KLASIFIKASI KOMENTAR PADA GRUP PRODI TEKNIK INFORMATIKA 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 :

RIO ARDINAS
12.1.03.02.0159

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA

UN PGRI KEDIRI

2016

Skripsi Oleh :

RIO ARDINAS

12.1.03.02.0159

Judul :

KLASIFIKASI KOMENTAR PADA GRUP PRODI TEKNIK INFORMATIKA MENGGUNAKAN METODE LVQ (LEARNING VECTOR QUANTIZATION)

Telah disetujui untuk diajukan Kepada

Panitia Ujian/Sidang Skripsi Jurusan Teknik Informatika

FAKULTAS TEKNIK UN PGRI Kediri

Tanggal : 28 Juli 2016

Pembimbing I	**Pembimbing II**
Resty Wulanningrum, M.Kom.	**Risky Aswi Ramadhani, M. Kom.**
NIDN 0719068702	**NIDN 0708049001**

Skripsi Oleh :

RIO ARDINAS

12.1.03.02.0159

Judul :

KLASIFIKASI KOMENTAR PADA GRUP PRODI TEKNIK INFORMATIKA MENGGUNAKAN METODE LVQ (LEARNING VECTOR QUANTIZATION)

Telah dipertahankan di depan Panitia Ujian / Sidang Skripsi
Program Studi Teknik Informatika FT UN PGRI Kediri
Pada tanggal : 12 Agustus 2016

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua Penguji: RESTY WULANNINGRUM, M.KOM.
2. Penguji I : Ardi Sanjaya, M.Kom.
3. Penguji II : RISKY ASWI RAMADHANI, M.KOM

Mengetahui,
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.Pd.
NIP. 19640202 199103 1 002

Klasifikasi Komentar pada Grup Prodi Teknik Informatika menggunakan Metode LVQ (Learning Vector Quantization)

Rio Ardinas
12.1.03.02.0159
Teknik –Teknik Informatika
rio.ardinas@gmail.com
Resty Wulanningrum, M.Kom. dan Risky Aswi Ramadhani, M. Kom.
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

RIO ARDINAS : Klasifikasi Komentar pada Grup Prodi Teknik Informatika Menggunakan Metode LVQ (Learning Vector Quantization), Skripsi, Teknik Informatika, Fakultas Teknik UN PGRI Kediri 2016

Komentar tanya dan jawab adalah hal yang biasa dilakukan oleh mahasiswa pada sebuah group. Seperti halnya group facebook Departement Of Informatics, Nusantara PGRI Kediri. Agar admin Group mengerti berapa Mahasiswa yang bertanya dan menjawab komentar pada postingannya. Untuk menanggulangi hal tersebut, diperlukan adanya mekanisme pelabelan data komentar secara otomatis, sehingga dapat meminimalisir kesalahan.

Metode data mining yang akan diterapkan dalam penelitian ini adalah klasifikasi. 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 dalam pengklasifikasian komentar.

Pada penelitian kali ini dilakukan klasifikasi komentar menggunakan metode *LVQ* (*Learning Vector Quantization*). Data komentar diklasifikasikan menjadi 2 yaitu komentar tanya dan komentar jawab. Dari berbagai ujicoba yang dilakukan didapatkan hasil metode *LVQ* berhasil mengenali 14 data komentar, dengan berhasil mengenali 8 komentar tanya dan 5 komentar jawab. Kurangnya data training dapat menyebabkan kesalahan saat proses klasifikasi.

Kata Kunci : Teks Mining, Klasifikasi, *LVQ*.

I. Latar Belakang

Media sosial merupakan sebuah media online, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Jejaring sosial pada masa ini lekat dengan budaya digital. Jejaring sosial merupakan sebuah website dengan antarmuka, aplikasi atau *platform* tertentu, yang didedikasikan serta memungkinkan penggunanya untuk dapat berkomunikasi melalui informasi, pesan, gambar, video, komentar dan lain sebagainya (Fauzan,

2015). Seperti halnya facebook. Facebook dapat digunakan untuk berkomunikasi melalui pesan, gambar, video, komentar dan lain sebagainya.

