

Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik

Ismail Sunni¹, Dwi Hendratmo Widyantoro²

*Program Studi Teknik Informatika, Institut Teknologi Bandung
Jl. Ganesha 10 Bandung 40132, Indonesia*

¹ismailsunni@yahoo.co.id

²dwi@stei.itb.ac.id

Abstraksi—Makin maraknya penggunaan jejaring sosial seperti Twitter menjadikan jejaring sosial tersebut sebagai data yang sangat besar. Salah satu pemanfaatan data ini adalah untuk mengetahui opini atau sentimen pengguna jejaring sosial terhadap suatu topik. Salah satu topik yang penting adalah tokoh publik, seperti calon gubernur, ketua partai, dan lain sebagainya yang akan dijadikan sebagai domain dalam makalah ini. Selain kebutuhan mengetahui sentimen tentang tokoh publik, ada juga kebutuhan untuk mengetahui kapan terjadi perubahan sentimen dan topik apa yang menyebabkan perubahan sentimen. Permasalahan utama yang muncul adalah model bahasa pengguna Twitter di Indonesia yang memiliki karakteristik yang berbeda dari pengguna di luar Indonesia. Solusi yang ditawarkan adalah F3 (F3 is Factor Finder) yang memiliki beberapa metode praproses yang diperkirakan mampu menangani permasalahan model bahasa yang ditemukan. F3 menggunakan Naïve Bayes untuk melakukan analisis sentimen karena telah teruji di berbagai penelitian. Sedangkan untuk mengetahui perubahan sentimen, F3 akan menampilkan perubahan sentimen dalam bentuk kurva. Terakhir, untuk mengetahui topik yang menyebabkan perubahan sentimen, F3 menggunakan metode Tf-Idf dengan discounted-cumulative untuk menangani karakter topik yang muncul di Twitter yang berkelanjutan. Hasil analisis dan pengujian menunjukkan tahapan praproses tidak memiliki pengaruh yang signifikan terhadap akurasi (69.4%-72.8%) klasifikasi sentimen. Sedangkan untuk mengekstraksi topik menunjukkan bahwa penggunaan Tf-Idf dengan discounted cumulative mampu meningkatkan jumlah topik terekstrak yang sesuai. Namun, memiliki kelemahan ketika menghadapi topik yang termuat di hampir seluruh selang waktu atau topik yang bukan bersumber dari berita di media internet.

Kata Kunci — Analisis Sentimen, *discounted-cumulative*, ekstraksi topik, Naïve Bayes, Tf-Idf, Twitter

I. PENDAHULUAN

Beberapa tahun terakhir, jejaring sosial telah tumbuh secara luar biasa cepatnya. Pertumbuhan mulai dari pengguna hingga fasilitas yang ditawarkan oleh jejaring sosial tersebut. Indonesia telah menjadi negara dengan pengguna Twitter terbesar di Asia dan ke enam di dunia (Sysomos Inc., 2010) dengan banyak pengguna Twitter di tahun 2010 adalah 2.41% dari seluruh pengguna di dunia. Angka ini naik dari 0.5% pada tahun 2009. Tentu saja, informasi yang terkandung dalam Twitter ini sangat berharga sebagai alat penentu kebijakan. Dan hal ini bisa dilakukan dengan menggunakan analisis sentimen. Namun, analisis sentimen ini mendapat tantangan berupa model bahasa tidak formal yang digunakan di Twitter.

Selain mengekstraksi sentimen, hal yang biasa orang ingin tahu adalah, kapan terjadi perubahan sentimen dan apa yang menyebabkan sentimen tersebut berubah. Hal ini menjadi penting, karena dengan mengetahui apa yang menyebabkan sentimen berubah, pihak yang bersangkutan bisa mengambil keputusan dengan lebih baik. Misal saja ketika mengetahui suatu topik atau kejadian menyebabkan sentimen turun, maka pihak yang bersangkutan akan menghindari kejadian serupa untuk meningkatkan sentimen.

Kebutuhan-kebutuhan tersebut biasanya muncul ketika suatu pihak ingin mendapatkan sentimen publik yang baik atau melakukan pencitraan. Kebutuhan seperti ini biasa dimiliki oleh tokoh-tokoh publik, atau lebih khusus lagi tokoh politik seperti calon gubernur, calon presiden, menteri, atau ketua partai. Oleh karena itu, makalah ini mengambil domain tokoh politik.

