

ANALISIS SENTIMEN PUBLIK PADA MEDIA SOSIAL TWITTER TERHADAP PELAKSANAAN PILKADA SERENTAK MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

PUBLIC SENTIMENT ANALYSIS ON TWITTER SOCIAL MEDIA TO PILKADA SERENTAK EVENT USING SUPPORT VECTOR MACHINE ALGORITHM

Atika Rahmawati¹

Aris Marjuni²

Junta Zeniarja³

Jurusan Teknik Informatika, Universitas Dian Nuswantoro Semarang^{1,2,3}
e-mail: 111201207272@mhs.dinus.ac.id¹, aris.marjuni@dsn.dinus.ac.id²,
junta@dsn.dinus.ac.id³

Diterima: 6 Juni 2017/ Disetujui : 13 Juni 2017

ABSTRACT

Pilkada Serentak is a very important event for the future viability regions and countries. Through this election people can cast their vote and elect representatives of the people according to their choice. Public respond can be expressed through twitter social media. Using twitter social media sentiment analysis can then be made about the public response to the implementation of the election simultaneously. The classification process can be detected via text tweeted by twitter users. In this study, the classification of responses detected by text because it is easily obtained and applied. This study determined the classification of the response to the Indonesian language text and increase accuracy by using SVM. Tweet classification method used by the categorical approach is divided into two classes tweet basic level: positive and negative. Data collected from Indonesian twitter tweet as much as 3000. The labeling is not done manually but using clustering method that divides the 3000 data into two groups. Cluster 1 as a group of positive tweets and Cluster 2 as a negative group tweet. 2700 for training data and 300 for the test data. The stage of pre-processing the data include tokenization, case normalization, stop word detection, and stemming. The process of classification using Support Vector Machine (SVM). Accuracy of SVM showed the highest yield that is 91% compared to the k-means clustering with the results of 82%.

Kata kunci : *Pilkada serentak, sentiment analysis, twitter, k-means, SVM*

ABSTRAK

Pilkada serentak merupakan kegiatan yang sangat penting untuk kelangsungan masa depan daerah dan negara. Melalui pilkada ini masyarakat dapat memberikan suaranya dan memilih wakil rakyat sesuai hati nurani mereka. Respon yang diberikan masyarakat dapat diungkapkan melalui media sosial twitter. Melalui media sosial twitter maka dapat dibuat analisis sentimen tentang respon masyarakat terhadap pelaksanaan pilkada serentak. Proses klasifikasi dapat dideteksi melalui teks yang ditweet oleh pengguna twitter. Dalam penelitian ini, klasifikasi respon dideteksi melalui teks karena lebih mudah didapat dan diterapkan. Penelitian ini melakukan klasifikasi respon untuk teks bahasa Indonesia dan meningkatkan akurasi dengan menggunakan metode klasifikasi svm. Tweet yang digunakan berdasarkan pendekatan kategorik yang dibagi menjadi 2 kelas tweet tingkat dasar yaitu positif dan negatif. Data dikumpulkan dari twitter dengan tweet bahasa Indonesia sebanyak 3000. Pemberian label tidak dilakukan

secara manual tetapi menggunakan metode clustering yang membagi 3000 data kedalam dua kelompok. Cluster 1 sebagai kelompok tweet positif dan Cluster 2 sebagai kelompok tweet negatif. 2700 untuk data pelatihan dan 300 untuk data pengujian. Tahap pre-processing pada data meliputi tokenisasi, casenormalization, deteksi stop word, dan stemming. Proses klasifikasi menggunakan metode Support Vector Machine (SVM). Akurasi SVM menunjukkan hasil tertinggi yaitu 91% dibanding clustering k-means dengan hasil 82%.

