ABSTRAK

Intrusion Detection System (IDS) adalah sebuah sistem yang dirancang khusus untuk mendeteksi aktivitas mencurigakan atau tidak sah dalam jaringan komputer atau sistem komputer. Keberadaan IDS menjadi sangat penting seiring dengan munculnya berbagai jenis serangan yang semakin berkembang. Untuk meningkatkan efektivitas IDS, banyak penelitian yang mengimplementasikan berbagai teknik *machine learning*, terutama teknik klasifikasi. Namun, penerapan *machine learning* pada IDS, masih ditemukan beberapa masalah, yaitu hasil deteksi yang kurang tepat dan seringnya terjadi *false positive*. Permasalahan tersebut terjadi dikarenakan beberapa dataset masih mengandung banyak atribut yang berlebihan (*high dimensionality*), *redundant* dan jumlah *class* yang *imbalance*. Selain itu, penggunaan algoritma berbasis *single classifier* juga dapat menghambat proses klasifikasi secara keseluruhan.

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan dua pendekatan yang digunakan secara bersamaan, yaitu pendekatan pada level data dan pada level algoritma. Pada level data, teknik seleksi fitur berbasis *Recursive Feature Elimination* (RFE) dan *Rough Set Theory* (RST) diterapkan untuk mendapatkan *subset* terbaik dari dataset, sehingga meningkatkan kinerja klasifikasi. Sedangkan pada level algoritma, menerapkan tiga model yaitu Bagging-SDN, S-SDN, dan B-DT. Ketiga model ini dibangun menggunakan teknik *bagging* dan *stacking* berbasis *ensemble classifier*. Pada *ensemble classifier* menggunakan tiga *single classifier* sebagai *base-learner*, yaitu *Support Vector Machine* (SVM), *Decision Tree* (DT), dan *Naïve Bayes* (NB). Pendekatan ini bertujuan untuk meningkatkan performa IDS dalam mendeteksi serangan. Dengan model yang diusulkan ini, diharapkan performa IDS dapat meningkat dibandingkan dengan menggunakan teknik lama (*single classifier*).

Hasil penelitian menunjukkan bahwa model yang diusulkan (Bagging-SDN, S-SDN, dan B-DT), mampu meningkatkan performa IDS dalam mendeteksi serangan. Namun, masih ditemukan kelemahan pada model B-DT ketika dikombinasikan dengan teknik RST. Teknik RST tidak mampu menghasilkan *subset* yang optimal. Evaluasi model dilakukan menggunakan empat dataset publik, yaitu NSL-KDD, UNSW-NB15, CIC-IDS2017, dan DoHBrw-2020. Standar evaluasi yang digunakan berupa *confusion matrix*, yaitu dengan mengukur nilai *accuracy, precision, recall, f1-score, kappa-score* dan *testing time*. Berdasarkan hasil eksperimen dan uji McNemar, terdapat peningkatan yang signifikan pada model yang diusulkan (*ensemble classifier*), terutama pada model B-DT.

Kata kunci—*Intrusion Detection System, Recursive Feature Elimination, Rough Set Theory, Single Classifier, Support Vector Machine, Decision Tree, Naïve Bayes, Ensemble Classifier, Bagging, Stacking*.

ABSTRACT

Intrusion Detection System (IDS) is a system specifically designed to detect suspicious or unauthorized activity in a computer network or computer system. The existence of IDS becomes increasingly important with the emergence of various types of attacks that are evolving. Many current studies implement machine learning techniques to improve IDS performance, especially classification techniques. However, several challenges remain in applying machine learning to IDS, such as inaccurate detection results and frequent false positives. This issue occurs because some datasets still contain many excessive attributes (high dimensionality), redundant and an imbalanced number of classes. Additionally, single-classifier-based algorithms can hinder the overall classification process.

To overcome these problems, this research proposes two approaches that can be used simultaneously at the data and algorithm levels. At the data level, feature selection techniques based on RFE (Recursive Feature Elimination) and RST (Rough Set Theory) are applied to obtain the best subset of the dataset, thereby improving classification performance. Meanwhile, three models are implemented at the algorithm level: Bagging-SDN, S-SDN, and B-DT. These three models were built using ensemble classifier-based bagging and stacking techniques. The ensemble classifier uses three single classifiers as base learners, namely SVM (Support Vector Machine), DT (Decision Tree), and NB (Naïve Bayes). This approach aims to improve IDS performance in detecting attacks. With this proposed model, it is hoped that IDS performance can be increased compared to using the old technique (single classifier).

The research results show that the proposed models (Bagging-SDN, S-SDN, and B-DT) can improve IDS performance in detecting attacks. However, weaknesses were still found in the B-DT model when combined with the RST technique. The RST technique is not able to produce optimal subsets. Model evaluation is conducted using four public datasets: NSL-KDD, UNSW-NB15, CIC-IDS2017, and DoHBrw-2020. The evaluation standard used is a confusion matrix by measuring accuracy, precision, recall, f1-score, kappa-score and testing time. Based on experimental results and McNemar tests, there is a significant improvement in the proposed model (ensemble classifier), especially in B-DT model.

Keywords—Intrusion Detection System, Recursive Feature Elimination, Rough Set Theory, Single Classifier, Support Vector Machine, Decision Tree, Naïve Bayes, Ensemble Classifier, Bagging, Stacking.