II. PENELITIAN TERKAIT

Penelitian mengenai klasifikasi sentimen telah dilakukan oleh Bo Pang (2002) Pada papernya, Bo Pang melakukan klasifikasi sentimen terhadap review film dengan menggunakan berbagai teknik pembelajaran mesin. Teknik pembelajaran mesin yang digunakan yaitu Naïve Bayes, Maximum Entropy, dan Support Vector Machines (SVM). Pada penelitian itu juga digunakan beberapa pendekatan untuk melakukan ekstraksi fitur, yaitu unigram, unigram+bigram, unigram+Part of Speech (POS), adjective, dan unigram+posisi. Hasil dari eksperimen yang dilakukan di penelitian ini menemukan bahwa SVM menjadi metode terbaik ketika dikombinasikan dengan unigram dengan akurasi 82.9%. [3]

Penelitian lain yang mengarahkan ke media jejaring sosial dilakukan oleh Ravi Parikh dan Matin Movassate (2009). Mereka melakukan analisis sentimen terhadap media jejaring sosial Twitter dengan menggunakan beberapa teknik klasifikasi. Metode yang digunakan adalah Unigram Naïve Bayes, Multinomial Naïve Bayes, dan Maximum Entropy Classification. Hasil yang diperoleh dari penelitian mereka dapat ditarik kesimpulan bahwa kedua Naïve Bayes memperlihatkan performa yang lebih baik dari Maximum Entropy. Masih dari penelitian ini, ditemukan bahwa terdapat kesulitan dalam mendapatkan tweet yang dapat dijadikan sebagai data latih. Hal ini disebabkan kebanyakan tweet hanya berisi tautan, tidak menunjukkan sentimen tertentu, dan ditulis dalam bahasa asing (bukan bahasa Inggris). Selain itu,

disarankan untuk memperkaya teknik preproses yang digunakan agar hasil diperoleh lebih baik.[4]

Suhaad Prasad (2011) mencoba untuk menggunakan Naïve Bayes dengan berbagai macam pendekatan yakni, Bernoulli, Bernoulli Chi Square, Multinomial Unigram, Linear Bigram, Backoff Bigram, Empirical Bigram, dan Weighted-Normalized Complement Naïve Bayes (WCNB). Dari hasil uji coba diketahui bahwa Multinomial Unigram, Bernoulli Chi Square, dan Linear Bigram menunjukkan hasil yang cenderung lebih baik dari pendekatan lain.[5]

Selain itu, Jonathon Read (2005) menyatakan pentingnya domain, topik, waktu, dan gaya bahasa dalam melakukan analisis sentimen. Oleh karenanya, Read melakukan pendekatan dengan memanfaatkan emoticon yang ada di tiap sentimen. Namun, hasil yang diperoleh tidak terlalu signifikan.[6]

Penelitian yang mencakup area ekstraksi topik dilakukan oleh Cuneyt Gorcan Ackora. Pada penelitiannya, digunakan prinsip Tf.Idf untuk mengekstrak topik dalam Twitter. Hasil yang diperoleh menunjukkan metode yang digunakan mampu menentukan topik dalam Twitter.[1]

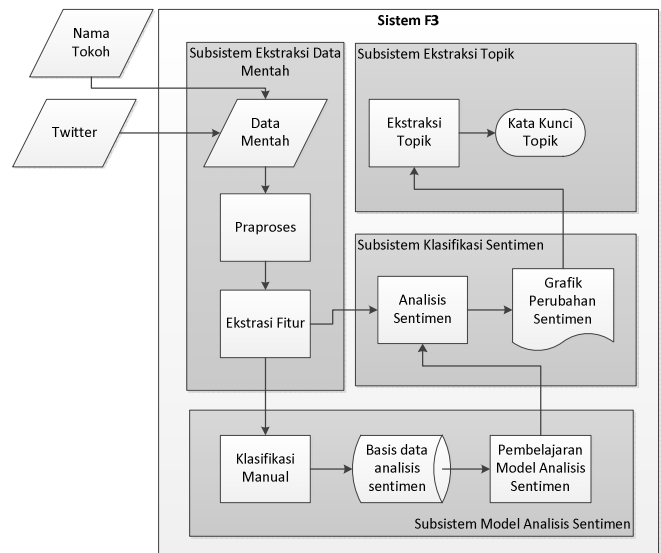
Sedangkan penelitian mengenai jejaring sosial berbahasa Indonesia telah dilakukan oleh Aqsath Rasyid Naradipha & Ayu Purwarianti (2011). Penelitian ini melakukan eksperimen untuk melakukan klasifikasi sentimen terhadap data yang diperoleh dari Facebook Page suatu perusahaan. Beberapa langkah preproses yang digunakan dalam penelitian ini adalah CleanNumber, ConvertNumber, RemoveRepeat, dan Translate dengan menggunakan KamusAlay. Hasil yang diperoleh menunjukkan penggunaan kamus dan formalisasi meningkatkan akurasi secara signifikan.[2]

III. ANALISIS SENTIMEN DAN EKSTRAKSI TOPIK PENENTU SENTIMEN PADA OPINI TERHADAP TOKOH PUBLIK

A. Deskripsi Sistem F3

Sistem F3 yang dibangun memiliki empat buah subsistem yaitu subsistem ekstraksi data mentah, subsistem model analisis sentimen, subsistem klasifikasi sentimen, dan subsistem ekstraksi topik. Masukan dari luar sistem adalah nama tokoh dan data dari Twitter. Nama tokoh dijadikan sebagai filter untuk mendapatkan tweet yang mengandung nama tokoh yang bersangkutan. Nama tokoh yang digunakan disarankan nama populer tokoh tersebut, misalnya “foke” untuk Fauzi Bowo atau “Gus Dur” untuk Kiai Haji Abdurrahman Wahid. Skema sistem F3 yang dibangun bisa dilihat pada Gambar 1.

