ABSTRACT

Farriz Abdillah Mukmin. 50422536

COMPARISON OF RANDOM FOREST, NAÏVE BAYES, AND SUPPORT VECTOR MACHINE ALGORITHMS FOR NETWORK ATTACK CLASSIFICATION USING THE NSL-KDD DATASET.

Thesis, Department of Informatics, Faculty of Industrial Technology, Gunadarma University, 2026.

Keywords: Machine learning, Naïve Bayes, Network Attack Classification, NSL-KDD, Random Forest, Support Vector Machine.

(xiv+71+Lampiran)

The rapid advancement of information technology has significantly increased the use of computer networks across various sectors. However, this growth has also led to a rise in cyberattacks that threaten the confidentiality, integrity, and availability of network systems. Therefore, an effective classification mechanism is required to accurately identify network activities and support intrusion detection. This study aims to compare the performance of the Random Forest, Support Vector Machine (SVM), and Naïve Bayes algorithms for network attack classification using the NSL-KDD dataset. The research was conducted based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which consists of the Business Understanding, Data Understanding, Data Preparation, modeling, Evaluation, and Deployment phases. The NSL-KDD dataset was grouped into two classes, namely normal and attack, and the same preprocessing procedures were applied to all algorithms. The models were developed using their default parameters (baseline models) without hyperparameter optimization to ensure that the comparison reflected the inherent performance of each algorithm. Model performance was evaluated using a Confusion Matrix with Accuracy, Precision, Recall, and F1-Score metrics. The experimental results indicate that Random Forest achieved the best performance, with an Accuracy of 99.66%, Precision of 99.72%, Recall of 99.64%, and F1-Score of 99.68%. Support Vector Machine (SVM) achieved an Accuracy of 98.36%, Precision of 98.22%, Recall of 98.72%, and F1-Score of 98.47%, while Naïve Bayes obtained an Accuracy of 88.32%, Precision of 87.80%, Recall of 90.76%, and F1-Score of 89.26%. Based on these findings, Random Forest outperformed Support Vector Machine (SVM) and Naïve Bayes, making it the most suitable algorithm for network attack classification using the NSL-KDD dataset.

References (2020-2026)