Kata kunci : *Pilkada serentak, analisis sentimen, twitter, k-means, SVM*

PENDAHULUAN

Pilkada serentak adalah pemilihan kepala daerah yaitu gubernur, bupati, dan walikota yang dilaksanakan secara bersamaan. Komisioner Komisi Pemilihan Umum (KPU) RI, Arief Budiman, dalam Acara Rapat Koordinasi (Rakor) Persiapan dan Pengelolaan Anggaran Pemilihan Serentak Tahun 2015 menjelaskan bahwa tujuan dilaksanakannya pemilihan kepala daerah serentak supaya tercipta efektivitas dan efisiensi anggaran analisis dan opini pilkada. Selain itu keuntungan dari pilkada serentak diantaranya perencanaan pembangunan lebih sinergi antara pusat dan daerah. Keuntungan bagi rakyat yaitu tidak perlu berulang kali datang ke bilik suara. Walaupun tujuan dari pilkada serentak sudah jelas, namun masih memunculkan pro dan kontra. Hal yang menyebabkan pro dan kontra adalah perpu pilkada Nomor 1 Tahun 2014 yang bertele-tele.

Dari pembahasan masalah yang muncul diatas, maka penting untuk memahami apa yang orang pikirkan atau pendapat mereka adalah fundamental untuk pengambilan keputusan, terutama saat di mana orang mengekspresikan komentar mereka secara sukarela, yang bertujuan untuk saling bekerja sama. Opini yang disampaikan masyarakat melalui media sosial twitter akan menjadi sumber data bagi penelitian ini. Pengguna twitter tweet tentang topik apapun dalam 140 karakter membatasi dan mengikuti orang lain untuk menerima tweets mereka. Tidak seperti layanan jaringan sosial lainnya yang mengharuskan pengguna untuk memberikan teman link ke pengguna lain berteman dengan mereka, twitter mempekerjakan model jaringan sosial yang disebut "*following*", dimana masing masing pengguna diperbolehkan untuk memilih siapa yang dia inginkan untuk mengikuti tanpa meminta izin apapun.

Analisis sentimen baru-baru ini menjadi fokus dari banyak peneliti, karena analisis online adalah menguntungkan dan menuntut untuk riset pasar, survei ilmiah dari perspektif psikologis dan sosiologis, jajak pendapat politik, intelijen bisnis, peningkatan secara online infrastruktur belanja, dan lain-lain. SA adalah proses penggalian polaritas rakyat pendapat subjektif dari teks bahasa alami biasa. Ada dua metode dasar yang ada, yang pertama adalah tingkat dokumen dan yang kedua adalah tingkat kalimat. Dalam tingkat dokumen, analisis didasarkan pada lengkap dokumen, sedangkan di tingkat kalimat, analisis adalah dilakukan pada tingkat kalimat.

Penelitian mengenai analisis sentiment telah banyak dilakukan. Penelitian sebelumnya yaitu dengan judul Perbandingan Teknik SVM Vs Naif-Bayes untuk Analisis sentimen Tweet: Studi Kasus dengan Piala Konfederasi FIFA 2013. Hasil Naif-Bayes dan SVM subjektivitas pengklasifikasi menunjukkan akurasi 82% dan 84%; dan F-ukuran 0,819 dan 0,821, berturut-turut. Hasil yang diperoleh oleh SVM sentimen classifier menunjukkan sebuah F-Measure of 0,873 dan akurasi 80,0% untuk mendeteksi sentimen polaritas. Naif-Bayes classifier sentimen disajikan F-ukuran dari 0,791 dan akurasi 72,7%. Penggunaan classifier SVM meningkatkan akurasi dalam 8%, ini adalah hasil yang baik.

Berdasarkan penjabaran diatas maka dilakukan penelitian tentang analisa sentiment publik terhadap pelaksanaan pilkada serentak dengan metode klasifikasi menggunakan algoritma support vectore machine (svm). Penelitian ini diharapkan dapat berguna bagi KPU

untuk mengetahui bagaimana respon masyarakat terhadap pelaksanaan pilkada serentak. Bagi masyarakat dapat terlibat langsung dengan memberikan pendapat mengenai pelaksanaan pilkada serentak.

