Analisis Sistem Deteksi Anomali Trafik Menggunakan Algoritma Clustering Isodata (Self-Organizing Data Analysis Technique) Dengan Euclidean Distance

Analysis Of Traffic Anomaly Detection System Using Isodata Clustering Algorithm (Self-Organizing Data Analysis Technique) With Euclidean Distance

Putu Ananda Kusuma Wiradharma[1], Yudha Purwanto[2], Tito Waluyo Purboyo[3]

[1,2,3]Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

[1]anandakusuma@student.telkomuniversity.ac.id, [2]omyudha@telkomuniversity.ac.id, titowaluyo@telkomuniversity@telkomuniversity.ac.id

Abstrak

Berkembangnya teknologi internet telah meningkatkan jumlah aktivitas masyarakat terhadap penggunaan internet. Seiring meningkatnya jumlah *user* mengakses internet memicu adanya fenomena *anomaly traffic.* Fenomena-fenomena *anomaly traffic* dapat berupa serangan *Distributed Denial of Service* (DDoS) dan *Flash crowd*. Menimbang dari dampak negatif yang diterima dari fenomena *anomaly traffic* tersebut, dirasa penting membangun metode deteksi yang dapat membedakan *flash crowd* dan serangan DDoS. Pada penelitian ini dibangun sebuah metode *Intrusion Detection System* (IDS) dengan teknik *unsupervised learning* yang menggunakan algoritma Isodata *clustering*. Pada penelitian ini menggunakan metode *euclidean distance* untuk rumus pengukuran jarak dan metode *dunn index* untuk melihat kualitas cluster yang dihasilkan. Hasil dari penelitian ini, sistem yang dibagun dapat bekerja dengan baik dalam deteksi dan membedakan antara *traffic* normal dan *traffic anomaly*. Dibuktikan dengan penggunaan metode *Euclidean distance* mengahasilkan performansi algoritma yang baik dibandingkan dengan penggunaan metode *manhattan distance.*

Kata Kunci : ***Anomaly Traffic*, DDoS, *flash crowd*, Isodata, *Clustering*, *Euclidean Distance***

Abstract

Development of Internet technology now has increased the number of community activities to the use of the internet. With the increasing number of users accessing the Internet triggered the phenomenon of traffic anomalies. Traffic anomaly phenomena such as Distributed Denial of Service (DDoS) attack and flash crowd. Considering the negative impacts received from menomena the traffic anomalies, deemed important to build detection method that can distinguish DDoS attacks and flash crowd. In this research constructed a method Intrusion Detection System (IDS) with unsupervised learning techniques that use ISODATA clustering algorithm. in this research using euclidean distance method for distance measurement formulas and dunn index methods to see the quality of the cluster. Results from this research, which was built systems can work well in detection and distinguish between normal traffic and anomaly traffic. Evidenced by using Euclidean distance methods get better result of performance compared system using manhattan distance methods.

Keywords: traffic anomalies, DDoS, flash-crowd, Isodata, Clustering, Euclidean Distance

1. Pendahuluan

Distributed Denial of Service (DDoS) adalah suatu jenis serangan terhadap sebuah komputer atau server dengan salah satu cara membanjiri lalulintas jaringan dengan banyak permintaan (*request flooding*) sehingga tidak dapat diakses oleh *user* yang berhak. *Flash crowd* merupakan situasi terjadinya sebuah peningkatan trafik yang sangat tinggi dalam suatu jaringan sehingga tidak dapat diakses dalam rentang waktu tertentu. Fenomena *anomaly traffic* ini sering merugikan dari segi penyedia jasa maupun masyarakat.

Dalam penelitian deteksi *anomaly traffic*, informasi 5-*tuple* dari *IP* menjadi karakteristik acuan analisis dalam pemrosesan, diantaranya *protocol type*, *source IP addres*, *destination IP address*, *source port*, dan *destination port*, serta jumlah paket dan besarnya paket menjadi salah satu acuan data pemrosesan [1] [2]. Dalam penelitian sebelumnya [2], dengan metode *data mining* dengan algoritma K – Mean dalam mendeteksi *anomaly traffic* dan berhasil membedakan *traffic* normal dan serangan DDoS akan tetapi belum menunjukan hasil yang signifikan. Kekurangan algoritma k – mean ini sangat sensitif terhadap *outlier* dalam *data* [3] [4] serta sering menghasilkan *cluster* yang memanjang [4].

