

Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter

Ghulam Asrofi Buntoro

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Ponorogo

Email: ghulamasrofibuntoro@gmail.com

Abstract. *Jakarta Governor Election 2017 discussed in society or internet, especially Twitter. Everyone is free opine on Jakarta governor candidate 2017 so many opinions, not only positive or neutral opinion but also negative. Social media, especially Twitter now become promotions or campaigns are effective and efficient. This research is expected be useful to conduct on public opinion containing sentiment positive, neutral or negative. The method used in this study, for data preprocessing using tokenisasi, cleansing and filtering, to define class sentiment with methods Lexicon Based. For classification using Naive Bayes classifier (NBC) and Support Vector Machine (SVM). The data is 300 tweet in Indonesian by keyword AHY, Ahok, Anies. The results of research is analysis sentiment Jakarta governor candidate 2017. The highest accuracy when using the method of classification Naïve Bayes Classifier (NBC), with average 95% accuracy, 95% precision, 95% recall, TP rate 96,8% and TN rate 84,6%.*

Keywords: *analisis sentimen, jakarta governor candidate 2017, lexicon based, naïve bayes classifier, support vector machine*

Abstrak. *Pemilihan Gubernur DKI Jakarta 2017 ramai diperbincangkan di dunia nyata maupun dunia maya, khususnya di media sosial Twitter. Semua orang bebas berpendapat atau beropini tentang calon Gubernur DKI Jakarta 2017 sehingga memunculkan banyak opini, tidak hanya opini yang positif atau netral tapi juga yang negatif. Media sosial khususnya Twitter sekarang ini menjadi salah satu tempat promosi atau kampanye yang efektif dan efisien. Penelitian ini diharapkan dapat bermanfaat membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif, netral atau negatif. Metode yang digunakan dalam penelitian ini, untuk preprocessing data menggunakan tokenisasi, cleansing dan filtering, untuk menentukan class sentimen dengan metode Lexicon Based. Untuk proses klasifikasinya menggunakan metode Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). Data yang digunakan adalah tweet dalam bahasa Indonesia dengan kata kunci AHY, Ahok, Anies, dengan jumlah dataset sebanyak 300 tweet. Hasil dari penelitian ini adalah analisis sentimen terhadap calon gubernur DKI Jakarta 2017. Akurasi tertinggi didapat saat menggunakan metode klasifikasi Naïve Bayes Classifier (NBC), dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6%.*

Kata Kunci: *analisis sentimen, calon gubernur dki jakarta 2016, lexicon based, naïve bayes classifier, support vector machine*

1. Pendahuluan

Pemilihan Gubernur DKI Jakarta yang akan dilangsungkan pada tahun 2017 keramaiannya sudah mulai bisa dirasakan. Jadwal rangkaian Pemilihan Gubernur DKI Jakarta sudah ditetapkan (KPU, 2016). Sejak tahapan pendaftaran dan ditetapkan calon Gubernur DKI Jakarta 2017 nama-nama calonnya mulai banyak diperbincang, baik di dunia nyata maupun dunia maya. Semua orang bebas berpendapat atau beropini tentang calon Gubernur DKI Jakarta 2017

sehingga memunculkan banyak opini, tidak hanya opini yang positif atau netral tapi juga yang negatif.

Perkembangan media yang begitu pesat, memunculkan banyak media online dari media berita sampai media sosial. Media sosial saja sudah begitu banyak, dari Facebook, Twitter, Path, Instagram, Google+, Tumblr, LinkedIn dan sebagainya masih banyak lagi (Evadollzz, 2014). Media sosial sekarang ini tidak hanya digunakan sebagai sarana pertemanan, mencari teman, tapi sudah banyak digunakan untuk kegiatan lain. Promo dagangan, jual beli apa saja sampai promo partai politik atau kampanye calon legislatif dan presiden,

Media sosial khususnya Twitter sekarang ini menjadi perangkat komunikasi yang sangat populer di kalangan pengguna internet. Pada konferensi resmi pengembang Twitter Chirp 2010, perusahaan tersebut menyampaikan statistik mengenai situs dan pengguna Twitter. Statistik tersebut menyebutkan bahwa pada bulan April 2010, Twitter memiliki 106 juta akun dan sebanyak 180 juta pengunjung unik setiap bulannya. Jumlah pengguna Twitter disebutkan terus meningkat 300.000 user setiap harinya (Yarrow, 2010). Digital Buzz blog suatu situs yang menyediakan statistik infografik menyebutkan data statistik yang sama.

