

PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGUNAKAN METODE *K-NEAREST NEIGHBOR*

Akhmad Deviyanto ⁽¹⁾, M. Didik R. Wahyudi ⁽²⁾
Teknik Informatika UIN Sunan Kalijaga Yogyakarta
Jl. Marsda Adi Sucipto No 1 Yogyakarta
e-mail : akhmad.deviyanto@gmail.com ⁽¹⁾, m.didik@uin-suka.ac.id ⁽²⁾

Abstract

This research is made to implement the KNN (K-Nearest Neighbor) algorithm for sentiment analysis Twitter about Jakarta Governor Election 2017. The object is 2000 data tweets in Indonesia collected from Twitter during Januari 2017 using Python package called Twtterscraper. The method used in sentiment analysis system is KNN with TF-IDF term weighting and Cosine similarity measure. As the test result, the highest accuracy is 67,2% when k=5, the highest precision is 56,94% with k=5, and the highest recall 78,24% with k=15.

Keywords : *K – Nearest Neighbor, Twtterscraper, TF-IDF, Cosine Similarity*

Abstrak

Penelitian ini dibuat untuk mengimplementasikan algoritma KNN (K - Nearest Neighbor) dalam analisis sentimen pengguna Twitter tentang topik Pilkada DKI 2017. Data tweet yang digunakan adalah sebanyak 2000 data tweet berbahasa Indonesia yang dikumpulkan selama bulan Januari 2017 menggunakan package Python bernama Twtterscraper. Menggunakan algoritma KNN dengan pembobotan kata TF-IDF dan fungsi Cosine Similarity, akan dilakukan pengklasifikasian nilai sentimen ke dalam dua kelas : positif dan negatif. Dari hasil pengujian diketahui bahwa nilai akurasi terbesar adalah 67,2% ketika k=5, presisi tertinggi 56,94% ketika k=5, dan recall 78,24% dengan k=15.

Kata Kunci : *K – Nearest Neighbor, Twtterscraper, TF-IDF, Cosine Similarity*

1. PENDAHULUAN

Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) telah melakukan survei pada tahun 2016. Ada sekitar 132,7 juta pengguna internet di Indonesia (naik secara signifikan dari tahun 2014 sebanyak 88 juta pengguna). Dari jumlah tersebut, sebanyak 97,4% (129,2 juta) adalah pengguna yang menggunakan internet untuk mengakses media sosial. Lima media sosial dengan pengguna terbanyak adalah Facebook, Instagram, Youtube, Google Plus, dan Twitter.

Social media Twitter memiliki 7,2 juta pengguna di Indonesia. Setiap harinya, paling tidak ada 4,1 juta tweet yang berasal dari Indonesia (CNN Indonesia, 2016). Jumlah yang cukup besar tersebut merupakan cuitan para penggunanya tentang banyak hal: pendidikan, hiburan, pekerjaan, dan termasuk juga politik.

Salah satu isu politik yang menjadi trending topic di Twitter pada tahun 2017 adalah tentang pemilihan gubernur Jakarta (Pilkada DKI). Terdapat tiga pasangan calon pada Pilkada DKI 2017. Ketiga paslon tersebut adalah Agus Harimurti Yudhoyono - Sylviana Murni, Basuki Tjahaja Purnama - Djarot Saiful Hidayat, Anies Rasyid Baswedan - Sandiaga Salahuddin Uno.

Sentimen Analisis atau opinion mining adalah jenis natural language yaitu pengolahan kata untuk melacak mood masyarakat tentang produk atau topik tertentu. Analisis sentimen, disebut opinion mining. (G.Vinodhini, M.Chandrasekaran 2012).

Penulis pada penelitian ini akan melakukan analisis sentimen para pengguna Twitter terhadap ketiga pasangan kandidat pada Pilkada DKI 2017. Dengan input berupa data tweet dalam Bahasa Indonesia, akan dilakukan klasifikasi dengan algoritma KNN (K-Nearest Neighbor) untuk menentukan apakah tweet tersebut bersentimen positif atau negatif.

