

Analisis Sentimen Menggunakan Metode Learning Vector Quantization

Sentiment Analysis Using Learning Vector Quantization Method

M Indra Halim Arsyah Dwi Akbari¹, Astri Novianty S. T., M. T.², Casi Setianingsih S. T., M. T.³

^{1,2,3}Prodi S1 Sistem Komputer, Fakultas Teknik Elektro, Telkom University
Bandung, Indonesia

¹indrahalm@students.telkomuniversity.ac.id, ²astrinov@telkomuniversity.ac.id, ³setiacasie@telkomuniversity.ac.id

Abstrak

Media sosial sebagai media komunikasi untuk menghubungkan semua orang menjadi mudah diterima oleh masyarakat seiring dengan perkembangan teknologi informasi. Data dari media sosial yang berisi opini dari masyarakat tersebut dapat diolah menjadi sebuah informasi yang lebih penting nilainya, salah satunya dengan menggunakan analisis sentimen. Dalam perkembangannya analisis sentimen digunakan dalam banyak kepentingan misal perusahaan, lembaga survei, dan pemerintah.

Di dalam penelitian Tugas Akhir ini, dilakukan analisis sentimen menggunakan algoritma *Learning Vector Quantization* (LVQ). Langkah pertama yang dilakukan adalah pengambilan data *Tweet* dari Twitter menggunakan Twitter API. Setelah itu melakukan *preprocessing* untuk mengolah data dan mengekstraksi fitur yang terkandung di dalamnya. Setelah itu list fitur diubah menjadi bentuk *vector* dengan menggunakan pembobotan TF-IDF. Performansi akurasi dari metode didapatkan dengan menguji *learning rate*, *epoch* dan jumlah data yang digunakan.

Dari hasil pengujian performansi algoritma *Learning Vector Quantization* (LVQ) diketahui bahwa semakin tinggi *learning rate* maka akurasi yang didapatkan semakin kecil, kemudian semakin tinggi jumlah *epoch* maka akurasi semakin besar. Akurasi rata-rata yang berhasil dicapai pada penelitian ini adalah sebesar 73.15%, 66.42%, dan 69.58% untuk setiap data. Kata kunci : *Learning Vector Quantization*, LVQ, Analisis Sentimen, *Machine Learning*

Abstract

Social media as a medium of communication to connect everyone becomes easily accepted by the people along with the development of information technology. There are many ways to process data from social media that contains opinions from people to become some more valued information, one of them is using sentiment analysis. In its development, sentiment analysis has been used in many interests e.g. corporations, survey agencies, and governments.

This final project research used Learning Vector Quantization (LVQ) algorithm for analysis sentiment. The first step is to take Tweet data from Twitter using Twitter API and then do some standard preprocessing methods to process the data and extract the data features. The feature list is converted into a vector form by using TF-IDF weighting. The accuracy performance of the algorithm is obtained by testing the learning rate, epoch and the amount of data used.

Testing shows that one of important parameters that gives significant performance is learning rate parameter. The higher the learning rate will result poor accuracy. In the end of testing, the accuracy achieved in this research is 73.15%, 66.42% and 69.58% for each data.

Keywords : *Learning Vector Quantization*, LVQ, Sentiment Analytic, Machine Learning

1. Pendahuluan

Pertumbuhan *social media* di Indonesia beberapa tahun terakhir ini sangat tinggi mengingat internet adalah kebutuhan penting dalam mendapatkan dan berbagi informasi. Dengan kata lain, Internet adalah media penyimpan informasi terbesar di dunia yang terbagi menjadi banyak bentuk seperti artikel, postingan, multimedia, komentar, dan *review*. Hampir semua masyarakat lebih aktif di media sosial daripada dunia nyata dalam menyampaikan opini dan buah pikiran mereka tentang suatu masalah tertentu[2]. Dengan berkembangnya internet, masyarakat sudah menemukan cara baru dalam mendapatkan informasi ataupun memberikan informasi. Opini masyarakat tentang suatu masalah yang berlangsung merupakan suatu informasi yang berharga. Sebagai contoh dalam dunia bisnis, masyarakat sudah tidak lagi membutuhkan toko untuk berjualan, mereka dengan mudahnya memanfaatkan media sosial untuk mempromosikan produk mereka. Dalam dunia politik juga, para politisi lebih mudah mengenalkan diri mereka serta

tindakan mereka melalui media sosial[2]. *Sentiment Analysis* bertujuan untuk mengidentifikasi dan mengklasifikasikan opini atau sentimen pada *social media* seperti *Facebook*, *Twitter* dan lain-lain yang diekspresikan dalam bentuk *text*. Dengan begitu kita bisa memahami pola pikir dan opini mereka dengan mengklasifikasikannya ke dalam kelas-kelas tertentu.