Dalam survey [4], algoritma Isodata dapat mengatasi kelemahan yang terdapat pada k – mean terhadap outlier, serta dengan proses pembelahan yang terdapat dalam algoritma isodata dapat mengatasi *cluster* yang memanjang yang sering dihasilkan oleh algoritma k – mean [4]. Algoritma Isodata ini merupakan perkembangan dari algoritma K-means dengan penambahan proses pengabungan *cluster* dan pembelahan *cluster* [5] [6] [7] [8] [9]. Dengan metode ini kepadatan suatu cluster dapat dikontrol dengan algoritma. Dalam Algoritma ini juga terdapat masukan – masukan yang mendukung [5] [6] untuk melakukan prosesnya.

Pada tugas akhir ini kami akan menggunkan algoritma Isodata dan menggunkan teknik pengukuran jarak *Euclidean Distance* [10] [11]. Fokus penelitian tugas akhir ini menerapkan algoritma Isodata ke dalam sistem deteksi *anomaly traffic* dan penggunaan metode pengukuran *euclidean distance* untuk mempercepat komputasi proses dalam sistem deteksi *anomaly traffic*, serta penggunaan metode *dunn index* untuk melihat kualitas *cluster* yang dihasilkan.

2. Dasar Teori dan Perancangan

2.1. Sistem Deteksi Anomaly

Dalam sistem deteksi *anomly traffic*, terdapat dua istilah yang sering muncul yaitu *Itrusion Detection System* (IDS) dan *Intrusion Prevesion System* (IPS). Pada sistem IDS, sistem harus memonitor atau mengawasi terlebih dahulu aliran trafik dan jika mengalami sererangan sistem akan memberikan tanda yang berupa alarm sehingga sistem akan menindak lanjuti secara manual oleh operator. Sistem IPS merupakan pengembangan dari sistem IDS dengan kemampuan sistem memonitor atau mengawasi aliran *traffic* dan jika mengalami serangan akan langsung ditindaklanjuti secara otomatis [1] [12].

2.2. Distributed-Denial of Service (DDoS)

Denial of Service (DoS) adalah salah satu serangan yang ditakutkan di dunia internet. Dalam sebuah serangan *Denial of Service* (DoS) penyerang (*hacker*) akan mencoba untuk mencegah komputer lain atau client untuk mengakses suatu komputer atau jaringan dengan beberapa cara atau teknik. Sebuah seranagn DDOS menggunkan banyak komputer unktuk melancarakan seranagn DOS yang terkoordinasi ke satu tarket atau lebih. Menggunakan metoda *client / server*, para hacker mempu memperbanyak efektifitas serangan DOS secara signifikan dengan memanfaatkan sumber daya / *resource* [1] komputer danpa disadari oleh sistem tersbut dan sumber daya komputasi (proses, *memory*, *buffer*) pada *server* atau *node* jaringan untuk membuat sistem pengolahan kehabisan sumber daya yang akhirnya membuat c*rash/down* sehingga tidak dapat melayani servis yg diminta *user* [1]. Sasarannya pada *bandwitdh* / *link* membuat sumber daya *bandwitdh* menjadi penuh . Biasanya program master DDOS diinstal pada satu komputer saja dengan menggunkan *account* curian. Pada waktu yang ditentukan program master akan berkomunikasi dengan agent program lainnya, yang di instal pada komputer lainnya di internet. Pada saat menerima perinta agent program akan memulai serangan. Dengan metologi *client / server*, program master akan dapat memlulai ratusan atau bahkan ribuan program *agent* dalam hitungan detik [13].