Penelitian menganalisis sentiment menggunakan algoritma SVM bukan yang pertama kali dilakukan. Beberapa penelitian yang relevan dengan membandingkan algoritma svm dengan algoritma lain telah dilakukan oleh peneliti sebelumnya diantaranya adalah sebagai berikut:

Peneliti terdahulu telah melakukan perbandingan algoritma svm dengan naïve bayes untuk analisis sentimen di twitter dengan studi kasus Piala Konfederasi FIFA 2013. Hasil yang dicapai dengan pengklasifikasi dikembangkan, hasil terbaik diperoleh oleh classifier yang mengimplementasikan teknik SVM. Hasil Naif-Bayes dan SVM subjektivitas pengklasifikasi menunjukkan akurasi 82% dan 84%; dan F-measure masing-masing 0,819 dan 0,821.

Penelitian lain yang membandingkan dua algoritma juga telah dilakukan, mengenai analisis sentimen terhadap wacana politik pada media masa online menggunakan algoritma svm dan naïve bayes [7]. Penelitian ini menjelaskan hasil analisa dengan uji data sebanyak 700 data dengan menerapkan algoritma svm menghasilkan jumlah akurasi yang lebih besar yaitu 90,50%. Sedangkan penerapan algoritma naïve bayes dengan uji data yang sama yaitu 700 data menunjukkan akurasi hanya sebesar 59,98%. Hasil ini dapat menyimpulkan bahwa svm memiliki tingkat akurasi yang lebih baik dari naïve bayes.

Penelitian selanjutnya yang melakukan analisis sentimen yang menggunakan lebih banyak perbandingan algoritma adalah studi banding dari seleksi fitur dan teknik pembelajaran mesin untuk analisis sentimen. Hasil penelitian menunjukkan bahwa Rasio Gain melakukan yang terbaik antara metode seleksi fitur sentimen dan SVM menunjukkan kinerja terbaik untuk klasifikasi sentimen tugas, sementara Naïve Bayes classifier memberikan hasil yang lebih baik bila digunakan dengan fitur yang lebih sedikit. Masing-masing penelitian memiliki kelemahan tidak terkecuali penelitian tentang menganalisis sentimen pengguna jejaring social menggunakan svm. Pada penelitian ini menghasilkan tingkat akurasi sebesar 81%. Akurasi sistem salah satunya dipengaruhi oleh jumlah Data Latih yang digunakan dalam proses pelatihan. menjelaskan bahwa svm memiliki kelemahan apabila tweet memiliki makna ganda atau ambigu.

Penelitian yang paling mendekati dengan kasus yang sedang peneliti lakukan sekarang ini adalah penelitian mengenai analisis sentimen melalui media twitter menggunakan svm dengan studi kasus kpk versus kapolri. Penelitian ini menggunakan opini public melalui tweet yang dikirim melalui akun twitter tentang kpk dan kapolri yang keduanya sama-sama memiliki daya tarik untuk dibicarakan. Hasil penelitian ini menunjukkan bahwa lebih banyak jumlah sentimen negatif dari pada positif untuk keduanya. Diantara keduanya, kpk menjadi yang paling banyak dibicarakan di media sosial twitter.

METODOLOGI PENELITIAN

2.1 Dataset

Pengumpulan data dalam penelitian ini berasal dari data twitter dengan mendaftarkan akun untuk dapat mengakses API twitter. Setelah mendaftarkan akun maka peneliti akan mendapatkan consumer key, consumer secret, acces token, dan acces token secret. Hal ini nantinya akan digunakan untuk memanggil library twitter sehingga peneliti bisa mendapatkan data sesuai kata kunci yang dimasukkan yaitu pilkada serentak. Pemanggilan library twitter menggunakan bahasa pemrograman R. Dalam hal ini peneliti melakukan satu kali crawling data dengan mendapatkan hasil 3000 data.

Tabel 1. Contoh Data tweet mentah

text	created	id	Status Source	Screen Name
Karangasem Jadi "Zona Merah" Pilkada Serentak - https://t.co/sVZNoBIU6A	16/12/2015 5:41	677000492 988424000	Google	Balitoday
Aher Berharap Pemenang Pilkada Serentak Bisa Segera Dilantik Bandung https://t.co/XFodkkISFM	16/12/2015 5:30	676997885 460930000	Hootsuite	sapa_jabar

2.2 Pre-Processing

Didalam *pre-processing* terdapat subproses yaitu *case normalization*, deteksi *stop word*, tokenisasi, *part-of-speech (pos) tagging*, *parsing*, *stemming*, dan *lemmatization*. Dalam penelitian ini, hanya menerapkan beberapa *pre-processing* keterangan sebagai berikut:

- Tokenisasi adalah proses untuk memotong teks ke karakter/ symbol, kata, frasa.
- *Case normalization* adalah merubah kalimat menjadi huruf kecil semua.
- Deteksi *stopword* adalah proses untuk memfilter kata yang sering digunakan seperti mereka, kenapa, jika, seperti, untuk, ke, dan.
- *Stemming* adalah proses mengurangi kata-kata untuk bentuk katadasar. Contohnya memakan, dimakan menjadi makan.

Tabel 2 merupakan contoh *pre-processing* yang digunakan dalam penelitian ini. Hasil dari *pre-processing* akan disebut atribut.

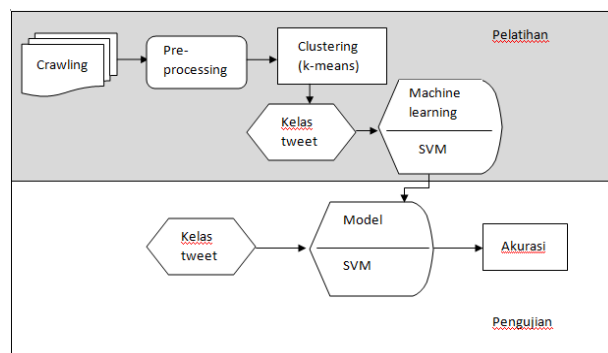
Tabel 2. contoh *pre-processing*

Tweet Mentah	Pre-pocessing
"Karangasem Jadi ""Zona Merah"" Pilkada Serentak - https://t.co/sVZNoBIU6A https://t.co/SdOZuHe4ax #karangasem #bali"	karangasem jadi zona merah pilkada serentak
Pilkada Serentak Karangasem - ANTARA	pilkada serentak karangasem

https://t.co/4Tzcyo4hND #karangasem #bali	
Pilkada Serentak Karangasem - ANTARA https://t.co/4Tzcyo4hND #karangasem #bali	pangdam udayana apresiasi pilkada serentak bali langsung aman
Pangdam Udayana Apresiasi Pilkada Serentak di Bali Berlangsung aman https://t.co/1jrRb03tSw #karangasem #bali	pilkada serentak ntb kondusif tingkat partisipasi jadi catat mataram global fm lombok situasi aman

2.3 Machine-Learning

Terdapat 2 teknik *machine learning*, yaitu pembelajaran terbimbing (*supervised learning*) dan pembelajaran tidak terbimbing (*unsupervised learning*). Perbedaan antara 2 teknik ini adalah ada atau tidaknya label pada data *training*. *Supervised learning* terdapat label tiap data yang akan digunakan pada penelitian ini, sedangkan *unsupervised learning* tidak. Label kelas yang digunakan tiap data meliputi 2 kategori yaitu positif dan negatif.



Gambar 1. Proses utama klasifikasi tweet

Pada penelitian ini menggunakan metode clustering yaitu k-means untuk pelabelan dan metode klasifikasi yaitu SVM untuk mendapatkan akurasi tertinggi.

2.4 K-means

Data tweet mentah yang telah melewati tahap preprocessing selanjutnya diberi label dengan menggunakan k-means. Langkah-langkah menghitung k-means :

1. Tentukan centroid awal
2. Menghitung jarak terdekat dari data ke titik

centroid dengan rumus

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2} \quad (1)$$

3. Pembaruan titik centroid menggunakan rumus:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

2.5 SVM

SVM adalah metode machinelearning yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space [HYPERLINK "file:///H:\tugas%2021%20juni\Coris\jurnal%20atika.docx" \l "Bas13" 11]. Secara sederhana, SVM berusaha mencari *hyperplane* terbaik untuk pemisah dua buah kelas. Metode SVM pada software Weka menggunakan *Sequential Minimal Optimization(SMO)*[12]}. SMO adalah sebuah algoritma optimasi yang digunakan untuk melatih SVM pada kumpulan data, sehingga terciptalah model dengan pembobotan. Proses SVM pada software Weka. Software yang dikembangkan pertama kali di Selandia Baru oleh Universitas Waikato. Weka merupakan *software* data mining yang berbasis Java [HYPERLINK "file:///H:\tugas%2021%20juni\Coris\jurnal%20atika.docx" \l "Wit02" 13]. Perhitungan manual svm menggunakan langkah - langkah berikut :