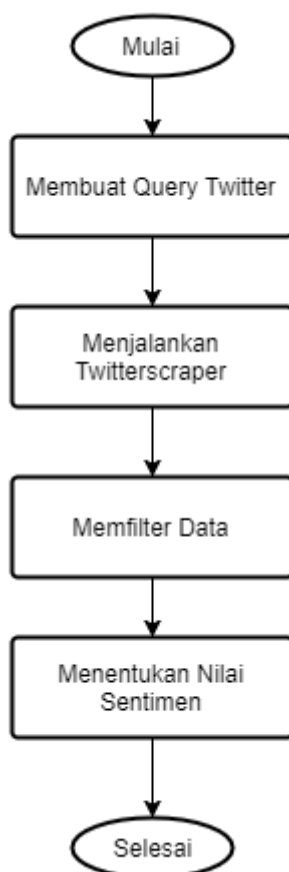
2. METODE PENELITIAN

Secara garis besar, ada dua proses utama dalam penelitian ini yaitu pengambilan data dari Twitter dan proses analisis sentimen.

Pengambilan data dari Twitter

Untuk mengambil data tweets yang diperlukan, digunakan package Twitterscraper yang merupakan salah satu package Python.

Berikut alur pengambilan data dari Twitter dengan menggunakan Twitterscraper :



Gambar 1. Alur Pengambilan Data Tweet

Agar dapat melakukan scraping data dari Twitter, harus ditentukan terlebih dahulu query untuk Twitterscraper. Query ini didapat dari halaman pencarian lanjutan Twitter (<https://twitter.com/search-advanced>).

URL untuk mencari data tweets yang mengandung kata 'AHY' dari mulai tanggal 1 Januari 2017 sampai dengan 31 Januari 2017 adalah :

```
https://twitter.com/search?l=id&q=AHY%20since%3A2017-01-01%20until%3A2017-01-31&src=typd
```

Query yang dibutuhkan Twitterscraper untuk melakukan scraping data dari URL tersebut adalah `AHY%20since%3A2017-01-01%20until%3A2017-01-31` sehingga *command* yang diinputkan ketika menjalankan Twitterscraper adalah sebagai berikut :

```
twitterscraper AHY%20since%3A2017-01-01%20until%3A2017-01-31 -o ahy.json
```

Nama output file dari perintah di atas adalah ahy.json dan otomatis tersimpan pada direktori C:\Users\Namauser.

```
C:\Users\Akhmad Deviyanto>twitterscraper AHY%20since%3A2017-01-01%20until%3A2017-01-31 -o ahy.json
INFO: Running iteration no 1, query is 'AHY%20since%3A2017-01-01%20until%3A2017-01-31'
INFO: Querying AHY%20since%3A2017-01-01%20until%3A2017-01-31
INFO: Got 0 tweets (20 new),
INFO: Got 20 tweets (20 new),
INFO: Got 40 tweets (20 new),
INFO: Got 60 tweets (20 new),
INFO: Got 80 tweets (20 new),
INFO: Got 100 tweets (20 new),
INFO: Got 120 tweets (20 new),
INFO: Got 140 tweets (17 new),
INFO: Got 157 tweets (20 new),
INFO: Got 177 tweets (20 new),
```