Learning Vector Quantization (LVQ) adalah sebuah metode klasifikasi dimana setiap unit *output* dari algoritma merepresentasikan sebuah kelas. LVQ digunakan untuk pengelompokan dimana jumlah kelompok sudah ditentukan arsitekturnya (target/kelas sudah ditentukan). Dengan kelas yang sudah ditentukan tersebut akan didapatkan hasil sentimen berdasarkan bobot data yang sudah diberikan sebelumnya.

2. Dasar Teori

2.1 Analisis Sentimen

Analisis sentimen merupakan sebuah interdisiplin ilmu yang melibatkan *natural language processing*, *artificial intelligence* dan *text mining*[6]. Definisi secara umumnya adalah mengenali secara umum opini orang yang diekspresikan dalam sebuah *text*. Opini memiliki beberapa bentuk penilaian, evaluasi, emosi, kepercayaan, kepuasan. Pengertian dari analisis sentimen itu sendiri merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi. Analisis sentimen biasa dilakukan untuk melihat pendapat terhadap sebuah masalah, atau dapat juga digunakan untuk identifikasi kecenderungan hal di pasar. Secara umum, analisis sentimen dibagi menjadi dua kategori.[7] :

a. *Coarse-grained sentiment analysis*

Kategori *coarse-grained* ini merupakan pengklasifikasian yang berorientasi pada isi dari dokumen secara keseluruhan. Orientasi itu sendiri dibedakan menjadi tiga jenis yaitu positif, netral, dan negatif. Namun ada juga yang menjadikan nilai orientasi tersebut ke dalam sifat yang kontinu. tidak diskrit[14].

b. *Fined-grained sentiment analysis*

Kategori *finer-grained* ini tidak berfokus pada klasifikasi di tingkat dokumen, melainkan pada sebuah kalimat yang ada dalam suatu dokumen[14].

2.2 Twitter

Twitter adalah situs mikroblogging dalam social media yang memungkinkan penggunanya mengirimkan tweet yang berisikan 140 karakter yang dikenal dengan kicauan (tweet). Twitter dioperasikan oleh Twitter, Inc yang didirikan pada bulan Maret 2006 oleh Jack Dorsey. Twitter mengalami pertumbuhan yang pesat dan dengan cepat meraih popularitas di seluruh dunia. Hingga bulan Januari 2013, terdapat lebih dari 500 juta pengguna terdaftar dan 200 juta diantaranya adalah pengguna aktif. Lonjakan penggunaan Twitter umumnya berlangsung saat terjadinya peristiwa-peristiwa populer. Pada awal 2013, pengguna Twitter mengirimkan lebih dari 340 juta kicauan per hari, dan Twitter menangani lebih dari 1,6 miliar permintaan pencarian per hari [16].

2.3 Data Mining

Data Mining adalah studi tentang mengumpulkan (*collecting*), pembersihan (*cleaning*), pengolahan (*processing*), menganalisis (*analyzing*), dan memperoleh (*gaining*) wawasan atau pandangan yang berguna dari data tersebut. Ada variasi yang luas dalam domain masalah, aplikasi, formulasi, dan representasi data yang dihadapi dalam aplikasi nyata. Oleh karena itu, *Data Mining* adalah istilah luas payung yang digunakan untuk menggambarkan aspek-aspek berbeda dari pengolahan data (*data processing*). Proses data mining memiliki beberapa fase seperti pembersihan data (*data cleaning*), ekstraksi fitur (*feature extraction*), dan desain algoritma (*algorithmic design*).

2.4 Ekstraksi Fitur

Feature extraction atau ekstraksi fitur adalah proses pengolahan data agar menjadi format data yang dapat diolah oleh algoritma. Proses ini bertujuan untuk melakukan pencarian apakah di dalam data tersebut terdapat fitur atau tidak. Proses ini dilakukan untuk mengidentifikasi entitas yang dituju dalam suatu kalimat atau teks opini, serta mencari tahu opini yang dibicarakan terhadap fitur atau entitas yang sedang dibicarakan [15]. Dalam penilitan ini, metode ekstraksi fitur yang digunakan adalah dengan mengidentifikasi label kata benda (noun) dalam tweet yang dianalisis. Semua kata benda yang didapatkan dari hasil ekstraksi akan dijadikan sebagai fitur.

2.5 Term Frequency – Inverse Document Frequency

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan

menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus [18]. Metode ini akan menghitung bobot setiap token t di dokumen d dengan rumus [18] :

$$W_{dt} = tf_{dt} * IDF_t$$

Term Frequency (TF) adalah jumlah kemunculan term dalam suatu dokumen. Semakin besar jumlah kemunculan term maka semakin tinggi nilai bobot (W) term tersebut.

