

Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan *Emoji*

Agnes Rossi Trisna Lestari¹, Rizal Setya Perdana², M. Ali Fauzi³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹agnesrossi.46@gmail.com, ²rizalespe@ub.ac.id, ³moch.ali.fauzi@ub.ac.id

Abstrak

Analisis sentimen merupakan cabang dari *text mining*, fokus utamanya adalah menganalisa dokumen teks. Dokumen teks terkadang mengandung unsur non-tekstual, salah satunya *emoji*. *Emoji* merupakan simbol grafis *Unicode* berupa gambar untuk mengekspresikan perasaan seseorang. Algoritme yang digunakan dalam penelitian ini adalah Naïve Bayes dengan pembaharuan berupa penambahan pembobotan non-tekstual (*emoji*). Hasil dari pembobotan tekstual dan non tekstual yang dinormalisasi dengan metode Min-max digabungkan disertai nilai konstanta tertentu sehingga menghasilkan sentimen positif maupun negatif. Data diambil dari Twitter tentang Pilkada DKI 2017 sebanyak 900 data *tweet*. Dari hasil pengujian akurasi, diperoleh 68,52% untuk kondisi pembobotan tekstual, 75,93% untuk pembobotan non-tesktual, dan 74,81% untuk kondisi penggabungan dengan nilai konstanta 0,5 untuk tekstual dan 0,5 untuk non-tekstual. Dari hasil pengujian pengaruh pembobotan non-tesktual disimpulkan bahwa pembobotan non-tekstual berpengaruh terhadap akurasi dan pengklasifikasian, dengan komposisi konstanta pengali terbaik ketika $\alpha=0,4$ dan $\beta=0,6$ sampai dengan $\alpha=0,1$ dan $\beta=0,9$.

Kata Kunci: analisis sentimen, Twitter, Naïve Bayes, pembobotan *emoji*

Abstract

Sentiment analysis is a part of text mining, the main focus is to analysis text documents. Sometimes text documents contain non-textual elements, e.g. emojis. Emoji is an Unicode graphic Symbol representation using pictures to express a person's feelings. The algorithm used in this research is Naïve Bayes with renewal in addition of non-textual weighting (emoticon). The results of normalised textual and non-textual weightings with Min-Max method will be combined with certain constant values that resulting in both positive and negative sentiments. Data taken from Twitter about 2017 DKI Jakarta elections as much as 900 data tweet. From the accuracy test results, 68,52% were obtained for textual weighting conditions, 74,81% for non-actual weighting, and 73,57% for merging conditions 0,5 for textual and 0,5 for non-textual. From the result of the examination non-textual weighting effect, can be conclude that the non-textual weighting had an effect on the accuracy and classification, with the best multiplier constants when $\alpha = 0,4$ and $\beta = 0,6$ to $\alpha = 0,1$ and $\beta = 0,9$.

Keywords: sentiment analysis, Twitter, Naïve Bayes, emoji weighting

1. PENDAHULUAN

Pada sistem politik demokrasi di Indonesia, media sosial memainkan peranan penting sebagai media kampanye pemilu (pemilihan umum). Hal ini terjadi karena media sosial dapat memudahkan penggunanya mengakses berita politik paling mutakhir serta opini masyarakat yang tidak dimuat dalam surat kabar nasional maupun televisi (Alexa, 2013). Twitter dilirik karena mudah diakses, jumlah pengikut tidak terbatas dan jumlah karakter pesan singkat (*tweet*) hanya 140 karakter sehingga mampu menyampaikan maksud dan tujuan dengan singkat, padat dan jelas. Indonesia menduduki

posisi ketiga sebagai Negara dengan pengguna Twitter terbanyak. Menurut CEO Twitter Dick Costlo, pada pertengahan tahun 2015 mencapai lima puluh juta (CNN Indonesia, 2016).

