

Deteksi Jenis Serangan pada Distributed Denial of Service Berbasis Clustering dan Classification Menggunakan Algoritma Minkowski Weighted K-Means dan Decision Tree

Detection of attack on Distributed Denial Of Service based on Clustering and Classification using MINKOWSKI WEIGHTED K-MEANS Algorithm and DECISION TREE

Riski Pristi Ananto¹, Yudha Purwanto², Astri Novianty³

^{1,2,3}Prodi S1 Sistem Komputer, Fakultas Teknik, Universitas Telkom

justriski@students.telkomuniversity.ac.id, omyudha@telkomuniversity.ac.id, astrinov@telkomuniversity.ac.id

Abstrak

Seiring dengan meningkatnya perkembangan dan peran internet saat ini, juga berdampak pada meningkatnya jenis gangguan atau serangan terhadap pengguna internet, salah satunya Distributed Denial of Service (DDoS). DDoS merupakan salah satu jenis serangan Denial of Service dimana serangan ini menggunakan banyak host penyerang secara bersamaan dengan mengirimkan data secara berulang-ulang dengan tujuan agar target tidak bisa menggunakan hak akses sebuah jaringan dengan baik. Banyak penelitian terkait metode deteksi pada serangan DDoS, namun kebanyakan penelitian hanya terbatas pada penggunaan algoritma yang sering kali tidak memberikan hasil deteksi yang optimal apabila dilakukan pengujian menggunakan berbagai macam bentuk dataset yang berbeda. Maka dibutuhkan suatu sistem deteksi yang terdiri dari beberapa algoritma yang mampu beradaptasi dengan berbagai macam bentuk dataset agar dapat memberikan hasil deteksi yang lebih optimal.

Pada tugas akhir ini sistem melakukan pengelompokan serangan dan sekaligus melabelkan tipe serangan tersebut. Pengelompokan data dilakukan menggunakan algoritma Minkowski Weighted K-means yang mengelompokkan data berdasarkan kemiripannya. Data yang sudah dikelompokkan selanjutnya diklasifikasi dan diberikan label menggunakan metode Decision Tree.

Berdasarkan kemampuan adaptasi algoritma Minkowski Weighted K-means dalam melakukan pengelompokan data menjadi beberapa kluster yang optimal dan ditambah efektifitas dari metode Decision Tree pada proses pelabelan masing – masing data yang diolah, didapatkan hasil berupa nilai akurasi sebesar 94.78% dengan *false positive rate* 0.26% dan *detection rate* 99.98%. Dengan perbandingan beberapa skenario didapatkan bahwa nilai *p* untuk *minkowski distance* yang memberikan hasil paling optimal adalah 2.5.

Kata kunci: *Minkowski Weighted K-means, Decision Tree, DDoS*

Abstract

Along with the increasing development and the role of the internet today, also resulted in increased type of intrusion or attacks to Internet users, like Distributed Denial of Service (DDoS). DDoS is a type of Denial of Service attack where attackers use multiple hosts simultaneously to transmit data repeatedly with the aim that the target can not use their right to access network service. Many studies related to the DDoS attack detection methods, but most studies are limited to use the algorithms that often do not provide optimal detection results when tested using various forms of different datasets. So we need a detection system that consists of several algorithms that can adapt to various forms of datasets in order to provide optimal detection results.

In this final project the system grouping the type of attack and simultaneously labeling that attack. Grouping data using algorithms Minkowski Weighted K-means that make data group based on the similarity. Data that have been grouped will be classified and labeled using Decision Tree.

Based on Minkowski Weighted K-means algorithm adaptability in grouping the data into several clusters optimized and augmented the effectiveness of Decision Tree method in process of labeling each processed data, accuracy obtained in this research was 94.78% with 0.26% false positive rate and 99.98% detection rate. By comparison of multiple scenarios shows that the *p*-value for the Minkowski distance that gives the most optimal result was 2.5.

Keyword: *Minkowski Weighted K-means, Decision Tree, DDoS*

1. Pendahuluan

Internet adalah sebuah sistem dari seluruh jaringan komputer yang saling terhubung. Internet telah mengalami banyak sekali perkembangan dan telah menjadi salah satu kebutuhan bagi masyarakat. Dalam sebuah survey yang dimuat dalam [1] terkait penggunaan internet di masyarakat menunjukkan bahwa pengguna internet di dunia telah mencapai 1,5 Milyar jiwa atau sekitar 20% dari 6,5 Milyar penduduk bumi.

