KLASIFIKASI DATA FORUM PENGADUAN PESERTA PELATIHAN DI DIKLAT PERTANIAN JAWA TIMUR DENGAN MENGGUNAKAN METODE LVQ (*LEARNING VECTOR QUANTIZATION*)

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Program Studi Teknik Informatika

OLEH :

MOH. DADANG HALIMI YUSUF

12.1.03.02.0226

FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

2016

Skripsi Oleh :

MOH. DADANG HALIMI YUSUF

12.1.03.02.0226

Judul :

KLASIFIKASI DATA FORUM PENGADUAN PESERTA PELATIHAN DI DIKLAT PERTANIAN JAWA TIMUR DENGAN MENGGUNAKAN METODE LVQ (*LEARNING VECTOR QUANTIZATION*)

Telah disetujui untuk diajukan kepada
Panitia Ujian/Sidang Skripsi Program Studi Teknik Informatika
Fakultas Teknik UN PGRI Kediri

Tanggal: 04 - 08 - 2016

Pembimbing I	Pembimbing II
FATKUR RHOHMAN, M.Pd.	**INTAN NUR FARIDA, M.KOM**
NIDN : 0728088503	NIDN: 0704108701

ii

Skripsi Oleh :

MOH. DADANG HALIMI YUSUF

12.1.03.02.0226

Judul :

KLASIFIKASI DATA FORUM PENGADUAN PESERTA PELATIHAN DI DIKLAT PERTANIAN JAWA TIMUR DENGAN MENGGUNAKAN METODE LVQ (*LEARNING VECTOR QUANTIZATION*)

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik UN PGRI Kediri
Pada tanggal : 11 - 08 - 2016

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua Penguji : Fatkur Rhohman, M.Pd.
2. Penguji I : Dr. Suryo Widodo, M.Pd
3. Penguji II : Intan Nur Farida, M.KOM

Mengetahui
Dekan FT

Dr. SURYO WIDODO, M. Pd
NIP. 196 10202 199103 1 002

iii

KLASIFIKASI DATA FORUM PENGADUAN PESERTA PELATIHAN DI DIKLAT PERTANIAN JAWA TIMUR DENGAN MENGGUNAKAN METODE LVQ (*LEARNING VECTOR QUANTIZATION*)

Moh. Dadang Halimi Yusuf
12.1.03.02.0266
Teknik - Informatika
Yusufhalim27@gmail.com
Fatkhur Rohman, M.Pd dan Intan Nur Farida, M.KOM.
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Proses pengelompokan data forum pengaduan peserta pelatihan di UPT Diklat Pertanian Jawa Timur masih dilakukan secara manual. Sehingga membutuhkan waktu yang lama dan sering terjadi kehilangan data serta kesalahan dalam pengelompokkan data pengaduan. Rumusan masalah dalam penelitian ini (1)Bagaimana mengelompokkan data forum pengaduan peserta pelatihan sesuai dengan isinya berupa kritik atau saran?. (2)Berapa akurasi yang dihasilkan oleh metode LVQ dalam mengklasifikasikan komentar?. Sehingga tujuan dari penelitian ini adalah (1) Mengelompokkan data forum pengaduan peserta menjadi dua kelas. yaitu, kritik atau saran sehingga mempermudah pegawai terkait dalam memprioritakan aspirasi peserta yang harus ditangani terlebih dahulu tentang sarana dan prasarana.(2) Mengetahui akurasi terbaik metode LVQ dalam mengklasifikasikan komentar.

Sistem ini menggunakan metode LVQ (*Learning Vector Quantization*), dimana masing-masing komentar akan dihitung dengan bobot dari hasil data *learning* serta membandingan komentar tersebut termasuk dalam kelas kritik atau saran. sehingga memberikan *output* label dari setiap komentar yang masuk. Pengelompokkan ini berdasarkan hasil dari penghitungan bobot dan komentar dengan mencari jarak terdekat dari hasil perhitungan tersebut.

Sistem klasifikasi ini membantu pegawai dalam melakukan pengelompokkan data pengaduan peserta sesuai dengan kelasnya yaitu kritik atau saran. Hal ini berguna untuk memudahkan pekerjaan pegawai dalam pengelompokkan sera mengetahui masalah apa yang harus segera ditangani sehingga dapat memberikan hasil pelayanan yang maksimal.