Tahun 2017 merupakan periode akhir masa jabatan Gubernur DKI Jakarta, sehingga perlu diselenggarakan pemilihan kepala daerah (pilkada) periode 2017-2022. Pada Pilkada ini terdapat tiga pasangan cagub (calon gubernur) beserta cawagub (calon wakil gubernur). Dalam pra-pelaksanaan maupun pelaksanaan Pilkada DKI Jakarta terdapat berbagai opini dan tanggapan dengan sentimen positif dan negatif pada Twitter. Permasalahan yang muncul ketika

menganalisa seluruh hasil sentimen dan pengklasifikasian *tweet* pada Twitter secara manual maka akan dibutuhkan waktu dan usaha yang banyak dalam pelaksanaannya. Pada periode sebelumnya, prediksi hasil pilkada dilakukan oleh lembaga survei dan *quick count* (Kompas, 2014). Hasil survei yang dilakukan oleh lembaga survei terkadang tidak sesuai dengan kenyataan, namun dengan adanya survei yang dilaksanakan pada pra-pemilu membuat cagub dan cawagub mengetahui parameter kemenangan dan kekalahan lebih dini sehingga bisa memperbaiki strategi kampanye. Lain halnya dengan penggunaan *quick count* yang lebih efektif dari lembaga survei karena menghasilkan akurasi yang tinggi, namun *quick count* dilaksanakan setelah pemilu terselenggara sehingga cagub dan cawagub tidak mampu melaksanakan evaluasi dan tindakan lanjut terhadap strategi kampanye yang mempengaruhi hasil pilkada. Dalam membangun *quick count*, juga harus melibatkan banyak pihak secara langsung dan membutuhkan waktu yang lama. Dari permasalahan tersebut, dibutuhkan sistem pengklasifikasian opini berupa teks sentimen pada Twitter untuk memprediksi kekuatan cagub dan cawagub lebih dini dengan cepat dan akurat. Untuk memecahkan permasalahan diajukan solusi berupa sistem Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia dengan harapan dapat dijadikan sebagai perhitungan sekunder setelah lembaga survei dan *quick count*, serta mampu memperbaiki strategi kampanye para cagub dan cawagub.

Algoritme pengklasifikasian yang digunakan adalah Naïve Bayes, karena dari penelitian yang dilakukan oleh Sandi dan Edi (2013) tentang klasifikasi *tweet* kemacetan lalu lintas di Bandung dengan hasil akurasi tertinggi mencapai 93,58% (Rodiyansyah & Winarko, 2013), oleh Ahmad dan Azhari (2014) tentang klasifikasi *tweet* yang berisi sentimen masyarakat pada tokoh publik tertentu didapatkan hasil berupa prosentase akurasi pengujian klasifikasi fitur *term frequency* mencapai 79,91% sedangkan fitur TF-IDF mencapai 79,68% dengan metode Naïve Bayes (Hidayatullah & Azhari, 2014) terbukti bahwa metode Naïve Bayes memiliki performansi yang cukup baik untuk melakukan klasifikasi *tweet* sehingga memiliki tingkat keberhasilan yang tinggi. Sebagai upaya pembaharuan ditambahkan proses pembobotan pada non-tekstual (*emoji*) karena pada penelitian yang

dilakukan Chin (2016) *emoji* sangat populer digunakan oleh masyarakat untuk mengungkapkan gagasan atau ide yang tidak dapat dituliskan dengan kata-kata (Chin, Zappone, & Zhao, 2016). Pada penelitian Sreeja (2016) terdapat opini sindiran, dimana sentimen yang terkandung dalam non-tekstual lebih mendominasi dari sentimen yang terkandung pada tekstual (kalimat) sehingga menyebabkan keambiguan opini. (Jayanthi & Sreeja, 2016). Dengan adanya pembobotan emoticon, diharapkan hasil sentimen menjadi benar – benar memiliki batas yang jelas antara sentimen positif dan negatif (Dayalani & Patil, 2014) sehingga akurasi yang dihasilkan oleh sistem ini mengalami peningkatan.