2.3. Flashcrowd

Flash Crowd bukanlah merupakan suatu serangan seperti *Denial of Service* (DoS) ataupun *Distributed Denial of Service*(DDoS) melainkan dimana situasi terjadinya sebuah peningkatan trafik yang sangat tinggi dalam suatu jaringan sehingga tidak dapat diakses dalam rentang waktu tertentu [1]. Peningkatan trafik ini terjadi karena *user* yang mengakses sebuah jaringan tersebut sangat banyak. Kejadian *flashcrowd* ini dapat terjadi kapan saja, karena peningkatan akses secara dramatis/tinggi ke suatu server dipengaruhi dari suatu kejadian seperti bencana alam, peluncuran produk, *breaking news*, dll [1].

2.4. Isodata Clustering

Clustering merupakan salah satu teknik pengelompokan data berdasarkan kesamaan karakteristik data. Clustering-based memiliki beberapa tipe penting, diantaranya *Partitional Clustering*. *Partitional Clustering* merupakan pembagian data ke dalam sebuah himpunan data (*cluster*) yang tidak overlap sedemikian setiap data berada dalam satu *cluster* saja. Terdapat beberapa Algoritma dalam *partitional clustering* ini, diantaranya Algoritma isodata Algoritma. Isodata (*Self-Organizing Data Analysis Technique*) deperkenalkan oleh Ball, Hall dkk pada sekitar tahun 1960an ialah *clustering* berbasis *unsupervised learning* algoritma yang pengembangan dari algoritma *K-Means*. Dalam algoritma isodata terdapat proses pembagian, penggabungan, dan penghapusan *cluster*, algoritma isodata mampu mengatur jumlah *cluster* fleksibel dan mengcontol kepadatan suatu *cluster* [6] [5] [8] [9]. Dalam algoritma terdapat masukan sistem yang mendukung kenerja algoritma ini, masukan tersebut ialah jumlah cluster awal (*k*), banyaknya iterasi (*i*), maksimum jarak untuk

melakukan pengabungan (*minjar*), maksimum variansi (*var*) dalam melakukan pembelahan, serta minimum anggota sebuah *cluster* (*minjum*), penjelasan masukan sistem dapat dilihat pada Tabel 1. Dalam mengunakan algoritma isodata data tidak perlu terdistribusi normal. Jika iterasi yang ditetapkan cukup, algoritma clustering isodata ini mudah untuk menemukan *cluster* yang benar dalam data. Namun, lebih banyak waktu komputasi yang dibutuhkan.

Tabel 1 Masukan Sistem

	Masukan	Keterangan
1	Jumlah *Cluster* (*k*)	Menentukan berapa banyak jumlah *cluster* awal untuk memulai sebuah algoritma
2	Iterasi (*i*)	Menentukan berapa banyak jumlah Iterasi
3	Minimal jumlah data dalam satu *cluster* (*minjum*)	Menentukan jumlah minimal data dalam satu *cluster*, parameter ini digunakan dalam pengoprasian proses penghapusan *cluster* yang mimiliki jumlah data yang kurang dari jumlah minimal.
4	Minimal jarak antara *Cluster* (*minjar*)	Menentukan jarak minimal *centroid* / titik pusat *cluster*, parameter ini digunakan dalam pengoprasian proses pengabngan *cluster* jika memiliki jarak dibawah minimal
5	Minimal Variansi (*var*)	Menentukan angka minimal variansi satu *cluster*, parameter ini digunakan dalam pengoprasian proses pembelahan satu *cluster* menjadi 2 buah *cluster* baru

2.5. Euclidean Distance

Pada pnelitian sebelumnya [2] mendeteksi anomal mendeteksi *anomaly traffic* menggunakan euclidean distance sebagai rumus pengukuran jaraknya, mendapatkan hasil cang cukup baik dalam perhitungannnya. Dalam matematika *Euclidean Distance* (ED), diibaratkan garis lurus yang terhubung antara dua buah objek. *Euclidean Distance* sering digunakan dalam pengukuran jarak atar dua buah objek [11]. Similaritas yang dihitung dengan menggunakan *Euclidean distance* diperoleh dengan mendapatkan nilai terendah. Dua objek yang dibandingkan dan dihitung dengan menggunakan *Euclidean distance* dapat dikatakan mirip jika nilai yang didapatkan adalah nilai paling rendah bahkan mendekati 0 [14].

