Klasifikasi Masalah Intrusion Detection System dengan Menggunakan Metode Fuzzy C-Means dan Laplacian Score

[1]Aini Suri Talita dan [2]Zuherman Rustam

[1]Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Gunadarma
[2]Departemen Matematika, FMIPA, Universitas Indonesia
[1]ainisuri@staff.gunadarma.ac.id, [2]rustam@ui.ac.id

Abstrak

Di era internet of things keamanan sistem maupun jaringan perlu ditingkatkan untuk mengantisipasi kemungkinan adanya serangan. Intrusion Detection System (IDS) adalah alat ataupun perangkat lunak untuk memantau sistem atau jaringan dari kemungkinan adanya serangan atau pelanggaran yang terjadi. Terdapat dua pendekatan umum pada masalah IDS yaitu masalah Misuse Detection dan Anomaly Detection. Pada pendekatan Anomaly Detection, pendeteksian serangan didasarkan pada keberadaan anomaly behavior dengan asumsi normal behavior diketahui sebelumnya. Masalah IDS yang berbasiskan Anomaly Detection dapat diselesaikan menggunakan metode klasifikasi maupun clustering. Akan tetapi besarnya jumlah feature data dalam suatu masalah klasifikasi dapat menyebabkan besarnya biaya perhitungan baik dalam kaitannya dengan waktu komputasi maupun sumber daya memori yang digunakan. Kemungkinan lain adalah adanya sebagian feature yang tidak mewakili data dengan baik ataupun feature yang redundant dan dapat menyebabkan penurunan performa klasifikasi. Untuk mengatasi masalah tersebut peneliti klasifikasi menggunakan metode pemilihan feature data dalam menyelesaikan masalah klasifikasinya. Dalam makalah ini, masalah IDS berbasiskan Anomaly Detection diselesaikan dengan menggunakan metode Fuzzy C-Means dan metode Laplacian Score digunakan sebagai metode pemilihan feature.

Kata Kunci : anomaly detection, fuzzy c-means, intrusion detection system, laplacian score, pemilihan feature

1. Pendahuluan

Internet Detection System (IDS) dapat digunakan untuk mendeteksi adanya serangan pada sistem atau jaringan yang dapat bertujuan untuk mengakses informasi terbatas maupun memanfaatkan sumber daya yang ada, memonitor dan menganalisa aktifitas pengguna, menjaga keutuhan sistam dan data, maupun memberikan respon apabila terjadi serangan. Berdasarkan sumber informasinya, IDS dapat dibagi menjadi dua jenis yaitu Host-based dan Network-based IDS [1]. Host-based IDS memonitor sumber daya seperti pada system logs dan file systems sedangkan network-based IDS memonitor data yang melewati jaringan. Secara umum teknik IDS dapat dibagi menjadi dua berdasarkan kriteria pendeteksiannya, yaitu Misuse Detection (MD) dan Anomaly Detection (AD). MD menggunakan informasi dari serangan terdahulu untuk menentukan apakah suatu kegiatan merupakan serangan atau bukan serangan. Sedangkan AD menggunakan informasi normal behavior yang telah diketahui sebelumnya. Ketika suatu kegiatan memiliki pola yang berbeda dengan normal behavior maka akan dikategorikan sebagai serangan. Masalah AD – IDS merupakan masalah pengenalan pola yang dapat diselesaikan dengan metode klasifikasi dan clustering. Pada penelitian terdahulu Chou et al. [2] menggunakan Decision tree (C4.5) dan metode Naïve Bayes sedangkan Rustam and Talita [3] menggunakan Fuzzy Kernel C-Means (FKCM) untuk menyelesaikan masalah AD-IDS. Pada penelitian ini, masalah IDS berbasiskan Anomaly Detection diselesaikan dengan menggunakan metode clustering Fuzzy C-Means dan untuk mengurangi dimensi data digunakan metode pemilihan feature Laplacian Score.