2. METODE USULAN

Alur proses pada penelitian ini diawali dengan memasukkan dokumen latih dan uji, dokumen diolah melalui pembobotan non-tekstual (*emoji*) maupun pembobotan tekstual (teks). Proses awal pada pembobotan non-tekstual ialah mendeteksi adanya *emoji* berdasarkan kamus *emoji* yang telah tersimpan dalam *file csv*, jika ditemukan maka langkah selanjutnya ialah memboboti *emoji* dengan menjumlahkan *emoji* yang ada dalam tiap kategori, kemudian dilakukan normalisasi *min-max*. Pada pembobotan tekstual proses yang dilakukan ialah praproses teks, kemudian dilakukan perhitungan dengan Naïve Bayes dan dilanjutkan dengan normalisasi tekstual sehingga menghasilkan nilai probabilitas positif maupun negatif. Hasil dari perhitungan pembobotan tekstual (Naïve Bayes) dan pembobotan non-tekstual (*emoji*) digabungkan disertai nilai konstanta pengali tertentu sehingga mampu mengkasifikasikan dokumen tersebut dalam kategori sentimen positif maupun negatif. Alur proses sistem ditunjukkan dalam Gambar 1.

2.1. Naïve Bayes Classifier

Naïve Bayes Classifier adalah konsep probabilitas penentuan kelompok kelas dokumen. Algoritme klasifikasi ini dapat mengolah data dalam jumlah besar dengan hasil akurasi yang tinggi. Persamaan umum metode Naïve Bayes, ditunjukkan dalam Persamaan 1 (C. Aggarwal, 2015).

$$P(C_j|W_i) = \frac{P(C_j) \times P(W_i|C_j)}{P(W_i)} \quad (1)$$

Keterangan:

$P(c_j|w_i)$: *Posterior*, adalah kemunculan peluang pada kategori j tertentu ketika terdapat kemunculan kata i

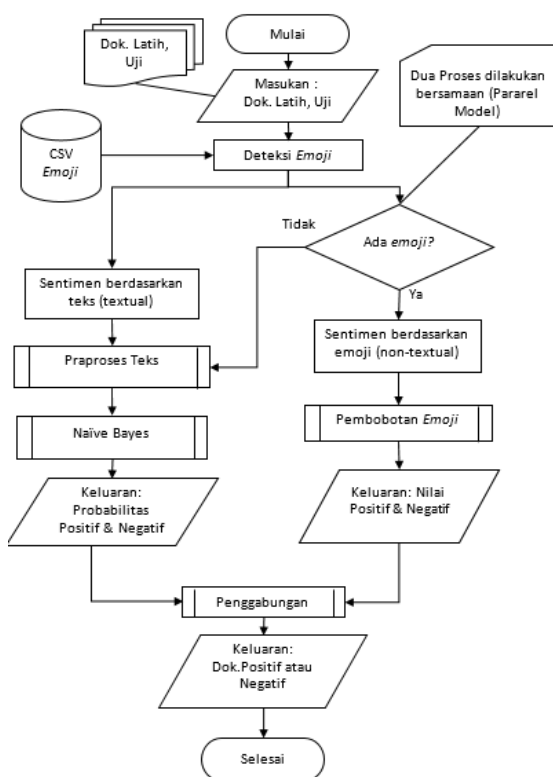
$P(c_j)$: *Prior*, adalah peluang kemunculan dokumen pada kategori j

$P(w_i|c_j)$: *Likelihood* atau *Conditional Probability*, adalah peluang sebuah kata i masuk ke dalam kategori j

$P(w_i)$: *Evidence*, adalah peluang kemunculan sebuah kata

i : indeks kata yang berawal dari 1 sampai dengan kata ke- k

j : indeks kategori yang berawal dari 1 sampai dengan kategori ke- n



Gambar 1. Diagram Alur Proses Sistem

Dalam Persamaan 2 merupakan penyederhanaan perhitungan posterior pada proses klasifikasi dimana peluang kemunculan kata/*evidence* dihilangkan, karena tidak akan berpengaruh pada perbandingan hasil klasifikasi pada tiap kategori (Manning, Raghavan, & Schutze, 2009).

$$P(c_j|w_i) = P(c_j) \times P(w_i|c_j) \quad (2)$$

Persamaan untuk menghitung peluang kemunculan dokumen ada kategori tertentu (*Prior*), ditunjukkan dalam Persamaan 3 (Manning, Raghavan, & Schutze, 2009).

$$P(c_j) = \frac{NC_j}{N} \quad (3)$$

Keterangan:

NC_j : Jumlah dokumen latih yang masuk ke dalam kategori c_j

N : Jumlah seluruh dokumen latih yang digunakan.