Tim sukses sebuah pasangan calon Gubernur atau kepala daerah sekarang ini bisa menghalalkan segala cara dalam mengkampanyekan calonnya. Khususnya pada setiap masa kampanye pemilihan Gubernur, atau kepala daerah ada istilah Black Campaign terutama di media sosial. Karena sekarang ini promosi, kampanye atau pencitraan tidak hanya di dunia nyata tapi juga di dunia maya. Media sosial khususnya Twitter sekarang ini menjadi salah satu tempat promosi atau kampanye yang efektif dan efisien.

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen (Huang, 2009). Pada penelitian ini Analisis sentimen dilakukan untuk melihat sebuah opini seseorang yang ditujukan kepada calon Gubernur DKI Jakarta 2017, opini itu bisa dimasukkan kategori opini positif, netral atau negatif. Besarnya sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017 bisa dijadikan sebuah parameter kemenangan atau kekalahan seorang calon.

2. Tinjauan Pustaka

Penelitian oleh (Mesut, 2012) menggunakan machine learning untuk mengklasifikasikan berita politik Turki. Penelitian ini melakukan klasifikasi sentimen terhadap berita politik Turki dan menentukan apakah berita politik Turki tersebut memiliki sentimen positif ataukah negatif. Fitur yang berbeda dari berita politik Turki tersebut diekstrak dengan algoritme machine learning Naïve Bayes Classifier (NBC), Maximum Entropy (ME) dan Support Vector Machine (SVM) untuk menghasilkan model klasifikasi. Penelitian ini memperoleh Akurasi 72,05% untuk Naïve Bayes Classifier (NBC), Akurasi 69,44% Maximum Entropy dan 66,81% untuk SVM pada penggunaan bigram.

(Pak, 2010) Menggunakan emoticon untuk membangun corpus berbahasa Inggris dari Twitter dengan sentimen positif, negatif dan netral. Untuk kelas netral Pak dan Paurobek mengambil data training dari tweet akun media berbahasa Inggris. Metode yang digunakan adalah Naïve Bayes dengan n-gram. Performasi terbaik dihasilkan ketika menggunakan bigram.

Penggunaan Twitter sekarang ini tidak hanya sebatas pada media sosial untuk pertemanan, akan tetapi Twitter juga digunakan sebagai alat promosi dan kampanye (Buntoro, 2016). Pengguna Twitter bebas untuk mengeluarkan pendapat dan opininya, termasuk opini tentang calon Presiden Indonesia 2014. Penelitian ini menampung pendapat dan opini masyarakat tersebut dengan membaginya menjadi lima class attribute, yaitu sangat positif, positif, netral, negative dan sangat negatif. Proses klasifikasi pada penelitian ini menggunakan metode klasifikasi Naïve Bayes Classifier (NBC) dengan preprocessing data menggunakan tokenisasi, cleansing dan filtering. Data yang digunakan adalah tweet dalam bahasa Indonesia tentang calon Presiden Indonesia 2014, dengan jumlah dataset sebanyak 900 tweet yang didistribusikan secara

merata kedalam lima class attribute. Hasil akurasi tertinggi didapatkan saat menggunakan tokenisasi n-gram, stopwords list WEKA dan emoticons, yaitu dengan nilai akurasi 71,9%, nilai presisi 71,6%, nilai recall 71,9% nilai TP rate 66,1% dan nilaiTN rate 65%.

Penelitian menggunakan machine learning untuk mengklasifikasikan movie reviews (Pang, 2002). Penelitian ini melakukan klasifikasi sentimen terhadap review film dan menentukan apakah review film tersebut memiliki sentimen positif ataukah negatif. Fitur yang berbeda dari review tersebut diekstrak dan digunakan algoritme machine learning Naïve Bayes dan Support Vector Machine (SVM) untuk menghasilkan model klasifikasi. Mereka memperoleh akurasi antara 78,7% ketika menggunakan Naïve Bayes pada penggunaan unigram. Akurasi yang diperoleh ketika menggunakan SVM dengan unigram adalah 72,8%.