Gambar 2. Tampilan Running Twitterscraper pada Command Prompt

1	timestamp	id	user	fullname	text
2	2016-12-30T00:00:54	814622251374907394	bongkar	Bongkar	Akan banyak akrobat dari #CikeasBandarMakar demi kelanc
3	2016-12-30T00:03:21	814622867669192704	rherehendri	rherehendri	Belajarlah sikap sopan dan perduli kepada warga dari AHY f
4	2016-12-30T00:03:47	814622976175841280	rherehendri	rherehendri	Status kita tetap pendukung AHY #omAHYmpic.twitter.com/Z
5	2016-12-30T00:04:20	814623114113925121	banniesoedarman	soedarman bannie	AHY menang....rakyat DKI bersiap-siap melongo....habis dikil
6	2016-12-30T00:07:53	814624011170656256	iskandarlu	Isk	kalau lompat ke gerombolan orang spt yg AHY lakukan? Sela
7	2016-12-30T00:07:59	814624034927177728	banniesoedarman	soedarman bannie	Program visoner AHY cuma satu :Kita bikin sibuk KPK dengar
8	2016-12-30T00:09:05	814624312854396928	buyangbinguang	Pergi Merantau	Puncaki Survei, AHY Terus Dihantam Kampanye Hitam http://
9	2016-12-30T00:11:02	814624802765819904	alex_hunter146	alexandi hunt s	Viral Ide Luar Biasa Rumah Apung, Itu Memejokan AHY
10	2016-12-30T00:11:08	814624826895646721	banniesoedarman	soedarman bannie	Yg gusur AHOK yang korupsi AHY kan asyiiiik
11	2016-12-30T00:11:10	814624837746294784	GunRomli	Mohamad Guntur Romli	Suami Sylviana Murni Akan Dimintai Keterangan Terkait Kas
12	2016-12-30T00:12:54	814625269969342465	GunRomli	Mohamad Guntur Romli	Ada Uang dari Suami Sylviana Murni ke Penyebar SARA 212 f
13	2016-12-30T00:14:30	814625672815443969	IndraYunaidi	Indra Yunaidi	Puncaki Survei, AHY Terus Dihantam Kampanye Hitam http://
14	2016-12-30T00:15:34	814625942110748672	hendraazmi	Indonesia Emas 2045	Pemberdayaan dan pelibatan RT RW yg akan dilakukan AHY j
15	2016-12-30T00:15:40	814625967339499521	FyanHooligan	Fyan Hooligan	Saya rasa menguntungkan AHY. Secara FPI lagi demikian dis
16	2016-12-30T00:17:20	814626388342755329	hendraazmi	Indonesia Emas 2045	Mereka yg di diserang pakai kampanye hitam, mereka lah yg
17	2016-12-30T00:18:14	814626613287489536	SartonoHutomo	Sartono Hutomo	"@IndraYunaidi: Puncaki Survei, AHY Terus Dihantam Kampa
18	2016-12-30T00:25:59	814628564020494396	agusnasrudin	djaki	fitnah bukan di pojok AHY di tengah2 koq ...
19	2016-12-30T00:30:21	814629665486893056	alphomega	Yosua Sianturi	Sekarang gv ngeri PR besar SBY untuk AHY. Public speaking
20	2016-12-30T00:31:05	814629847603367936	noninanto	Noni Marianto Liat	Hmmm....itu dia mbak @RinjanjB , saya jd gak yakin apaka
21	2016-12-30T00:32:56	814630315184427008	Mentimoen	Daemoen	Gde Sardjana ada di bagan jaringan pendanaan dr gurita cil

Gambar 3. Tampilan File JSON pada MS Excel

Kemudian, ulangi lagi running Twitterscraper untuk mengambil data tweets yang mengandung keyword 'ahok' dan 'anies'. Perintahnya adalah sebagai berikut :

- twitterscraper ahok%20since%3A2017-01-01%20until%3A2017-01-31 -o ahok.json
- twitterscraper Anies%20since%3A2017-01-01%20until%3A2017-01-31 -o Anies.json

Setelah data selesai diambil, selanjutnya dilakukan filter data dengan bantuan fitur 'Remove Duplicate' dari Microsoft Excel. Pertama, hilangkan duplikasi data berdasarkan kolom 'user'. Selanjutnya hilangkan duplikasi data berdasarkan kolom 'text'. Lalu, pilih data dan tentukan nilai sentimen setiap data tweets yang akan dimasukkan ke dalam sistem secara manual. Data yang dipakai dalam penelitian ini berjumlah 2000 data tweets dengan perincian 1500 sebagai data latih dan 500 sebagai data uji. Setelah terpilih, data kemudian disimpan dalam bentuk CSV untuk kemudian diimport ke dalam database.

Proses Analisis Sentimen

Dalam proses sentimen analisis ada dua tahap yaitu preprocessing dan proses analisis sentimen.