Seiring dengan perkembangan internet yang semakin pesat, berdampak pula pada tipe serangan yang semakin beragam. Namun dengan adanya suatu sistem keamanan pada jaringan, salah satunya Firewall yang dapat mendeteksi dan mengklasifikasi tipe serangan tersebut sehingga dapat memudahkan untuk menentukan solusi yang paling tepat mengatasi tiap tipe serangan.

Pada penelitian ini lebih memfokuskan pada pendeteksian dan pengklasifikasian tipe serangan pada trafik jaringan. Metode yang digunakan dapat membedakan antara trafik normal dan trafik serangan serta mengklasifikasikan tipe dari serangan tersebut. Fitur yang dijadikan acuan dari penelitian ini adalah fitur pada KDDCUP 1999 yang telah sering digunakan dalam berbagai penelitian. Metode yang digunakan untuk pendeteksian adalah clustering, sedangkan metode untuk menentukan tipe serangan adalah classification. Algoritma yang digunakan untuk clustering adalah Minkowski Weighted K-Means, sedangkan algoritma untuk classification adalah Decision Tree.

2. Dasar Teori

2.1 Intrusion Detection System

P. Jongsuebsuk et al. [3] menjelaskan bahwa Intrusion Detection System (IDS) merupakan sebuah sistem yang dapat mendeteksi serangan yang terjadi dalam jaringan. Algoritma Intrusion Detection System dibagi menjadi dua jenis yaitu supervised learning yang memiliki detection rate yang tinggi karena menggunakan database yang menyimpan pola serangan namun memiliki kekurangan yaitu database harus selalu update agar dapat mengenali serangan dan unsupervised learning yang dapat melakukan deteksi jenis serangan tanpa informasi dari data training namun biasanya menghasilkan detection rate yang rendah [3].

2.2 Distributed Denial of Service

Rui Zhong et al.[4] Distributed Denial of Service (DDoS) adalah jenis serangan yang dilakukan secara masif dengan tujuan mengganggu hak akses pengguna jaringan. DDoS merupakan serangan flooding trafik yang dilakukan dengan sengaja untuk mengganggu QoS dari sistem jaringan yang bertujuan untuk membuat sumber daya server habis. Serangan DDoS pada dasarnya sama dengan serangan DoS namun serangan dilakukan dengan banyak sumber secara serentak. Untuk melancarkan serangan DDoS, penyerang biasanya mengumpulkan pasukan dengan cara mengambil alih komputer-komputer yang kemudian dijadikan zombie yang merupakan komputer yang siap diperintah dan dikendalikan oleh botnet [6].

2.3 Network Simulator – 3 (NS-3)

Network Simulator (Version 3) merupakan network simulator yang pada pengembangannya ditargetkan untuk riset dan pendidikan. Proyek NS-3 mulai dikembangkan pada tahun 2006 dan merupakan proyek yang bersifat open source yang diatur oleh komunitas peneliti dan pengembang. Saat ini penggunaan NS-3 hanya terbatas pada sistem operasi Linux dan MacOS. NS-3 selain dikembangkan dalam bentuk C++, dapat juga menggunakan script python dan juga dapat diintegrasikan dengan software wiresharks dan pcap untuk melakukan analisis pada trace file. Berdasarkan fungsinya sebagai tools simulasi jaringan, NS-3 mempunyai model-model untuk elemen jaringan yang biasanya terdapat pada jaringan real [11].

2.4 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [12]. Salah satu metode data mining yang digunakan adalah metode klasifikasi untuk menyatakan suatu data yang kelas atau kategorinya belum diketahui ke salah satu kelas yang telah didefinisikan sebelumnya berdasarkan atribut-atribut dari data tersebut[15].