1. Tentukan vector di \mathbb{R}^2 . Definisi setiap titik $x = (x_1, x_2), x \neq 0$ (3)

2. Definisi arah vector $w(w_1, w_2)$ adalah vektor

$$w \left(\frac{w_1}{\|w\|}, \frac{w_2}{\|w\|} \right) \quad (4)$$

3. Menghitung dot produk dengan rumus $x \cdot y = x_1y_1 + x_2y_2 = \sum_{i=1}^2 (x_iy_i)$ (5)

4. Mencari hyperplane dengan rumus $\mathbb{W} = \sum_i \alpha_i s_i$ selanjutnya dihitung dengan rumus $hyperplane = w^T x + b = 0$ (6)

5. $p = (w \cdot a)w$ dengan $\|p\| = \sqrt{(w \cdot a)w}$ dan rumus akhir menghitung margin adalah $2\|p\|$ (7)

2.6 Evaluasi

Confusion Matrix digunakan untuk mengetahui efektifitas pemodelan klasifikasi, terdiri dari baris dan kolom yang membentuk sebuah tabel berisi label dari data uji yang diprediksi.

Confusion matrix yang terdiri dari *true positives*(TP), dan *true negatives*(TN) adalah klasifikasi yang benar. *False positives*(FP) merupakan suatu kesalahan dimana data yang aslinya berlabel no namun dideteksi label yes, dan *false negatives* (FN) merupakan suatu kesalahan dimana data yang aslinya berlabel yes namun dideteksi label no. Perhitungan evaluasi dengan menggunakan *tabel confusion matrix* adalah sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (11)$$

Akurasi dihitung secara global, sedangkan presisi, *recall*, dan *f-measure* dihitung per kelas tweet. Dengan demikian nilai TP, TN, FP, FN berubah tiap kelas tweet.

HASIL DAN PEMBAHASAN

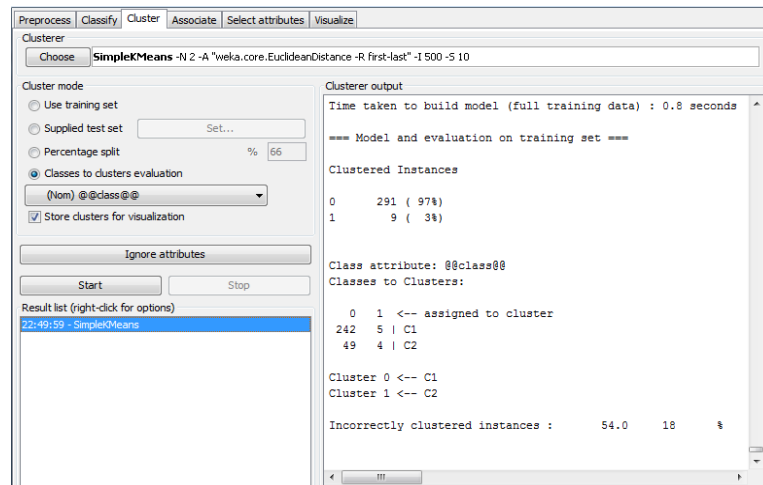
Proses clustering k-means yang dilakukan menggunakan program java menghasilkan pengelompokan sebagai berikut :

Tabel 3. Hasil Pengelompokan Data K-means

Kelas	Hasil
C1 (Positif)	2482
C2 (Negatif)	518

Jumlah	3000
---------------	-------------

Karena metode clustering menggunakan k-means digunakan sebagai dasar pelabelan untuk klasifikasi menggunakan svm maka harus diuji dengan melakukan evaluasi.



Gambar 2. Hasil Pengujian K-means

Gambar 4.1 menunjukkan bahwa persentase kesalahan saat melakukan clustering menggunakan k-means untuk pelabelan adalah 18% dari 100%. Jadi tingkat keakuratan data yang dilabeli adalah 82%.