Secara umum, subsistem ekstraksi data mentah bertugas untuk melakukan pengambilan *tweet* dan melakukan preproses serta ekstraksi fitur sehingga *tweet* yang diambil siap digunakan. Subsistem model analisis sentimen bertugas untuk melakukan pembelajaran model analisis sentimen. Model ini dipergunakan untuk melakukan analisis sentimen pada subsistem klasifikasi sentimen. Subsistem ekstraksi topik memiliki fungsi utama untuk mengekstrak topik yang diwakili sebagai kata kunci.



Gambar 1 Skema Sistem F3

B. Subsistem Ekstraksi Data Mentah

Seperti yang telah dijelaskan di atas, subsistem ini berfungsi untuk mengambil data dari Twitter menjadikan data tersebut siap pakai. Pengambilan data dari Twitter menggunakan bantuan *library* 140dev¹ yang bersifat *open source*. Data ini kemudian dikenai praproses. Ada beberapa langkah praproses yang dirancang. Di bawah ini dijelaskan masing-masing metode praproses.

Praproses merupakan salah satu langkah penting dalam melakukan analisis sentimen. Dengan melihat karakteristik *tweet* sebagai data mentah dari Twitter, maka perlu dilakukan langkah-langkah praproses sebagai berikut :

1) *Tokenizer*: Tokenizer merupakan proses untuk memisahkan kata atau word. Proses tokenizer dalam Twitter memiliki perbedaan dengan proses tokenizer pada teks lain. Hal ini dikarenakan adanya emoticon yang sering digunakan oleh pengguna Twitter dalam mengungkapkan perasaannya. Tahapan *tokenizer* dimulai dari memisah-misah bagian *tweet* yang dipisahkan dengan karakter spasi. Selanjutnya, bagian yang hanya memiliki satu karakter non alfabet dan angka akan dibuang. Bagian yang termasuk dalam daftar emotikon akan dikonversi. Terakhir, bagian yang memiliki karakter selain alfabet, angka, dan garis bawah akan dipecah sesuai posisi karakter tersebut.

2) *Normalisasi Fitur*: Ada beberapa komponen khas yang biasa ada di *tweet* yakni, username, URL, “RT” (tanda retweet), dan hashtag. Karena username, URL, dan “RT” tidak memiliki pengaruh apapun terhadap nilai sentimen, maka ketiga komponen di atas akan dibuang. Komponen username diidentifikasi dengan kemunculan karakter ‘@’, sedangkan komponen URL dikenali melalui ekspresi regular (regex). Untuk kasus hashtag, yang ditandai dengan munculnya karakter ‘#’, akan diubah menjadi string hashtag.

¹www.140dev.com

3) *Casefolding*: Model kesalahan orthographic model, akan dikoreksi dengan cara mengubah semua huruf menjadi lowercase atau huruf kecil

4) *CleanNumber*: Menurut Aqsath & Ayu (2011), kemunculan angka di depan dan diakhir kata memiliki pengaruh yang tidak signifikan dari nilai sentimen suatu tweet.[2]

5) *ConvertNumber*: Masih berdasarkan Aqsath & Ayu (2011), kemunculan angka di tengah-tengah kata perlu diubah menjadi karakter serupa. Konversi angka akan dilakukan berdasarkan *Tabel 1* **Error! Reference source not found.**[2]

Tabel 1 Konversi Angka

Angka	Konversi	Angka	Konversi
0	o	5	s
00	u	6	g
1	i	7	t
2	Sama dengan karakter sebelum angka 2	8	b
3	e	9	g
4	a		

6) *ConvertEmoticon*: Emoticon (emotion icon) merupakan salah satu cara pengungkapan perasaan secara tekstual. Hal ini tentu akan membantu dalam menentukan sentimen suatu tweet. Setiap emoticon akan dikonversikan ke dalam string yang bersesuaian sesuai dengan **Error! Reference source not found.** Namun, tidak semua akan diimplementasikan, karena tidak semua emoticon sering digunakan oleh pengguna Twitter.