2. Tinjauan Pustaka

2.1. Metode Seleksi Feature Laplacian Score

Secara teori, dengan menambah banyaknya feature yang digunakan akan meningkatkan tingkat representasi data yang dalam masalah klasifikasi diharapkan meningkatkan performa classifier. Namun pada praktiknya besarnya jumlah feature dapat menimbulkan masalah baru yaitu besarnya biaya proses klasifikasi baik waktu komputasi maupun besar sumber daya memori yang digunakan. Banyaknya jumlah feature juga secara signifikan memperlambat proses pembelajaran dan menyebabkan overtraining pada classifier serta dapat mempengaruhi pembangunan model [4]. Salah satu cara untuk mengatasi hal ini adalah dengan menggunakan metode pemilihan feature. Beberapa keuntungan dari metode pemilihan feature diantaranya adalah dapat mengurangi biaya komputasi dan kapasitas penyimpanan yang diperlukan, menangani munculnya penurunan keakuratan klasifikasi yang diakibatkan adanya himpunan data training yang tidak efisien, serta mempermudah dalam visualisasi dan pemahaman data. Untuk mengestimasi individual feature relatif lebih mudah dibandingkan mengestimasi himpunan bagian feature yang memberikan keakuratan maksimal dalam masalah klasifikasi. Masalah ini tergolong NP- hard problem yang solusi optimalnya tidak dijamin didapatkan kecuali dilakukan pencarian exhaustive di ruang solusi untuk menentukan kandidat himpunan bagian feature terbaik. Secara umum, feature dikategorikan sebagai relevan, tidak relevan, dan redundant. Feature yang relevan adalah feature yang berpengaruh pada output dan peranan mereka tidak dapat digantikan dari feature lainnya. Feature yang tidak relevan adalah feature yang tidak memiliki pengaruh terhadap output. Sedangkan feature yang redundant adalah feature yang berpengaruh terhadap output namun perannya dapat digantikan oleh feature lainnya. Tujuan dari pemilihan feature adalah menentukan himpunan bagian optimal yang terdiri dari m feature terpilih dari n total feature. Berbeda dengan masalah ekstraksi feature yang mentransformasikan feature asli data kemudian memilih sebagian dari feature hasil transformasi untuk digunakan, dalam metode seleksi feature, feature yang digunakan adalah feature asli data. Secara teori, seleksi feature dapat dianggap sebagai jenis khusus dari ekstraksi feature dengan fungsi transformasi yang digunakan adalah fungsi identitas. Relevansi dari feature dapat ditentukan baik secara individu maupun multivariable. Pendekatan univariat lebih sederhana dan cepat namun mengabaikan kemungkinan adanya korelasi dan ketergantungan diantara feature. Sehingga pada umumnya pendekatan multivariate lebih berguna dalam menentukan himpunan bagian feature optimal. Pendekatan ini memiliki beberapa keterbatasan diantaranya adalah masalah overtraining, khususnya pada kasus dimana banyaknya feature jauh lebih besar dari banyaknya data contoh. Keterbatasan lainnya adalah pendekatan ini memiliki biaya komputasi yang besar untuk data berdimensi tinggi. Salah satu metode seleksi feature yang sering digunakan adalah metode forward (backward) sequential yang selalu menghasilkan solusi suboptimal. Penelitian [5] menggunakan Sequential Feature Selection (SFS) untuk menyelesaikan masalah klasifikasi Wisconsin Diagnostic Breast Cancer. Metode Sequential Forward Floating Selection (SFFS) meningkatkan performa Forward Sequantial Feature Selection dengan menambahkan langkah backward setelah langkah forward selama kriteria fungsi objektif tercapai [6]. Metode pemilihan feature lainnya adalah metode berbasiskan Genetic Algorithm (GA) yang merupakan metode pencarian kombinatoris berdasarkan pada ukuran probabilistik maupun acak (random). Himpunan bagian dari feature dievaluasi menggunakan fitness function dan dikombinasikan dengan operator mutasi dan cross-over untuk menghasilkan generasi berikutnya. Penelitian [7] oleh Huang et al. mennggunakan Hybrid GA untuk menentukan himpunan bagian feature yang paling relevan dengan masalah klasifikasi yang berkaitan. Dua tahapan optimisasi dilakukan pada penelitian ini, outer optimization dan inner optimization. Outer optimization menyelesaikan pencarian global menggunakan pendekatan wrapper sedangkan inner optimization menjalankan pencarian lokal berbasiskan pendekatan filter. Gheyas and Smith [8] mengajukan algoritma hybrid (SAGA) yang didasarkan pada Simulated Annealing, GA, Generalized Neural Network dan Greedy Algorithm. Metode Spectral Feature Selection (SPEC) merupakan salah satu metode pemili-