Proses SVM menghasilkan akurasi global sebesar 98% yang didapat dari *confusion matrix* pada tabel 4.2 untuk hasil presisi, *recall*, dan *f-measure* tiap kelas tweet dapat dilihat pada tabel 4.3

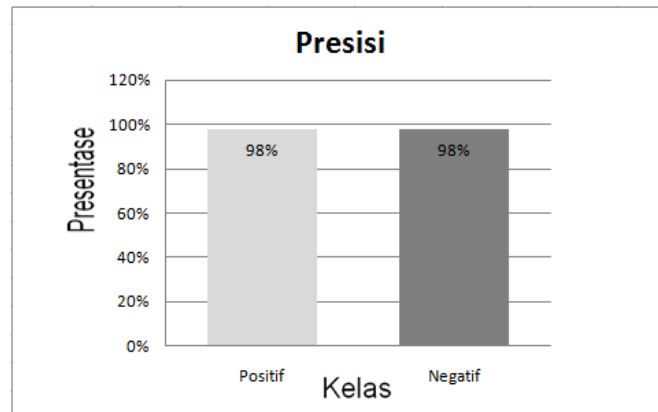
Tabel 4. *Confussion Matrix SVM*

A	B	Label
246	1	A = C1(positif)
5	48	B = C2(negatif)

Tabel 5. Hasil presisi, *recall*, *f-measure* pada SVM

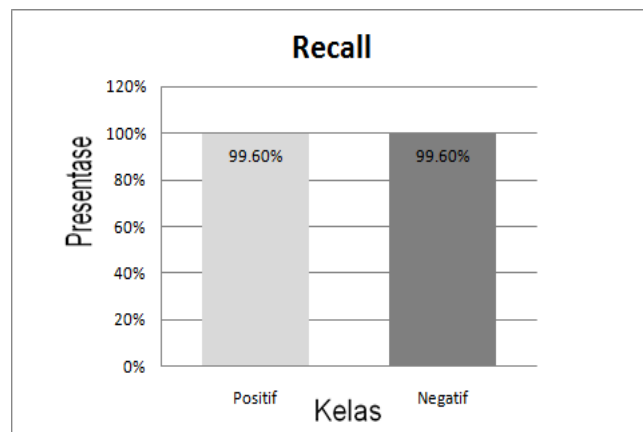
Kelas tweet	Presisi (%)	Recall (%)	F-Measure (%)
Positif	98	99.6	98.8
Negatif	98	90.6	94.1

Pada gambar 3, 4, dan 5 tergambar grafik presisi, *recall*, dan *f-measure* pada tiap kelas dengan SVM. Adanya grafik agar kita dapat lebih mudah mengetahui perbedaan presentase tiap kelas.

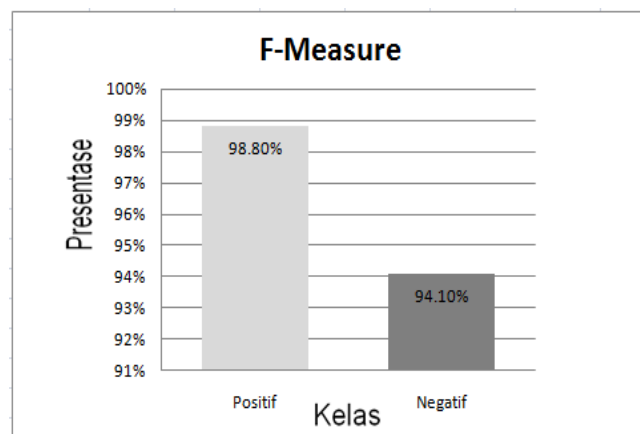


Gambar 3. Grafik Presisi SVM

Hasil presisi uji data antara kelas positif dan negative sama sama tinggi yaitu 98%. Ini terjadi karena semua data uji berlabel positif dan negatif terdeteksi hampir tepat mendekati 100%.