Perhitungan peluang kemunculan kata i masuk ke dalam kategori j (*likelihood/conditional probability*) dilakukan dengan menggunakan Multinomial Naïve Bayes. Dalam kasus ini *conditional probability* yang berawal dari $i=1$ sampai dengan $i=k$ sehingga perhitungan nilai *Posterior* dapat dilakukan dalam Persamaan 4.

$$P(c_j|w_i) = P(c_j) \times P(w_1|c_j) \times \dots \times P(w_k|c_j) \quad (4)$$

2.1.1. Multinomial Naïve Bayes

Multinomial Naïve Bayes diterapkan tanpa memperhitungkan urutan kata dan informasi yang ada dalam kalimat atau dokumen secara umum. Dalam melakukan perhitungan peluang sebuah kata i masuk ke dalam kategori j dapat dilakukan dengan menggunakan Persamaan 5, dimana persamaan tersebut merupakan jumlah dari hasil perhitungan *likelihood/conditional probability* data uji yang ditambahkan dengan angka satu untuk menghindari nilai nol (Manning, Raghavan, & Schutze, 2009), kemudian dibagi dengan jumlah seluruh hasil perhitungan *likelihood/conditional probability* seluruh data yang ada pada kategori tertentu dan ditambah dengan jumlah seluruh kata unik yang terdapat pada seluruh kategori.

$$P(w_i|c_j) = \frac{\text{count}(w_i, c_j) + 1}{(\sum_{w \in V} \text{count}(w, c_j)) + |V|} \quad (5)$$

Keterangan :

$P(w_i|c_j)$: *Likelihood* atau *conditional probability*

$\text{count}(w_i, c_j) + 1$: Jumlah kemunculan dari kata uji i yang muncul dalam kategori c_j ditambah dengan 1 untuk menghindari nilai nol.

$\sum_{w \in V} \text{count}(w, c_j)$: Jumlah kemunculan seluruh kata yang ada pada kategori c_j .

$|V|$: Jumlah seluruh kata unik yang ada pada seluruh kategori.

2.2. Pembobotan Emoji

Emoji simbol grafis *Unicode* berupa gambar untuk mengekspresikan sentimen tertentu pada dokumen, oleh karena itu pembobotan *emoji* diperlukan supaya sentimen positif ataupun negatif dapat dilihat secara jelas hasilnya. Pembobotan emoji (non-tekstual) dilakukan bersamaan dengan pengklasifikasian tekstual. Data berupa teks (tekstual) tersebut akan

diidentifikasi terlebih dahulu, apakah teks mengandung emoji atau tidak berdasarkan kamus emoji pada tracker emoji Twitter yang tersimpan dalam database (Chin, Zappone, & Zhao, 2016) selanjutnya dilakukan pembobotan pada emoji (non-tekstual) dengan cara menjumlahkan seluruh emoji yang terkandung dalam sentimen, baik positif ataupun negatif.

2.3. Normalisasi *min-max*

Normalisasi merupakan bentuk dari transformasi (Han, Kamber, & Pei, 2012), dimana variabel berupa angka diposisikan dalam skala dengan jangkauan yang lebih kecil (Junaedi, Budianto, Maryati, & Melani, 2011). Salah satu metode pada normalisasi adalah *min-max* yang merupakan bentuk transformasi linear terhadap data asli (Han, Kamber, & Pei, 2012). Dengan adanya normalisasi ini nilai perbandingan data sebelum ataupun sesudah diproses menjadi seimbang atau terstandarisasi. Rumus normalisasi *min-max* ditunjukkan dalam Persamaan 6.