han feature yang dapat menangani generic data set. SPEC [9] mengestimasi relevansi dari feature dengan mengestimasi konsistensi feature dengan menggunakan matriks yang diturunkan dari matriks similaritas S. SPEC menggunakan Radial-Bases Function (RBF) sebagai fungsi similaritas antara dua vektor xi dan xj:

$$S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$

Graf G dibangun dengan menggunakan S dan matriks adjacency W. Selanjutnya degree matrix $\bar{D}$ direkonstruksi dengan menggunakan W. $\bar{D}$adalah matriks diagonal dengan $\bar{D_{ij}} = \sum_{j=1}^{n} W_{ij}$. Apabila diberikan $\bar{D}$ dan W, matriks Laplacian L dan normalized Laplacian matrix L dihitung dengan menggunakan formula:

$\mathcal{L} = \bar{\mathcal{D}} - W$; $\mathcal{L} = \bar{D}^{-\frac{1}{2}} L \bar{D}^{-\frac{1}{2}}$

Bobot dari setiap *feature* f_i pada SPEC dievaluasi menggunakan tiga fungsi Ψ_1, Ψ_2, Ψ_3. Fungsi ini diturunkan dari *normalized cut function* dengan *spectrum* dari graf. Diasumsikan ketiga fungsi ini memetakan *feature vector* $f_i$dan mengembalikan nilai bobot berdasarkan normalized Laplacian matrix L. Algoritma SPEC diberikan pada Tabel 1.

Tabel 1: Algoritma SPEC

INPUT
D: Dataset
$\psi \in \{\psi_1, \psi_2, \psi_3\}$; Fungsi pembobotan *feature*
n banyaknya data
OUTPUT
F : Daftar urutan *feature*
1. Bangun matriks similaritas *S* dari *D*
2. Bangun graf G dari *S*
3. Bangun *W* dari *S*
4. Bangun $\bar{D}$ dari *W*
5. Definisikan *L* dan $\mathcal{L}$ menggunakan Persamaan (2)
6. For each *feature vector* f_i do
7. $\hat{f}_i \leftarrow \frac{\bar{D}^{-\frac{1}{2}} f_i}{\|\bar{D}^{-\frac{1}{2}} f_i\|}$
8. $F_i \leftarrow \psi(f_i)$
9. End for
10. Urutkan *F* berdasarkan ψ

Laplacian Score [10] merupakan metode pemilihan feature yang merupakan kasus khusus dari SPEC, yaitu ketika fungsi ranking yang digunakan adalah:

$F_i \leftarrow \frac{\hat{f}_i^T L \hat{f}_i}{\hat{f}_i^T \bar{D} f_i}$ dengan $\hat{f}_i = f_i - \frac{f_i^T \bar{D}_1}{1^T \bar{D}_1} 1$

2.2 Metode Fuzzy C-Means

Clustering adalah suatu teknik untuk mengelompokkan beberapa objek dengan karakteristik yang similar ke dalam suatu cluster yang sama. Teknik ini umum digunakan dalam masalah mesin pembelajaran maupun pengenalan pola. Pada mesin pembelajaran, clustering tergolong ke dalam unsupervised learning method yang bertujuan untuk mengklasifikasi sejumlah objek ke dalam beberapa cluster tertentu sedemikian sehingga objek-objek yang berada dalam cluster yang sama memiliki derajat similaritas yang tinggi dan objek-objek pada cluster berbeda memiliki derajat similaritas yang rendah.