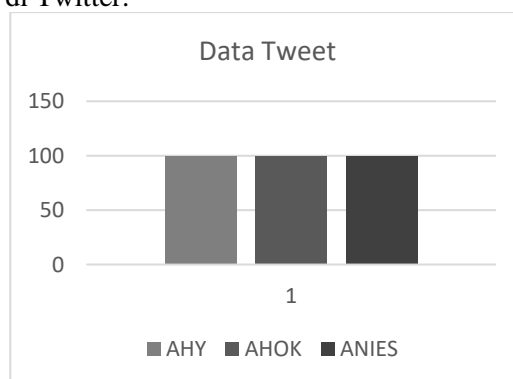
(Frangky, 2008) mencoba untuk mengulangi eksperimen klasifikasi sentiment movie review oleh Pang untuk Bahasa Indonesia. Berkaitan dengan ketidakterdediaannya training corpora untuk Bahasa Indonesia, maka diaplikasikan machine translation tools untuk mentranslasikan corpus Bahasa Inggris yang dibuat Pang yang asli ke Bahasa Indonesia dan hasil translasinya digunakan untuk train klasifikasi. Berbagai pilihan machine translation digunakan mulai dari commercial tool hingga translasi sederhana kata demi kata dan metode klasifikasi teks dicoba. Hasil akurasi rata-rata yang diperoleh untuk metode Naïve Bayes adalah 74,6% dan 75,62% untuk metode SVM. Hasil terbaik diperoleh sama dengan yang diperoleh ketika menggunakan eksperimen dalam Bahasa Inggris.

3. Metode Penelitian

Langkah-langkah penelitian sesuai dengan alur penelitian adalah sebagai berikut:

1. Mengumpulkan data tweet

Data tweet diambil dengan metode Crawling dari media sosial Twitter. Data yang diambil hanya tweet dalam bahasa Indonesia, yaitu 100 tweet dengan kata kunci AHY, 100 tweet dengan kata kunci Ahok, dan 100 tweet dengan kata kunci Anies. Data diambil secara acak baik dari user biasa ataupun media online di Twitter.



Gambar 1. Data tweet

Tabel 1. Contoh tweet

No.	Data Tweet	Sentimen
1.	@Raywailersrasta: Yang tua, yang remaja, yang muda semua dukung AHY #SemuaDukungAgus @AYOJakarta2017	Positif
3.	@rumahkusatu: Jakarta akan dibangun Anies dengan gerakan, bukan program\nhttps://t.co/jrvaMsgBdH #SalamBersama #terpopuler #news @inilah	Positif
4.	@TeukuDiciawi: AHY : Saya Beruntung Dapat Mpok Sylvi... #MpokSylviPelayanMasyarakat https://t.co/A7lF4WQqCS.	Netral
6.	@jitunews: Nasional Unik, Sandiaga Uno Dikira Anies Baswedan Saat Kampanye di Tanjung Priok https://t.co/1zwns6rsI9 #BeritaJitu	Netral

No.	Data Tweet	Sentimen
7.	@edo_macho: @cagubnyinyir @anto_winarno @Yuli52389173 @arifin_henry @marcel_prawira @GhazaliSil mnusia buta yg yobloz ahy bs krj apa klo cuma jnji mnis	Negatif
8.	@benysalim11: Kalau Ahok Menista Agama, Mayatnya Udah Tidak utuh https://t.co/BsOqtGU5NF	Negatif

2. Preprocessing Data

Pada tahap preprocessing, dilakukan 4 langkah sebagai berikut.

1. Seleksi komentar

Pada tahap ini, dilakukan seleksi komentar yang mengandung kata atau hashtag(#) AHY, Ahok atau Anies, karena Twitter terdapat fungsi retweet, yaitu memberikan komentar terhadap tweet komentar seseorang, karena komentar tweet akan mengganggu dalam proses Analisis Sentimen tweet. Jadi dalam preprocessing ini komentar tweet dihapus.

2. Cleansing

Kalimat yang didapat biasanya masih terdapat noise, yaitu kesalahan acak atau varian dalam variable terukur (Ariawan, 2014), untuk itu, kita harus menghilangkan noise tersebut. Kata yang dihilangkan adalah karakter HTML, kata kunci, ikon emosi, hashtag (#), username (@username), url (<http://website.com>), dan email (nama@website.com).

3. Parsing

yaitu proses memecah dokumen menjadi sebuah kata dengan melakukan analisa terhadap kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut (Putro, 2011).