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (new_{\max_A} - new_{\min_A}) + new_{\min_A} \quad (6)$$

Keterangan:

v'_i : Nilai baru data hasil dari normalisasi *min-max*

v_i : Nilai data yang akan di normalisasikan

\min_A : Nilai minimum dari data

\max_A : Nilai maksimum dari data

new_{\min_A} : Nilai minimum yang diharapkan dari normalisasi(0,1)

new_{\max_A} : Nilai maksimum yang diharapkan dari normalisasi(0,9)

2.4. Penggabungan Pembobotan Tekstual dan Non-Tekstual

Hasil/keputusan akhir klasifikasi untuk menentukan sentimen positif ataupun negatif pada analisis sentimen yang mengandung fitur pembobotan tekstual dan pembobotan non-tekstual, diperoleh dari penggabungan dua pembobotan tersebut (Perdana & Pinandito, 2017). Rumus persamaan penggabungan pembobotan tekstual dan non-tekstual ditunjukkan dalam Persamaan 7 (Perdana & Pinandito, 2017).

$$PenggabC_j = (\alpha \times TeksC_j) + (\beta \times NonTeksC_j) \quad (7)$$

Keterangan:

α : Konstanta tertentu, pengali pembobotan tekstual

$TeksC_j$: Nilai probabilitas perhitungan posterior yang telah dinormalisasi pada kelas sentimen positif atau negatif

β : Konstanta tertentu, pengali pembobotan non-tekstual

$NonTeksC_j$: Nilai pembobotan *emoji* yang telah dinormalisasi pada kelas sentimen positif atau negatif

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Akurasi Sistem

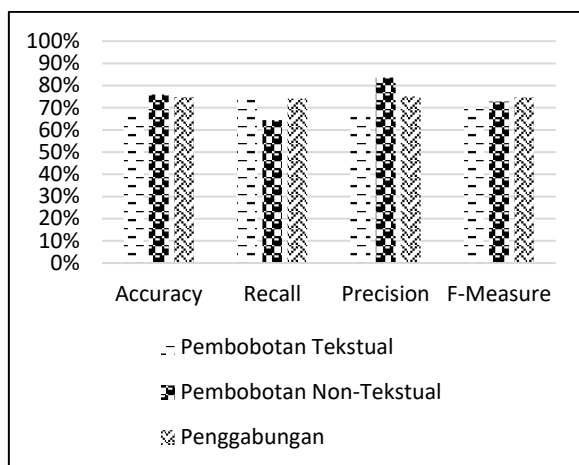
Pengujian akurasi sistem didapat dari perbandingan kelas sentimen yang dilabeli secara manual terhadap kelas sentimen yang dihasilkan oleh pengolahan sistem. Dengan adanya pengujian hasil akurasi ini dapat diketahui tingkat keberhasilan sistem. Karena data yang digunakan hanya memiliki dua kelas (positif atau negatif), maka evaluasi dilakukan dengan menggunakan *confusion matrix* dengan empat parameter.

Jumlah perbandingan data latih dan data uji yang digunakan dalam pengujian ini adalah 70% (630 *tweet*) untuk data latih dan 30% (270 *tweet*) untuk data uji dengan sentimen positif dan negatif yang seimbang. Terdapat 3 hasil *accuracy*, *recall*, *precision* dan *f-measure* yang diuji, yaitu dengan pembobotan tekstual saja, dengan pembobotan non-tekstual saja, dan ketika pembobotan tekstual digabungkan bersama pembobotan non-tekstual dimana perbandingan konstanta pengali α untuk tekstual 0,5 dan konstanta pengali β untuk non-tekstual 0,5. Hasil pengujian akurasi sistem, ditunjukkan dalam Tabel 1.