Kata Kunci : Data mining, klasifikasi, *Learning Vector Quantization.*

I. LATAR BELAKANG

Perkembangan teknologi telah mengakibatkan meningkatnya data dalam jumlah besar. Data berukuran besar yang sudah disimpan jarang digunakan secara optimal karena manusia seringkali tidak memiliki waktu dan kemampuan yang cukup untuk mengelolanya (Hermadi, 2007). Data bervolume besar seperti data teks, jauh melampaui kapasitas pengolahan manusia yang sangat terbatas (Toyota, 2012).

Kasus yang diangkat adalah data pengaduan atau evaluasi peserta pelatihan di UPT Diklat Pertanian Jawa Timur. Pengaduan atau evaluasi peserta pelatihan di UPT Diklat Pertanian Jawa Timur saat ini masih menggunakan cara manua, Masing masing peserta diminta untuk mengisi kritik maupun saran di kertas yang sudah disediakan. Kertas tersebut akan di kumopulkan oleh panitia untuk dilakukan pengelompokan. Pengelompokan data tersebut dilakukan secara manual sehingga sering terjadi kesalahan. Untuk menanggulangi hal tersebut, diperlukan adanya mekanisme pelabelan data pengaduan secara otomatis, sehingga dapat meminimalisir kesalahan.

Untuk memmberikan label pada data pengaduan peserta tersebut salah satunya bisa dengan menggunkan data mining sebagaimana penelitian yang telah dilakukuan oleh Hariri (2014) dalam Mengklasifikasikan dokumen abastrak tugas akhir sesuai dengan kategorinya.

Metode data mining yang akan diterapkan dalam penelitian ini adalah klasifikasi. Berbeda dengan *clustering* yang hanya mengelompokkan, proses klasifikasi, yaitu proses *learning*, dan juga adanya kasus baru yang akan dimasukkan ke dalam suatu kelas. Klasifikasi komentar adalah proses memberi label komentar sesuai dengan kategori yang dimilikinya. Klasifikasi komentar merupakan masalah yang mendasar namun sangat penting karena manfaatnya cukup besar mengingat jumlah komentar yang ada setiap hari semakin bertambah. Sebuah komentar dapat dikelompokkan ke dalam kategori tertentu berdasarkan kata-kata dan kalimat-kalimat yang ada di dalam komentar tersebut. Kata atau kalimat yang terdapat di dalam sebuah komentar memiliki makna tertentu dan dapat digunakan sebagai dasar untuk menentukan kategori dari komentar tersebut sehingga meminimalkan waktu dan sumber daya manusia dalam pengklasifikasian.

Berdasarkan permasalahan diatas maka penelitian ini akan membahas mengenai klasifikasi dokumen teks menggunakan metode LVQ (*Learning Vector Quantization*). Algoritma ini dikenal dengan kemampuannya dalam klasifikasi yang mempunyai tinggkat akurasi tinggi dan

kecepatan dalam hal proses (Hadnanto, 1996).

Penelitian ini fokus dalam melihat performa LVQ dalam mengklasifikasikan komentar dan untuk *stemming* dokumennya menggunakan algoritma nazief & andriani karena algoritma ini memiliki akurasi yang lebih baik dibandingkan algoritma *porter* (Agusta, 2009).

A. IDENTIFIKASI MASALAH

Berdasarkan uraian latar belakang yang telah dipaparkan, maka pokok identifikasi masalah adalah sebagai berikut:

1. Proses evaluasi pelatihan yang dilakukan di UPT Diklat Pertanian Jawa Timur masih dilakukan secara manual dan belum terkomputerisasi.
2. Pengelompokan data evaluasi / forum sering kali terjadi kesalahan dalam pemahaman data tersebut apakah termasuk kritik atau saran.
3. Sering terjadi kehilangan data karena faktor pengumpulan yang masih dilakukan secara manual.