Metode hierarki dan partisi merupakan teknik clustering yang paling sering digunakan. Metode partisi bertujuan untuk mempartisi suatu himpunan data menjadi beberapa cluster dengan cara mengoptimalkan suatu fungsi objektif tertentu. Metode partisi dapat dibagi menjadi dua yaitu hard dan soft (fuzzy) partition based method. Pada hard partitition method, setiap objek secara terbatas diklasifikasikan pada suatu cluster, sedangkan pada metode fuzzy dengan menggunakan konsep fuzzy membership, setiap objek memiliki membership degree pada setiap cluster [11]. Salah satu metode berdasarkan fuzzy adalah Fuzzy C-Means (FCM) yang diajukan oleh Dunn pada [12] dan selanjutnya dikembangkan oleh Bezdek pada [11]. FCM mengklasifikasikan objek yang masing-masing objek memiliki nilai keanggotaan pada setiap cluster. Pusat cluster ditentukan dengan beberapa aturan dan memberikan nilai keanggotaan antara 0 dan 1 untuk setiap objeknya. Berikut ini adalah langkah-langkah pada metode FCM.

Diberikan himpunan data $X = \{x_1, x_2, .., x_m\} \subseteq \mathbb{R}^k$, definisikan matriks keanggotaan $U_{nxc} = [U_{ij}], 1 \leq i \leq n, 1 \leq j \leq c$ dan hipunan pusat *cluster* $V = \{v_1, v_2, ..., v_c\}$ dengan setiap anggota pada V adalah vektor di ruang Euclid dimensi k. Model matematis dari FCM adalah:

$J(U,V) = min \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^m d^2(x_i, v_j)$

Dengan kenala:

$\sum_{j=1}^{c} u_{ij} = 1;\ j = 1, 2, ..., c$

$\sum_{i=1}^{n} u_{ij} > 0;\ i = 1, 2, ..., n$

$u_{ij} \in [0, 1]$

Dimana d merupakan fungsi disimilaritas atau fungsi jarak, dan $m \in [1, \propto]$ merupakan nilai fuzziness degree. Pusat cluster dan nilai keanggotaan diperbaharui menggunakan:

$v_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m}$

Dan

$u_{ij} = \left(\sum_{j=1}^{c} \left(\frac{d(x_i, v_i^t)}{d(x_i, v_j^t)} \right)^{\frac{2}{m-1}} \right)^{-1}$

3. Hasil dan Pembahasan

KDDCUP'99 data set [13] yang merupakan salah satu benchmark data untuk masalah IDS digunakan sebagai data percobaan dalam penelitian ini. Data yang didasarkan pada DARPA 98 data set ini merupakan satu-satunya data set terlabel yang dapat diakses secara bebas untuk para peneliti IDS. Data ini dapat dikategorikan dalam empat jenis:

1. Denial of Service Attacks (DoS), dimana pengguna sistem tidak bisa mengakses sistem dikarenakan jaringan atau sumber daya tidak tersedia. Contoh serangan bertipe DoS adalah Back, Land, SYN Flood, dan Smurf.

2. User to Super user atau User to Root Attacks (U2R), dimana penyerang mengakses sistem menggunakan akun pengguna normal dan mendeteksi kelemahan sistem untuk menyelusup pada root sistem. Contoh serangan jenis ini diantaranya adalah Eject, Perl, Cterm, Ps, dan Fdformat.

3. Probing Attacks (Probe), penyerang mencari informasi pada sistem atau kelemahannya dengan cara melakukan scanning terhadap jaringan computer, seperti pada Satan, Saint, Ipsweep, dan Mscan.

4. Remote to Local Attacks (R2L), dimana penyerang mengirimkan packages pada sistem melalui jaringan dan mendeteksi kelemahan sistem sehingga dapat bertindak sebagai pengguna. Serangan yang berjenis R2L diantaranya adalah Guest, Ftp Write, Imap, Dictionary, Xlock, dan Xsnoop.