$$d = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (1)$$

2.6. Dunn Index

Pada penelitian sebelumnya [15] [16], *Dunn Index* digunakan untuk validasi dan pelabelan sebuah *cluster*. Parameter yang digunakan dalam proses dunn index nilai jarak *inter cluster* dan nilai jarak *intra cluster*. Validasi *cluster* dikatakan baik jika nilai dari *dunn index* yang didapatkan tinggi. *Dunn index* mempunyai renta nilai dari 0 samapi ∞.

Formulasi dari *dunn index* $D = \min_{i=1,\dots,n} \left\{ \min_{j=i+1,\dots,n} \left(\frac{d(i,j)}{\max_{k=1,\dots,n} d'(k)} \right) \right\}$ (2)

2.7. Dataset DARPA 1999 dan World Cup 1998

Pada penelitian ini menggunakan *dataset* yang sudah sering digunakan pada peneletian serupa sebelumnya. Untuk pengujian *traffic* DDoS digunakan *dataset* DARPA 1998 [17] dan untuk pengujian *traffic flash crowd* [18] digunakan dataset World Cup 1998, kedua *dataset* tersebut sudah menjalani proses *preprocessing* agar mudah dianalisia. Untuk pengujian metode IDS yang dirancang, dilakukan simulasi menggunakan bahasa pemrograman Java

2.8. Parameter Uji

Beberapa parameter yang digunakan digunakan untuk mengetahui sebereapa akuratnya algoritma isodata *clustering* dengan menggunakan *euclidean distance* dalam melakukan pembedaan antara *traffic* normal dan *traffic anomaly*. Beberapa parameter awal yang digunakan dalam mengukur keakuratan algoritma sebagai berikut :

Tabel 2 Matching matrix

ACTUAL	*PREDICTION*	
	ATTACK	*NORMAL*
ATTACK	*TRUE POSITIVE (TP)*	*FALSE NEGATIVE (FN)*
NORMAL	*FALSE POSITIVE (FP)*	*TRUE NAGATIVE (TN)*

True positive (TP) adalah kondisi dimana algoritma mendeteksi data sebagai serangan dan kelanjutan sebenarnya memang data tersebut merupakan serangan. *True nagative* (TN) adalah dimana algoritma mendeteksi data sebagai kondisi normal dan kenyataannya memang data tersebut merupakan data normal. *False positive* (FP) adalah dimana kondisi algoritma mendeteksi data dengan kondisi normal tetapi disebut sebuah serangan. *False negative* (FN) adalah kondisi dimana algoritma melakukan salah deteksi yang menyatakan data dengan kondisi serangan disebut segabagi kondisi normal.

Detection Rate (DR)

Detection rate merupakan presentase yang menyatakan seberapa besar algoritma dapat memberikan true alarm terhadap serangan yang terjadi. Formulasi untuk *detection rate* (DR) sebagai berikut :

$$DR = \frac{TP}{TP + FN} \quad (3)$$

False Positive Rate (FPR)

False positive rate merupakan presentase yang menyatakan seberapa besar kesalahan algoritma memberikan false alarm, dimana algoritma mendeteksi sebuah kondisi serangan yang sebenarnya adalah kondisi normal. Formulasi untuk *false positive rate* (FPR) sebagai berikut:

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

Akurasi (Acc)

Akurasi merupakan presentase yang menyatakan seberapa benarnya algoritma melakukan pendekteksian, serta seberapa besar memisahkan data normal dan data serangan. Formulasi untuk akurasi algoritma sebagai berikut :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

3. Pembahasan

Preprocessing adalah suatu proses untuk normalisasi sebuah data trafik agar mudah/cocok untuk digunakan pada proses pendeteksian. Pada penelitian sebelumnya [19] menggunakan metode *preprocessing* memudahkan menganalisis dan meningkatkan hasil analisis yang dilakukan. Tujuan dalam proses *preprocessing* untuk melakukan mendapatkan fitur yang relevan dari *raw data*, dalam penelitian ini dilakukan pada dataset DARPA 1998. Fitur yang digunakan dapat dilihat pada Tabel 3. Algoritma Isodata dapat dilihat pada Algoritma 1. *Dataset* dalam penelitian ini diberi label untuk mempermudah analisa hasil akhir yang didapatkan.