Inverse Document Frequency (IDF) digunakan untuk menangani term yang banyak muncul di berbagai dokumen (tweet) yang dianggap sebagai term yang umum (common term) sehingga tidak penting nilainya. Sebaliknya, faktor kelangkaan kata (uncommon term) dalam koleksi dokumen merupakan hal yang penting untuk diperhatikan dalam pemberian bobot. Oleh karena itu, pembobotan akan dihitung berdasarkan faktor kebalikan frekuensi dokumen yang mengandung suatu term (inverse document frequency). Untuk menghitung nilai IDF didapatkan dari persamaan [18] :

$$IDF = \log\left(\frac{N}{dft}\right)$$

2.6. Machine Learning

Teknik *machine learning* adalah salah satu teknik dalam *artificial intelligence* yang menggunakan training set dan test set untuk klasifikasi. Training set memuat inputan fitur vektor untuk memberikan label yang cocok pada tiap kelas. Test set digunakan untuk memvalidasi model dengan memprediksi label kelas[3]. Ada banyak algoritma dalam machine learning yang terbagi menjadi 3 tipe:

a. Supervised Learning

Supervised learning adalah teknik *machine learning* yang terawasi dimana hasil akhirnya sudah diketahui dengan membuat suatu fungsi dari data latihan. Data latihan terdiri dari pasangan input dan output yang diharapkan dari input yang bersangkutan[3]. Metode ini bekerja dengan memberikan data input dan menggambarkan output secara langsung.

b. Unsupervised Learning

Unsupervised learning adalah teknik *machine learning* yang menggunakan prosedur yang berusaha mencari partisi dari sebuah pola. *Unsupervised learning* mempelajari bagaimana sistem dapat belajar untuk merepresentasikan pola input dengan cara menggambarkan struktur statistik dari keseluruhan input. Metode ini tidak memiliki target output yang eksplisit atau tidak ada pengklasifikasian input[3]. Metode ini lebih bersifat *self-learning* dari mesin yang hampir meniru pola belajar otak manusia.

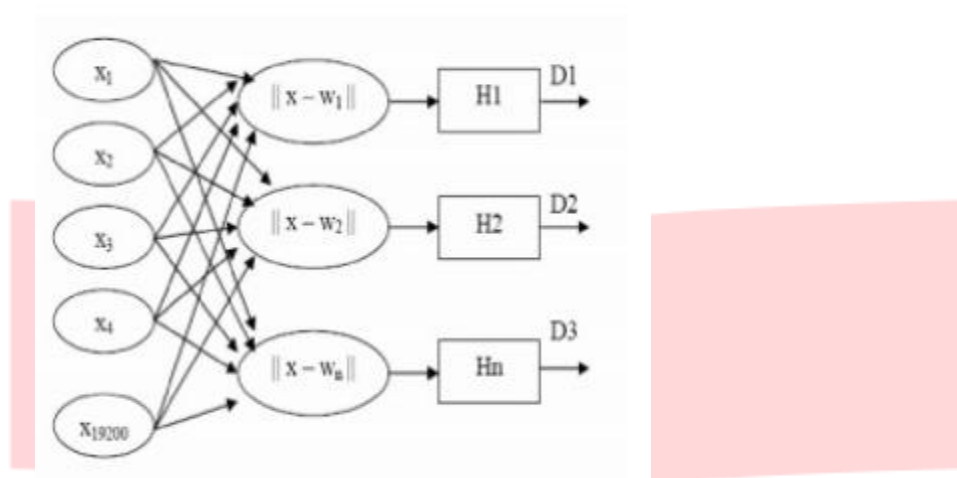
c. Hybrid Learning

Gabungan pembelajaran terawasi dan tak terawasi (*hybrid*), merupakan kombinasi dari kedua pembelajaran tersebut. Sebagian dari bobot-bobotnya ditentukan melalui pembelajaran terawasi dan sebagian lainnya melalui pembelajaran tak terawasi.

2.7. Learning Vector Quantization

Learning Vector Quantization adalah metode klasifikasi dimana setiap unit output merepresentasikan sebuah kelas. LVQ digunakan untuk pengelompokan dimana jumlah kelompok sudah ditentukan arsitekturnya (target/kelas sudah ditentukan). LVQ adalah salah satu Jaringan Syaraf Tiruan (JST) yang merupakan algoritma pembelajaran kompetitif terawasi versi dari algoritma Kohonen Self-Organizing Map (SOM). Tujuan dari algoritma ini adalah untuk mendekati distribusi kelas vektor untuk meminimalkan kesalahan dalam pengklasifikasian[4].

Konsep dasar dari LVQ adalah *competitive learning neural networks*, dimana semua sel akan dilatih untuk membentuk lapisan input, dan semua sel tersebut juga akan mendapatkan output yang sama. Lapisan kompetitif akan secara otomatis belajar untuk mengklasifikasikan vektor-vektor input. Berdasarkan konsep "*winner-take-all*", kelas-kelas yang didapat sebagai hasil dari lapisan kompetitif ini hanya tergantung pada jarak antara vektor-vektor input[20]. Jika vektor input mendekati sama maka lapisan kompetitif akan mengklasifikasikan kedua vektor input tersebut kedalam kelas yang sama..