1. Text preprocessing

Merupakan tahapan pemrosesan data agar menjadi data yang siap untuk dianalisis. Setelah data terstruktur maka dapat diolah lebih lanjut. Beberapa proses yang dilakukan pada tahap ini adalah sebagai berikut :

a. *Case folding*

Merupakan proses perubahan semua huruf pada dokumen tweet menjadi huruf kecil. Hanya huruf a sampai z yang diproses. Karakter selain huruf akan dibiarkan.

b. *Tokenizing*

Pada tahap ini, kalimat dipotong atau dipecah berdasarkan tiap kata yang menyusunnya.

c. *Stopword Removal*

Merupakan tahap pembuangan kata-kata yang dianggap tidak penting. Langkah ini dilakukan supaya perhitungan lebih berfokus pada kata-kata yang jauh lebih penting.

d. *Stemming*

Tahap stemming adalah tahap mencari root (bentuk dasar) dari tiap kata. Pada tahap ini, dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama.

2. Analisis Sentimen dengan K-Nearest Neighbor

Setelah data melalui tahap pre-processing maka data telah siap untuk diolah menggunakan metode K-Nearest Neighbor. Metode K-Nearest Neighbor adalah proses untuk mengelompokkan data ke dalam kelas-kelas yang telah ditentukan sebelumnya berdasarkan jarak terdekat / tingkat kemiripan data tersebut dengan dataset / data latih yang ada. Nantinya data akan dikelompokkan ke dalam suatu kelas dengan melihat sejumlah "k" nilai jarak terdekat nya dengan data latih. Dalam penelitian ini, proses K-Nearest Neighbor meliputi 3 proses, yaitu :

a. Menghitung Bobot Kata (TF-IDF)

Lakukan penghitungan bobot kata (term) dari data ada menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). *Term frequency* menyatakan frekuensi (tingkat keseringan) munculnya suatu *term* dalam suatu dokumen. Sedangkan *document frequency* adalah banyaknya jumlah dokumen dimana sebuah *term* itu muncul.

Rumus dari TF-IDF adalah sebagai berikut :

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

Dimana *tf* menyatakan nilai term frequency dan $\log (N/df)$ menyatakan nilai IDF dengan *N* adalah jumlah banyaknya data.

Sebagai contoh, misal ada 7 data *tweet*. Akan dicari bobot kata "politik" pada dokumen pertama. Kata politik muncul sebanyak satu kali pada dokumen pertama.

Dari ketujuh dokumen tersebut, kata politik terulang pada dua dokumen. Maka hasil perhitungannya adalah :

$$tf = 1 \quad df = 2 \quad idf = \log\left(\frac{7}{2}\right) = 0,544$$

Sehingga : $w = 1 * 0,544 = 0,544$

b. Menghitung Tingkat Kemiripan (*Cosine Similarity*)

Dari langkah sebelumnya telah diketahui bobot tiap kata. Langkah selanjutnya adalah menghitung jarak atau tingkat kemiripan data dengan setiap data latih yang ada menggunakan rumus jarak Cosine Similarity. Lalu, sistem akan mengurutkan nilai jarak dari yang tertinggi sampai terendah.

Tahapan pada Cosine similarity adalah sebagai berikut :

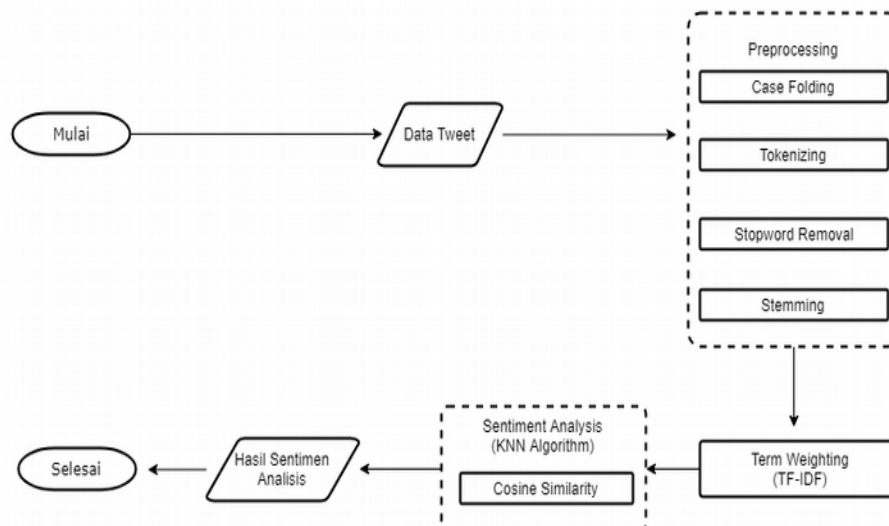
- i. Kalikan bobot dari setiap term pada D_1 dengan setiap term dari semua dokumen data latih yang ada..
- ii. Hasil perkalian D_1 dengan setiap dokumen kemudian dijumlahkan.
- iii. Hitung hasil kuadrat dari masing-masing term dalam setiap dokumen (termasuk D_1) kemudian jumlahkan lalu diakarkan.
- iv. Lakukan pembagian antara hasil dari langkah nomor 2 dengan langkah nomor 3. Maka, didapatkan nilai Cosine Similarity.