Tabel 3 Ekstraksi Fitur

Nama Fitur	**Jenis Koneksi**	**Penjelasan**
Count	-	Jumlah *traffic* dalam satu window
IP_source	IP Source dan IP Destination sama	Jumlah *traffic* dari IP Source ke IP Destination yang sama
Protocol		Jumlah protocol yang sama
SYN		Jumlah *traffic* "SYN"
ACK		Jumlah *traffic* "ACK"
Port_Out		Jumlah *traffic* menuju ke port out yang sama
Length		Jumlah *traffic* dengan length yang sama
Different_Source	IP Destination sama	Jumlah *traffic* dengan IP Source berbeda
New_IP	-	Jumlah kemunculan IP baru

Algoritma 1 : Proses Clustering dengan Isodata (*extracted dataset, k, i, minjum, minjar, var*)

1: Masukkan *extracted dataset*
2: Masaukan *k* sebagai jumlah *cluster* awal
3: Masukan *i* sebagai jumlah iterasi
4: Masukan *minjum* sebagai minimal jumlah anggota *cluster*
5: Masukan *minjar* sebagai minimal jarak antara centroid
6: Masukan *var* sebagai minimal variansi
7: Bentuk *k-centroid* secara acak sebanyak *k* buah
8: ***for*** 1 to *x* **do**
9: Hitung jarak x_n ke *k-centroid*
10: Tetapkan x_n ke *k-centroid* terdekat
11: ***end for***
12: Hapus k_{kosong}
13: ***repeat***
14: ***if*** *i* = ganjil ***do***
15: ***for*** 1 to *k* **do**
16: ***if*** jumlah *x* k_n < *minjum* ***do***
17: Hapus k-centroid
18: *x* anggota *k-centroid* tetapkan ke *cluster* terdekat
19: ***end if***
20: Hiting jarak k_n-*centroid* ke k_n-*centroid* lainnya
21: ***if*** jarak k_n-*centroid* < *minjar* ***do***
22: Gabungkan kedua *k-cluster*
23: ***end if***
24: ***end if***
25: ***end if***
26: ***if*** *i* = genap ***do***
27: ***for*** 1 to *k* **do**
28: ***if*** jumlah *x* k_n < *minjum* ***do***
29: Hapus k-centroid
30: *x* anggota *k-centroid* tetapkan ke *cluster* terdekat
31: ***end if***
32: Hiting variansi k_n
33: ***if*** variansi k_n > var ***do***
34: Belah kn menjadi dua buah cluster baru
35 hitung *x* anggota k_n dengan *k* baru
35 Tetapka *x* anggota k_n ke *cluster* baru yang terdekat
23: ***end if***
24: ***end if***
25: ***end if***
26: ***until iterasi terakhir***
27: Menghitung *DR*, *FPR*, dan Akurasi