Tabel 2 Konversi Emotikon

Emoticon	Konversi
>:] :-) :o) :] :3 :c) :> =] 8)	_e_happy_
=) :] :^)	
>:D :-D :D 8-D 8D x-D xD X-D XD =-D =D =-3 =3	_e_laugh_
>:[:- (: (:-c :c :-< :< :- [: {>.><.<.>.<	_e_sad_
D:< D: D8 D: D= DX v.v D-':	_e_horror_
>:P :-P :P X-P x-p xp XP :-p :p =p :-P :P :-b :b	_e_tongue_
>:o >:O :-O :O °o° °O° :O o_O o.O 8-0	_e_shock_
>:\>:/ :-/ :- :/ :\ =/ =\ :S	_e_annoyed_
: :-	_e_straight_face

7) *ConvertWord*: Beberapa kata yang muncul di tweet cenderung sangat tidak baku, dan tidak ada dalam kamus. Oleh karenanya, untuk kata yang tidak ada di kamus, akan ditranslasikan ke kata terdekat dengan menggunakan kamus yang dibuat dengan melihat pola kemunculan kata-kata tidak baku tersebut

8) *ConvertNegation*: Beberapa kata yang bersifat negasi, akan merubah nilai sentimen suatu tweet. Oleh karenanya, kata-kata ini akan dihilangkan, dan akan diberikan penanda saja. Sehingga, tidak perlu diproses lebih lanjut. Ketika banyak kata negasi adalah ganjil, maka sentimen tweet tersebut akan diubah. Kata yang bersifat negasi adalah “bkn”, “bukan”, “tidak”, “enggak”, “g”, “ga”, “jangan”, “nggak”, “tak”, “tdk”, dan “gak”.

9) *StopWordRemoval*: Proses ini akan menghilangkan kata-kata yang sering muncul namun tidak memiliki pengaruh apapun dalam ekstraksi sentimen suatu tweet. Kata yang termasuk kata ini adalah kata “gaul” seperti “cc”, “gw”, atau “ente”; kata hubung seperti “di”, “dengan”, atau “ke”, penunjuk waktu, dan kata tanya.

Setelah dilakukan praproses, langkah selanjutnya adalah melakukan ekstraksi fitur. Dilihat dari struktur dari tweet yang terbatas dengan 140 karakter, pemilihan unit fitur menjadi penting. Pemilihan frasa sebagai unit fitur tentu diperkirakan tidak memberikan hasil yang optimal karena sangat tidak bakunya struktur kalimat di Twitter. Selain itu, penggunaan n-gram dengan gram sebagai karakter atau kalimat jelas tidak sesuai. Karena, karakter tidak mencerminkan sentimen suatu tweet dandalam satu tweet seringkali hanya memiliki satu kalimat.

Pilihan yang diperkirakan memberikan hasil yang paling baik adalah *word (unigram)* dan *bigram*, dimana *gram* adalah kata. Hal ini juga didukung oleh penelitian yang telah dijelaskan di bab dua yang menyebutkan Bigram dan Unigram memberikan hasil yang cenderung lebih baik dari yang lain. Selain pemilihan unit terkecil fitur, isu yang perlu diperhatikan adalah pemilihan fitur. Hal ini dimotivasi dari banyaknya *word* dan *bigram* yang muncul dari *tweet* yang dijadikan data mentah. Oleh karenanya, perlu diberikan syarat untuk membatasi unit fitur yang dijadikan sebuah fitur. Pemilihan fitur pada makalah ini menggunakan *frequency based selection*. *Frequency based selection* dipilih karena melihat isi *tweet* yang seringkali mengandung kesalahan ejaan yang cukup parah. Kesalahan ini biasanya tidak diulangi di *tweet* lain, sehingga hanya dalam satu kali *tweet* saja kesalahan itu muncul. Akibatnya, tidak bisa diketahui pengaruhnya terhadap kelas sentimen. Dengan menggunakan *frequency based selection* kata-kata salah ejaan tersebut bisa dihilangkan.

C. Subsistem Model Analisis Sentimen

Dalam subsistem ini, dipergunakan 2000 tweet yang dilabeli sebagai positif, negatif, dan netral. Tweet yang mengandung judul berita saja akan dikategorikan sebagai tweet yang bersifat netral. Dalam domain ini, banyak sekali tweet yang hanya mengandung judul berita beserta link berita tersebut. Hal ini menjadi salah satu karakteristik tweet berbahasa Indonesia dengan domain tokoh publik. Pengguna Twitter di luar negeri, khususnya yang menggunakan Bahasa Inggris. Pengguna Twitter di luar cenderung jarang melakukan retweet terhadap tweet yang hanya berisi judul berita.

Langkah selanjutnya adalah melakukan pembentukan model analisis sentimen. Pembentukan model analisis

sentimen ini menggunakan metode pembelajaran mesin. Sistem mengambil data tweet yang tersimpan di basis data analisis sentimen. Tweet yang tersimpan ini telah melalui praproses dan ekstraksi fitur serta telah dilabeli dengan kelas sentimennya secara manual. Data inilah yang digunakan sebagai data latih untuk membentuk model analisis sentimen. Model inilah yang dipergunakan untuk mengklasifikasikan sebuah sentimen. Pada makalah ini dipergunakan dua buah metode pembelajaran mesin yakni Naïve Bayes. Naïve Bayes dipilih karena sudah teruji di beberapa penelitian mampu menghasilkan akurasi yang baik. Selain itu, Naïve Bayes memang cocok untuk klasifikasi dengan atribut homogen seperti pada analisis sentimen ini yang semua atributnya diskrit. Alasan lain pemilihan Naïve Bayes ini adalah implementasinya yang relatif lebih mudah daripada SVM karena pada makalah ini menekankan pada praproses.

