ISSN: 2302 -2805

SISTEM DETEKSI INTRUSI PADA JARINGAN DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN TEORI DEMPSTER SHAFER

Akhmad Alimudin [1] , Waskitho Wibisono [2] , Diana Purwitasari [3]
Teknik Informatika Institut Teknologi Sepuluh Nopember Surabaya
Kampus ITS Keputih Sukolilo
email : alioke@eepis-its.edu [1], waswib@its-sby.edu [2], diana@its-sby.edu [3]

Abstrak

Penelitian mengenai Intrusion Detection System(IDS) telah banyak dilakukan untuk mendapatkan hasil deteksi intrusi yang baik dan akurat dari IDS. Untuk membangun sistem IDS, salah satu komponen utamanya adalah komponen yang mampu melakukan proses klasifikasi terhadap data log paket jaringan. Pada penelitian ini, kami akan melakukan penggunaan metode KNN dan Dempster Shafer (KNN-DS) untuk diterapkan pada Sistem Deteksi Intrusi pada jaringan. Data yang akan digunakan pada penelitian kali ini adalah data KDDCUP 99 yang merupakan data tcpdump dari DARPA98 yang sudah dilakukan preprocessing. Dengan menggunakan KNN-DS ini akan didapatkan hasil yang lebih maksimal untuk mendukung nilai kebenaran dari output yang dihasilkan oleh KNN.

Kata kunci :
IDS, Dempster Shafer, KNN, klasifikasi

1. Pendahuluan

Permasalahan yang sering muncul pada IDS yang ada saat ini adalah tingkat akurasi dari *signature* yang dihasilkan oleh IDS. Terlalu banyaknya muncul *false positive* dan *false negative* pada IDS membuat *network administrator* kesulitan dalam memutuskan atau menangani laporan dari IDS[6]. Mengingat beban pekerjaan dari seorang network administrator, kecil kemungkinan bagi mereka untuk melakukan update jenis-jenis serangan baru dan membuat signature baru setiap saat. Maka dari hal tersebut, muncullah ide untuk membuat suatu sistem IDS yang mampu mengenali pola serangan baru dan akurat dalam melakukan pelaporan.

Salah satu cara untuk meningkatkan akurasi dari IDS adalah dengan cara melakukan kombinasi terhadap beberapa klasifikasi data. Karena pada setiap jenis metode klasifikasi memiliki kelebihan dan kelemahan, maka untuk meningkatkan akurasi dari hasil klasifikasi adalah dengan cara mengambil kelebihan dari masing-masing klasifikasi dan menghindari kelemahan dari masing-masing metode[26]. Salah satu metode untuk mengkombinasikan beberapa klasifikasi adalah dengan menggunakan teori Dempster-Shafer. Salah satu keuntungan menggunakan teori Dempster-Shafer adalah mampu memberikan metode yang efektif untuk menggabungkan informasi yang diperoleh dari beberapa sumber data.

Hal inilah yang melatar belakangi penelitian ini. Pada penelitian ini akan dibuat klasifikasi IDS dengan menggunakan KNN dengan menggunakan Teori Dempster-Shafer

2. Tinjauan Pustaka

Penelitian mengenai *Intrusion Detection System*(IDS) telah banyak dilakukan untuk mendapatkan hasil terbaik dan akurat dari sistem yang dibuat. Salah satunya adalah dengan cara melakukan klasifikasi pada data IDS. Untuk mendapatkan hasil yang akurat dari klasifikasi sistem, telah dilakukan penelitian dengan cara mengkombinasikan beberapa *output* dari klasifikasi dengan menggunakan Dempster-Shafer[26] pada kasus kategorisasi teks. Penelitian tentang klasifikasi IDS telah dilakukan dengan membandingkan 3 metode klasifikasi[4] yaitu dengan metode *association rules*, *decision trees*, dan *neural network* dengan menggunakan *dataset training* KDD CUP 99[15]. Penelitian tentang penggunaan *extended* Dempster-Shafer yang diterapkan pada IDS[10] menunjukkan bagaimana cara meningkatkan akurasi dari IDS, dan pada hasilnya menunjukkan bahwa *extended* Dempster-Shafer 30% lebih akurat dibandingkan dengan Dempster-Shafer. Penelitian ini mencoba mengkombinasikan data IDS dengan menggunakan dataset training dari DARPA 99 yang telah dilakukan klasifikasi menggunakan 2 metode klasifikasi, dengan terlebih dahulu melakukan *clustering* terhadap data training guna mempersingkat proses perhitungan[25], lalu kemudian dikombinasikan dengan menggunakan teori Dempster-Shafer.