Gambar - 1 Pola Algoritma LVQ

- a. x_1 sampai dengan x_{19200} = nilai input
- b. $\|x - w_1\|$ sampai dengan $\|x - w_n\|$ = jarak bobot
- c. H_1 sampai dengan H_n = lapisan output
- d. D_1 sampai dengan D_n = nilai output
- e. n = jumlah data karakter (jumlah kelas) dimana:
 - x_1 sampai dengan x_{19200} merupakan elemen matriks dalam setiap pola karakter yang akan dijadikan sebagai nilai input.
 - $\|x - w_1\|$ sampai dengan $\|x - w_n\|$ merupakan perhitungan jarak bobot terkecil dengan w_1 sampai dengan w_n adalah nilai data inisialisasi.
 - H_1 sampai dengan H_n adalah lapisan output.
 - D_1 sampai dengan D_n adalah bobot akhir yang nantinya akan dipakai dalam proses pengujian dengan data karakter baru yang dimasukkan.

Ada beberapa langkah yang dilakukan untuk mendapatkan nilai kuantisasi vector (LVQ Algorithm) yang dijabarkan seperti berikut :

1. Tetapkan bobot (w), maksimum iterasi/epoch (MaxEpoch), Error minimum (Eps) dan learning rate α .
2. Masukkan :
 - i. Input : $x(m,n)$; dimana m = jumlah input dan n = jumlah data
 - ii. Target : $T(1,n)$
3. Tetapkan kondisi awal : $Epoch=0$
4. Kerjakan jika: ($epoch < MaxEpoch$) atau ($\alpha > Eps$)
 - i. $Epoch = epoch + 1$
 - ii. Kerjakan untuk $i = 1$ sampai n
 - tentukan j sehingga $\|x - w_j\|$ adalah minimum
 - iii. Perbaiki w_j dengan ketentuan:

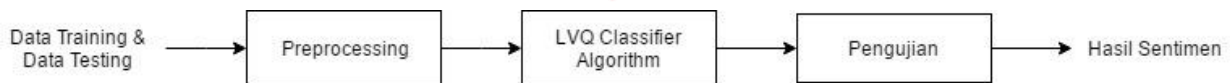
- Jika $T = J$, maka:
 $w_j(\text{baru}) = w_j(\text{lama}) + \alpha(x - w_j(\text{lama}))$
- Jika $T \neq J$, maka:
 $w_j(\text{baru}) = w_j(\text{lama}) - \alpha(x - w_j(\text{lama}))$

iv. Kurangi nilai α

3. Pembahasan

3.1 Deskripsi Sistem

Penelitian ini melakukan analisis terhadap data *tweet* dari Twitter untuk menganalisa sentimen terhadap objek pemelukada menggunakan algoritma Learning Vector Quantization pada machine learning. Analisis sentimen pada penelitian ini berada pada kategori *Fined-grained sentiment analysis* dimana sentimen berupa suatu kalimat (*tweet*) bukan suatu dokumen. Aplikasi yang dibuat untuk mengetahui performansi klasifikasi dari algoritma LVQ dalam mengklasifikasi data. Hasil keluaran dari aplikasi adalah persentase kecocokan klasifikasi dari data uji terhadap data training berupa hasil sentimen positif dan negatif terhadap objek yang diteliti.



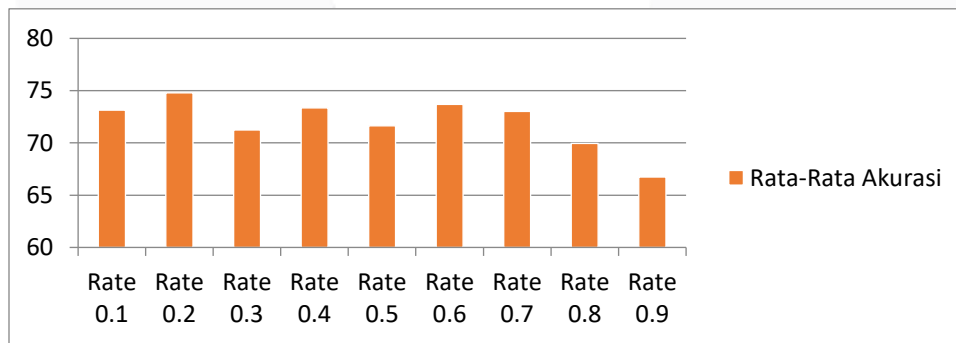
Gambar - 3.1 Blok Diagram Sistem

3.2 Pengujian

Pengujian performansi dilakukan dengan menguji perubahan yang terjadi apabila nilai learning rate dan jumlah iterasi berubah. Nilai Learning Rate adalah sebuah parameter yang umum dalam proses belajar algoritma Jaringan Syaraf Tiruan yang mempengaruhi kecepatan dimana JST tiba disolusi minimum. Sedangkan jumlah iterasi (*epoch*) adalah jumlah perulangan pelatihan yang dilakukan dalam suatu jangka waktu sehingga tingkat kesalahan menjadi nol. Pengujian performansi dilakukan sebanyak 10 kali percobaan terhadap masing-masing kondisi untuk melihat rata-rata hasil performansi dari setiap kondisi. Hasil rata-rata dari tiap percobaan menunjukkan kecenderungan hasil klasifikasi data.