4. Normalisasi Kalimat

Bertujuan untuk menormalkan kalimat sehingga kalimat gaul menjadi normal (Adiyasa, 2013), sehingga bahasa gaul tersebut dapat dikenali sebagai bahasa yang sesuai dengan KBBI. Yang harus dilakukan untuk normalisasi kalimat adalah:

1. Meregangkan tanda baca (punctuation) dan symbol selain alphabet
2. Meregangkan tanda baca adalah memberikan jarak terhadap tanda baca dari kata-kata sesudah atau sebelumnya, tujuannya agar tanda baca dan symbol selain alphabet tidak menjadi satu dengan kata-kata pada saat proses tokenisasi.
3. Mengubah menjadi huruf kecil semua
4. Normalisasi kata

Tabel 2. Aturan normalisasi kata

Tidak Normal / Gaul	Normal
Akhiran -ny	Akhiran -nya
Akhiran -nk	Akhiran -ng
Akhiran -x	Akhiran -nya
Akhiran -z	Akhiran -s
Akhiran -dh	Akhiran -t
Kata berulang: sama2	Kata berulang: sama-sama
Ejaan: oe	Huruf: u
Ejaan: dj	Huruf: j

5. Menghilangkan huruf yang berulang

Ketika sedang senang atau kesal, seseorang bebas menuliskan opini berdasarkan emosinya, biasanya seseorang menuliskan dengan mengulang huruf yang sama. Contohnya: “kereeen” untuk mengekspresikan kesenangan. Kata berulang seperti “kereeen” akan di normalisasi menjadi “keren”.

6. Menghilangkan emoticon

Ketika sedang menulis status (*tweet*) seseorang kadang salah atau kurang tepat dalam penggunaan emoticon, entah disengaja atau tidak banyak yang melakukannya. Contohnya: Mereka hanya bisa memfitnah karena tidak bisa ketemu fakta buruk :), kata opini fitnah tapi

emoticonnya senyum :), dengan begitu emoticon akan mengganggu dalam proses *Sentiment Analysis tweet*, jadi dalam proses ini emoticon dihapus atau diabaikan.

Tabel 3. Emoticon, Feeling and Sentiment

<i>Emoticon</i>	<i>Feeling</i>	<i>Sentiment</i>
:) :-)	Happy	Positive
:(:-(-	Sad	Negative
:D :-D	Very Happy	Positive
D: D=	Very Sad	Negative
*_*_*_*_*_*	Fascinated	Positive
D:< D: D8	Horror, disgust, sadness	Negative
xD XD	Laughing, big, grin	Positive
: = :-	Straight face no expression	Neutral

3. Tokenisasi

Setelah normalisasi kalimat, selanjutnya kalimat tersebut dipecah kedalam token-token menggunakan pembatas atau delimiter spasi. Token yang digunakan dalam penelitian ini yaitu: N-gram: token yang terdiri dari tiga kata setiap satu token, contohnya: Daerah Khusus Ibukota.

4. Part of Speech (POS) Tagger

POS tagger adalah sebuah proses untuk memberikan kelas pada sebuah kata. Dalam proses POS tagger dilakukan dengan cara parsing, kemudian ditentukan kelas tiap kata dengan menggunakan bantuan kamus yang di buat sendiri berdasarkan Kamus Besar Bahasa Indonesia (KBBI) menggunakan metode Maximum Entropy. Proses POS tagging terbagi ke dalam tiga proses yaitu pemisahan setiap token dalam dokumen dengan pengecekan setiap kata dalam dokumen, mengidentifikasi setiap kata dalam dokumen dengan pemberian jenis kata, pengecekan kata yang belum teridentifikasi terhadap bentuk imbuhan dan akhiran sehingga diperoleh kata dasar.

Berdasarkan aturan linguistik pada kata diperoleh sentimen sementara. Penentuan sentimen dilakukan dengan melihat adanya kata yang mengandung opini baik yang memiliki polarity positif maupun negatif dari tweet yang sudah dilabeli kelas katanya. Kelas kata yang dipilih adalah kata sifat (adjective), kata keterangan (adverb), kata benda (noun) dan kata kerja (verb), sesuai dengan penelitian (Azhar, 2013) bahwa keempat jenis kata di atas merupakan jenis kata yang paling banyak mengandung sentimen. Dalam sistem ini jika suatu tweet terdapat kata benda (NN) pada sebelum atau setelah kata sifat (JJ) atau kata keterangan (RB) dan kata benda (memiliki polarity berlawanan dengan kata sifat atau kata keterangan maka polarity yang diperoleh berdasarkan kata sifat atau kata keterangan, karena kata sifat atau kata keterangan memberikan penegasan terhadap kata benda. (Putranti, 2014)

5. Penentuan Class Attribute

Data Twitter yang sudah dilakukan Preprocessing kemudian akan ditentukan class attribute, class attribute yang dimunculkan dalam penelitian ini ada 3, diantaranya positif, netral, dan negatif. Dengan 3 class attribute ini diharapkan mampu memberi penilaian masyarakat secara akurat terhadap objek tertentu.