2.1. Sistem Deteksi Intrusi (IDS)

Karena mekanisme keamanan mengalami banyak kegagalan, perlindungan tambahan untuk suatu jaringan sangatlah dibutuhkan, untuk memberikan lapisan pertahanan yang lebih, maka dibuatlah sistem deteksi intrusi (IDS). IDS akan mengamati setiap data dengan tujuan untuk mencari tanda-tanda atau perilaku yang mencurigakan pada sebuah jaringan. Ketika keanehan ditemukan oleh IDS, maka IDS akan melaporkan kejadian tersebut kepada *Administrator* sehingga *Administrator* dapat melakukan tindakan untuk menangani laporan dari IDS.

2.2. Teori Dempster-Shafer

Pendekatan Data-Fusion yang telah terintegrasi dengan model CS(*Context Spaces*) adalah pembuktian teori Dempster-Shafer(D-S)[24]. Salah satu kelebihan dari teori D-S adalah dapat menangani informasi yang tidak tepat dan tidak pasti[2, 17]. Pada pembuktian teori D-S, rangkaian hipotesis eksklusif (Θ={h_1,..,h_n }, yang mewakili n merupakan situasi yang mungkin di pertimbangkan) yang disebut sebagai *frame of discernment*[24].

Misalkan 2^{Θ} menyatakan himpunan kekuatan dari Θ. *Basic Probability Assignment* (BPA) (yang juga dikenal sebagai fungsi massa m) didefinisikan sebagai fungsi m: $2^{\Theta} \rightarrow [0,1]$ dimana $m(\emptyset) = 0$ dan $\sum_{\subseteq} h = 1$

Nilai massa dari situasi hipotesis yang diberikan h , (h) menunjukkan kekuatan bukti yang relevan yang mendukung proposisi dari h . Nilai massa yang dimasukkan kedalam *frame of discernment* (Θ) merupakan ketidaktahuan atau ketidakpastian dari situasi yang diberikan sesuai dengan bukti. Selanjutnya *Belief*(Bel) dan *Plaussability(Pls)*, sebagai batas atas dan batas bawah interval dapat didefinisikan dari fungsi massa sebagai berikut[14-16,23]:

$$( ) = \sum_{\subseteq} ( ) \quad 2.1$$

$$( ) = \sum_{\cap \emptyset} ( ) \quad 2.2$$

Fungsi-fungsi ini memiliki properti sebagai berikut:

$$( ) \leq ( ) \quad 2.3$$

$$( ) = 1 - ( ) \quad 2.4$$

Misalkan dua sensor yang berbeda melaporkan pengamatan mereka dengan menetapkan satu set nilai massa melalui kemungkinan hipotesis situasi. Teori D-S menyediakan metode yang dikenal sebagai peraturan kombinasi D-S, untuk menggabungkan nilai massa dari sumber informasi yang berbeda[24].

Mengingat dua nilai massa yang berbeda berasal dari laporan informasi dari dua sensor yang berbeda(yaitu dan) kombinasi dari nilai-nilai massa diperoleh sebagai berikut[24,22,9,12]

$$, h \quad \frac{\sum_{\cap} ( ) ( )}{,} \quad 2.5$$

Koefisien normalisasi , , menunjukkan nilai massa terkait dengan konflik diberikan dua nilai massa yang berbeda. Nilai ini diperoleh dengan menjumlahkan produk dari *mass assignment* dari semua hipotesis situasi dengan *null intersection*[23]. Persamaan berikut menunjukkan bagaimana , dihitung:

$$, = \sum_{\cap \emptyset} ( ) ( ) \quad 2.6$$

Dalam [2], pekerjaan awal mengintegrasikan pendekatan model CS dengan toeri D-S. Pada pendekatan ini, teori D-S diaplikasikan pada model CS dengan cara melakukan komputasi distribusi massa untuk setiap wilayah dapat diterima, di ruang situasi yang telah didefinisikan. Hal ini terjadi ketika nilai atribut konteks berada dalam wilayah terkait dalam ruang situasi yang telah didefinisikan [2]. Dengan menggunakan pendekatan ini, setiap kali *context-attribute* memenuhi predikat tiap region, nilai-nilai massa yang sesuai untuk semua hipotesis dihitung. Nilai massa menunjukkan tingkat dukungan terhadap terjadinya hipotesis yang sesuai.

2.3. K-Nearest Neighbor

Algoritma k-Nearest Neighbor[18] (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

KNN merupakan metode klasifikasi *instancebased*, memilih satu objek latih yang memiliki sifat ketetanggaan (*neighborhood*) yang paling dekat. Sifat ketetanggaan ini didapatkan dari perhitungan nilai kemiripan ataupun ketidakmiripan.

KNN menggunakan metode perhitungan nilai ketidakmiripan (Euclidian, Manhattan, Square Euclidian ,dll). KNN akan memilih tetangga terdekat untuk menentukan hasil klasifikasi dengan melihat jumlah kemunculan dari kelas dalam tetangga yang terpilih. Kelas yang paling banyak muncullah yang akan menjadi kelas hasil klasifikasi.

Rumus Euclidian

$$= \sqrt{(-) + (-)} \quad 2.7$$

2.3.1. Algoritma KNN

Langkah-langkah untuk menghitung algoritma KNN[18]:

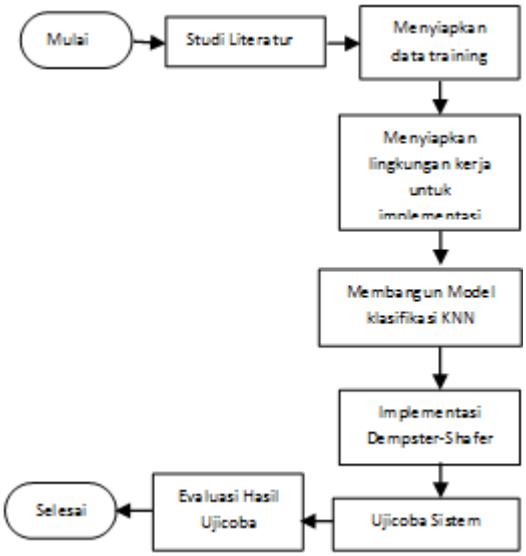
- Tentukan parameter K (*positive integer* yang menunjukkan jumlah tetangga terdekat).
- Hitung jarak antara *query instance* dan semua sampel *training*.
- Urutkan jarak dan tentukan tetangga terdekat berdasar jarak minimum K-th.
- Kumpulkan kategori Y dari tetangga terdekat.
- Gunakan mayoritas kecil dari kategori tetangga terdekat sebagai nilai prediksi *query instance*.

3. Metode Penelitian

Pada penelitian ini akan dilakukan dalam beberapa tahapan. Tahap yang pertama adalah studi literatur untuk mempelajari dan memahami materi-