Kasus yang diangkat adalah data komentar Mahasiswa pada postingan admin di group Facebook Department Of Informatics, Nusantara PGRI Kediri. Ada 2 jenis komentar yang akan diproses yaitu komentar Tanya dan komentar Jawab. Agar admin Group mengerti berapa Mahasiswa yang bertanya dan menjawab komentar pada postingannya. Untuk menanggulangi hal tersebut, diperlukan adanya mekanisme pelabelan data komentar secara otomatis, sehingga dapat meminimalisir kesalahan.

Data mining sangat sesuai untuk diterapkan pada data berukuran besar (Turban, 2005). Penerapan data mining pada data komentar mahasiswa di group facebook Department Of Informatics, Nusantara PGRI Kediri diharapkan bisa menambah ilmu pengetahuan dan informasi yang penting dan berguna untuk pengambilan keputusan di masa depan dan untuk mengevaluasi prodi teknik informatika. Selain itu kometar juga dapat digunakan wadah penampungan saran dan kritik Mahasiswa.

Metode data mining yang akan diterapkan dalam penelitian ini adalah klasifikasi. 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 dalam pengklasifikasian komentar.

Adapun beberapa penelitian mengenai klasifikasi komentar yang menginspirasi dilakukan penelitian pada skripsi kali ini yaitu: 1. Penelitian Klasifikasi Data Forum dengan menggunakan *Metode Naïve Bayes Classifier* oleh aida indriani pada tahun 2014. Pada penelitian ini di diperoleh hasil yaitu, dari 15 data uji diperoleh akurasi kecocokan kelas sebenarnya terhadap kelas prediksi, dengan NBC sebesar 73%. 2.

Klasifikasi Dokument Teks Menggunakan Algoritma Naive Bayes dengan Bahasa Pemograman Java oleh Silfia Andini pada tahun 2013, menggunakan artikel berita, dengan metode naïve bayes bisa membantu user dalam memilih dan mengkategorikan dokumen. Pada penelitian ini diperoleh hasil akurasi kebenaran pengklasifikasian terhadap 15 data uji sebesar 73%. Dengan menggunakan pengukuran efektifitas Confusion Matrix. 3. Penelitian tentang klasifikasi dokumen pernah dilakukan oleh Fajar Rahman Hariri untuk mengklasifikasikan dokumen abstrak pada tahun 2014 dengan judul *Learning Vector Quantization untuk Klasifikasi Abstrak Tesis* Untuk Klasifikasi Dokumen. Pada penelitian ini diperoleh hasil metode LVQ dalam mengklasifikasi dokumen abstrak mencapai 90% untuk keseluruhan data, dengan berhasil mengenali abstrak tugas akhir bidang minat SI RPL dan CAI dengan akurasi 100% , namun untuk bidang minat multimedia hanya menghasilkan akurasi sebesar 70%.

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 bagaimana metode *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), dan akan digunakan untuk mengklasifikasi data komentar mahasiswa di Department Of Informatics, Nusantara PGRI Kediri.

II. Metode *Klasifikasi Learning Vector Quantization (LVQ)*

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 (Effendy, 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 Euclidean Distance dapat dirumuskan : (Kohonen, 1988)

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

Dimana :

dj2 = distance

Xt = Node data input

Wij = 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 diupdate bobot, update bobot node winner yang dirumuskan sebagai berikut :

Jika sesuai target memakai rumus:

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

dan jika tidak

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

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

Dimana :

x = Input pixel

w = bobot

Ne = Nilai neighborhood

t = waktu

i = index node input

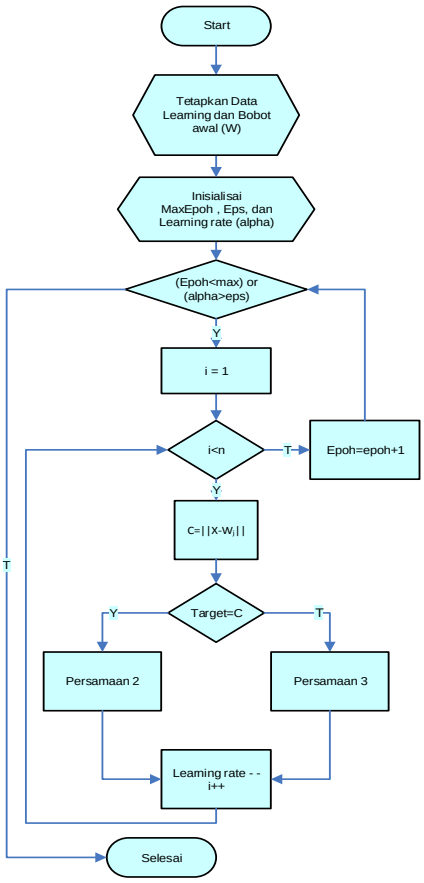
j = index node output

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)

$$\alpha(t+1) = 0.1 * \alpha(t)$$

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



Gambar 2.1. *Flowchart LVQ*

Algoritma dari model LVQ adalah sebagai berikut :

a. Siapkan data learning, x (m,n) dan target T (1,n).
b. Inisialisasi bobot (W), maksimum epoh (Max Epoh), error minimum yang diharapkan (Eps), learning rate (α). Max Epoh dan learning rate digunakan untuk menentukan batas ambang komputasi.
c. Melakukan proses sebagai berikut selama (epoh<makEpoh) atau (α>eps).
   1) epoh = epoh+1
   2) Kerjakan untuk i=1 sampai n
      - Tentukan j sedemikian rupa sehingga ||X-Wj|| minimum (Sebut sebagai Cj)
      - Perbaiki Wj dengan ketentuan
      - Jika T=Cj maka : Wj(baru)=Wj(lama) +α(X-Wj(lama))
      - Jika T≠Cj maka : Wj(baru)=Wj(lama) -α(X-Wj(lama))
   3) Kurangi nilai α
d. Kembali ke langkah ke-3, jika (epoh < makEpoh) 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.

III. KESIMPULAN

Dari semua analisis data yang sudah disajikan pada bab sebelumnya, maka dapat diambil kesimpulan mengenai bagaimana cara mengelompokkan komentar sesuai dengan isinya. Berdasarkan penelitian dan pembahasan, maka dapat diambil kesimpulan sebagai berikut :

1. Berdasarkan penelitian yang telah dilakukan oleh penulis maka cara yang digunakan untuk mengelompokkan komentar sesuai

dengan isinya yaitu dengan cara mengumpulkan data komentar yang terdahulu untuk dijadikan sebagai data training/data latih.

2. Cara mengimplementasikan metode LVQ pada aplikasi sistem klasifikasi komentar berbasis web yang dilakukan peneliti maka diambil langkah-langkah sebagai berikut, pertama user harus memasukkan data komentar terdahulu untuk dijadikan sebagai data training/data latih, setelah itu masukkan komentar baru pada data uji untuk mengetahui komentar tersebut masuk kelas apa.

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.

Balza, A., Kartika, F. 2005. *Teknik Pengolahan Citra Digital*. Yogyakarta: Andi Publishing.

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 Januari 2014.

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

Hadnanto, M.A. 1996. *Perbandingan Beberapa Metode Algoritma JST untuk Pengenalan Pola Gambar*. Tugas Akhir.Surabaya : Lab. Teknik Elektronika ITS Surabaya.

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

Huang, W. Y. and R. P. Lippmann, "*Comparisons between neural net and conventional classifiers*," IEEE 1st Int. Conf. Neural Networks, San Diego, CA, 1987, 485–493.

Indriani, Aida. 2014. *Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier*.

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.

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

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