6. Load Dictionary

Setelah dilakukan tokenisasi dan ditentukan class attribute, langkah selanjutnya adalah load dictionary. Banyak jenis kamus yang dapat digunakan, contohnya: kamus kata kunci sentimen positif (positif keywords), kamus kata kunci sentimen negatif (negatif keywords), kamus kata negasi (negation keywords), dan kamus normalisasi bahasa gaul atau alay.

Berikut adalah contoh kamus dan isinya (Ariawan, 2014):

1. Positif keywords: baik, hebat, jujur, cerdas, keren.
2. Negatif keywords: bohong, korupsi, jahat, jelek.
3. Negation keywords: nggak, tidak, bukan, jauh.

4. Kamus konversi bahasa gaul ke KBBA: sp = siapa, spt = seperti, brp = berapa, hrg = harga, ciyus = serius.

7. Determine Sentiment

Pada proses ini metode Lexicon Based digunakan, yaitu saat menentukan sentimen suatu kalimat opini, penentuan dilakukan dengan penjumlahan n skor polaritas kata opini p yang mengomentari fitur f . Skor polaritas suatu kata opini p akan bernilai 1 jika kata tersebut adalah kata opini positif, dan bernilai -1 jika kata tersebut adalah kata opini negatif (Liu, 2005).

Setelah diketahui kata yang mengandung positif, negatif dan netral di dalam sebuah kalimat, selanjutnya dihitung bobot nilai yang terkandung dalam kalimat tersebut yang dilakukan dengan menjumlahkan nilai kata opini. Jika jumlah nilai opini dalam kalimat tersebut 1, maka nilai sentimen dari kalimat tersebut adalah positif, jika nilai opini dalam kalimat tersebut = 0, maka nilai sentimen dari kalimat tersebut adalah netral, jika nilai opini dalam kalimat tersebut = -1, maka nilai sentimen dari kalimat tersebut adalah negatif.

Tabel 4. Determine Sentiment

Sentimen	Nilai
Positif	1
Netral	0
Negatif	-1

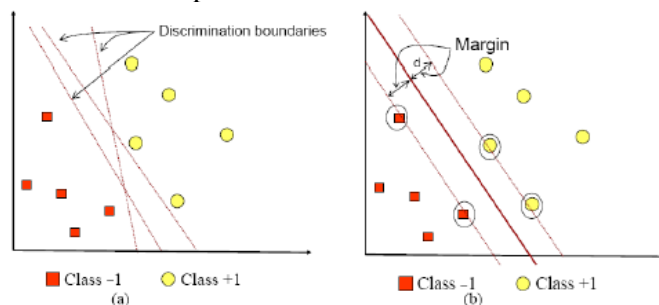
8. Klasifikasi

Masuk pada proses klasifikasi, proses ini dilakukan untuk menguji akurasi metode Lexicon Based dalam menentukan sentimen sebuah tweet opini. Proses klasifikasi menggunakan WEKA 3.8.1. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*.

Naïve Bayes Classifier (NBC) adalah metode *classifier* berdasarkan probabilitas dan *Teorema Bayesian* dengan asumsi bahwa setiap variabel X bersifat bebas (*independence*). Dengan kata lain, *Naïve Bayesian Classifier (NBC)* mengansumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan beradaan atribut (variabel) yang lain. Berikut adalah Rumusnya.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Support Vector Machine (SVM) adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. Algoritma SVM asli diciptakan oleh Vladimir Vapnik dan turunan standar saat ini (margin lunak) diusulkan oleh Corinna Cortes dan Vapnik Vladimir



Gambar 2. Ilustrasi Metode Support Vector Machine (SVM)

Pada proses klasifikasi data diuji dengan menggunakan metode *10 fold cross validation* (Witten, 2013). Jadi dataset akan dibagi menjadi dua, yaitu 10 bagian dengan 9/10 bagian digunakan untuk proses *training* dan 1/10 bagian digunakan untuk proses *testing*. Iterasi

berlangsung 10 kali dengan variasi data *training* dan *testing* menggunakan kombinasi 10 bagian data.