c. Menentukan Nilai Sentimen

Setelah diketahui jarak yang tertinggi sampai terendah, akan diambil sebanyak k data tertinggi. Dari k data tersebut akan dilihat nilai sentimen mana yang paling banyak muncul. Kelas sentimen yang paling banyak muncul kelas/nilai sentimen untuk data yang sedang dihitung.

3. HASIL DAN PEMBAHASAN

Proses analisis sentimen dalam penelitian ini dapat digambarkan dengan diagram sebagai berikut :



Gambar 4. Flowchart Sistem Analisis Sentimen

Misal dari tahap pengambilan data dari Twitter didapatkan data sebagai berikut :

Tabel 1. Data Tweet

No	Tweet	Sentimen
1	Anies optimis Partai Gerindra akan menang di Pilkada 2017.	?? (akan dicari)
2	Tokoh politik dari berbagai partai mengadakan rapat untuk membahas koalisi baru menjelang pilkada 2017 dan pilpres 2019.	Positif
3	Partai politik sudah tidak dapat dipercaya. Sebagian besar partai mengutamakan kepentingan partai daripada kebutuhan rakyat.	Negatif
4	PDIP memenangkan Pilkada 2012 karena figur Jokowi. Partai Gerindra berusaha menang pada 2017. Pertandingan 2 partai ini akan seru.	Positif
5	Mengejek Jokowi yang produk demokrasi rakyat, tapi menjilat AHY yang jelas2 produk dinasti politik, malumu ditaruh di mana kawan? #kamiahok	Negatif
6	Suap menyuap sudah lazim di negeri Ini. Pemilu ada suap. Pilkada juga suap. Mungkin pula saat Pilpres.	Negatif

Setelah melalui tahap *preprocessing*, akan diperoleh data sebagai berikut :

Tabel 2. Hasil Tahap Preprocessing

No	Tweet	Hasil
1	Anies optimis Partai Gerindra akan menang di Pilkada 2017.	anies optimis partai gerindra menang pilkada
2	Tokoh politik dari berbagai partai mengadakan rapat untuk membahas koalisi baru menjelang pilkada 2017 dan pilpres 2019.	tokoh politik partai rapat bahas koalisi jelang pilkada pilpres
3	Partai politik sudah tidak dapat dipercaya. Sebagian besar partai mengutamakan kepentingan partai daripada kebutuhan rakyat.	partai politik percaya partai utama penting partai butuh rakyat
4	PDIP memenangkan Pilkada 2012 karena figur Jokowi. Partai Gerindra berusaha menang pada 2017. Pertandingan 2 partai ini akan seru.	pdip menang pilkada figur jokowi partai gerindra usaha menang tanding partai seru
5	Mengejek Jokowi yang produk demokrasi rakyat, tapi menjilat AHY yang jelas2 produk dinasti politik, malumu ditaruh di mana kawan? #kamiahok	ejek jokowi produk demokrasi rakyat jilat ahy produk dinasti politik malu taruh kawan
6	Suap menyuap sudah lazim di negeri Ini. Pemilu ada suap. Pilkada juga suap. Mungkin pula saat Pilpres.	suap suap lazim negeri pemilu suap pilkada pilpres

Kemudian setiap term dari hasil preprocessing di atas akan dihitung bobotnya dengan TF-IDF. Berikut adalah hasilnya :