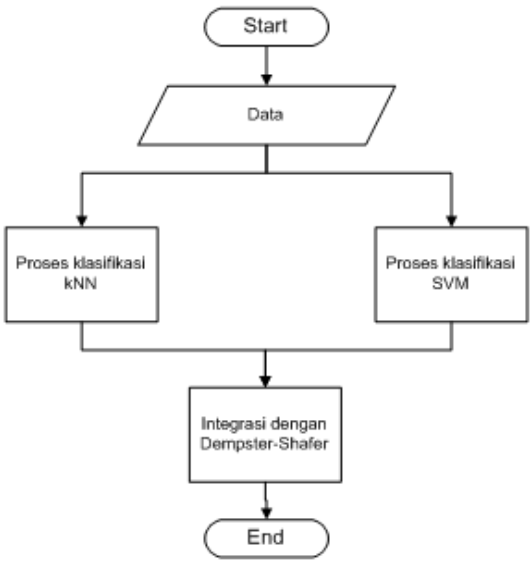
ISSN: 2302 -2805

materi yang terkait dengan penelitian, seperti teori Dempster-Shafer dan KNN, serta mempelajari tentang data dari KDDCUP99 yang merupakan data hasil *preprocessing* dari DARPA 98. Tahap kedua adalah melakukan klasifikasi data terhadap data serangan atau data test dengan menggunakan metode KNN. Selanjutnya, tahap ketiga yaitu implementasi teori Dempster-Shafer dari hasil klasifikasi, dan tahap keempat melakukan ujicoba hasil implementasi dengan sejumlah data tes (*unlabelled*),. Dan tahap yang terakhir adalah melakukan evaluasi hasil dari hasil ujicoba, berikut dapat dilihat pada gambar 1 untuk flowchart tahapan penelitian



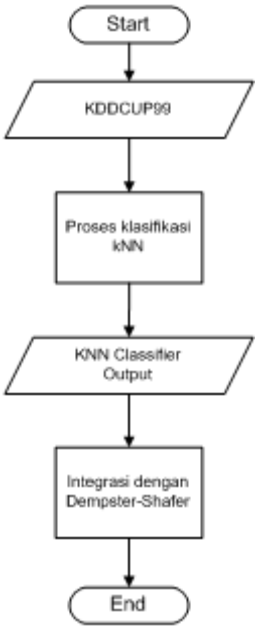
Gambar 1 : Flowchart Tahapan Penelitian

Pada penelitian sebelumnya[26] telah diimplementasikan kombinasi antara 2 sumber klasifikasi data yaitu KNN dan SVM pada kasus kategorisasi text dengan menggunakan teori Dempster-Shafer. *Flowchart* dari sistem implementasi dapat dilihat pada gambar 2.



Gambar 2 : Flowchart Integrasi Dempster-Shafer

Sedangkan untuk rancangan sistem yang dibuat pada penelitian ini dapat dilihat pada gambar 3. Pada penelitian ini, akan digunakan KDDCUP 99 dataset sebagai data utama yang merupakan data hasil preprocessing dari data DARPA 98, lalu kemudian data akan diklasifikasikan dengan menggunakan metode yaitu KNN. Dan kemudian akan dilanjutkan dengan proses Dempster-Shafer.



Gambar 3 : Flowchart Sistem

4. Hasil dan Pembahasan

Tahapan ujicoba yang dilakukan pada penelitian kali ini adalah dengan menggunakan 10% dari data KDDCUP99 secara keseluruhan. Pada pengujian kali ini, akan dilakukan perbandingan antara *Voting* KNN (KNN), *Distance Weighted* KNN (DWKNN), dan KNN dengan menggunakan teori Dempster Shafer (DSKNN). Pada tabel berikut ini, didapatkan hasil dari salah satu proses klasifikasi dengan menggunakan KNN dan didapatkan pula *mass function* untuk Dempster Shafer. Data serangan diatas merupakan data serangan dengan nama serangan "multihop".

Tabel 1 : Mass Function dari klasifikasi KNN

No	Jarak	m1	m2	kelas
1	1.732	0.929	0.071	Multihop
2	24.474	0.696	0.304	normal.
3	86.394	0.317	0.683	rootkit
4	106.419	0.246	0.754	normal.
5	108.517	0.239	0.761	normal.
6	111.700	0.230	0.770	multihop
7	111.897	0.229	0.771	normal.

Pada tabel 1 diatas, didapatkan jarak hasil dari klasifikasi KNN, dan hasil dari perhitungan tersebut

digunakan untuk menentukan *mass function* dari masing-masing nilai. Jika dilihat dari contoh tabel diatas, kita dapat menentukan hasil dari KNN dan DWKNN. Untuk menentukan hasil dari DSKNN, diperlukan perhitungan dengan menentukan *Global Mass Value* untuk menentukan hasil dari masing-masing kelas.

Dari perhitungan yang telah dilakukan, didapatkan data sebagai berikut dapat dilihat pada tabel 2 :

Tabel 2 : *Global Mass Value* untuk masing-masing kelas

Multihop	Normal	Rootkit	Θ (Unknown)
0.6800	0.3197	0.0002	0.0001

Dari data yang telah didapatkan diatas, kita telah bisa menentukan hasil dari masing-masing metode yang digunakan.