Pengujian	Dataset									
1	■									
2		■								
3			■							
4				■						
5					■					
6						■				
7							■			
8								■		
9									■	
10										■

Gambar 3. Ilustrasi 10 fold cross validation

9. Evaluasi Hasil

Melakukan evaluasi performa Akurasi, Presisi dan *Recall* dari eksperimen yang telah dilakukan. Evaluasi dilakukan dengan menggunakan *Confusion Matrix* yaitu *true positive rate* (*TP rate*), *true negative rate* (*TN rate*), *false positive rate* (*FP rate*) dan *false negative rate* (*FN rate*) sebagai indikator. *TP rate* adalah persentase dari kelas positif yang berhasil diklasifikasi sebagai kelas positif, sedangkan *TN rate* adalah persentase dari kelas negatif yang berhasil diklasifikasi sebagai kelas negatif. *FP rate* adalah kelas negatif yang diklasifikasi sebagai kelas positif. *FN rate* adalah kelas positif yang diklasifikasi sebagai kelas negatif (Kohavi, 1998).

Tabel 5. *Confusion Matrix*

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positif	c	d

4. Uji Coba dan Pembahasan

Dataset pada penelitian ini menggunakan format ARFF yang dikumpulkan dari Twitter dengan metode *Crawling* dari media sosial Twitter. Data yang diambil hanya *tweet* dalam bahasa Indonesia, yaitu *tweet* dengan kata kunci AHY untuk Calon Gubernur DKI Jakarta 2017 Agus Yudhoyono, Ahok untuk Calon Gubernur DKI Jakarta 2017 Basuki Tjahja Purnama dan Anies untuk Calon Gubernur DKI Jakarta 2017 Anies Baswedan. Data diambil secara acak baik dari user biasa ataupun media *online* di *Twitter*.

Dataset yang digunakan sebanyak 300 Tweet, data dibagi secara seimbang (*balanced*) setiap kelasnya, karena dengan data yang tidak seimbang (*imbalanced*), klasifikasi yang dibangun memiliki kecenderungan untuk mengabaikan *minority class* (Kohavi, 1998). Data dibagi menjadi AHY 100 Tweet, Ahok 100 Tweet dan Anies 100 Tweet. Pemberian label dilakukan dengan metode *Lexicon Based* dan bantuan ahli Bahasa Indonesia.

Hasil Analisis Sentimen calon Gubernur DKI Jakarta 2017 menggunakan metode *Lexicon Based* dengan tiga class atribut yaitu positif, netral dan negatif.

Tabel 6. Hasil Analisis Sentimen metode *Lexicon Based*

Sentimen	AHY	Ahok	Anies
Positif	62	33	62
Netral	25	10	26
Negatif	13	57	12

Untuk mengetahui akurasinya, Analisis Sentimen calon Gubernur DKI Jakarta 2017 dengan metode *Lexicon Based* diklasifikasi menggunakan metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) dengan perangkat lunak WEKA versi 3.8.1. WEKA

menggunakan tipe dokumen *Atribut-Relation File Format (ARFF)* sebagai masukan untuk melakukan klasifikasi data.

Hasil dari proses klasifikasi kemudian diuji dengan menggunakan metode *10 fold cross validation*, data dibagi mejadi 10 bagian dengan 9/10 bagian digunakan untuk proses *training* dan 1/10 bagian digunakan untuk proses *testing*. Iterasi berlangsung 10 kali dengan variasi data *training* dan *testing* menggunakan kombinasi 10 bagian data.

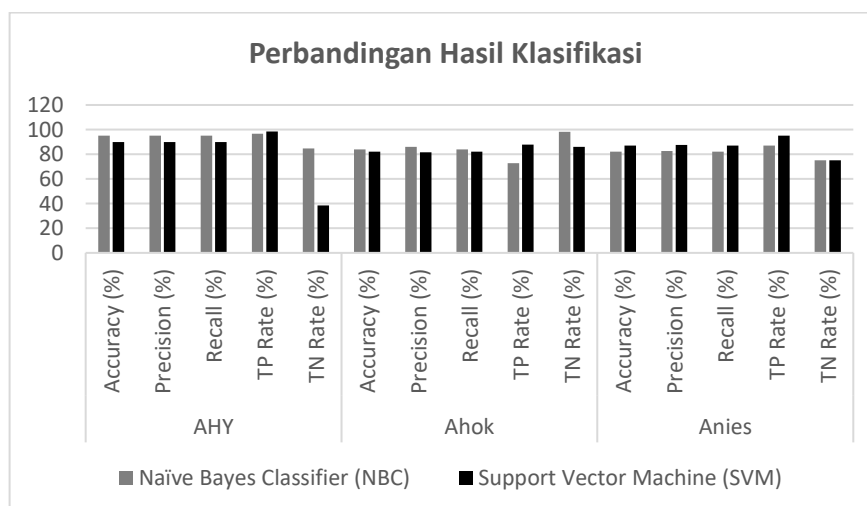
Perbandingan hasil dari metode klasifikasi *Naïve Bayes Classifier (NBC)* dengan *Support Vector Machine (SVM)*.