Tabel 3. Hasil Perhitungan TF-IDF

term	Tf						df	Idf	Bobot (W) = tf * idf					
	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆			D ₁	D ₂	D ₃	D ₄	D ₅	D ₆
anies	1						1	0,77815125	0,77815125	0	0	0	0	0
optimis	1						1	0,77815125	0,77815125	0	0	0	0	0
partai	1	1	3	2			4	0,176091259	0,176091259	0,528273777	0,352182518	0	0	0
gerindra	1			1			2	0,477121255	0,477121255	0	0	0,477121255	0	0
menang	1			2			2	0,477121255	0,477121255	0	0	0,954242509	0	0
pilkada	1	1		1		1	4	0,176091259	0,176091259	0	0	0,176091259	0	0,176091259
tokoh		1					1	0,77815125	0	0,77815125	0	0	0	0
politik		1	1		1		3	0,301029996	0	0,301029996	0,301029996	0	0,301029996	0
rapat		1					1	0,77815125	0	0,77815125	0	0	0	0
bahas		1					1	0,77815125	0	0,77815125	0	0	0	0
koalisi		1					1	0,77815125	0	0,77815125	0	0	0	0
jelang		1					1	0,77815125	0	0,77815125	0	0	0	0
pilpres		1				1	2	0,477121255	0	0,477121255	0	0	0	0,477121255
percaya			1				1	0,77815125	0	0	0,77815125	0	0	0
utama			1				1	0,77815125	0	0	0,77815125	0	0	0
penting			1				1	0,77815125	0	0	0,77815125	0	0	0
butuh			1				1	0,77815125	0	0	0,77815125	0	0	0
rakyat			1		1		2	0,477121255	0	0	0,477121255	0	0,477121255	0
pdip				1			1	0,77815125	0	0	0	0,77815125	0	0
figur				1			1	0,77815125	0	0	0	0,77815125	0	0
jokowi				1	1		2	0,477121255	0	0	0	0,477121255	0,477121255	0
usaha				1			1	0,77815125	0	0	0	0,77815125	0	0

tanding			1		1	0,77815125	0	0	0	0,77815125	0	0
---------	--	--	---	--	---	------------	---	---	---	------------	---	---

Tabel 3. Hasil Perhitungan TF-IDF (Lanjutan)

term	Tf						df	Idf	Bobot (W) = tf * idf					
	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆			D ₁	D ₂	D ₃	D ₄	D ₅	D ₆
seru				1			1	0,77815125	0	0	0	0,77815125	0	0
ejek					1		1	0,77815125	0	0	0	0	0,77815125	0
produk					2		1	0,77815125	0	0	0	0	1,556302501	0
demokrasi					1		1	0,77815125	0	0	0	0	0,77815125	0
jilat					1		1	0,77815125	0	0	0	0	0,77815125	0
ahy					1		1	0,77815125	0	0	0	0	0,77815125	0
dinasti					1		1	0,77815125	0	0	0	0	0,77815125	0
malu					1		1	0,77815125	0	0	0	0	0,77815125	0
taruh					1		1	0,77815125	0	0	0	0	0,77815125	0
kawan					1		1	0,77815125	0	0	0	0	0,77815125	0
suap						3	1	0,77815125	0	0	0	0	0	2,334453751
lazim						1	1	0,77815125	0	0	0	0	0	0,77815125
negeri						1	1	0,77815125	0	0	0	0	0	0,77815125
pemilu						1	1	0,77815125	0	0	0	0	0	0,77815125

$\sum W_{D1}^*$ WDn	0,06201626 3	0,09302439 5	0,77595847	0	0,03100813 2
------------------------	-----------------	-----------------	------------	---	-----------------

Tabel 5. Hasil Cosine Similarity (Langkah 3)