2.5 Minkowski Weighted K-Means

Algoritma Minkowski Weighted K-Means merupakan algoritma yang mengelompokkan data yang memiliki kemiripan dan memetakannya ke dalam cluster yang sama. Algoritma Minkowski Weighted K-means merupakan metode clustering dengan kategori partitioning method [13]. Clustering pendekatan partisi merupakan metode yang mempartisi data ke dalam satu atau lebih kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kelompok yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [14]. Proses pada Algoritma Minkowski Weighted K-means hampir sama dengan proses clustering pada algoritma K-means, namun dibedakan pada metode untuk menentukan jarak antara entitas dengan centroid menggunakan Minkowski Distance, update centroid dan ditambahkan step untuk update weighted pada fitur [9]. Beberapa modifikasi yang dilakukan terletak pada penambahan formulasi untuk menghitung nilai bobot setiap fitur pada masing-masing cluster dan bebasnya penentuan nilai dari eksponen (p)

pada perhitungan jarak antara data dan perhitungan feature weighting. Berikut formulasi distance space dari minkowski weighted k-mean, dengan y dan c merupakan data yang akan dihitung jaraknya, w merupakan nilai feature weighted pada masing-masing cluster, dan p merupakan eksponen minkowski [10].

$$D(y, c) = \sum_{i=1}^V w_i |y_i - c_i|^p \quad (1)$$

Dimana V merupakan fitur untuk masing-masing data yang diolah, y_i merupakan nilai masing-masing data yang dihitung jaraknya dengan pusat cluster, c_i merupakan nilai centroid masing-masing cluster, $D(y, c)$ merupakan nilai jarak antara data dengan centroid masing-masing cluster, w_i merupakan nilai bobot setiap fitur data pada masing-masing kluster dengan menggunakan nilai pangkat minkowski distance metric, dan $|y_i - c_i|^p$ merupakan perhitungan untuk jarak antara data dengan centroid masing-masing kluster menggunakan nilai pangkat minkowski distance metric.

2.6 Minkowski Distance

Minkowski distance merupakan bentuk umum dari rumus untuk menghitung distance space atau jarak antar dua titik. Faktor yang membedakan antara Minkowski Distance dengan perhitungan distance space lainnya terletak pada nilai eksponen (p) yang digunakan dalam perhitungan distance space.

Beberapa perhitungan distance space yang digunakan diantaranya apabila nilai eksponen (p) = 1 disebut dengan Manhattan Distance, apabila nilai eksponen (p) = 2 disebut dengan Eclidean Distance [7]. Apabila nilai eksponen (p) ditentukan oleh pengguna sesuai dengan kebutuhan data yang akan diolah biasa disebut Minkowski Distance dan juga merupakan algoritma yang digunakan pada penelitian ini.

$$D(y, c) = \sqrt[p]{\sum_{i=1}^P |y_i - c_i|^p}, p > 0 \quad (2)$$

Dimana P merupakan nilai pangkat untuk minkowski distance metric dengan syarat $p > 0$, i merupakan variabel untuk menyatakan data ke- i , y_i merupakan nilai untuk masing-masing pusat cluster, c_i merupakan nilai dari masing-masing data yang dihitung jaraknya dengan pusat cluster, $|y_i - c_i|^p$ merupakan nilai jarak antara data dengan pusat cluster menggunakan nilai pangkat minkowski distance metric.

2.7 Feature Weighting

Salah satu modifikasi yang dilakukan pada algoritma k-means yang juga akan digunakan sebagai algoritma clustering pada penelitian tugas akhir ini adalah dengan menambahkan feature weighting untuk tiap cluster. Feature weighting dapat diformulasikan seperti pada gambar [on init]. Dengan D_{kv} merupakan persebaran data pada setiap cluster berdasarkan fiturnya, w_{kv} merupakan nilai bobot tiap fitur untuk masing-masing cluster, dan p merupakan eksponen minkowski [10].

$$w_{kv} = \frac{1}{\sum_{i=1}^V [D_{kv}]^{(p-1)}} \quad (3)$$

Dimana V merupakan fitur pada masing-masing data yang diolah, P merupakan nilai pangkat untuk minkowski distance metric, w_{kv} merupakan nilai bobot untuk setiap fitur data pada masing-masing cluster, D_{kv}

merupakan persebaran data untuk setiap fitur pada masing-masing kluster, σ_k merupakan persebaran semua data pada masing-masing kluster.

$$\sigma_k = \sum_{x \in S_k} |x - \mu_k|^p \quad (4)$$

Dimana σ_k merupakan persebaran data untuk setiap fitur pada masing-masing kluster, k merupakan variabel untuk menyatakan kluster, p merupakan variabel untuk menyatakan data ke- i , $|x - \mu_k|^p$ merupakan pangkat minkowski distance metric.