Tabel 7. Perbandingan Hasil Klasifikasi

<i>Naïve Bayes Classifier (NBC)</i>					
Calon Gubernur	Accuracy (%)	Precision (%)	Recall (%)	TP Rate (%)	TN Rate (%)
AHY	95	95	95	96,8	84,6
Ahok	82	81,6	82	87,9	86
Anies	82	82,7	82	87,1	75
<i>Support Vector Machine (SVM)</i>					
Calon Gubernur	Accuracy (%)	Precision (%)	Recall (%)	TP Rate (%)	TN Rate (%)
AHY	90	89,9	90	98,4	38,5
Ahok	84	86,1	84	72,7	98,2
Anies	87	87,5	87	95,2	75

*) Nilai *Precision* dan *Recall* merupakan nilai rata-rata dari nilai kelas positif dan kelas negatif.

Tabel 6. berisi informasi mengenai nilai akurasi, presisi, *recall*, *TP rate* dan *TN rate* dari masing-masing uji coba yang telah dilakukan. Bagian kolom berisi informasi mengenai calon Gubernur DKI Jakarta 2017. Sedang bagian baris berisi nilai akurasi, presisi, *recall*, *TP rate* dan *TN rate* dari masing-masing uji coba yang telah dilakukan. Dari proses data *preprocessing* menghasilkan sejumlah token yang kemudian digunakan sebagai input sebuah proses klasifikasi. Proses klasifikasi dilakukan menggunakan metode *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*. Dari proses klasifikasi diperoleh nilai akurasi, presisi, *recall*, *TP rate* dan *TN rate* dari masing-masing uji coba.



Gambar 4. Grafik Perbandingan Hasil Klasifikasi

Dari Gambar 4. dapat dilihat bahwa nilai akurasi, presisi, *recall*, *TP rate* dan *TN rate* dicek dengan dua metode klasifikasi yaitu *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*. Nilai akurasi tertinggi didapat metode klasifikasi *Naïve Bayes Classifier (NBC)*.

untuk klasifikasi data AHY, dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai *recall* 95% nilai *TP rate* 96,8% dan nilai *TN rate* 84,6%. Metode klasifikasi *Naïve Bayes Classifier (NBC)* mendapatkan akurasi paling tinggi, metode ini cenderung lebih stabil karena berbasis probabilitas kemunculan kata dalam sebuah kalimat. Akurasi terendah saat metode klasifikasi *Naïve Bayes Classifier (NBC)* digunakan untuk klasifikasi data Ahok dan Anies. Nilai akurasi adalah salah satu parameter penilaian dari metode yang telah digunakan, nilai akurasi didapat dari jumlah banyaknya data yang berhasil diklasifikasi dengan benar sesuai kelas sentimennya dari seluruh jumlah data yang diklasifikasi.

Dari penelitian yang telah dilakukan, diketahui bahwa hasil akurasi tertinggi didapatkan saat menggunakan metode klasifikasi *Naïve Bayes Classifier (NBC)* untuk klasifikasi data AHY, sedangkan untuk nilai akurasi terendah didapat saat metode klasifikasi *Naïve Bayes Classifier (NBC)* digunakan untuk klasifikasi data Ahok dan Anies. Meskipun menghasilkan akurasi yang cukup baik, tapi model yang dibangun masih melakukan sedikit kesalahan pada saat proses klasifikasi data yang pembagian sentimennya tidak seimbang. Karena dengan menggunakan data yang tidak seimbang akan menyebabkan data *minority class* yang salah diklasifikasi sebagai data *majority class* (Kohavi, 1998). Pada akhirnya menjadikan selisih nilai menjadi besar.