KDDCUP'99 terdiri atas 494.021 records dimana 97.277 diantaranya dikategorikan Normal (19.69%), 391.458 serangan berjenis DoS (79.24%), 1.126 bertipe R2L (0.23%) dan 52 diklasifikasikan sebagai U2R (0.01%). Setiap records terdiri atas 41 features yang diberikan pada Tabel 2.

Tabel 2: 41 Feature dari Data KDDCUP'99

No	Nama *Feature*
1	duration
2	protocol_type
3	Service
4	flag
5	src_bytes
6	dst_bytes
7	land
8	wrong_fragment
9	urgent
10	hot
11	num_failed_logins
12	logged_in
13	num_compromised
14	root_shell
15	su_attemted
16	num_root
17	nu_file_creations
18	num_shells
19	num_access_file
20	num_outbond_cmds
21	is_host_login
22	is_guest_login
23	count
24	srv_count
25	serror_rate
26	srv_serror_rate
27	rerror_rate
28	srv_rerror_rate
29	same_srv_rate
30	diff_srv_rate
31	srv_diff_host_rate
32	dst_host_count
33	dst_host_srv_count
34	dst_host_same_srv_rate
35	dst_host_diff_srv_rate
36	dst_host_same_src_port_rate
37	dst_host_srv_diff_host_rate
38	dst_host_serror_rate
39	dst_host_srv_serror_rate
40	dst_host_rerror_rate
41	dst_host_srv_rerror_rate
42	attack_type

Pada penelitian ini, keakuratan classifier dievaluasi dengan mengkombinasikan hasil klasifikasi yang benar dan salah seperti diberikan pada Tabel 3.

Tabel 3: Aturan Evaluasi dari Performa Classifier

Nilai Sebenarnya	Nilai Hasil Klasifikasi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN
Keakuratan $= \frac{(n_T P + n_T N)}{(n_T P + n_T N + n_F P + n_F N)}$		

True Positive (TP) menyatakan banyaknya serangan yang terdeteksi sebagai serangan, True Negative (TN) menyatakan banyaknya normal behavior yang diklasifikasikan dengan tepat, False Positive (FP) menyatakan banyaknya serangan yang diklasifikasikan sebagai normal behavior, dan False Negative (FN) menyatakan banyaknya normal behavior yang dideteksi sebagai serangan.

Hasil klasifikasi data KDDCUP'99 dengan menggunakan metode FCM dan pemilihan fea-

ture menggunakan Laplacian Score untuk data normal vs serangan Dos diberikan pada Tabel 4. Algoritma diaplikasikan pada komputer pribadi dengan prosesor i-7, hard disk 1 TB, dan menggunakan software Matlab 2012a. Pada penelitian ini digunakan 70% data training.

Tabel 4: Hasil Klasifikasi Normal vs DoS

Banyak *Feature*	Keakuratan (%)	*Running Time* (detik)
5	99,21	543,60
10	98,65	5652,17
15	97,58	4324,18
20	96,26	5480,00
25	98,83	5468,46
30	98,64	5472,82
35	98,90	5515,71
Semua *feature*	98,70	5825,20

Dapat dilihat pada Tabel 4 bahwa hanya dengan menggunakan 5 buah feature keakuratan klasifikasi telah mencapai 99.21 % dimana hasil ini lebih baik daripada menggunakan semua feature yaitu 98.70 %. Running time ketika menggunakan hanya lima buah feature juga lebih kecil daripada menggunakan semua feature yaitu hanya 543,60 detik. Hasil klasifikasi data KDDCUP'99 dengan menggunakan metode FCM dan pemilihan feature menggunakan Laplacian Score untuk data normal vs serangan Probe diberikan pada Tabel 5 dimana hasil klasifikasi menggunakan 5 buah feature 94.04 % memberikan keakuratan yang lebih baik daripada menggunakan semua feature 90.33 %, hal ini menandakan ada feature yang tidak relevan atau redundant, sehingga ketika feature tersebut turut diperhatikan dalam tahap klasifikasi menyebabkan penurunan keakuratan klasifikasi. Pada kasus ini, running time dengan menggunakan 5 feature lebih besar daripada dengan menggunakan semua feature. Hal ini dimungkinkan karena proses pemilihan feature dengan Laplacian Score membutuhkan waktu lebih sehingga mengakibatkan bertambahnya waktu komputasi keseluruhan algoritma.