3.2.1 Pengujian 100 Data

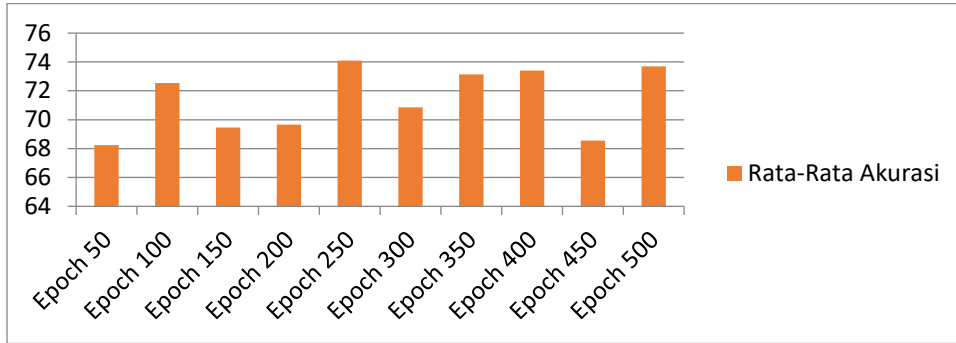
Hasil uji coba klasifikasi metode LVQ pada 100 data training dan testing dengan mengubah learning rate didapatkan hasil sebagai berikut :



Gambar – 3.2 Grafik Hasil Uji Akurasi 100 data dengan Learning Rate

Dengan menguji *learning rate* dengan menggunakan jumlah iterasi sebanyak 500 kali didapatkan hasil seperti diatas. Dari hasil pengujian di atas didapatkan akurasi tertinggi pada saat *learning rate* 0.2 dengan nilai akurasi sebesar 74.8%, sedangkan akurasi terendah didapatkan pada saat *learning rate* 0.9 dengan nilai akurasi sebesar 66.75%.

Hasil uji coba klasifikasi metode LVQ pada 100 data training dan testing dengan mengubah jumlah *epoch* didapatkan hasil sebagai berikut :

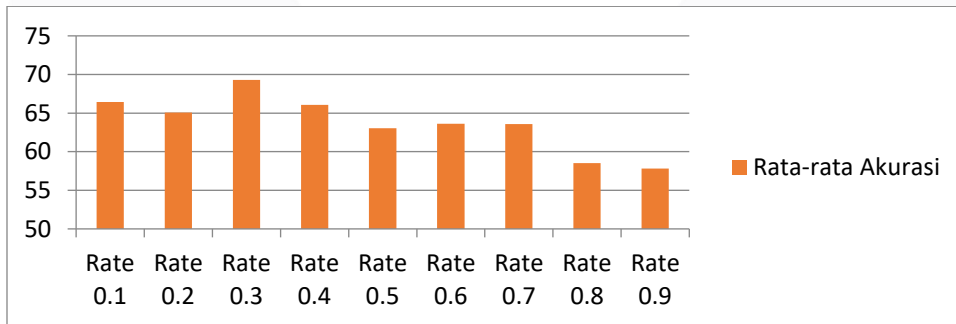


Gambar - 3.3 Grafik Uji Akurasi 100 data dengan Epoch

Dengan menguji *epoch* atau jumlah iterasi dengan menggunakan *learning rate* sebesar 0.1 didapatkan hasil seperti diatas. Dari hasil pengujian di atas didapatkan akurasi tertinggi pada epoch 250 kali dengan nilai akurasi sebesar 74.1%, sedangkan akurasi terendah didapatkan pada saat epoch 50 kali dengan nilai akurasi sebesar 68.25%.

3.2.2. Pengujian 200 Data

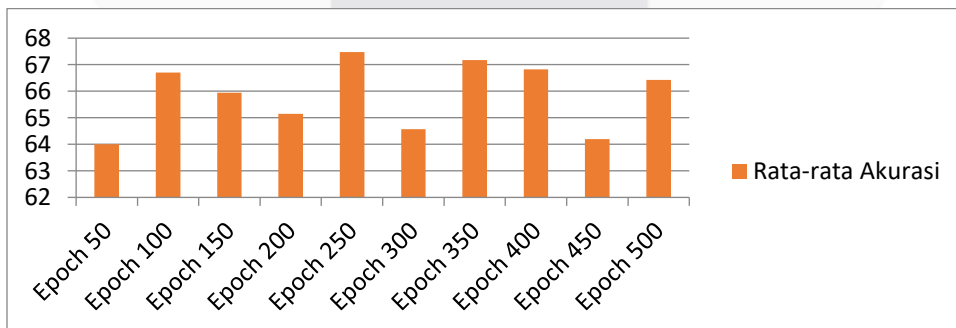
Hasil uji coba klasifikasi metode LVQ pada 200 data training dan testing dengan mengubah *learning rate* didapatkan hasil sebagai berikut :