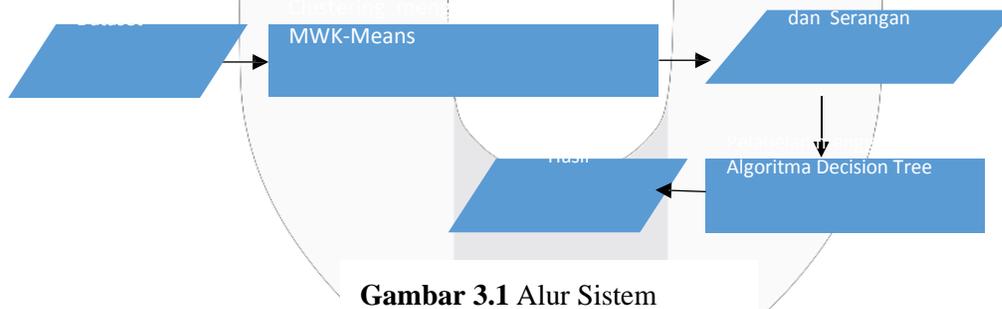
2.8 Decision Tree

Hangga Buwana et al.[5] memaparkan bahwa Decision Tree merupakan salah satu metode data mining yang banyak diterapkan sebagai solusi untuk mengklasifikasikan masalah. Decision Tree merupakan salah satu klasifikasi pada data mining [18]. Metode klasifikasi pada Decision Tree menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas [17]. Pembuatan Decision Tree sendiri menggunakan metode supervised learning yaitu proses pembelajaran dimana data baru diklasifikasikan berdasarkan training samples yang ada [5]. Sebuah Decision Tree terdiri dari root, internal node, dan leaf [2]. Konsep yang sering digunakan dalam penentuan root, internal node dan leaf pada Decision Tree berupa konsep entropy dan konsep gini [16]. Pada penelitian ini dalam menentukan root, internal node dan leaf pada Decision Tree menggunakan perhitungan Gini index dan Gini splitting index [19].

3. Pembahasan

3.1 Deskripsi Sistem

Pada penelitian tugas akhir mengikuti alur sistem seperti gambar berikut:



Gambar 3.1 Alur Sistem

3.2 Dataset Hasil Preprocessing PCA Dan Sampling

Dataset ini merupakan hasil dari preprocessing raw data menggunakan metode PCA dan sampling. Fitur yang dijadikan acuan pada dataset ini adalah beberapa fitur pada KDDCup 1999