B. PEMBATASAN PENELITIAN

Permasalahan yang akan dibahas di dalam penelitian ini dibatasi sebagai berikut:

1. Data komentar yang digunakan adalah data kalimat komentar dari peserta pelatihan yang sebelumya dituangkan dalam evaluasi pelatihan sebanyak 4 angkatan pelatihan.
2. Komentar yang diolah hanya yang berbahasa Indonesia saja.
3. Isi komentar yang diolah hanya yang berupa kata-kata, sedangkan simbol dan angka tidak diolah.
4. Proses *stemming* hanya berlaku pada kata-kata berbahasa Indonesia saja.
5. Data akan dikelompokkan dalam 2 kelas, yaitu Kritik Atau Saran.
6. Untuk data *learning* dan *testing* jumlah masing-masing kelas sebanding.
7. Untuk *stemming* dokumen menggunakan Algoritma Nazief & Adriani.
8. Akurasi ditentukan berdasarkan hasil asli yang dibandingkan dengan hasil dari algoritma LVQ.
9. Parameter yang diubah-ubah dan diamati hasilnya untuk mendapatkan kondisi terbaik adalah nilai dimensi.
10. Pengecekan parameter reduksi dimensi dilakukan antara 40% sampai 80% dengan penambahan parameter sebesar 20%.
11. *Prototype* yang akan dibuat hanya akan mensimulasikan proses pembentukan bobot oleh algoritma LVQ dan proses pengklasifikasian komentar.
12. *Database* menggunakan *mySQL* sebagai *database* pasif yang kedudukannya hanya digunakan menyimpan data.
13. *Prototype* dibangun menggunakan bahasa pemprograman PHP.

II. METODE

Jaringan saraf tiruan *Learning Vector Quantization* (LVQ) telah banyak dimanfaatkan untuk pengenalan pola baik berupa citra, suara, dan lain-lain. Jaringan LVQ sering pula digunakan untuk ekstraksi ciri (*feature*) pada proses awal pengenalan pola. Metode Jaringan Syaraf LVQ termasuk dengan *Supervised Learning* dalam penentuan bobot / model pembelajarannya, dimana pada metode LVQ ditentukan hasil seperti apa selama proses pembelajaran. Selama proses pembelajaran nilai bobot disusun dalam suatu *range* tertentu tergantung pada nilai *input* yang diberikan. Tujuan pembelajaran ini adalah pengelompokan unit-unit yang hampir sama dalam satu area tertentu. Pembelajaran seperti ini sangat cocok untuk pengelompokan (*klasifikasi*) pola (Efendi,2009)

Prinsip kerja dari algoritma LVQ adalah pengurangan node-node tetangganya (*neighbour*) (Kusumadewi,2003), sehingga pada akhirnya hanya ada satu *node output* yang terpilih (*winner node*). Pertama kali yang dilakukan adalah melakukan inisialisasi bobot untuk tiap-tiap *class.* Setelah diberikan bobot, maka jaringan diberi *input* sejumlah dimensi *node/neuron input*. Setelah input diterima jaringan, maka jaringan mulai melakukan perhitungan jarak vektor yang didapatkan dengan menjumlah selisih/jarak antara vektor *input* dengan vektor bobot menggunakan *Euclidean distance*. Secara matematis *Ecluidean Distance* dapat dirumuskan : (Kohonen,1988)

$$d_j^2 = \sum_{i=0}^{n-1} (X_1(t) - W_{ij}(t))^2 \ldots..(1)$$

dimana :

d_j^2 = *distance*

X_t = Node data *input*

W_{ij} = Bobot ke-ij

Setelah diketahui tiap-tiap jarak antara *node output* dengan *input* maka dilakukan perhitungan jumlah jarak selisih *minimum.* Dimana *node* yang terpilih (*winner*) berjarak minimum akan di *update* bobot, *update* bobot *node winner* yang dirumuskan sebagai berikut :