Tabel 3 : Hasil dari beberapa metode KNN

KNN	DWKNN	DSKNN
normal.	multihop	multihop

Dari tabel 3 didapatkan hasil dari masing-masing metode KNN yang digunakan, dari 3 metode yang digunakan, 2 metode menyatakan kelas yang dihasilkan dari data ujicoba diatas adalah "multihop".

Dari 100 ujicoba yang dilakukan dengan menggunakan ketiga metode KNN tersebut, didapatkan tingkat kebenaran/akurasi dari masing-masing metode seperti ditampilkan pada tabel 4 berikut :

Tabel 4 : Tingkat kebenaraan/akurasi masing-masing metode

KNN	DWKNN	DSKNN
79%	87%	88%

Dari beberapa ujicoba yang dilakukan, didapatkan hasil untuk metode teori Dempster Shafer KNN mendapatkan tingkat akurasi yang lebih tinggi dibandingkan dengan metode *Distance Weighted* dan *Voting.*

5. Kesimpulan dan Saran

Dari beberapa ujicoba yang dilakukan, didapatkan hasil untuk metode teori Dempster Shafer KNN mendapatkan tingkat akurasi yang lebih tinggi dibandingkan dengan metode *Distance Weighted* dan *Voting.* Namun dari beberapa ujicoba yang dilakukan, tingkat akurasi DSKNN juga dipengaruhi oleh banyaknya K dan jarak yang dihasilkan.

Untuk penelitian selanjutnya, diharapkan untuk bisa mengimplementasikan teori Dempster Shafer dengan menggabungkan beberapa metode klasifikasi seperti KNN dengan metode SVM yang dikombinasikan dengan menggunakan toeri Dempster Shafer guna mendapatkan hasil yang lebih akurat.

Daftar Pustaka

[1] A. O. Boudraa, e. a. (2004). Dempster-Shafer's Basic Probability Assignment Based on Fuzzy Membership Functions. Electronic Letters on Computer Vision and Image Analysis , vol. 4, 1-9.

[2] [2] A. Padovitz, e. a. (2006). A Unifying Model for Representing and Reasoning About Context under Uncertainty. 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems,IPMU, Paris, France , 1983 - 1989.

[3] [3] Adel N. Toosi, M. K. (2006). Network Intrusion Detection Based on Neuro-Fuzzy Classification. In Proceeding of IEEE International Conference on Computing and Informatics .

[4] [4] Adnan M.A. Brifcani, A. S. (2011). Intrusion Detection and Attack Classifier Based on Three Techniques:A Comparative Study. In Eng. & Tech. Journal'2011 .

[5] [5] Alief Habibiy, I. U. (2009). Rancang Bangun Perangkat Lunak Klasifikasi Intrusi pada Jaringan Menggunakan Metode Support Vector Machine(SVM). Seminar Proyek Akhir PENS .

[6] [6] Bambang Wijanarko, E. M. (2009). Algoritma Fuzzy Sebagai Metode Pendeteksi Pola Serangan Pada Jaringan Berbasis Snort IDS. Seminar Proyek Akhir PENS .

[7] [7] Berbers, D. P. (2006). Quality Extensions and Uncertainty Handling for Context Ontologies. Workshop on Context and Ontologies: Theory, Practice and Applications, Riva del Garda, Italy , 62-64.

[8] [8] Bloch, I. (1996). Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. Pattern Recognition Letters, vol. 17 , 905 - 919.

[9] [9] Denoeux, T. (1995). A k-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory. IEEE Transaction on Systems, Man and Cybernetics , vol. 25, 804-813.

[10] [10] Dong Yu, D. A. (2005). Alert confidence fusion in intrusion detection systems with extended Dempster-Shafer theory. In Proceedings of ACM Southeast Regional Conference (2)'2005 , 142-147.

[11] [11] Golshani, E. C.-R. (1990). Uncertain reasoning using the Dempster-Shafer method: an application in forecasting and marketing management. Expert Systems , vol. 7, 9-18.