3.1. Pengujian Dataset DDos

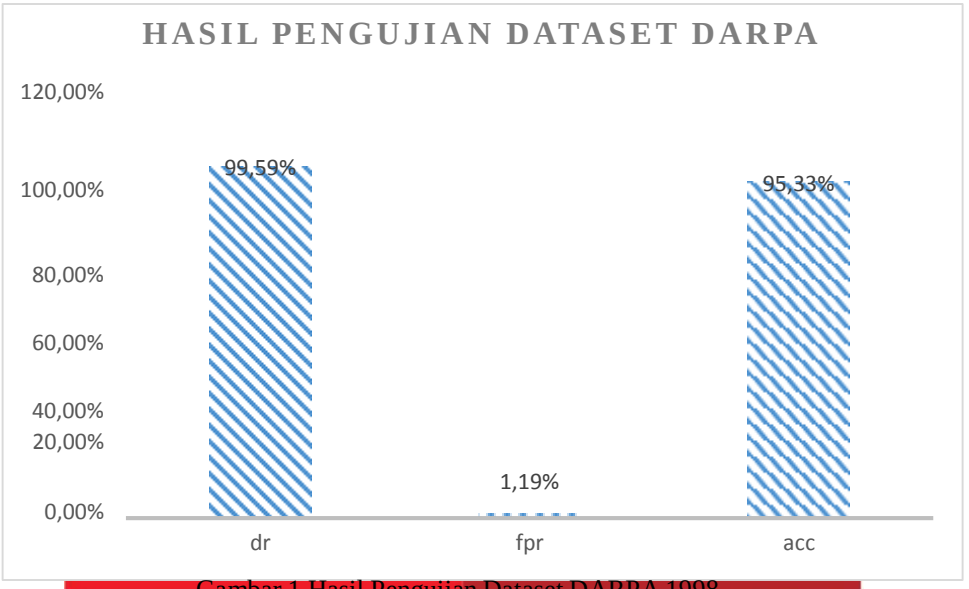
Proses *preprocessing* dahulu dilakukan pada dataset DARPA 1998, dikarenakan dataset darpa masih berupa *raw data*. Tujuan dari proses *preprocessing* untuk medapatkan karekteristik dari *traffic* DDoS sehingga hasil performansi deteksi yang dihasilkan lebih baik. Pengujian dilakukan pada *dataset* normal dan serangan mendapati hasil yang beragam. Hasil pengujian dari *dataset* DARPA 1999 dapat dilihat pada tabel 5 dan gambar 1 dibawah

Tabel 4. Masukan Sistem

Masukan sistem				
k	*i*	*minjum*	*minjar*	*var*
20	2	5000	1000	0,5

Tabel 5. Hasil Analisa *matching matrix* dari dataset Darpa 1998

Aktual	**Prediksi**	
	Serangan	**Normal**
Serangan	Total TP = 1	Total FN = 40866
Normal	Total FP = 248	Total TN = 834356