Naïve bayes merupakan metode pembelajaran mesin yang memiliki model dalam bentuk probabilitas atau peluang. Data latih yang berupa pasangan tweet dan kelas atau sentimen dijadikan sebagai sumber pembentukan model analisis. Setiap fitur yang merepresentasikan tweet dihitung probabilitas kemunculannya di kelas positif maupun negatif. Ketika ada tweet baru yang ingin diklasifikasikan, maka nilai probabilitas dari setiap fitur untuk masing-masing kelas dikalikan. Nilai probabilitas yang paling besar diantara kelas positif dan negatif dijadikan kelas tweet baru tersebut. Formula untuk menentukan kelas dari tweet baru ditunjukkan oleh persamaan (1) berikut ini:

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_{i=1}^n P(a_i | v_j) \quad (1)$$

Dimana :

a_i = atribut atau fitur ke- i

v_j = kelas ke- j (positif atau negatif)

V = himpunan kelas target

V_{MAP} = kelas sentimen suatu tweet

Bisa dilihat, pada persamaan (1), bahwa untuk setiap atribut atau fitur diasumsikan tidak memiliki keterhubungan satu sama lainnya.

Pada makalah ini, digunakan implementasi Naïve Bayes pada NLTK². NLTK atau Natural Language Tool Kit yang biasa digunakan untuk melakukan eksperimen di bidang bahasa alami. NLTK ditulis menggunakan bahasa pemrograman python dan bersifat open source.

D. Subsistem Klasifikasi Sentimen

Subsistem ini bertugas mengklasifikasikan sentimen untuk tweet baru dengan menggunakan model analisis sentimen yang dihasilkan oleh subsistem model analisis sentimen. Kemudian, setiap tweet yang telah diklasifikasikan akan dimasukkan ke dalam beberapa kelompok interval waktu berdasarkan kapan tweet tersebut dipos. Selanjutnya, untuk setiap interval waktu dihitung rata-rata sentimen dengan sentimen positif bernilai 1, netral bernilai 0, dan negatif bernilai -1.

Untuk memudahkan pengguna dalam melihat kapan terjadi perubahan sentimen, hasil klasifikasi sentimen tersebut ditampilkan dalam bentuk grafik. Karena Twitter merupakan media yang kontinu, maka sentimen untuk suatu interval tertentu tidak hanya dipengaruhi oleh rata-rata sentimen pada interval tersebut saja. Rataan sentimen interval sebelumnya dianggap memiliki pengaruh. Pengaruh ini dimasukkan dalam menghitung sentimen dalam suatu selang dengan perhitungan sebagai berikut.

Misalkan T_n merupakan nilai rata-rata sentimen pada selang waktu ke- n . Serta didefinisikan pula α sebagai *discounted cumulative factor*. Maka nilai sentimen untuk selang n , disimbolkan dengan τ_n , didefinisikan secara rekursif seperti pada persamaan **Error! Reference source not found.**

$$\tau_n = (1 - \alpha)T_n + \alpha\tau_{n-1} \quad (2)$$

E. Subsistem Topic Extractor

Pada subsistem ini diekstraksi topik untuk setiap selang waktu. Metode pengekstrakan ini menggunakan nilai Tf.Idf. Idf diperlukan untuk mengurangi bobot dari kata yang muncul di banyak interval. Kata-kata ini biasanya kata yang sedang populer muncul dan tidak menunjukkan topik apapun namun tidak termasuk dalam daftar stopwords. Atau bisa juga stopwords baru yang belum dimasukkan ke dalam daftar stopwords.

Untuk setiap kata atau fitur yang dimiliki tweet-tweet pada selang waktu tersebut dihitung nilai Td.Idf-nya. Selanjutnya diambil beberapa kata atau fitur yang memiliki nilai paling tinggi. Hasil akhir dari sistem F3 ini adalah kata atau term yang memiliki nilai paling tinggi yang merepresentasikan topik pemicu nilai sentimen di selang waktu tersebut.