Jika sesuai target memakai rumus:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t).(x_i(t) - w_{ij}(t)), j \in N_e \ldots..(2)$$

dan jika tidak

$$w_{ij}(t+1) = w_{ij}(t) - \alpha(t).(x_i(t) - w_{ij}(t)), j \in N_e \ldots..(3)$$

$$0 < \alpha(t) < 1 \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Dimana :

x = *Input pixel*

w = bobot

Ne = Nilai *neighborhood*

t = waktu

i = *index node input*

j = *index node output*

α = *alpha learning rate*

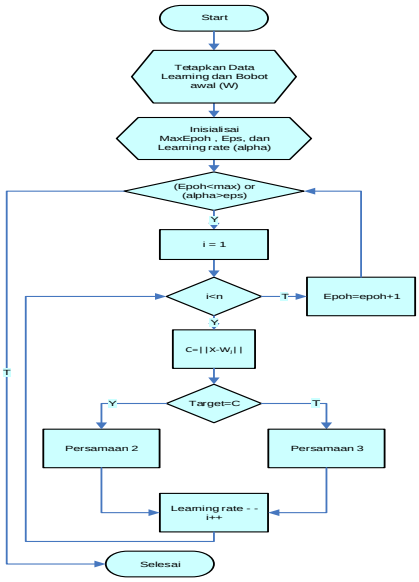
α(t) merupakan *alpha/learning rate* yaitu faktor pengali pada perubahan bobot yang berubah terhadap perubahan *error*. Perubahan *alpha* ini sesuai dengan banyaknya *input* yang masuk. Faktor pengali *alpha/learning rate* ini akan selalu berkurang bila tidak ada perubahan *error*. Dalam penelitian ini *alpha learning rate* akan berubah berkurang secara geometris sebagai berikut: (Balza,2005)

$$\alpha(t+1) = 0.1 * \alpha(t) \text{.....}(5)$$

Secara garis besar algoritma *Learning Vector Quantization* (LVQ) sebagai berikut:

1. Siapkan data learning, x (m,n) dan target T (1,n).
2. Inisialisasi bobot (W), maksimum *Epoch* (*Max Epoch*), error minimum yang diharapkan (Eps), *learning rate* (α). *Max Epoch* dan *learning rate* digunakan untuk menentukan batas ambang komputasi
3. Melakukan proses sebagai berikut selama (*Epoch*<*maxEpoch*) atau (α>eps).
   a. *Epoch* = *Epoch*+1
   b. Kerjakan untuk i=1 sampai n
      - Tentukan j sedemikian rupa sehingga $||X-W_j||$ minimum (Sebutsebagai C_j)
      - Perbaiki W_j dengan ketentuan
      - Jika $T=C_j$ maka : $W_j(baru)=W_j(lama) + \alpha(X-W_j(lama))$
      - Jika $T \neq C_j$ maka : $W_j(baru)=W_j(lama) - \alpha(X-W_j(lama))$
   c. Kurangi nilai α
4. Kembali ke langkah ke-3, jika (Epoch < *maxEpoch*) atau (α > eps) tidak terpenuhi, selesai.

Setelah dilakukan pelatihan, akan diperoleh bobot akhir (W). Bobot-bobot ini nantinya akan digunakan untuk melakukan klasifikasi terhadap data baru. Gambar 2.1 dibawah mengilustrasikan proses perhitungan metode LVQ.



Gambar 2.4 *Flowchart* LVQ

III. HASIL DAN KESIMPULAN

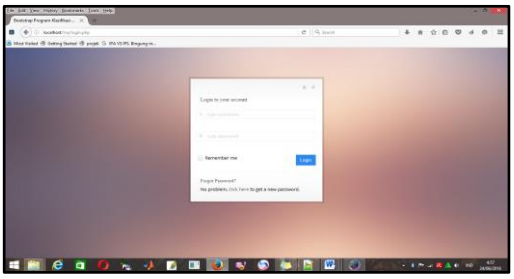
Dalam tahapan awal membuat sebuah aplikasi, tentu diperlukan sebuah rancangan awal sebagai patokan dan sebagai dasar fondasi agar nantinya aplikasi tersebut bisa berjalan sesuai dengan tujuan awal dibangunnya aplikasi tersebut. Maka dari itu, dalam sub bab Desain Rancangan Aplikasi ini, penulis berusaha menguraikan tahapan-tahapan dalam merancang sistem ini yang nantinya meliputi desain sketsa *input* serta *output*, DFD (*Data Flow Diagram*), dan terakhir adalah daftar tabel-tabel yang digunakan untuk menyimpan data dari sistem yang dibuat.