[12] [12] H. Wu, e. a. (2003). Sensor Fusion Using Dempster-Shafer Theory II: Static Weighting and Kalman Filter-like Dynamic Weighting. IEEE Instrumentation and Measurement Technology Conferencev (IMTC 2003), Colorado, USA .

[13] [13] Hadjiefthymiades, C. A. (1153 - 1168). Enhancing Situation-Aware Systems Through Imprecise Reasoning. IEEE Transactions On Mobile Computing, vol. 7 , 2008.

[14] [14] JC Burgess, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Bell Laboratories, Lucent Technologies .

[15] [15] KDD Cup 1999 Intrusion detection dataset. (1999). Retrieved from http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[16] [16] luthfi, E. T. (2007). Fuzzy C-Means untuk Clustering Data(Studi Kasus:Data Performance Mengajar Dosen). Seminar Nasional Teknologi .

ISSN: 2302 -2805

[17] [17] M. C. Florea, e. a. (2007). Dempster-Shafer Evidence Theory Through The Years: Limitations, Practical Examples, Variants Under Conflict and a New Adaptive Combination Rule. Advances and Challenges in Multisensor Data and Information Processing, ed: IOS Press , 148-156.
[18] [18] Novi Anisyah, Z. S. (2011). Aplikasi Mobile untuk Metode K-Nearest Neighbor pada Intrusion Detection System Berbasis Snort. Seminar Proyek Akhir PENS .
[19] [19] P. D. Haghighi, e. a. (2009). Situation-Aware Adaptive Processing (SAAP) of Data Streams. Pervasive Computing, Innovations in Intelligent Multimedia and Applications, ed: Springer-Verlag , 318-356.
[20] [20] Padovitz, A. (2006). Context Management and Reasoning about Situations in Pervasive Computing. Ph.D Thesis, Caulfield School of Information Technology, Monash University .
[21] [21] Qi Chen, U. A. (2006). Anomaly Detection Using the Dempster-Shafer Method. In Proceedings of DMIN'2006 , 232-240.
[22] [22] S. L. H´egarat-Mascle, e. a. (1997). Application of Dempster–Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing. IEEE Transactions on Geoscience and Remote Sensing , vol. 35, 1028-1037.
[23] [23] Sentz, K. (2010, 3 9). Combination of Evidence in Dempster-Shafer Theory. Retrieved from www.sandia.gov: http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf
[24] [24] Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton University Press .
[25] [25] V. Venkatachalam, S. S. (2008). Clustering and Sample Selection to Enhance the Performance of the Lamstar Intrusion Detection System. In I.J of Simulation'2008 , 13-20.
[26] [26] Yaxin Bi, D. A. (2004). Combining Multiple Classifiers Using Dempster's Rule of Combination for Text Categorization. In Proceedings of MDAI'2004 , 127-138 .

Biodata Penulis

Akhmad Alimudin, memperoleh gelar Ahli Madya (A.Md), program studi Teknik Informatika Politeknik Elektronika Negeri Surabaya pada tahun 2008. Tahun 2010 memperoleh gelar Sarjana Sains Terapan (S.ST), Program Teknik Informatika Politeknik Elektronika Negeri Surabaya. Saat ini sedang menjalani kuliah program pasca sarjana di Teknik Informatika, Institut Teknologi Sepuluh Nopember

Waskitho Wibisono, memperoleh gelar Sarjana Komputer (S.Kom) di Teknik Informatika Institut Teknologi Sepuluh Nopember, dan memperoleh gelar M.Eng di Jepang. Pada tahun 2011 memperoleh gelar PhD dari Monash university. Saat ini bertugas sebagai pengajar di jurusan Teknik Informatika Institut Teknologi Sepuluh Nopember Surabaya dan merangkap menjadi kepala program studi magister di jurusan teknik informatika

Diana Purwitasari, memperoleh gelar Sarjana Komputer (S.Kom) di Teknik Informatika Institut Teknologi Sepuluh Nopember, dan memperoleh gelar M.Sc di Jepang. Saat ini bertugas menjadi pengajar di jurusan Teknik Informatika Institut Teknologi Sepuluh Nopember Surabaya.

ISSN: 2302 -2805