Gambar - 3.4 Grafik Uji Akurasi 200 data dengan Learning Rate

Dengan menguji *learning rate* dengan menggunakan jumlah iterasi sebanyak 500 kali didapatkan hasil seperti diatas. Dari hasil pengujian di atas didapatkan akurasi tertinggi pada saat *learning rate* 0.3 dengan nilai akurasi sebesar 69.3%, sedangkan akurasi terendah didapatkan pada saat *learning rate* 0.9 dengan nilai akurasi sebesar 57.8331%.

Hasil uji coba klasifikasi metode LVQ pada 200 data training dan testing dengan mengubah jumlah iterasi (*epoch*) didapatkan hasil sebagai berikut :

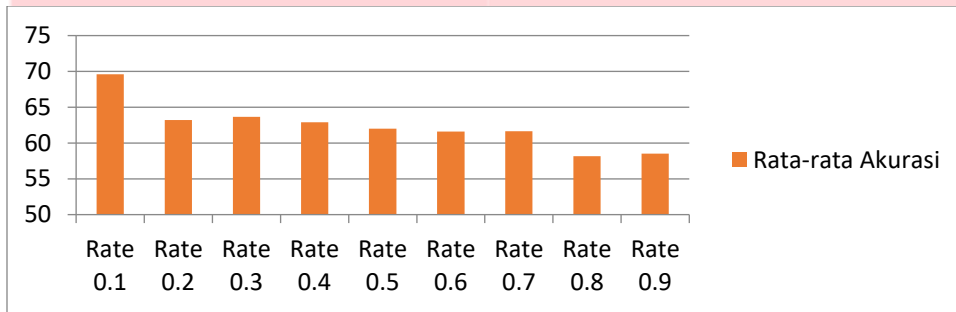


Gambar - 3.5 Grafik Hasil Uji Akurasi 200 data dengan Epoch

Dengan menguji *epoch* atau jumlah iterasi dengan menggunakan *learning rate* sebesar 0.1 didapatkan hasil seperti diatas. Dari hasil pengujian di atas didapatkan akurasi tertinggi pada epoch 250 kali dengan nilai akurasi sebesar 67.475%, sedangkan akurasi terendah didapatkan pada saat epoch 50 kali dengan nilai akurasi sebesar 64%.

3.2.3. Pengujian 300 Data

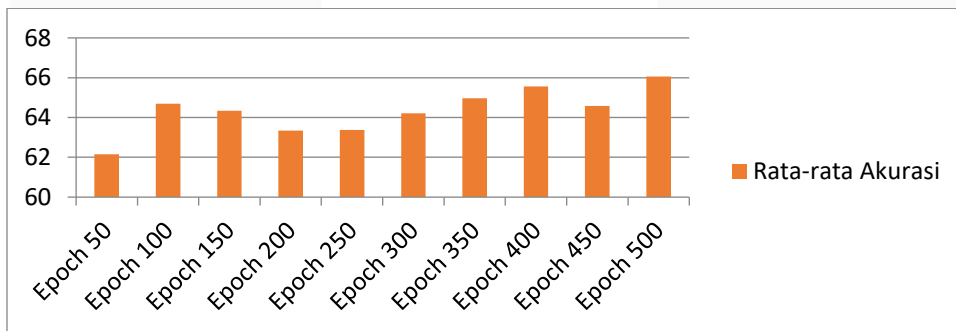
Hasil uji coba klasifikasi metode LVQ pada 300 data training dan testing dengan mengubah *learning rate* didapatkan hasil sebagai berikut :



Gambar - 3.6 Grafik Hasil Uji Akurasi 300 data dengan Learning Rate

Dengan menguji *learning rate* dengan menggunakan jumlah iterasi sebanyak 500 kali didapatkan hasil seperti diatas. Dari hasil pengujian di atas didapatkan akurasi tertinggi pada saat *learning rate* 0.1 dengan nilai akurasi sebesar 69.5864%, sedangkan akurasi terendah didapatkan pada saat *learning rate* 0.8 dengan nilai akurasi sebesar 58.148%.

