samples. In iteration 1 information tests, ELM provides better recall compared to SVM and RF. The second iterative SVM, ELM, is more noticeable than the RF. Iteration 3 information verification of SVM recall is better than RF and ELM. The audit and SVM mentioned earlier do note that the small data sample improves, while the ELM overcomes the different methodologies in large datasets.

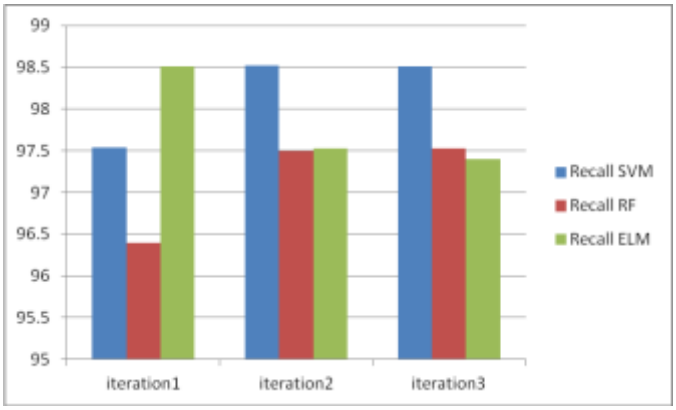

Chart 3: The detailed view of Recall.

DISCUSSION

The objective of this research work is to compare the performance of ELM, RF and SVM with accuracy, precision and recall for learning and recall. Using the NSL-KDD dataset, an evaluation of the intrusion detection system, which randomized and divided into three parts, i.e. iteration 1, iteration 2, and iteration 3 dataset other attacks that do not exist in the training data set include the NSL-KDD test, which has some rules, such as a test dataset, a test dataset from another probability distribution, and a training dataset.

5. CONCLUSIONS

This paper proposes several analyzes and attempts to evaluate the productivity and the performance of the accompanying ML classifiers: Random Forest, Extreme Learning Machine, and Support Vector Machine. All tests rely on the NSL-KDD infiltration identification database. Strategy proposed using a three ML method for effective implementation of the infiltration detection framework. Although many methods have been used in infiltration detection systems, ML methods have become important in recent literature. In addition, there are different ML methods used, but some methods are more suitable for analyzing big data to detect network and infiltration systems. The efficiency of the approach is monitored by dividing the NSL-KDD data into 3 parts, namely, iteration 1, iteration 2, and iteration 3. To solve this, the problem is researched and compared with various machine learning methods, namely, ELM, SVM, and RF. ELM overcomes various approaches in terms of accuracy, recall, and precision in iteration 1 information samples comprising 125973 records of operations. Dataset SVM shows better results in iteration 2 data samples and iteration 3 in the data samples. Therefore, ELM is an ideal method for a penetration detection system designed to analyze a large amount of data.

REFERENCES

[1] Iftikhar Ahmad 1, Mohammad Basheri1, Muhammad Javed Iqbal2, And aneel Rahim3 "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection" Received April 15, 2018, accepted May 18, 2018, date of publication May 30, 2018, date of current version July 6, 2018.

[2] Kuang,W. Xu, and S. Zhang, \A novel hybrid KPCA and SVM with GA model for intrusion detection", Appl. Soft Comput.,vol. 18, pp. 178184, May 2014.

[3] S. Teng, N.Wu, H. Zhu, L. Teng, and W. Zhang,\SVM-DT-based adaptive and collaborative intrusion detection", IEEE/CAA J. Automatica Sinica,vol. 5, no. 1, pp. 108118, Jan. 2018,

[4] N. Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system", Proc. Comput. Sci., vol. 89, pp. 213-217.

[5] R. M. Elbasiony, E. A. Sallam, T. E. Eltobely, and M. M. Fahmy, \A hybrid network intrusion detection framework based on random forests and weighted k-means", Ain Shams Eng. J., vol. 4, no. 4, pp. 753762, 2013.

[6] H.Wang, J. Gu, and S.Wang,\An e ective intrusion detection framework based on SVM with feature augmentation", Knowl.-Based System., vol. 136, pp. 130 : 139, Nov. 2017.

[7] Kuang,W. Xu, and S. Zhang, \A novel hybrid KPCA and SVM with GA model for intrusion detection", Appl. Soft Comput., vol. 18, pp. 178184, May 2014.

[8] A. Aburomman and M. B. I. Reaz, \A novel SVM-kNN-PSO ensemble method for intrusion detection system", Appl. Soft Comput., vol. 38, pp. 360372, Jan. 2016.