Karena Twitter merupakan media yang kontinu, sehingga tidak bisa melakukan ekstraksi topik secara terpisah untuk setiap interval waktu. Oleh karena itu, dipergunakan *discounted cumulative* dalam melakukan pencarian topik. Dengan *discounted cumulative* ini, kata kunci yang sering muncul di interval sebelumnya dihitung dalam penentuan bobot kata pada interval yang akan dihitung. Penentuan bobot ini bisa diformulasikan seperti persamaan (3) yang dimotivasi dari penelitian yang dilakukan oleh Cuneyt Gorcan Ackora[1].

$$w_{x,n} = TfIdf_{x,n} + \alpha w_{x,n-1} \quad (3)$$

Dimana,

$w_{x,n}$ = bobot kata x pada interval ke- n

$TfIdf_{x,n}$ = nilai Tf-Idf kata x pada interval ke- n

α = *discounted cumulative factor*

Nilai α yang disarankan berkisar antara 0 hingga 1. Hal yang perlu diperhatikan dalam pemilihan nilai α ini adalah panjang interval waktu yang digunakan untuk membagi tweet. Sedangkan pengekstrakan topik akan dibagi dua yakni dari tweet bersentimen positif dan tweet bersentimen negatif. Hal ini dilakukan untuk mengetahui topik apa yang memiliki pengaruh positif dan topik apa yang memiliki pengaruh negatif.

IV. EVALUASI DAN ANALISIS

² www.nltk.org

A. Data Set Eksperimen

Data set eksperimen yang digunakan adalah Twitter yang diambil selama 1 Minggu, mulai dari tanggal 10 April 2012 hingga 18 April 2012. Data set ini diambil menggunakan Stream API yang disediakan oleh Twitter dengan menggunakan kata kunci sebagai berikut : 'dahlan iskan', 'alex noerdin', 'faisal basri', 'fauzi bowo', 'foke', 'hendardji', 'hidayat nur wachid', 'jokowi', 'rachel maryam', 'ahmad heryawan', dan '@iskan_dahlan'. Karena kata kunci tersebut merupakan tokoh-tokoh Indonesia, tweet yang terambil pun hampir semua dalam bahasa Indonesia, jadi tidak perlu lagi dilakukan penyaringan bahasa. Dari data tersebut, diambil secara acak sebanyak 1000 tweet dan dilabeli secara manual menjadi tiga kelas: negatif (210 *tweet*), netral (566 *tweet*), dan positif (224 *tweet*).

B. Prosedur Eksperimen

1) Eksperimen Praproses

Dimulai dengan melakukan klasifikasi tanpa menggunakan tahapan praproses apapun, kecuali praproses yang wajib yakni *normalize_character*, *RemoveOneChar*, dan *RemoveURL*. Selanjutnya, untuk masing-masing metode praproses akan dicoba secara terpisah. Percobaan juga dilakukan untuk batas kemunculan minimal 1, 2, 3, dan 4 dalam data latih. Setelah mendapatkan hasil, dilakukan perbandingan akurasi. Terakhir adalah mengkombinasikan metode yang diperkirakan akan menghasilkan akurasi terbaik

2) Eksperimen Ekstraksi Topik

Melakukan klasifikasi terhadap *tweet* dalam rentang 15 April 2012 pukul 02.00 hingga 18 April 2012 pukul 10.00 dan dikelompokkan masing-masing 4 jam. Menentukan topik pemicu sentimen dengan *discounted cumulative factor* sebesar 0, 0.25, dan 0.5. mengecek topik yang berhasil diekstrak dengan cara melakukan pencarian kata kunci yang berhasil diekstrak dengan mesin pencari (Google) untuk memperoleh kejadian atau berita yang berhubungan dengan kata kunci tersebut. Menganalisis apakah kata kunci tersebut benar-benar menjadi sumber perubahan sentimen atau bukan

C. Analisis Hasil dan Diskusi Kesalahan Eksperimen Praproses

Hasil eksperimen terhadap masing-masing metode praproses bisa dilihat pada Tabel 3. Pada tabel tersebut, hampir semua praproses tidak meningkatkan akurasi, kecuali metode *ConvertNumber* dan *ConvertEmoticon* yang meningkatkan akurasi masing-masing sebesar 0.1% dan 0.3%. Sedangkan *RemoveHashtag* dan *ConvertNegation* memiliki akurasi yang sama dengan akurasi yang dimiliki oleh *Baseline*. Sedangkan metode *Casefolding* dan menggunakan semua metode menurunkan nilai akurasi paling tinggi. Hingga menyentuh nilai 69.4 % dan 69.5%.

Sedangkan untuk manipulasi nilai minimal kemunculan suatu fitur di data latih untuk bisa dijadikan fitur adalah seperti yang tercantum pada Tabel 4

found.. Bisa dilihat bahwa dengan merubah nilai minimal kemunculan, bisa menambah akurasi

Tabel 3 Hasil Eksperimen Praproses

No	Metode	Akurasi
1	<i>Baseline</i>	71.5%
2	<i>Casefolding</i>	69.4%
3	<i>RemoveRT</i>	71.1%
4	<i>RemoveHashtag</i>	71.5%
5	<i>ConvertEmoticon</i>	71.8%
6	<i>RemovePunctuation</i>	71.3%
7	<i>CleanNumber</i>	71.4%
8	<i>StopWordRemoval</i>	71.3%
9	<i>RemoveUsername</i>	70.6%
10	<i>ConvertNumber</i>	71.6%
11	<i>ConvertNegation</i>	71.5%
12	<i>ConvertWord</i>	71.4%