Tabel 1. Pengujian Akurasi Sistem *Confusion Matrix*

	Tekstual saja	Non-tekstual saja	Penggabungan tekstual ($\alpha=0,5$) dan non-tekstual ($\beta=0,5$)
<i>True Positive</i>	100	87	100
<i>False Positive</i>	50	17	33
<i>False Negative</i>	35	48	35
<i>True Negative</i>	85	118	102
Pengujian Akurasi Sistem			
	Tekstual saja	Non-tekstual saja	Penggabungan tekstual ($\alpha=0,5$) dan non-tekstual ($\beta=0,5$)
<i>Accuracy</i>	68,52%	75,93%	74,81%
<i>Recall</i>	74,07%	64,44%	74,07%
<i>Precision</i>	66,67%	83,65%	75,17%
<i>F-measure</i>	70,18%	72,80%	74,63%

Dari hasil tersebut dapat digambarkan dengan grafik yang ditunjukkan dalam Gambar 2.



Gambar 2. Grafik Pengujian Akurasi Sistem

Dari grafik pengujian akurasi sistem dapat dianalisa bahwa :

1. Pengujian yang dilakukan pada kondisi pembobotan teksstual saja mendapat hasil akurasi paling rendah yaitu 68,52%. Kondisi ini disebabkan karena sebagian kata pada data uji yang dibandingkan dengan data latih memiliki frekuensi kemunculan lebih tinggi pada kelas sentimen positif.
2. Pada kondisi pembobotan non-teksstual saja memperoleh hasil akurasi lebih tinggi dari pembobotan teksstual yaitu 75,93%. Kondisi ini terjadi karena terdapat *tweet* yang tidak memiliki *emoji* positif maupun negatif ataupun *tweet* yang memiliki jumlah *emoji* positif maupun negatif seimbang. Dalam kondisi tersebut dilakukan normalisasi dengan pemberian bobot 0,1 untuk positif dan 0,1 untuk negatif sehingga hasil akhir pengklasifikasian kelas sentimen masuk ke dalam kondisi: jika hasil sentimen positif lebih besar maka masuk ke dalam kelas sentimen positif atau dalam kondisi sebaliknya masuk dalam sentimen negatif, namun jika keduanya tidak terpenuhi maka masuk ke dalam kondisi lainnya, yaitu negatif. Selain itu juga dikarenakan terdapat *tweet* yang berisi sindiran, dimana teksstual masuk dalam sentimen positif sedangkan non-teksstual masuk dalam sentimen negatif atau sebaliknya.
3. Pada kondisi penggabungan antara pembobotan teksstual dan non-teksstual dengan konstanta α dan β yang seimbang, diperoleh hasil akurasi lebih tinggi dari akurasi pembobotan teksstual dan akurasi pembobotan *emoji* yaitu 74,81%. Kondisi ini terjadi karena terdapat 3 keadaan yang mempengaruhi, yaitu :

1. *Tweet* yang diposting berupa sindiran, dimana teksstual masuk dalam kategori sentimen negatif sedangkan non-teksstual masuk dalam kategori sentimen positif atau sebaliknya.
2. Sebagian kata pada data uji yang dibandingkan dengan data latih memiliki frekuensi kemunculan lebih tinggi pada kelas sentimen negatif.
3. Terdapat *tweet* yang tidak memiliki *emoticon* positif maupun negatif ataupun *tweet* yang memiliki jumlah *emoticon* positif maupun negatif seimbang. Dalam kondisi tersebut dilakukan normalisasi dengan pemberian bobot 0,1 untuk positif dan 0,1 untuk negatif sehingga hasil akhir pengklasifikasian kelas sentimen masuk ke dalam kondisi: jika hasil sentimen positif lebih besar maka masuk ke dalam kelas sentimen positif atau dalam kondisi sebaliknya masuk dalam sentimen negatif, namun jika keduanya tidak terpenuhi maka masuk ke dalam kondisi lainnya, yaitu negatif.