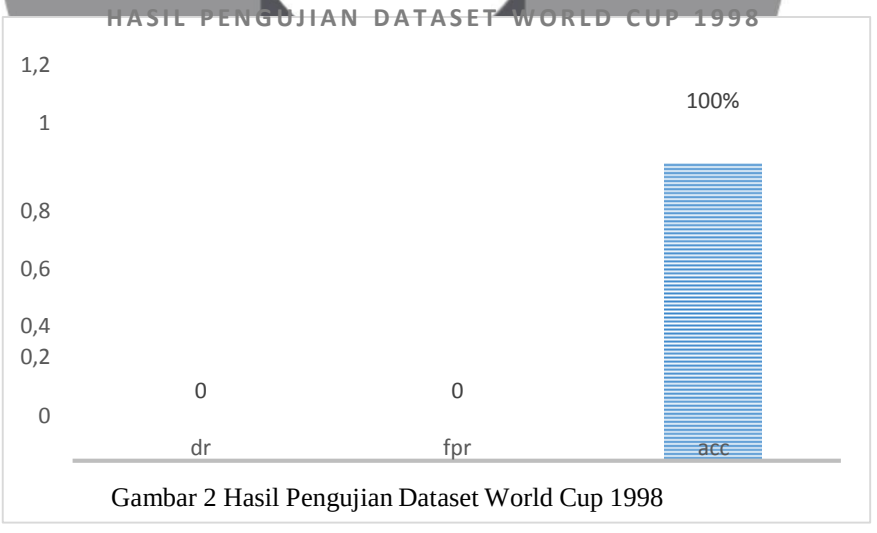


Gambar 1 Hasil Pengujian Dataset DARPA 1998

Pada *dataset* DARPA 1998, setiap pengujian dilakukan dengan masukan sistem yang beragam. Dilihat dari hasil performansi sistem pada gambar 1 diatas dilihat hasil terbaik didapat dengan percobaan dengam masukan sistem yang dapat dilihat pada tabel 4. Percobaan dilakukan berulang ulang pada dataset DARPA 1998 dengan masukan sistem yang bervariansi mendapat perubahan tidak terlalu tinggi, dari percobaan yang dilakukan mendapati semkin tinggi nilai *k* dan *minjar* yang dimasukan dan semakin rendah nilai *minjum*, *i*, dan *var* yang dimasukan ke sistem membuat parameter uji yang dihasilkan menjadi lebih baik walaupun hanya sedikit kenaikannnya, akan tetapi waktu yang dibutuhkan sistem semakin banyak. Rata rata waktu pemrosesan yang dataset DARPA 1998 ialah sekitar 10 detik. Pengujian kualitas *cluster* menggunakan *dunn index* mendapatkan hasil 0.1195, hasil tersebut masih kecil tetapi sistem sudah dapat membedakan *traffic* normal dan *traffic anomaly*.

3.2. Pengujian Dataset *Flash Crowd*

Skenario pengujian dan analisis pada *dataset* World Cup 1998 dilakukan dengan memberi *label* pada *dataset* sehingga lebih mudah untuk mendapatkan hasil akhir. *Dataset* ini hanya memuat dua jenis *traffic* yaitu, *traffic* normal dan *traffic flash crowds*, pelabelan *flash crowds* dilakukan pada saat *traffic* mulai menunjukan kenaikan jumlahnya. Pada pengujian beragamnya niali masukn tidak mempengaruhi hasil akhir, akan tetapi mempengarihi waktu eksekusi dataset oleh sistem deteksi. Dari hasil berbagai percobaan didapat hasil yang digambarkan pada gambar 2. Dibawah



Gambar 2 Hasil Pengujian Dataset World Cup 1998

Dapat dilihat dari gambar 2 diatas, hasil performansi sistem deteksi *anomaly traffic* dengan isodata clustering dan *euclidean distance* dapat bekerja dengan baik, dimana maka *traffic flash crowds* dan *traffic*

normal akan berada dalam satu *cluster* yang sama, sehingga tidak akan membentuk *cluster* dengan jumlah lebih dari satu. Pengujian klualitas cluster menggunakan *dunn index* mengasilkan nilai 0, nilai tersebut berarti terdapat hanya 1 *cluster* yang dihasilkan.

4. Kesimpulan

Kesimpulan dari penelitian ini, sistem deteksi *anomaly traffic* meggunakan metode clustering dengan algoritma isodata dan *euclidean distance* dapat diterapkan, dilihat dari sistem dapat membedakan antara *traffic* normal dan *traffic anomaly*, serta dibarengi dengan performansi algoritma menggunakan *euclidean distance* yang lebih baik dibandingkan dengan penggunaan metode *manhattan distance*. Masukan sistem yang beragam mempengaruhi nilai parameter uji, waktu pemrosesan *data* dan kualitas *cluster* yang dihasilkan. Penggunaan metode *preprocessing* memiliki peran penting dalam mengenali serangan DDoS sehingga dapat menghasilkan hasil yang optimal dinilai dari ketiga parameter uji.

Untuk penembangan sistem deteksi *anomaly traffic* pada penenlitian selanjutnya, modifikasi sistem deteksi *anomaly traffic* bisa dengan menambahkan metode *windowing* seperti *landmark window* atau *sliding window* untuk mempermudah pemrosesan *data* dengan cara pemotongan *data* agar minimnya kesalahan deteksi, serta penggunaan rumus jarak lainnya seperti *mahalobis distance* diharapkan mendapatkan hasil performansi yang lebih baik dan waktu pemrosesan yang lebih singkat.

Daftar Pustaka

[1] Y. Purwanto, Kuspriyanto, Hendrawan dan B. Rahardjo, "Traffic Anomaly Detection in DDoS Flooding," *International Conference on Telecommunication Systems Services and Applications (TSSA)*, vol. 8, pp. 313-318, 2014.

