Novel approach for network intrusion detection using machine learning

Prof. Sachin Pandey

Assistant Professor
Information Technology Department
Pune Institute of Computer Technology, Pune

***Abstract*: Advancement in the technologies has led to increment in the massive amount of data that massive amount of generated data has to be secured in such way that third party should not be able to take control over them The online platforms such as face book which has large number of users are the main sources of generating large amount of data ,each users activity on the internet is being captured in one or the other ways, security of network has become a great challenge in this modern era. Hence it has become very important to build an effective intrusion detection system. We have implemented a compressive survey of some of the major machine learning techniques based on Naïve Bays Classifier, K Nearest Neighbors Classifier, Decision Tree Classifier and the Logistic Regression in this paper.**

Index Terms: IDS, artificial neural network, NSL-KDD dataset, Feature selection.

Introduction

An Intrusion detection system is used to detect the malicious activity over the network to make the network, server lines secure and free from any intrusions. We can use intrusion detection system to detect the malicious activity related to some specific devices (host intrusion detection system) or to detect the malicious activity occurred in the entire network (network intrusion detection system) which is the common type used. The two main challenges in building the efficient intrusion detection system is first, the feature selection from the dataset is very difficult as it will tell us how important a feature can be. The feature selection changes with the change in attack type. Secondly, there does not exist a labeled traffic real-time networking. Intrusions in the network are mainly caused by unauthorized users trying to access the system and the authorized users who attempt to gain additional privileges given to them.

Machine learning is the field of study that allows the computer to learn automatically without being explicitly programmed therefore machine learning mainly focuses on the development of programmers that are able to learn themselves and perform the task, the information gained by different machine learning techniques is different for each set of input . It has become difficult for traditional network protection techniques to distinguish the normal traffic and the network traffic since the new emerging attacks are having similar behavior and characteristics to that of normal traffic. In this paper following machine learning techniques are implemented Naïve Bayer Classifier, Neighbors Classifier, Decision Tree Classifier and the Logistic Regression Model then later comparison is done based on their 5 major parameters accuracy, precision, recall, f1-score.

LITERATURE SURVEY

The system proposed by **Wei-Chao Lin, et al.** [18] uses the k-NN classifier to predict the state of each network packet, whether to be from a normal or attack traffic. This system is trained and evaluated using the KDD CUP'99 dataset, where the evaluation measures show a good prediction accuracy of 99.89% accurate predictions. However, as the k-NN classifier is a lazy classifier, the knowledge is extracted each time a prediction is required, i.e., the training dataset is scanned every time a new packet enters the network, which is a very resource-consuming process. **Neha G Relan and Dharmaraj R Patil [19] ,** which perform an intrusion detection system with the use of the decision tree classifier. The performance of the proposed system has scored a highest of 95.09%, using the KDDCUP'99 dataset for both training and testing stage. **Malek Al-Zewairi, et al.** [20] suggest an intrusion detection system depend on deep learning that include of five hidden layers with ten neurons in each layer. The deeper the neural network, the more complex attribute can be discover based on the input data, while raising the number of neurons in a layer rising the number of attributes that the layer can detect. The accuracy of the deep learning model is compared to other classifiers, such as logistic regression, decision tree, Naïve Bays 9+and neural network, where the experimental results show that the deep learning model has scored the highest with 98.99% accuracy when tested with the UNSW-NB15 dataset.

Proposed system

There are so many techniques and algorithms are available to improve the monitoring of network intrusion detection system in recent years. In this paper, we have presented a complete survey of some main methods of machine learning applied on intrusion Detection. In this paper following machine learning techniques are implemented by us: Naïve Bayer Classifier, Neighbors Classifier, Decision Tree Classifier and the Logistic Regression Model then later comparison is done based on their 5 major parameters.

Here implementation is done on two level of approach firstly here we implemented the machine learning algorithms and at next level we have taken the level 1 results to train the next level where in 5 major parameters accuracy, precision, recall, f-score are compared

3. Modules

1. Train dataset and test data set

Collecting data for training the system is one of the main challenges in developing a network-based intrusion detection system. Although the KDD 1999 data set is widely being used but these data sets are unreliable for building a system in reality, accuracy on NSL-KDD dataset is significantly higher than KDDCup99 dataset. Hence in this paper NSL KDD dataset rather than KDDCup99 dataset, accuracy on NSL-KDD dataset is significantly higher than KDDCup99 dataset.

2. **Feature selection:** Here we are removing the number of input variables while developing the predictive model reducing the input variables reduces the computational cost and as well as improve the performance. Once this is done we fit the data frame of testing and training dataset

Since 41 features are there in NSL KDD hence it impossible to take all those features and plot the feature selection graph hence we take decision of dropping the last 9 where there are no variations in the value some of the columns such as su_attempted, srv_error rate have only 0 as their values

3. Build the model: We build two models on the basis of two classifiers, a K Nearest Neighbors Classifier and a logistic regression classifier. Both of the above mentioned models are present in the Scikit-learn module.