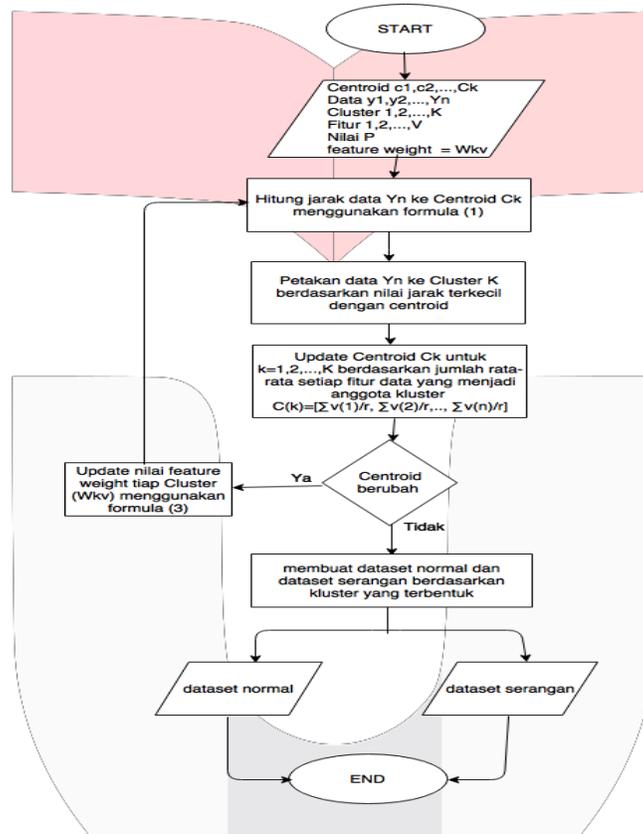
3.3 Dataset KDDCup 1999

Pada penelitian tugas akhir ini dataset 10% KDDCup 1999 hanya akan digunakan sebagai data training pada proses klasifikasi. Pada dataset ini masing – masing paket sudah memiliki label. Label yang akan difokuskan pada penelitian tugas akhir ini adalah label data normal dan label data serangan (Neptune, Back dan Smurf). Penggunaan KDDCup 1999 sebagai referensi untuk dataset dikarenakan dataset ini memiliki fitur dan label yang dibutuhkan untuk membangun sistem deteksi dari penelitian yang dilakukan serta dataset ini sudah banyak digunakan dalam berbagai penelitian. Pada dataset KDDCup 1999 hanya beberapa fitur saja yang akan diambil

dan dijadikan acuan pada penelitian ini dianggap fitur yang paling optimal untuk dijadikan sebagai acuan dari sistem yang dibangun.

3.4 Minkowski Weighted K-Means

Pada tahap clustering, data yang diolah oleh algoritma merupakan data yang sudah melalui preprocessing PCA dan sampling. Proses clustering dimulai dengan melakukan inialisasi centroid sesuai dengan banyak cluster yang telah ditentukan. Setiap data yang masuk akan dihitung jaraknya dengan setiap centroid menggunakan minkowski distance. Setelah semua data selesai dihitung jaraknya, maka jarak terkecil antara data dengan salah satu centroid akan menjadikan data sebagai anggota dari centroid tersebut. Setelah itu centroid tiap cluster akan diperbarui dan bobot fitur pada tiap cluster juga diperbarui. Semua proses akan diulang kembali sampai nilai dari tiap centroid tidak berubah.



Gambar 3.2 Flowchart Minkowski Weighted K-means

3.5 Decision Tree

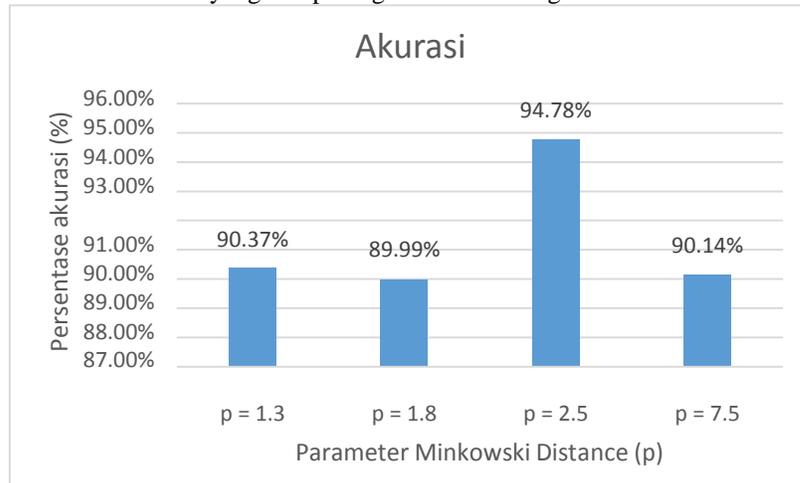
Pada metode klasifikasi Decision Tree, sebelum melakukan klasifikasi terhadap dataset hasil clustering dibutuhkan dataset untuk membuat pola metode klasifikasi yang optimal [8]. Pada penelitian ini akan menggunakan KDDCup 1999 sebagai acuan fitur metode klasifikasi Decision Tree. Decision Tree yang terbentuk dari dataset training akan menggambarkan pola dari metode klasifikasi yang nantinya akan di uji. Pola metode klasifikasi yang telah terbentuk akan diuji menggunakan dataset hasil clustering yang berupa sekumpulan data yang belum memiliki label atau kelas.

4. Analisis

Pada penelitian tugas akhir ini untuk mengukur seberapa akurat sistem deteksi yang dibuat dalam mengelompokkan dan mengklasifikasikan sekumpulan data yang belum memiliki label atau kelas akan menggunakan confusion matrix untuk menghitung tingkat *accuracy*, *false positive rate*, dan *detection rate*. Data

yang diuji yaitu 295.722 data yang diuji menggunakan beberapa skenario. Hasil dapat dilihat pada grafik dibawah ini.