Tabel 4 Hasil Eksperimen Minimal Kemunculan

No	Minimal Kemunculan	Akurasi
1	1	71.5%
2	2	72 %
3	3	72.8 %
4	4	71.5%

Dari melihat hasil eksperimen di atas, kemudian dilakukan berbagai kombinasi yang diperkirakan akan memiliki nilai akurasi paling tinggi. Karena ada 10 metode praproses dan ada 3 alternatif nilai minimal kemunculan, total kombinasi adalah $2^{10} \times 3 = 3072$. Angka tersebut adalah angka yang sangat besar. Sehingga, hanya akan dilakukan kombinasi yang diperkirakan menambah akurasi dilihat dari hasil percobaan sebelumnya. Pada Tabel 5 bisa dilihat hasil dari eksperimen kombinasi.

Tabel 5 Hasil Eksperimen Kombinasi

No	Metode	Min Kemunculan	Akurasi
1	<i>Baseline</i>	1	71.5%
2	<i>Baseline</i>	2	72 %
3	<i>Baseline</i>	3	72.8 %
4	<i>ConvertEmoticon</i>	2	72%
5	<i>ConvertNumber</i>	2	72%
6	<i>ConvertNumber, ConvertEmoticon</i>	2	72%
7	<i>ConvertNumber, ConvertEmoticon</i>	3	71.8%
8	Semua metode kecuali <i>Casefolding</i>	1	70.1%
9	Semua metode kecuali <i>Casefolding</i>	2	69.8%
10	<i>RemoveHashtag, ConvertNumber, ConvertEmoticon, ConvertNegation</i>	2	71.9%
11	<i>RemoveHashtag, ConvertNumber, ConvertEmoticon, ConvertNegation</i>	1	71.7%

Dari Tabel 5, hampir semua kombinasi menghasilkan nilai akurasi yang lebih tinggi dari baseline kecuali kombinasi nomer 8 dan 9. Dan bisa dilihat juga, rata-rata kenaikan akurasi adalah 0.5%. Dengan akurasi tertinggi dihasilkan pada kombinasi nomer 3, yakni baseline dan dengan menggunakan minimal kemunculan sebanyak 3.

Jika dilihat pada hasil-hasil sebelumnya, penggunaan metode praproses yang tepat bisa meningkatkan akurasi. Namun, jika dilihat lebih lanjut lagi, kenaikan yang diperoleh

maksimal 1.3 % dan ini bukan merupakan angka yang terlalu signifikan. Bahkan, ada metode praproses yang justru menurunkan nilai akurasi, seperti metode Casefolding.

Setelah dilihat dari model tweet yang berisikan nama seorang tokoh, lebih dari 50% dari data tes bersifat netral yang didominasi oleh judul berita berikut pranala atau link-nya. Jika dilihat lebih mendalam lagi, judul berita menggunakan huruf kapital pada setiap awal kata. Hal ini berbeda dengan kebiasaan pengguna Twitter yang jarang sekali menggunakan huruf kapital, kecuali di awal kalimat itupun karena *fitur auto-capitalize* yang ada di perangkat mobile.

Dari fakta tersebut, bisa dipahami jika Casefolding justru menurunkan akurasi karena program menyamakan kata yang muncul sebagai judul berita dan opini asli dari pengguna Twitter.

Fakta mengenai banyaknya pengguna Twitter yang hanya berbagi link berita juga bisa menjelaskan kenapa metode ConvertEmoticon dan ConvertNumber mampu menaikkan akurasi. Hal ini tak lain dan tak bukan adalah model bahasa yang digunakan pada judul berita sangat teratur. Belum pernah ditemukan ada judul berita yang menggunakan emoticon atau menggunakan angka sebagai huruf sehingga perlu di-convert.

Fakta lain dari Twitter adalah sering muncul kata-kata unik yang hanya sekali muncul dalam satu tweet. Hal ini menyebabkan nilai probabilitas kata tersebut untuk suatu sentimen menjadi sempurna, atau satu. Oleh karenanya, dengan membatasi nilai minimal kemunculan bisa mengatasi hal ini. Dan diketahui bahwa 3 merupakan batas dari nilai minimal ini. Karena, jika nilai minimal diset empat, akan menyebabkan banyak kata hilang walaupun bisa menunjukkan sentimen dengan lebih baik.

Dengan menggunakan kombinasi terbaik, yakni baseline dan nilai minimal kemunculan 3, ada beberapa kesalahan yang muncul ketika melakukan klasifikasi. Kesalahan ini muncul ketika dilakukan pengklasifikasian 100 tweet baru. Berikut contoh-contohnya :

Sirik ajh RT @detikfinance: Anggota DPR Galang Kekuatan 'Goyang' Dahlan Iskan <http://t.co/UmUhtWmH> Netral

Sebenarnya, tweet ini bernilai positif. Namun, kata “sirik” dan “ajh” jarang digunakan dan akhirnya terbuang karena tidak mencapai batas nilai minimal kemunculan. Akibatnya, hanya judul berita saja yang dianggap sebagai fitur, dan akhirnya diklasifikasikan sebagai tweet netral. Dengan alasan yang sama dengan contoh sebelumnya, kesalahan pada tweet di bawah ini pun dapat dijelaskan.