Nantinya, desain sketsa *input* meliputi rancangan *input* data komentar, login. Lalu desain sketsa *input* meliputi rancangan *ouput* beranda admin, komentar, data komentar, di dalam preporsesing terdapat beberapa *output* meliputi pembobotan TF, *Stemming Indexing*, dn terhir penghitungan metode LVQ.

Setelah selesai dengan tahapan perancangan sebuah aplikasi pada sub bab sebelumnya, langkah selanjutnya adalah tahapan implementasi dan evaluasi yang menampilkan tampilan (*interface*), *input-output*, serta fungsi-fungsi dari aplikasi yang sedang dibangun secara keseluruhan. Terutama yang akan dibahas dalam hal ini adalah tampilan keseluruhan dari dua (2) halaman pokok selain halaman *login* yang digunakan oleh *user*. Yaitu halaman admin, serta halaman peserta.

1. **Halaman *Form Login***

   Fungsi : sebagai halaman login untuk pegawai, Ka.Subag Tata Usaha, dan Kepala UPT.



Gambar 5.13 halaman *login*

2. **Halaman *admin***

   Pada halaman *admin* terdapat :

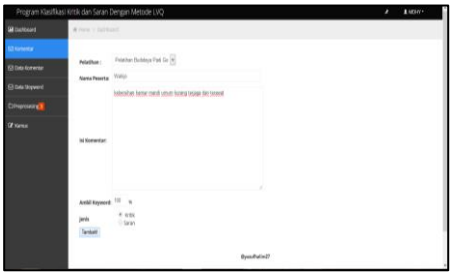
   a). Beranda *admin*

   Fungsi : Setelah berhasil login akan muncul tampilan seperti di bawah ini, yang menampilkan komentar kritik dn saran sesuai tabelnya, yang telah di kelompokkan oleh *system*. Tampilan *Home/ Dasboard* dapat dilihat pada gambar 5.14 :

Gambar 5.14 beranda *admin*

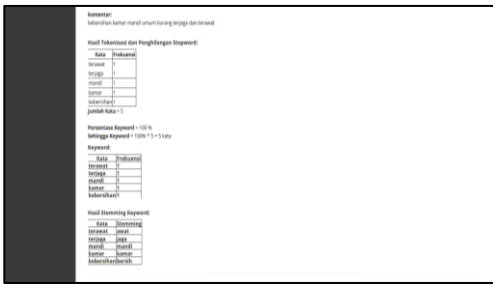
   b). Menu Komentar

Fungsi : menu ini untuk memasukkan komentar baru



Gambar 5.15 Form Komentar

Setelah komentar di isikan dan seluruh form di isi dan di klik tambah maka akan muncul tampilan dari komentar yang sudah di *tokenisasi* dan di *Stemming* seperti gambar 5.16 :



Gambar 5.16 Hasil tambah

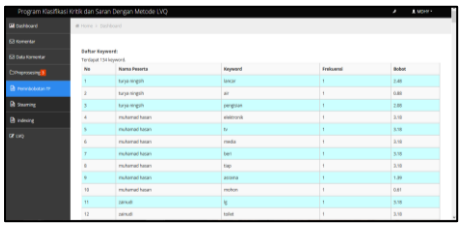
c). Menu Data Komentar

Fungsi : Halaman ini berisikan data-data komentar yang telah dimasukkan oleh *User*

Gambar 5.17 Data Komentar

d). Menu Pembobotan

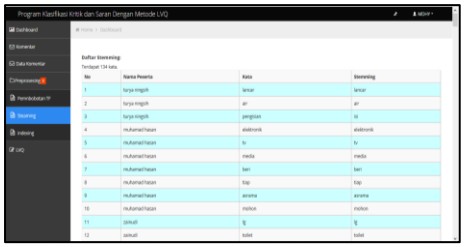
Fungsi : Halaman ini berisikan data-data komentar hasil dari proses pembobotan TF.



Gambar 5.18 Tampilan data Pembobotan

e). Menu *Stemming*

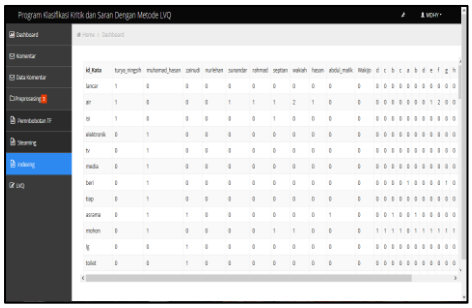
Fungsi : Halaman ini berisikan data-data hasil dari proses *Stemming*.