Proses dilakukan dengan *me-generated* data dari *ns-3* dan dipreprocessing dengan acuan fitur dari dataset KDDCup1999. Pada tahap pertama dalam pendeteksian adalah proses *clustering*, dan selanjutnya tahap klasifikasi dan pelabelan. Hasil akurasi yang didapat digambarkan oleh gambar 4.2.



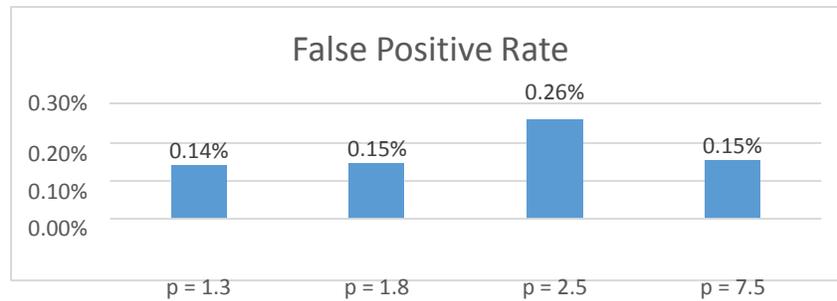
Gambar 4.2 Akurasi

Akurasi yang diperoleh berbeda untuk setiap skenario yang dibedakan pada nilai parameternya. Parameter yang berbeda sangat berpengaruh pada proses pendeteksian menggunakan metode *clustering*.



Gambar 4.3 Detection rate

Pada grafik tingkat deteksi dapat terlihat semua skenario menghasilkan tingkat deteksi yang relative sama, dikarenakan proses pendeteksian dan pelabelan setiap skenario sudah mampu dengan baik dalam melakukan pendeteksian walaupun masih adanya kesalahan dalam mendeteksi paket normal sebagai paket serangan.



Gambar 4.4 False Positive Rate

Dari perbandingan antara hasil yang didapat dari masing - masing skenario, tingkat akurasi skenario pertama yang memiliki nilai paling tinggi sebesar 94.78 %. Nilai akurasi tersebut didapatkan karena sistem deteksi sudah mampu dengan baik mendeteksi dan melabelkan data inputan. Tingkat pendeteksian sistem juga yang tinggi dan relatif sama dengan skenario yang lain terlihat dari nilai TP yang lebih tinggi dari nilai FP dengan persentase sebesar 99.98 %. Namun masih terdapat kesalahan dalam pendeteksian data aktual normal yang dideteksi sebagai data serangan yang dapat dilihat pada nilai FP yaitu sebanyak 546 data dan *false positive rate*. Dari keempat skenario didapatkan analisis akhir bahwa sistem mendapatkan hasil deteksi yang optimal saat nilai $p = 2.5$ untuk parameter minkowski distance.

5. Kesimpulan dan saran

5.1 Kesimpulan

Dari hasil yang didapatkan pada penelitian ini, dapat ditarik beberapa kesimpulan sebagai berikut :

1. Akurasi deteksi yang diperoleh dengan menggunakan algoritma Minkowski Weighted K-means dan Decision Tree adalah 94,78%. Jadi dapat disimpulkan kedua algoritma ini sudah mampu dengan baik dalam mendeteksi dan melabelkan tiap data normal dan data serangan
2. Pemilihan fitur yang tepat sangat mempengaruhi hasil dari deteksi ini, karena setiap serangan memiliki fitur yang mencirikan serangan tersebut.
3. Pemilihan nilai p yang juga menjadi fokus analisa pada minkowski distance juga mempengaruhi hasil dari deteksi menggunakan algoritma Minkowski Weighted K-means.

5.2 Saran

Saran untuk penelitian selanjutnya adalah :

1. Penggunaan fitur sebagai inputan sebaiknya disesuaikan dengan tipe serangan yang akan dideteksi agar dapat mengoptimalkan hasil deteksi.
2. Pada penelitian ini hanya mengambil 295.722 data uji yang belum bersifat data stream. Diharapkan penelitian selanjutnya mampu mengolah data stream sebagai inputan pada proses deteksi.
3. Pada penelitian ini, hanya menggunakan 3 tipe skenario dengan fokus skenario pada perubahan nilai p untuk Minkowski Weighted K-means. Diharapkan pada penelitian selanjutnya dapat memmperbanyak tipe skenario atau mengubah fokus pengujian.