Hasil uji coba klasifikasi metode LVQ pada 300 data training dan testing dengan mengubah jumlah iterasi (*epoch*) didapatkan hasil sebagai berikut :

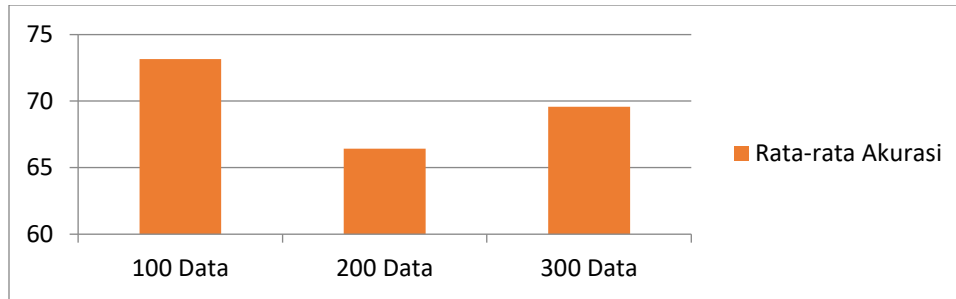


Gambar - 3.7 Grafik Hasil Uji Akurasi 300 data dengan Epoch

Dengan menguji *epoch* atau jumlah iterasi dengan menggunakan *learning rate* sebesar 0.1 didapatkan hasil seperti diatas. Dari hasil pengujian di atas didapatkan akurasi tertinggi pada epoch 500 kali dengan nilai akurasi sebesar 66.04%, sedangkan akurasi terendah didapatkan pada saat epoch 50 kali dengan nilai akurasi sebesar 62.149%.

3.2.4. Pengujian Jumlah Data Berbeda

Hasil yang pengujian dengan mengubah jumlah data tetapi menggunakan nilai *learning rate* sama dan jumlah iterasi yang sama didapatkan hasil seperti berikut :



Gambar - 3.8 Grafik Hasil Uji Akurasi dengan Jumlah Data berbeda

Dengan mengganti jumlah data didapatkan hasil seperti diatas didapatkan hasil akurasi tertinggi pada 100 data dengan nilai rata-rata akurasi sebesar 73.15%, dan nilai akurasi yang terendah didapatkan pada 200 data dengan nilai rata-rata akurasi sebesar 66.425%.

Dari hasil pengujian tersebut didapatkan hasil jumlah rata-rata sentimen dari 10 kali pengujian. Untuk 100 data didapatkan hasil rata-rata jumlah sentiment positif sebanyak 56 *tweet*, dan hasil rata-rata jumlah sentimen negatif sebanyak 44 *tweet*. Untuk 200 data didapatkan hasil rata-rata jumlah sentimen positif sebanyak 84 *tweet* dan hasil rata-rata jumlah sentimen negatif sebanyak 116 *tweet*. Untuk 300 data didapatkan hasil rata-rata jumlah sentiment positif sebanyak 147 *tweet* dan hasil rata-rata sentimen negatif sebanyak 153 *tweet*.

Dari hasil pengujian di atas dapat dianalisis bahwa akurasi yang paling baik terdapat pada pengujian 100 data dengan hasil kecenderungan akurasi sebesar 73.15% dan sistem berhasil menebak 56 *tweet* positif dan 44 *tweet* negatif. Dalam hal ini learning rate dan jumlah iterasi berpengaruh terhadap nilai akurasi. Semakin besar nilai learning rate yang digunakan maka hasil akurasi yang didapatkan cenderung semakin kecil. Semakin besar nilai epoch yang digunakan maka akurasi akan cenderung semakin meningkat. Sedangkan jumlah data uji mempengaruhi tingkat variasi data sehingga sistem akan memiliki semakin banyak fitur data yang akan digunakan yang akan mempengaruhi akurasi, semakin banyak fitur yang tersaring maka akurasi juga akan semakin bagus.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Berdasarkan hasil uji coba sistem pada penelitian ini, didapatkan beberapa kesimpulan sebagai berikut :

1. Akurasi performansi yang diperoleh menggunakan metode LVQ dalam menganalisis sentimen dari masing-masing 100, 200, dan 300 data adalah 73.15%, 66.425%, dan 69.5864. Dari hasil akurasi tersebut dapat kita simpulkan bahwa penggunaan algoritma LVQ baik dalam menganalisis sentimen. Sentimen yang dihasilkan sistem untuk 100 data sebanyak 56 *tweet* positif dan 44 *tweet* negatif, untuk 200 data sebanyak 84 *tweet* positif dan 116 *tweet* negatif dan untuk 300 data sebanyak 147 *tweet* positif dan 153 *tweet* negatif.
2. *Learning rate* berpengaruh pada performansi akurasi algoritma. Semakin besar nilai *learning rate* yang digunakan, maka hasil akurasi yang didapatkan semakin kecil. Hasil paling baik didapatkan dengan menggunakan *learning rate* 0.2 pada 100 data dengan hasil 74.8%, *learning rate* 0.1 pada 200 data dan 300 data dengan hasil masing-masing 66.425% dan 69.5864%.
3. Jumlah iterasi (*epoch*) juga berpengaruh pada akurasi algoritma. Semakin besar jumlah iterasi yang digunakan maka hasil akurasi yang didapatkan semakin besar. Hasil paling baik didapatkan dengan menggunakan jumlah iterasi 250 pada masing-masing 100 data dan 200 data dengan hasil masing – masing 74.1% dan 67.475%, sedangkan untuk 300 data hasil paling bagus didapatkan dengan menggunakan jumlah iterasi 500 dengan hasil akurasi sebesar 66.04%.