4. Prediction and Assessment (Authentication): With the use of the test data to make predictions of the model. Multiple scores are considered here like accuracy score, then recall, the f-measure, confusion matrix. Perform a 10-fold cross-validation.

EXPERIMENTAL RESULT

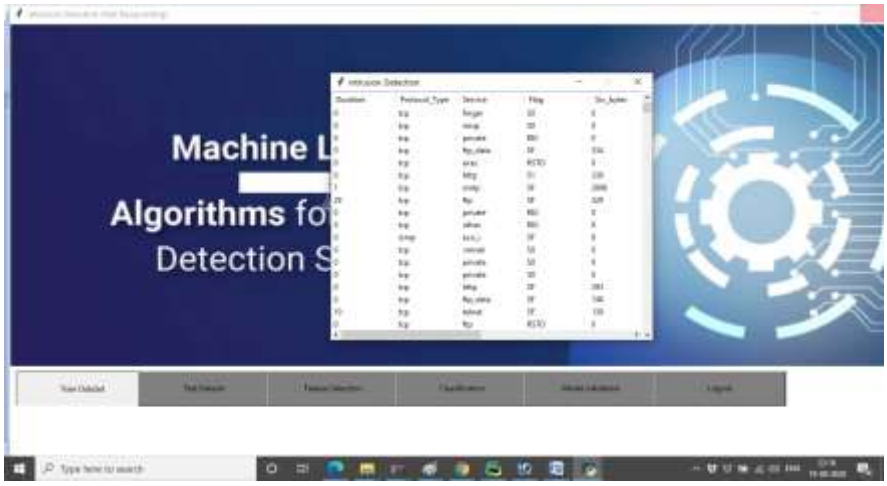

Fig 1: Loading Dataset

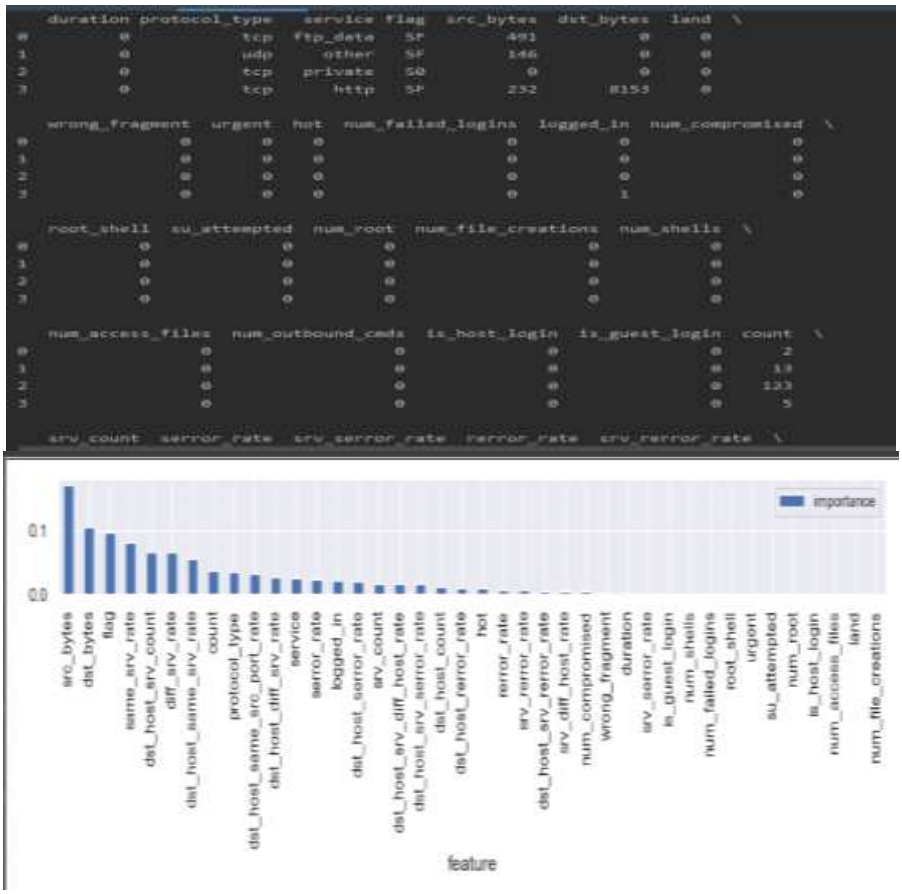

Fig 2: Feature selection graph

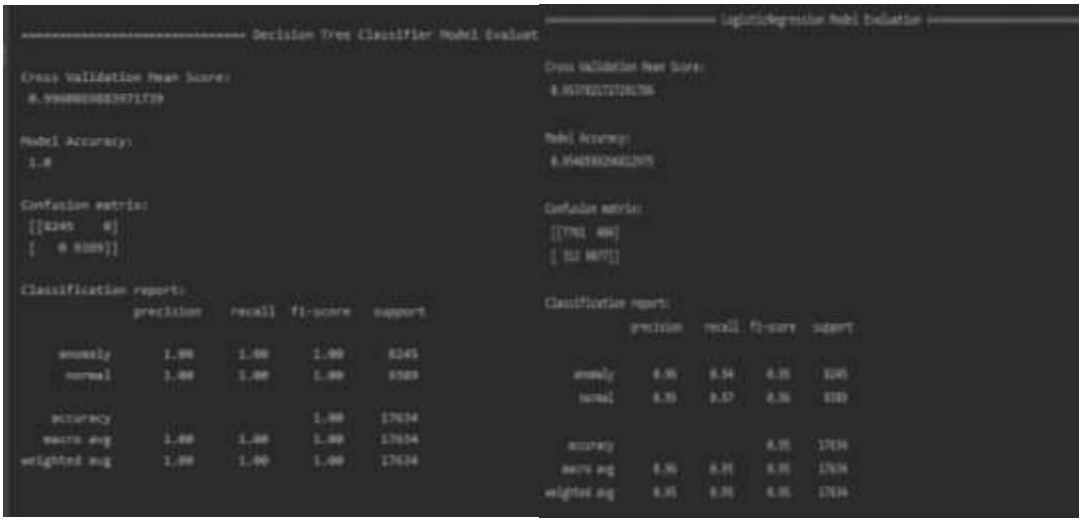

Fig: 3 Model Validation process

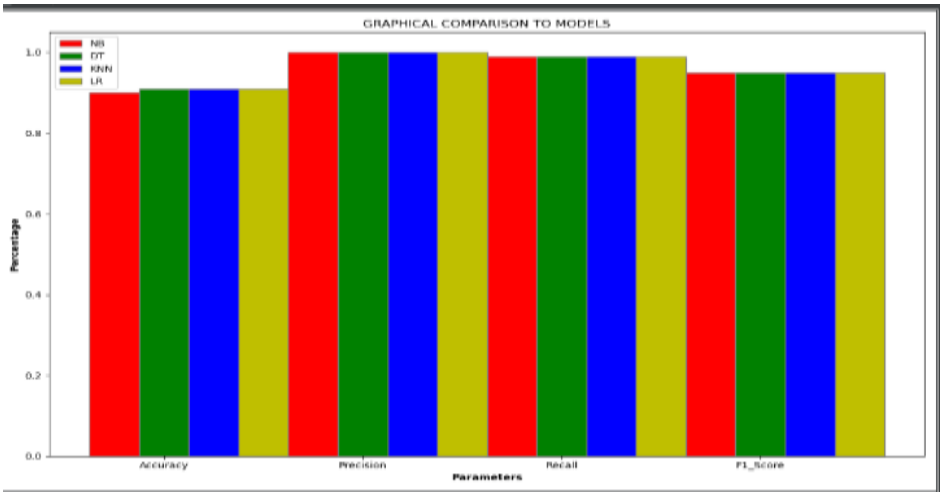

Fig: 4 Graphical Comparisons to Models

CONCLUSION

In the domain of network security, Network Intrusion Detection System is the most important and useful defense technology. So many of the available techniques have been implemented for the intrusion detection system in recent years. A detailed survey of major techniques implemented on intrusion Detection is presented in this research paper. Techniques based on classification algorithms such as Naive Bayes algorithm, Decision Tree algorithm, KNN, Linear Regression. We do a comparison of experimental values a system where we try to increase the efficiency of the parameters in the intrusion detection system compare the parameters such Accuracy, Precision, Recall, F-score.

REFERENCES

1. Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches",wiley.com, https://doi.org/10.1002/ett.4150
2. Javed Asharf, Nour Moustafa, Hasnat Khurshid, Essam Debie, Waqas Haider, Abdul Wahab, "A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions", mdpi.com, https://doi.org/10.3390/electronics9071177
3. K. A. I. PENG, V. C. M. LEUNG, and Q. HUANG, "Clustering Approach Based on Mini Batch K means for Intrusion Detection System Over Big Data," SPECIAL SECTION ON CYBERPHYSICAL- SOCIAL COMPUTING AND NETWORKING, 2018.[Online]. Available: 0.1109/ACCESS.2018.2810267
4. AHMAD, M. BASHERI, M. J. IQBAL, and A. RAHIM, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection."[Online]. Available: 0.1109/ACCESS.2018.2841987
5. Q. Niyaz, M. Alam, W. Sun, and A. Y. Javaid, "A Deep Learning Approach for Network Intrusion Detection System," in Conference Paper in Security and Safety, 2015.