3.2. Pengujian Pengaruh Pembobotan *Emoticon*

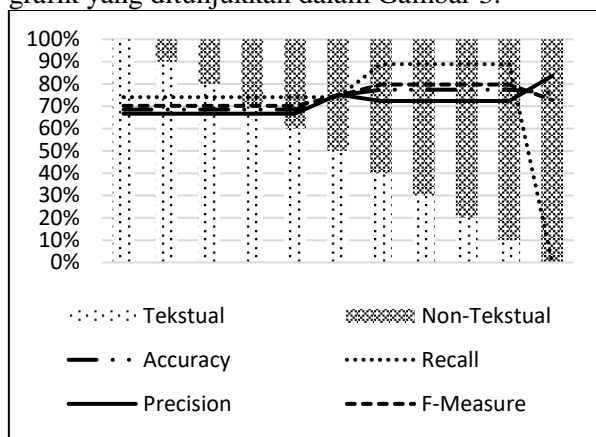
Pengujian ini dilakukan untuk mengetahui apakah pembobotan *emoticon* berpengaruh terhadap pembobotan teksstual dan hasil akurasi sistem, serta mampu mendapatkan komposisi perbandingan nilai pembobotan teksstual dan non-teksstual yang tepat. Jumlah perbandingan data latih dan data uji yang digunakan dalam didasari oleh penelitian Gunawan(2017), yaitu 70% (630 *tweet*) untuk data latih dan 30% (270 *tweet*) untuk data uji dengan sentimen positif dan negatif yang seimbang (Gunawan, Fauzi, & Adikara, 2017). Terdapat 11 hasil *accuracy*, *recall*, *precision* dan *f-measure* dengan perbandingan konstanta pengali α untuk pembobotan teksstual dan β untuk non-teksstual yang berbeda nilainya. Hasil pengujian pengaruh pembobotan *emoticon*, ditunjukkan dalam Tabel 2.

Tabel 2. Pengujian Pengaruh Pembobotan *Emoticon*

Konstanta		Ac	Rec	Pre	F-m
A	B (1- α)				
1	0	68,52%	74,07%	66,67%	70,18%
0,9	0,1	68,52%	74,07%	66,67%	70,18%
0,8	0,2	68,52%	74,07%	66,67%	70,18%
0,7	0,3	68,52%	74,07%	66,67%	70,18%
0,6	0,4	68,52%	74,07%	66,67%	70,18%
0,5	0,5	74,81%	74,07%	75,17%	74,63%
0,4	0,6	77,41%	88,89%	72,29%	79,73%

0,3	0,7	77,41%	88,89%	72,29%	79,73%
0,2	0,8	77,41%	88,89%	72,29%	79,73%
0,1	0,9	77,41%	88,89%	72,29%	79,73%
0	1	75,93%	64,44%	83,65%	72,80%

Dari hasil tersebut dapat digambarkan dengan grafik yang ditunjukkan dalam Gambar 3.



Gambar 3. Grafik Pengujian Pengaruh Pembobotan Emoticon

Dari grafik pengujian pengaruh pembobotan *emoji* dapat dianalisa bahwa:

1. Pembobotan *emoji* memiliki pengaruh terhadap pengklasifikasian dan akurasi sistem, terbukti dengan meningkatnya *accuracy*, *recall*, *precision* dan *f-measure* pada hasil penggabungan pembobotan tekstual dan non-tekstual. Dengan adanya pembobotan *emoji*, sentimen yang terdapat pada pembobotan tekstual semakin terdefinisi secara jelas dan nilai probabilitasnya meningkat sehingga memperjelas dokumen untuk masuk kedalam kategori kelas sentimen tertentu, selain itu juga terdapat beberapa *tweet* sindiran yang terklasifikasi secara kurang tepat akibat pembobotan *emoji* dengan perbandingan konstanta pengali tertentu.
2. Komposisi konstanta pengali yang tepat, yaitu α untuk pembobotan tekstual dan β untuk pembobotan non-tekstual (*emoji*) adalah 0,4 dan 0,6 sampai dengan 0,1 dan 0,9 karena dari hasil yang didapat perbandingan inilah yang memiliki nilai *accuracy*, *recall*, *precision* dan *f-measure* tinggi sehingga dapat disimpulkan bahwa pembobotan non-tekstual sangat mempengaruhi hasil dari klasifikasi sentimen.