4.2. Saran

Saran untuk penelitian selanjutnya :

1. Disarankan mencoba menggunakan sistem pembobotan yang berbeda.
2. Menambahkan teori mengenai penanganan negasi untuk lebih menganalisa kevalidan *tweet* untuk mengurangi tingkat ambiguitas dalam *tweet*.
3. Menambah frekuensi kata dalam kamus sehingga kosakata dari kata-kata tidak baku (*slang*) bisa tertangani dengan baik.
4. Untuk masalah optimasi akurasi LVQ bisa ditambahkan algoritma yang lain seperti algoritma genetika dan SOM.

Cek lagi perbaikannya di semua bagian, pastikan sama persis dengan perbaikan terakhir di buku TA. Tata penulisannya, blok diagram/flowchart, sitasi rumus, gambar, dsb..

DAFTAR PUSTAKA

- [1] Enny Itje Sela, S. H. (2005). **Pengenalan Jenis Penyakit THT Menggunakan Jaringan**. Yogyakarta: STMIK Akakom, Universitas Gajah Mada.
- [2] Troussas. C, Espinosa. J, Llaguno. Kevin, Caro Jamie, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning", Department of Computer Science
- [3] Zelfia.Nita, "Machine Learning", 4 April 2012, 7 Maret 2016, <[http://nita_zelfiafst09.web.unair.ac.id/artikel_detail-44883-Umum Machine%20Learning.html](http://nita_zelfiafst09.web.unair.ac.id/artikel_detail-44883-Umum%20Machine%20Learning.html)>
- [4] Al-Ikhs. Dwi, "Konsep Learning Vector Quantization (LVQ)", 3 Juli 2011, 6 Maret 2016, <<https://ikhs.wordpress.com/2011/07/03/konsep-learning-vector-quantization-lvq/>>
- [5] Tiara, Mira Kania Sabariah dan Veronikha Effendy. 2015. *Sentiment Analysis on Twitter Using the Combination of Lexicon-Based and Support Vector Machine for Assesing the Performance of a Television Program*. International Conference on Information Communication Technology (ICoICT).
- [6] Semiocast, 2012, "Twitter reaches half a bilion accounts More than 140 millionin the U.S.", 4Maret2016, <http://semiocast.com/en/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US>;
- [7] Pang, B., Lee, L., & Vithyanathan, S. (2002). *Thumbs Up ? Sentiment Classification Using Machine Learning Techniques*. Proceedings of The ACL-02 conference on Empirical methods in natural language processing (pp. 79-86). Stroudsburg: Association for ComputationalLinguistic
- [8] Siqueira, H., Barros, F. "A Feature Extraction Process for Sentiment Analysis of Opinions on Services".Universidade Federal de Pernambuco, Brazil
- [9] M.S Neethu, R Rajasree. 2013. *Sentiment Analysis in Twitter using Machine Learning Techniques*. Tiruchengode, India
- [10] Wicaksono, A.F., Purwarianti, A. (2010). "HMM Based POS Tagger for Bahasa Indonesia". On Proceedings of 4th International MALINDO (Malay - Indonesian Language) Workshop. 2nd August 2010
- [11] Ronald M.Kaplan.2005. *A Method for Tokenizing Text..* Penulis Individu : 55.
- [12] Wang D.D, Xu Jinwu. "Fault Detection Based On Evolving LVQ Neural Networks", University of Science and Technology Beijing. China

- [13] Rista Nur.K, Y.S. Christian ,dkk, "**Materi Mata Kuliah Jaringan Syaraf Tiruan Learning Vector Quantization**", Universitas Gadjah Mada, Yogyakarta
- [14] Tiara. (2014). "**Analisis Sentimen pada Twitter untuk Menilai Performansi Program Televisi dengan Kombinasi Metode Lexicon-based dan Support Vector Machine**". Tugas Akhir, Universitas Telkom,Bandung.
- [15] Siqueira, H., Barros, F. "**A Feature Extraction Process for Sentiment Analysis of Opinions on Services**", Universidade Federal de Pernambuco, Brazil
- [16] Aggarwal, Charu.C. "**Data Mining: The Textbook**", Springer, IBM T.J Watson Research Center, New York
- [17] Jurafsky, D., Martin, J.H. (2000). "**Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognitio**", PrenticeHall, Inc. New Jersey, USA.
- [18] Maarif, Abdul Azis. "**Penerapan Algoritma TF-IDF Untuk Pencarian Karya Ilmiah**". Teknik Informatika Universitas Dian Nuswantoro, Semarang