DAFTAR PUSTAKA

- [1] Musriha, Gilang R, "Pengaruh Intensitas Pemakaian Internet Terhadap Penggunaan Internet Untuk Berbelanja Online yang Dimoderasi Oleh Consumer Innovativeness Di Surabaya", 2014.
- [2] S.A Ali, N.Sulaiman, A. Mustapha, N. Mustapha, "K-Means Clustering to Improve the Accuracy of Decision Tree Response Classification", Information Technology Journal 8(8): 1256-1262, 2009 ISSN 1812-5638, 2009.

- [3] P. Jongsuebsuk, N. Wattanapongsakorn, C.Charnsripinyo, "Network Intrusion Detection with Fuzzy Genetic Algorithm for Unknown Attacks", 978-1-4673-5742-5/13/\$31.00 ©2013 IEEE, 2013.
- [4] Rui Zhong, Guangxue Yue, "DDoS Detection System Based on Data Mining", ISBN 978-952-5726-09-1 (Print) Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10) Jingtangshan, P. R. China, 2-4, April. 2010, pp. 062-065, 2010.
- [5] Hangga Buwana B.W, Dylan Keke, Trastoto P, Nanang Arfandi, Ariefan, Rian Chikita, Wahyudi Wibowo, Hendri Kurniawan, Agus Dwi, "Credit Scoring Menggunakan Metode Hybrid Kombinasi K-Means Cluster Dan Algoritma C4.5", Universitas Gadjah Mada, Yogyakarta, 2012.
- [6] Y. Purwanto, Kuspriyanto, Hendrawan dan B. Rahardjo, "Traffic Anomaly Detection in DDoS Flooding Attack," International Conference on Telecommunication, System, Services, and Application, vol. 8, pp. 313-318, 2014.
- [7] Made Indra Wira Pramana, Yudha Purwanto, Fiky Yosef Suratman, "Perancangan dan Analisis Deteksi Anomaly Berbasis Clustering Menggunakan Algoritma Modified K-Means dengan Random Initialization pada Landmark Window," Universitas Telkom, 2015.
- [8] Trinita S.P., Yudha Purwanto, Tito Waluyo Purboyo, "Analisis Metode Covariance Matrix Menggunakan Teknik Sliding Window Untuk Sistem Deteksi Anomali Trafik," Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom, 2015.
- [9] R.C. de Amorim and B. Mirkin. "Minkowski Metric, Feature Weighting and Anomalous Cluster Initializing in K-Means Clustering". In: Pattern Recognition 45.3, pp. 1061–1075, 2012.
- [10] R.C. de Amorim and P. Komisarczuk. "On Initializations for the Minkowski Weighted K-Means". In: Lecture Notes in Computer Science 7619 , pp. 45–55, 2012.
- [11] Dedy Irawan dan Rusdianto Roestam, "Simulasi Model Jaringan Mobile Ad-hoc (MANET) Dengan NS-3," 2011.
- [12] Mujib Ridwan, Hadi Suyono, dan M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," 2013.
- [13] Angger Kartayasa Pribadi Putra, Yudha Purwanto, Astri Novianty, "Analisis Sistem Deteksi Anomali Trafik Menggunakan Algoritma CURE (Clustering Using Representatives) dengan Koefisien Silhouette dalam Validasi Clustering," 2015.
- [14] Yudi Agusta, "K-Means – Penerapan, Permasalahan dan Metode Terkait," 2007.
- [15] Sunjana, "Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree," 2010.
- [16] Dwi Widiastuti, "Analisa Perbandingan Algoritma SVM, NAIVE BAYES, dan DECISION TREE dalam Mengklasifikasikan Serangan (ATTACKS) Pada Sistem Pendeteksi Intrusi," 2007.
- [17] Anik Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," 2012.
- [18] Tan, Steinbach, Kumar, "Introduction to Data Mining," 2006.
- [19] N.Suneetha, CH.V.M.K.Hari, V.Sunil Kumar, "Modified Gini Index Classification: A Case Study On Heart Disease Dataset". In: International Journal on Computer Science and Engineering Vol. 02, No. 06, 2010, 1959-1965, 2010.