Tabel 5: Hasil Klasifikasi Normal vs. Probe

Banyak *Feature*	Keakuratan (%)	*Running Time* (detik)
5	94,04	1092,20
10	97,60	1089,95
15	98,42	1090,81
20	89,76	664,81
25	88,17	429,29
30	89,38	538,12
35	87,17	540,70
Semua *feature*	90,33	314,17

Hasil klasifikasi untuk data normal vs serangan U2R dapat dilihat di Tabel 6 sedangkan untuk data normal vs serangan R2L dapat dilihat di Tabel 7. Hasil terbaik pada klasifikasi untuk data normal vs serangan U2R didapatkan pada penggunaan 35 feature yaitu 99.13%, sedangkan pada saat digunakannya 5 buah feature saja keakuratan menurun menjadi 89.11 %. Hal ini menunjukkan bahwa pada kasus serangan U2R sebagian besar feature adalah feature yang relevan. Running time pada pemilihan 35 feature pun merupakan running time terkecil yaitu sebesar 756,31 detik.

Tabel 6: Hasil Klasifikasi Normal vs U2R

Banyak *Feature*	Keakuratan (%)	*Running Time* (detik)
5	89,11	1094,50
10	94,64	1128,04
15	92,25	1141,93
20	92,35	1131,62
25	96,18	1144,43
30	96,32	1141,26
35	99,13	765,31
Semua *feature*	98,74	1188,48

Tabel 7: Hasil Klasifikasi Normal vs R2L

Banyak *Feature*	Keakuratan (%)	*Running Time* (detik)
5	55,10	1106,85
10	83,07	1104,01
15	89,82	1103,29
20	86,86	1110,82
25	87,18	1112,23
30	87,24	1114,65
35	87,09	1115,96
Semua *feature*	85,92	1155,32

Untuk kasus data normal vs serangan R2L klasifikasi dengan menggunakan 5 buah feature memberikan hasil yang kurang akurat 55.10 %. Keakuratan hasil meningkat pesat pada penggunaan 10 feature, yaitu keakuratan mencapai 83.07 % sedangkan penggunaan semua feature berdampak tidak begitu besar terhadap peningkatan keakuratan yaitu mencapai 85.92 %. Sedangkan running time pada beberapa kasus pemilihan jumlah feature tidak berbeda jauh sekitar 1100 detik.

Pada Tabel 8 diberikan urutan 5 feature terpilih untuk masing-masing kasus normal dan keempat jenis serangan. Pada kasus Normal vs Dos, ketika klasifikasi dilakukan menggunakan hanya 5 feature dari 41 feature yang ada keakuratan meningkat menjadi 99.21% dari 98.70 % bahkan running time jauh lebih singkat yaitu dari 5825,20 detik menjadi 543,60 detik saja. Hal ini menunjukkan bahwa feature 24, 23, 2,

33 dan 29 (lihat Tabel 1. Untuk nama feature) cukup mewakili data dan merupakan feature relevan sehingga proses klasifikasi tidak perlu menggunakan semua feature. Berbeda dengan kasus Normal vs R2L dimana ketika digunakan 5 buah feature saja, keakuratan menurun drastis menjadi 55,10 %, namun ketika digunakan 10 buah feature keakuratan meningkat menjadi hal ini 83.07 %. Hal ini menandakan bahwa selain 5 feature terpilih masih ada feature lain yang relevan dengan data sehingga ketika feature ini diabaikan keakuratan menurun.