[2] G. Münz, S. Li dan G. Carle, "Traffic Anomaly Detection Using KMeans Clustering," *In GI/ITG Workshop MMBnet*, 2007.

[3] F. A. Hermawati, Data Mining, Yogyakarta: Penerbit Andi, 2013.

[4] R. Xu dan D. Wunsch, "Survey of Clustering Algorithms," *Neural Networks, IEEE Transactions*, vol. 16, no. 3, pp. 645 - 678, 2005.

[5] A. Kohei dan B. XianQianhg, "ISODATA clustering with parameter (threshold for merge and split) estimation based on GA: Genetic Algorithm," *Reports of the Faculty of Science and Engineering, Saga University*, vol. 36, pp. 17-23, 2007.

[6] M. Merzougui, M. Nasri dan B. Bouali, "Image Segmentation using Isodata Clustering with Parameters Estimated by Evolutionary Approach: Application to Quality Control," *International Journal of Computer Applications*, vol. 66, pp. 25-30, 2013.

[7] P. Berkhin , "A Survey of Clustering Data Mining Techniques," dalam *Grouping Multidimensional Data*, Springer Berlin Heidelberg, 2006, pp. 25-71.

[8] N. MEMARSADEGHI, D. M. MOUNT, N. S. NETANYAHU dan J. L. MOIGNE, "A FAST IMPLEMENTATION OF THE ISODATA CLUSTERING ALGORITHM," *Int. J. Comput. Geometry Appl*, pp. 71-103, 2007.

[9] Seok-Woo Jang, Gye-Young Kim dan Siwoo Byun, "Clustering-Based Pattern Abnormality Detection in Distributed Sensor Networks," *International Journal of Distributed Sensor Networks*, 2014.

[10] S. Aggarwal dan J. Singh, "Outlier Detection Using K-Mean and Hybrid Distance Technique on Multi-Dimensional Data Set," *International Journal of Advanced Research in Computer Engineering &Technology (IJARCET)*, vol. 2, no. 9, 2013.

[11] D. Sinwar dan R. Kaushik, "Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering," *NTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)*, vol. 2, no. 5, 2014.

[12] K. Ramadhani, M. Yusuf dan H. E. Wahanani, "Pendeteksian Dini Sserangan UDP Flood Berdasarkan Anomali Perubahan Ttraffic Jaringan Berbais Cusum Algoritm," *Computer security*, 2013.

[13] F. Kargl, J. Maier dan M. Weber, "Protecting Web Servers from Distributed Denial of Service Attacks," dalam *Proceedings of the 10th International Conference on World Wide Web*, ACM, 2001, pp. 514-524.

[14] W. S. J. Saputra dan F. Muttaqin, "PENGENALAN KARAKTER PADA PROSES DIGITALISASI DOKUMEN MENGGUNAKAN COSINE SIMILARITY," *Seminar Nasional Teknik Informatika (SANTIKA) 2013*, pp. 55-56, 2013.

[15] S. Saitta, B. Raphael dan I. F. Smith, "A Bounded Index for Cluster Validity," *MLDM '07 Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition*, pp. 174 - 187, 2007.

[16] F. Kovács, C. Legány dan A. Babos, "Cluster Validity Measurement Techniques," *AIKED'06 Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, pp. 388 - 393 , 2006.

[17] "Cyber System and Technology," Lincoln Laboratory Messachusetts Institute of Technology, 4 December 1998. [Online]. Available: http://www.ll.mit.edu/ideval/data/. [Diakses 23 October 2014].

[18] P. Danzig, J. Mogul, V. Paxson dan M. Schwartz, "WorldCup98," ACM SIGCOMM, [Online]. Available: http://ita.ee.lbl.gov/html/contrib/WorldCup.html.

[19] Made Indra Wira Pramana, Yudha Purwanto dan Fiky Yosef Suratman, "DDoS Detection Using Modified K-Means Clustering with Chain Initialization Over Landmark Window," *International Confrence on Control, Electronics, Renewable Energy, and Communication*, 2015.