IkutRT @yesalicious: Klo bliau capres, gw pilih'a!! RT @Metro_TV: Dahlan Iskan Anggap Pangan (cont) <http://t.co/GwOzCL5i> Netral

D. Analisis Hasil dan Diskusi Kesalahan Eksperimen Ekstraksi Topik

Dari tiga buah nilai *discounted cumulative factor* sebesar 0, 0.25, dan 0.5 diperoleh presisi hasil ekstraksi topik sebesar 43.5%, 51.2%, dan 43.2%. Hal ini membuktikan bahwa memang ada pengaruh sentimen dan topik antar selang. Namun, tentu saja pengaruh yang disimbolkan dengan faktor

peluruhan ini tidak boleh terlalu besar, karena malah akan mengaburkan topik yang sebenarnya terjadi.

Dari data yang dipergunakan, hampir semua selang membicarakan tentang topik “Foke dan Jari Tengahnya”. Namun, ternyata tidak semua selang mampu mengekstraksi topik tersebut. Salah satu alasan yang dirasa mampu menjelaskan adalah karena topik ini terlalu sering muncul, sehingga mengurangi nilai Idf kata kunci yang bersangkutan. Dan lebih lanjut lagi, tentu akan mengurangi nilai Tf.Idf secara keseluruhan.

Kesalahan lain yang muncul adalah topik pembicaraan di Twitter yang tidak selalu bersumber dari berita. Misal saja seperti pada selang nomor urut tujuh yang memberikan kata kunci “kreen, blah, omabdi, _e_sad_, dikenal”. Jelas, kata-kata tersebut bukan dari berita, dan berasal dari tweet salah satu pengguna yang akhirnya di-retweet. Tentu saja, dengan metode pencarian kata kunci di mesin pencari, susah menemukan kata-kata kunci tersebut.

Kesalahan ke tiga yang terlihat adalah pola kesalahan yang mirip pada selang waktu 02.00-06.00. Dan di waktu-waktu ini, kebanyakan orang Indonesia sedang tertidur atau mengantuk. Jadi, lebih sering terjadi kesalahan ketik dan terjadi berulang-ulang sehingga memunculkan kata kunci seperti “mreka, kreen” dan kata-kata asing yang mungkin berasal dari belahan dunia lain yang masih siang hari.

V. KESIMPULAN

Kesimpulan yang bisa diambil dari penelitian ini adalah metode praproses yang dicoba dalam penelitian ternyata tidak memiliki pengaruh yang signifikan terhadap akurasi sistem. Penyebab utamanya adalah domain *tweet* tokoh masyarakat yang sering berisi judul berita.

Sedangkan penggunaan faktor peluruhan dalam pencarian topik ternyata mampu menaikkan jumlah topik yang berhasil diekstrak. Namun masih tidak terlalu baik ketika berhadapan dengan topik yang muncul di hampir semua waktu.

REFERENCES

- [1] Akcora, C. G., Bayir, M. A., Demirbas, M., & Ferhatosmanoglu, H. (2010). Identifying Breakpoints in Public Opinion Using Twitter. *SOMA '10 Proceedings of the First Workshop on Social Media Analytics* (pp. 62-66). New York: ACM New York.
- [2] Naradipha, A. R., & Purwarianti, A. (2011). Sentiment Classification for Indonesian Message in Social Media. *International Conference on Electrical Engineering and Informatics*, (pp. 1-4). Bandung.
- [3] 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 Computational Linguistic.
- [4] Parikh, R., & Movassate, M. (2009). *Sentiment Analysis of User Generated Twitter Updates using Various Classification*. Retrieved November 12, 2011, from CS224N Final Projects 2008-9: <http://nlp.stanford.edu/courses/cs224n/2009/fp/19.pdf>
- [5] Prasad, S. (n.d.). *Micro-blogging Sentiment Analysis Using Bayesian Classification Methods*. Retrieved November 29, 2011, from The Stanford NLP (Natural Language Processing) Group: <http://www-nlp.stanford.edu/courses/cs224n/2010/reports/suhaasp.pdf>
- [6] Read, J. (2005). Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification. *ACLstudent '05 Proceedings of the ACL Student Research Workshop* (pp. 43-48). Stroudsburg: Association for Computational Linguistic.

- [7] Sysomos Inc. (2010, January). *Twitter Usage Statistics by Geography, Countries, Cities, Around the World*. Retrieved December 29, 2011, from Sysomos: <http://www.sysomos.com/insidetwitter/geography/>