4. Kesimpulan dan Saran

Pada makalah ini permasalahan klasifikasi pada Anomaly Detection-Intrusion Detection System diselesaikan dengan menggunakan metode Fuzzy C-Means. Metode Laplacian Score digunakan sebagai metode pemilihan feature. Data yang digunakan pada implementasi algoritma adalah KDDCUP'99 data set yang merupakan benchmark data terlabel untuk masalah IDS. Hasil klasifikasi menunjukkan bahwa untuk kasus Normal vs serangan DoS hanya dengan menggunakan 5 buah feature dari total 41 buah feature yang ada berakibat pada meningkatnya keakuratan classifier dari 98,70% menjadi 99,21% dan terjadi percepatan running time dari 5825,20 detik menjadi 543,60 detik. Hal ini menunjukkan bahwa pada kasus Normal vs DoS, 5 feature terpilih cukup menjadi perwakilan dari semua 41 features yang ada. Sedangkan pada kasus Normal vs R2L, ketika klasifikasi dilakukan dengan menggunakan 5 feature terpilih, keakuratan menurun menjadi 55,10% dari keakuratan dengan menggunakan 10 feature yaitu 83,07%. Hal ini menunjukkan bahwa selain 5 feature terpilih masih ada feature lain yang merupakan feature relevan. Penelitian ini dapat dilanjutkan dengan menggunakan metode pemilihan feature lainnya khususnya pada kasus Normal vs serangan R2L untuk mendapatkan himpunan bagian feature yang optimal sekaligus menurunkan running time dari algoritma.

Daftar Pustaka

[1] S. Ganaphanty, K. Kulothungan, P. Yogesh, and A. Kannan. "A Novel Wighted Fuzzy C-Means Based on Immune Genetic Algorithm for Intrusion Detection". In Proc. Engineering SciVerse Scince Direct, pp. 1750-1757, Elsevier, 2012.

[2] T. S. Chou, K. K. Yen, and J. Luo, "Network Intrusion Detection Design Using Feature Selection of Soft Computing Paradigms". International Journal of Computational Intelligence, 4, 3, pp. 196-208, 2008.

[3] Z. Rustam and A. S. Talita. "Fuzzy Kernel C-Means Algorithm for Intrusion Detection Systems". Journal of Theoretical and Applied Information Technology, 81, 1, pp. 161-165, 2015.

[4] M.A. Hall and G. Holmes. "Benchmarking Attribute Selection Techniques for Discrete Class Data Set Mining". IEEE Trans Knowl Data Set Eng, 15, 3, pp. 1-16, 2003.

[5] A. S. Talita. "Klasifikasi Wisconsin Diagnostic Breast Cancer Data dengan Menggunakan Sequential Feature Selection dan Possibilistic C-Means". Jurnal Ilmiah Komputasi Komputer & Sistem Informasi, 15, 1, pp. 47-52, 2016.

[6] P. Somol, P. Pudil, and J. Kittler. "Fast Branch and Bound Algorithms for Optimal Feature Selection". IEEE Trans Pattern Anal Mach Intell, 26, 7, pp. 900-912, 2004.

[7] J. Huang, Y. Cai, and X. Xu. "A Hybrid Genetic Algorithm for Feature Selection Wrapper Based on Mutual Information". Pattern Recognit Lett, 28, pp. 1825-1844, 2007.

[8] I.A. Gheyas and L. S. Smith. "Feature Subset Selection in Large Dimensionality Domains". Pattern Recognit, 43, pp. 5-13, 2010.

[9] Z. Zhao and L. Huan. "Spectral Feature Selection for Supervised and Unsupervised Learning". In Proc. Of the 24th International Conference on Machine Learning,

pp. 1151-1157, ACM, New York, USA, 2007.

[10] X. He, D. Cai, and P. Niyogi. "Laplacian Score for Feature Selection". Advances in Neural Information Processing Systems, 18, 507, 2006.

[11] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Springer-Verlag, New York, 1981.

[12] J. C. Dunn. "A Fuzzy Relative of The ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". J. Cybernetics, 3, pp. 32-57, 1973.

[13] S. Hettich and S. D, Bay, The UCI KDD Archive, http://kdd.ics.uci.edu. Irvine, CA: University of California, Department of Information and Computer Science, 1999. 29 September 2016.

Halaman ini sengaja dikosongkan