Gambar 0. Grafik Recall SVM



Gambar 5. Grafik F-measure SVM

Sama halnya dengan presisi, *recall* dan *f-measure* kelas positif lebih tinggi dari pada kelas negatif. Hal ini bisa terjadi karena jumlah data pada kelas negatif lebih sedikit dibanding kelas positif.

KESIMPULAN

Berdasarkan penelitian yang dilakukan, data berasal dari twitter dengan kata kunci pilkada serentak berjumlah 3000 data. 2700 data pelatihan dan 300 data pengujian. Proses diawali dengan *pre-processing* yang terdiri dari tokenisasi, *casenormalization*, deteksi *stop word*, dan *stemming*. Proses selanjutnya dilakukan proses pemberian label dengan metode clustering yaitu membagi data kedalam dua kelompok. Cluster 1 sebagai kelompok data berlabel positif dan Cluster 2 sebagai kelompok data berlabel negatif. Klasifikasi dilakukan dengan metode SVM menggunakan weka.

Penulis menyimpulkan bahwa klasifikasi tweet tentang pilkada serentak bahasa Indonesia memiliki tingkat akurasi yang tinggi dengan hasil 98% dari pengujian label menggunakan metode clustering dengan algoritma k-means yaitu 82% .

SARAN

Data yang digunakan dalam penelitian ini diambil dengan proses crawling satu kali saja. Peneliti selanjutnya diharapkan bisa mengambil data secara realtime. Penelitian selanjutnya juga kedepannya dapat menggunakan metode lain untuk pelabelan data agar tidak melakukan pelabelan secara manual. Proses preprocessing dalam penelitian ini masih belum sempurna dalam penyaringan kata. Peneliti lain bisa meningkatkan hasil preprocessing karena proses ini sangat menentukan hasil proses klasifikasi. Penelitian ini hanya menggunakan dua kelas, selanjutnya peneliti lain dapat menggunakan multiclass. Peningkatan akurasi pada SVM dapat dikombinasi dengan Partical Swarm Optimization (PSO).

DAFTAR PUSTAKA

- [1] <http://www.kpu.go.id/index.php/post/read/2015/3829/Arief-Tujuan-Pilkada-Serentak-Untuk-Terciptanya-Efektivitas-dan-Efisiensi-Anggaran/berita>, "Arief: Tujuan Pilkada Serentak Untuk Terciptanya Efektivitas dan Efisiensi Anggaran," 2015.
- [2] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, "What is Twitter, a Social Network or a News Media?," 2010.
- [3] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He, "TwitterRank: Finding Topic-sensitive Influential Twitterers," 2010.
- [4] Seyed-Ali Bahrainian and Andreas Dengel, "Sentiment Analysis using Sentiment Features," 2013.
- [5] Umesh Rao Hodeghatta, "Sentiment Analysis of Hollywood Movies on Twitter," 2013.
- [6] André L. F. Alves, Cláudio de S. Baptista, A. Anderson Firmino, Maxwell G. de Oliveira, and Anselmo C. de Paiva, "A Comparison of SVM Versus Naive-Bayes Techniques for," 2014.
- [7] Andi Nurul Hidayat, "ANALISIS SENTIMEN TERHADAP WACANA POLITIK PADA," *JURNAL ELEKTRONIK SISTIM INFORMASI DAN KOMPUTER (JESIK)*, vol. 1, p. 1, Januari-Juni 2015.
- [8] Anuj Sharma and Shubhamoy Dey, "A Comparative Study of Feature Selection and Machine Learning Techniques for Sentiment Analysis," 2012.
- [9] M. Fachrurrozi M.T. and Novi Yusliani M.T., "ANALISIS SENTIMEN PENGGUNA JEJARING SOSIAL MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," 2015.
- [10] Fatimah Ilona Asa Sabsono and Endang Ripmiatin, "Sentimen Analisis di Media Sosial Twitter Menggunakan Support Vectore Machine (SVM)," 2015.
- [11] Abd. Samad Hasan Basari, Burairah Hussin, Gede Pramudya Ananta, and Junta Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," Desember 2013.

- [12] WEKA. Calcuration by SMO weights. [Online].
<http://weka.8497.n7.nabble.com/Calcuration-by-SMO-weights-td33491.html>
- [13] H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," vol. 31, No. 1, March 2002.