5. Kesimpulan

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen Twitter terhadap calon Gubernur DKI Jakarta 2017. Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada twit opini Bahasa Indonesia yang ada di Twitter. Setelah dilakukan analisis sentimen, terlihat berapa banyak sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017. Nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi *Naïve Bayes Classifier (NBC)* untuk klasifikasi data AHY, dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai *recall* 95% nilai *TP rate* 96,8% dan nilai *TN rate* 84,6%. Dalam penelitian ini juga dapat diketahui metode klasifikasi *Naïve Bayes Classifier (NBC)* lebih tinggi akurasinya untuk klasifikasi sentimen Tweet Bahasa Indonesia dibandingkan dengan metode klasifikasi *Support Vector Machine (SVM)*. Untuk penelitian selanjutnya perlu dicoba dikembangkan dengan menggunakan data yang lebih banyak dan Real Time. Perlu dikembangkan juga *stopword list* dan *stemmer* Bahasa Indonesia yang mampu meningkatkan akurasi dalam analisis sentimen Bahasa Indonesia.

Referensi

ARFF files from Text Collections.

<http://weka.wikispaces.com/ARFF+files+from+Text+Collections.html>.

Class *StringToWordVector*.

<http://weka.sourceforge.net/doc.de.v/weka/filters/unsupervised/attribute/StringToWordVector.html>.

Franky dan Manurung, R., (2008). *Machine Learning-based Sentiment Analysis of Automatic Indonesia n Translations of English Movie Reviews*. In Proceedings of the International Conference on Advanced Computational Intelligence and Its Applications.

G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, (2014). "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," in *The 6th Conference on Information Technology and Electrical Engineering (CITEE)*, 2014, pp. 39-43

Go, A., Huang, L., & Bhayani, R. (2009). *Twitter Sentiment Analysis*. Final Project Report, Stanford University, Department of Computer Science.

G. A. Buntoro, (2016). " *Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute*" in *International Journal of Computer Applications* (0975 – 8887), Volume 136 – No.2, February 2016.

Ian H. Witten. (2013) *Data Mining with WEKA*. Department of Computer Science University of Waikato New Zealand.

- J. Ariawan, "Data Preprocessing." [Online]. Available: <https://www.google.com/search?q=apa+itu+noise+dalam+data+mining&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a&channel=sb>. [Accessed: 10-Des-2016].
- Jennifer Yang Hui (2014). *Indonesian Presidential Election: Will Social Media Forecasts Prove Right?*
- Kohavi, & Provost. (1998) *Confusion Matrix*
http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html.
- KPUD DKI Jakarta (2016) Agenda Pemilihan Gubernur DKI Jakarta
<http://kpujakarta.go.id/agenda/>
- Liu, B., Hu, M., dan Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. In Proceedings of the 14th international conference on World Wide Web. ACM, hal. 342-351
- Marian Radke Yarrow, John A. Clausen and Paul R. Robbins (2010). *The Social Meaning of Mental Illness*. Journal of Social Issues. Volume 11, Issue 4, pages 33–48, Fall 1955.
- Mesut Kaya, Guven Fidan, Ismail H. Toroslu (2012). *Sentiment Analysis of Turkish Political News*. IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.
- M. Yusuf Nur Sumarno Putro, (2011). "Analisis Sentimen pada Dokumen Berbahasa Indonesia dengan Pendekatan Support Vector Machine," masters, binus.
- Noviah Dwi Putranti, Edi Winarko (2014). "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan *Maximum Entropy* dan *Support Vector Machine*"
- N. Adiyasa, "Analisis Sentimen Pada Opini Berbahasa Indonesia Menggunakan Pendekatan Lexicon-Based," *Catatan Kecil*, 2011. [Online]. Available: <http://adiyasan.wordpress.com/2013/02/08/sentiment-analysis-menggunakan-pendekatan-lexicon-based/>.
- Olson, David L.; & Delen, Dursun (2008); *Advanced Data Mining Techniques*, Springer, 1st edition (February 1, 2008), page 138, ISBN 3-540-76916-1.
- Pak, A., dan Paurobek, P., (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*, Universite de Paris-Sud, Laboratoire LIMSI-CNRS.
- 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).
- Read, J. (2005). *Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification*. Meeting of the Association for Computational Linguistics - ACL, 43.1-6.
- Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. M.S. thesis. M.Sc. Thesis. Master of Logic Project. Institute for Logic, language and Computation. Universiteti van Amsterdam The Netherlands.
- Top Media Sosial <http://www.evadollzz.com/2014/09/top-10-social-networkings-terpopuler.html> [Accessed: 10-Des-2016].
- Yufis Azhar, Agus Zainal Arifin, Diana Purwitasari. 2013. "Otomatisasi Perbandingan Produk Berdasarkan Bobot Fitur pada Teks Opini".