W_n^2					
W_{D1}^2	W_{D2}^2	W_{D3}^2	W_{D4}^2	W_{D5}^2	W_{D6}^2
0,60551 9	0	0	0	0	0
0,60551 9	0	0	0	0	0
0,03100 8	0,03100 8	0,27907 3	0,12403 3	0	0
0,22764 5	0	0	0,22764 5	0	0
0,22764 5	0	0	0,91057 9	0	0
0,03100 8	0,03100 8	0	0,03100 8	0	0,03100 8
0	0,60551 9	0	0	0	0
0	0,09061 9	0,09061 9	0	0,09061 9	0
0	0,60551 9	0	0	0	0
0	0,60551 9	0	0	0	0
0	0,60551 9	0	0	0	0
0	0,60551 9	0	0	0	0
0	0,22764 5	0	0	0	0,22764 5
0	0	0,60551 9	0	0	0
0	0	0,60551 9	0	0	0
0	0	0,60551 9	0	0	0
0	0	0,60551 9	0	0	0
0	0	0,22764 5	0	0,22764 5	0
0	0	0	0,60551 9	0	0
0	0	0	0,60551 9	0	0
0	0	0	0,22764 5	0,22764 5	0
0	0	0	0,60551 9	0	0
0	0	0	0,60551 9	0	0
0	0	0	0,60551 9	0	0
0	0	0	0,60551 9	0	0
0	0	0	0	0,60551 9	0
0	0	0	0	2,42207 7	0
0	0	0	0	0,60551 9	0
0	0	0	0	0,60551 9	0

	0	0	0	0	0,60551 9	0
	0	0	0	0	0,60551 9	0
	0	0	0	0	0,60551 9	0
	0	0	0	0	0,60551 9	0
	0	0	0	0	0,60551 9	0
	0	0	0	0	0	5,44967 4
	0	0	0	0	0	0,60551 9
	0	0	0	0	0	0,60551 9
	0	0	0	0	0	0,60551 9
$\Sigma Wn2$	1,72834 4	3,40787 7	3,01941 4	4,54850 6	7,81214 1	7,52488 5
$\sqrt{\Sigma Wn2}$	1,31466 5	1,84604 4	1,73764 6	2,13272 3	2,79502 1	2,74315 2

Tabel 6. Hasil Cosine Similarity (Langkah 4)

Cosine Similarity				
Cos (D ₁ ,D ₂)	Cos (D ₁ ,D ₃)	Cos (D ₁ ,D ₄)	Cos (D ₁ ,D ₅)	Cos (D ₁ ,D ₆)
0,02555338 9	0,04072118	0,27675085 2	0	0,008598259

Dari hasil tersebut dapat diurutkan nilai similarity D₁ dari yang tertinggi ke yang terendah yaitu :

1. D₄ (Positif)
2. D₃ (Negatif)
3. D₂ (Positif)
4. D₆ (Negatif)
5. D₅ (Negatif)

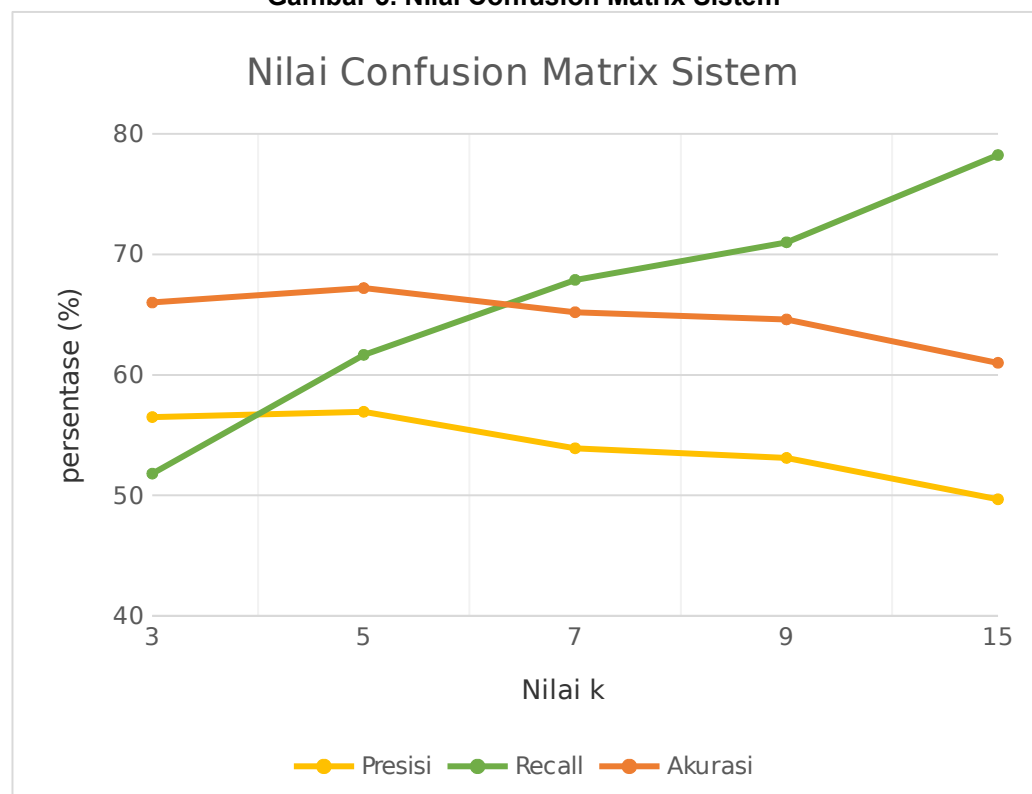
Jika dipilih nilai k untuk KNN adalah 3 maka akan dipilih 3 nilai *similarity* yang tertinggi. Dari 3 nilai tertinggi tersebut, kelas sentimen yang paling banyak muncul adalah positif sehingga sistem akan mengklasifikasikan D₁ ke dalam sentimen Positif.