Gambar 5.19 Menu *Stemming*

f). Menu Indexing

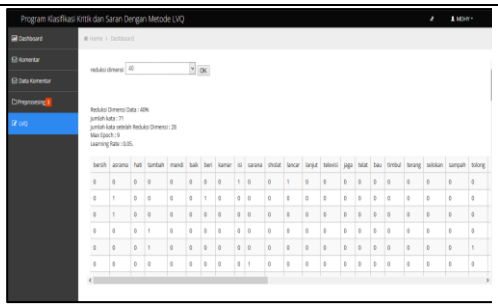
Fungsi : Halaman ini berisikan data-data hasil dari proses *Stemming* dan pembobotan lalu di *index*.



Gambar 5.20 Tampilan *indexing*

g). Menu LVQ

Fungsi : Halaman ini untuk menampilkan proses penghitungan LVQ.



Gambar 5.21 Tampilan Hasil LVQ

h). Menu Hasil *Update*

Fungsi : Halaman ini untuk menampilkan proses *update* komentar setelah dihitung dengan metode LVQ .

Gambar 5.22 Tampilan Data *Update*

IV. DAFTAR PUSTAKA

Agusta, Ledi. 2009 . *Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia*. Konferensi Nasional Sistem dan Informatika. 106-39

Alfiyanto, Dian ,Ratnawati ,Ridok. 2010, *Pengklasifikasian Kritik Dan Saran Dengan Menggunakan Metode Fuzzy K-Nearest Neighbour (F-Knn).* Program Studi Ilmu Komputer, Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.

Andini, Silfia. 2013, Klasifikasi Dokument Teks Menggunakan Algoritma Naive Bayes Dengan Bahasa Pemograman Java. *Jurnal Teknologi Informasi & Pendidikan* Vol.6 NO.2.

Effendy, N., Imanto, R., 2009. *Deteksi Pornografi Pada Citra Digital Menggunakan Pengolahan Citra Dan Jaringan Saraf Tiruan*, <URL:http://www.undip.ac.id/files/2010/02/deteksi-pornografi.pdf>. Diakses pada tanggal 3 Oktober 2015.

Fayyat, Usama. 1996. *Advanced in Knowledge and Discovery and Data Mining*. MTI Press.

Hamid, Nihal M. dkk. 2013, *Bees Algorithm-Based Document Clustering.* International Conference on Information Technology.

Hariri, Fajar Rahman., 2014. *Implementasi Learning Vector Quantization Untuk Klasifikasi Dokumen.* Tugas Akhir.Yogyakarta : STMIK AMIKOM Yogyakarta.

Hermadi, I.2007, Clustering Menggunakan Self-Organizing Maps (Studi Kasus: Data PPMB IPB). *FMIPA Institut Pertanian Bogor*. 5: 2.

Kadir, Abdul. 2010. *Algoritma dan Pemrograman Menggunakan C & C++*. Yogyakarta : Andi.

Kohonen, T.,1988, *An introduction to neural network, Neural Network*.,1:3-16

Kusumadewi, Sri. 2003. *Artificial Intelligence : Teknik & Aplikasinya*. Yogyakarta: Graha Ilmu.

Nazief, Bobby dan Mirna Adriani. 1996. *Confix-Stripping: Approach to Stemming Algorithm for Bahasa Indonesia*, Fakulty of Computer Science University of Indonesia.

Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikais menggunakan MATLAB*, Yogyakarta. Penerbit ANDI.

Toyota, T., Nobuhara, H. 2012, *Visualization of the Internet News Based on Efficient Self-Organizing Map Using Restricted Region Search and Dimensionality Reduction*. *JACIII*. 16: 222.

Turban, E, .2005.*Decision Support System and Intelligent System*, edisi Bahasa Indonesia jilid 1.Yogyakarta.Penerbit ANDI.

Yakub. 2012. *Pengantar Sistem Informasi*. Yogyakarta : Graha Ilmu