4. KESIMPULAN

Berdasarkan hasil pengujian dan analisa maka dapat disimpulkan :

1. Algoritme Naïve Bayes dan pembobotan *emoji* dapat diterapkan pada penelitian ini. Pada data latih dan data uji dilakukan deteksi *emoji* dimana jika ditemukan maka dilakukan pembobotan *emoji* dan normalisasi min-max, kemudian dilanjutkan praproses teks, pembobotan tekstual yang juga harus dinormalisasi menggunakan min-max dan penggabungan pembobotan diantara keduanya sehingga mendapatkan hasil berupa sentimen positif atau negatif.
2. Hasil akurasi sistem pada pembobotan tesktual 68,52%, pada pembobotan non-tesktual 75,93%, dan pada penggabungan kedua pembobotan 74,81% sehingga disimpulkan bahwa penggabungan kedua pembobotan meningkatkan hasil akurasi sistem.
3. Dari pengujian sebanyak 11 kali didapat hasil penambahan akurasi pada perbandingan konstanta pengali tertentu. Perbandingan konstanta α untuk pembobotan tekstual dan β untuk pembobotan non-tekstual adalah 0,4 dan 0,6 sampai dengan 0,1 dan 0,9 sehingga dapat disimpulkan bahwa pembobotan non-tekstual sangat mempengaruhi hasil dari klasifikasi sentimen.

5. DAFTAR PUSTAKA

- Alexa. (2013). *The Top 500 sites on the web*. Diakses tanggal: 21 Februari 2017, Tersedia di: <http://www.alex.com/topsites>
- C. Aggarwal, C. (2015). *Data Classification : Algorithms and Applications*. Berilustrasi peyunt. New York: CRC Press.
- Chin, D., Zappone, A., & Zhao, J. (2016). *Analyzing Twitter Sentiment of the 2016 Presidential Candidates*. Diakses tanggal: 27 Juli 2017, Tersedia di: <https://web.stanford.edu/~jesszhao/files/twitterSentiment.pdf>
- CNN Indonesia. (2016). *Berita Teknologi Informasi : Twitter Rahasiakan Jumlah Pengguna di Indonesia*. Diakses tanggal: 17 Februari 2017, Tersedia di: <http://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia>
- Dayalani, G. G., & Patil, P. B. (2014). Emoticon-based unsupervised

- sentiment classifier for polarity analysis in tweets. *International Journal of Engineering Research and General Science*, II(6), 438-445.
- Gunawan, F., Fauzi, M. A., & Adikara, P. P. (2017). Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JTIK)*, 330-141.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts And Techniques* (3rd ed.). USA: Elsevier.
- Hidayatullah, A. F., & Azhari. (2014). Analisis Sentimen Dan Klasifikasi Katrgori Terhadap Tokoh Publik Pada Twitter. Yogyakarta: UPN Veteran.
- Jayanthi, & Sreeja. (2016, April). Sentiment Analysis of Twitter Data containing Emoticons: A Survey. *International Journal of Innovations in Engineering and Technology (IJJET)*, VI(4), 147-153.
- Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data Transformation Pada Data Mining. Surabaya: IdeaTech2011.
- Kompas. (2014, Juli 11). *Kompas Nasional*. Diakses tanggal: 27 Februari 2017, Tersedia di: <http://nasional.kompas.com/read/2014/07/11/06330421/Bagaimana.Cara.Kerja.Quick.Count>.
- Manning, C. D., Raghavan, P., & Schutze, H. (2009). In *Introduction to Information Retrieval*. New York: Cambridge University Press.
- Perdana, R. S., & Pinandito, A. (2017). Combining Likes-Retweet Analysis and Naive Bayes Classifier Within Twitter for Sentiment Analysis. *International Conference On Communication And Computer Engineering (ICOCOE)*. Penang, Malaysia: Journal of Telecommunication, Electronic and Computer Engineering (JTEC).
- Rodiyansyah, F. S., & Winarko, E. (2013, Januari 1). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayes Classification. *IJCSS*, 7, 13-22.