4. KESIMPULAN

Dari hasil pengujian sistem didapatkan hasil sebagai berikut :

Data Rekap Akurasi Sistem			
Nilai k	Confusion Matrix		
	Akurasi	Presisi	Recall
3	66 %	56.5 %	51.81 %
5	67.2 %	56.94 %	61.66 %
7	65.2 %	53.91 %	67.88 %
9	64.6 %	53.11 %	71 %
15	61 %	49.68 %	78.24 %

Gambar 5. Nilai Confusion Matrix Sistem



Gambar 6. Grafik Nilai Confusion Matrix Sistem

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan bahwa penelitian tentang analisis sentimen pengguna Twitter terhadap topik Pilkada DKI 2017 dengan menggunakan metode *K Nearest Neighbor* telah berhasil dilakukan. Hasil akurasi terbesar adalah 67,2% dengan nilai $k=5$. Sedangkan nilai presisi tertinggi sebesar 56,94% saat $k=5$ dan *recall* terbesar 78,24 % ketika $k=15$.

DAFTAR PUSTAKA

- Bukhari, Varian Habbie. 2015. *Sentiment Analysis Menggunakan K-Nearest Neighbor dengan Perbandingan Fungsi Jarak (Studi Kasus : Twitter Indosat dan Telkomsel)*, Skripsi Universitas Widyatama Bandung.
- Ilyas, Husni. 2017. *Cosine Similarity Antar Dokumen: Sebuah Contoh*. <https://komputasi.files.wordpress.com/2011/01/cosine-similarity-antar-dokumen-sebuah-contoh.pdf>.
- Liu, B. (2012). *Sentiment Analysis and Subjectivity. Synthesis Lectures on Human Language Technologies*. USA: editor: Graeme Hirst Morgan & Claypool Publishers.
- Marpaung, Fransiska Humida. 2017. *Analisis Sentimen Opini Masyarakat Indonesia Mengenai Calon Gubernur DKI Jakarta 2017 Menggunakan Naïve Bayes Classifier*, Skripsi Universitas Atma Jaya Yogyakarta.
- Rosdiansyah, Defri. 2014. *Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor dan Pendekatan Lexicon*, Skripsi, Universitas Islam Negeri Sultan Syarif Kasim Pekanbaru.
- Saridewi, Ayu Indah. 2015. *Perancangan Dan Implementasi Sistem Peminatan Siswa Sma Dengan Algoritma C4.5 Pada Smk Harapan Denpasar*, Skripsi Universitas Udayana Bali.
- Taspinar, Ahmet dan Lasse Schuirmann. 2017. *Twitterscraper 0.2.7 : Python Package Index*. <https://pypi.python.org/pypi/twitterscraper/0.2.7>
- Vinodhini, G., & Chandrasekaran, RM., 2012. *Sentiment Analysis and Opinion Mining: A Survey*, International Journal of Advanced Research in Computer Science and Software Engineering Vol. 2, Issue 6, pp. 1