

ANALISIS KONTEN JEJARING SOSIAL TWITTER DALAM KASUS PEMILIHAN GUBERNUR DKI 2017

Muhammad Habibi

Program Studi Teknik Informatika
Sekolah Tinggi Manajemen Informatika dan Komputer
Jenderal Achmad Yani

habibi@stmikayani.ac.id

Abstrak

Pilkada DKI merupakan salah satu pilkada yang menjadi sorotan di Indonesia, hal ini disebabkan DKI Jakarta merupakan Ibu kota negara Indonesia. Perbincangan mengenai pilkada DKI Jakarta serta nama-nama calonnya tidak hanya terasa di dunia nyata bahkan sangat terasa di dunia maya. Masyarakat bebas berpendapat atau beropini mengenai calon gubernur DKI Jakarta 2017. Twitter merupakan salah satu platform media sosial yang memiliki basis pengguna yang sangat besar di Indonesia. Dalam penelitian ini pengumpulan dan analisis sederhana terhadap tweet yang terkait dengan calon gubernur dalam pilkada DKI Jakarta 2017. Tercatat, tidak kurang dari 79.856 tweet berhasil dikumpulkan dalam kurun waktu 7 hari, mulai dari tanggal 31 Januari sampai 6 Februari 2017. Hasil dari penelitian ini adalah mengetahui frekuensi kemunculan kata pada tweet dan membahas mengenai analisis co-occurrence, untuk mengetahui pasangan kata yang sering muncul dalam satu tweet. Serta mencari keterkaitan antara satu dengan kata yang lain.

Kata Kunci: Media Sosial, Twitter, Pilkada DKI Jakarta 2017, Analisis Konten, Text Mining

1. Pendahuluan

Pemilihan kepala daerah (Pilkada) merupakan agenda lima tahunan suatu daerah dalam menentukan kepala daerah dan wakil kepala daerah. Pilkada DKI merupakan salah satu pilkada yang menjadi sorotan di Indonesia, Hal ini disebabkan DKI Jakarta merupakan Ibu kota negara Indonesia. Perbincangan mengenai pilkada DKI Jakarta serta nama-nama calonnya tidak hanya terasa di dunia nyata bahkan sangat terasa di dunia maya. Masyarakat bebas berpendapat atau beropini mengenai calon Gubernur DKI Jakarta 2017.

Twitter merupakan salah satu media sosial yang banyak digunakan di Indonesia. Tercatat pada tahun 2015, Twitter memiliki sekitar 50 juta pengguna aktif di Indonesia¹. Hal ini menempatkan Indonesia pada peringkat ke-5

¹. <http://tekno.kompas.com/read/2015/03/26/16465417/Pengguna.Twitter.di.Indonesia.Capa.i.50.Juta> [diakses tanggal 22 Oktober 2017]

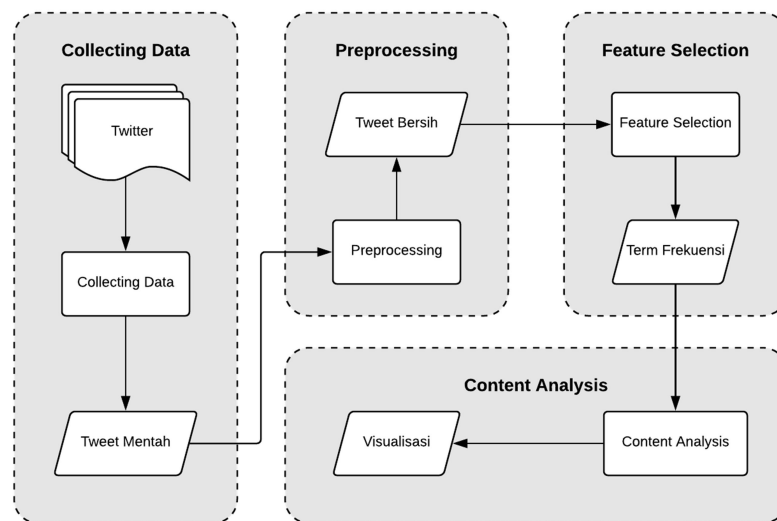
pengguna Twitter terbesar di dunia setelah USA, Brazil, Jepang dan Inggris (Takhyetev, dkk., 2012). Dalam beberapa tahun terakhir sejak kemunculannya di tahun 2006, Twitter menjadi salah satu platform jejaring media dengan perkembangan yang paling pesat. Lebih dari 75% perusahaan yang masuk dalam The Fortune Global 100 memiliki satu atau lebih akun Twitter korporasi maupun untuk branding produk secara spesifik (Malhotra, dkk., 2012).

Para pendukung sebuah pasangan calon gubernur atau kepala daerah biasanya gencar mengkampanyekan calonnya di media sosial, khususnya Twitter. Namun tidak hanya para pendukung, pihak oposisi biasanya juga sering meluapkan ketidak sukaan mereka pada sebuah pasangan calon di media sosial. Media sosial khususnya Twitter sekarang ini menjadi salah satu tempat promosi atau kampanye yang efektif dan efisien.

Analisis konten dapat digunakan untuk menginterpretasikan struktur isi pesan Twitter secara kualitatif terkait topik pilkada gubernur DKI 2017. Struktur dimaksud adalah membahas mengenai hasil dari frekuensi kemunculan kata pada *tweet* yang diunggah dan membahas mengenai analisis *co-occurrence*, untuk mengetahui pasangan kata yang sering muncul dalam satu *tweet*. Serta mencari keterkaitan antara satu dengan kata yang lain.

2. Metode Penelitian

Penelitian ini dilaksanakan dalam empat tahapan. Tahapan tersebut diantaranya *collecting data*, *preprocessing*, *feature selection* dan *content analysis*. Tahapan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 *Collecting Data*

Tahap pertama pada penelitian ini yaitu *Collecting data*, *tweet* dikumpulkan dengan memanfaatkan fasilitas API (*Application Programming Interface*) yang disediakan oleh Twitter. Proses pengumpulan data dimulai dari tanggal 31 Januari 2017 sampai 6 Februari 2017. Pengumpulan data *tweet* dilakukan sebelum pemilihan putaran pertama gubernur DKI Jakarta yang dilaksanakan pada tanggal 15 Februari 2017. *Tweet* yang diambil adalah *tweet* yang mengandung kata-kata “@basuki_btp”, “@aniesbaswedan” dan “@agusyudhoyono”. Dalam masa 7 hari tersebut terkumpul kurang lebih 79.856 *tweet*.

2.2 *Preprocessing*

Preprocessing data merupakan proses dimana teks yang akan digunakan dibersihkan dan dipersiapkan terlebih dahulu sebelum teks dianalisis (Haddi et. al, 2013). *Preprocessing* terhadap data dilakukan untuk menghindari adanya data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Hemalatha, dkk., 2012). Adapun tahapan *preprocessing* yang akan dilakukan dalam penelitian ini diantaranya:

1. **Menghapus URL.** URL (<http://www.situs.com>) dan email (nama@situs.com) dihapus karena tidak digunakan dalam analisis teks.
2. **Menghapus tanda *retweet* pada Twitter.** Proses ini dilakukan dengan menghapus karakter khusus Twitter yakni tanda *retweet* (RT)
3. **Menghapus simbol, *emoticon* dan tanda baca.** Proses ini berfungsi untuk menghapus tanda-tanda baca serta *emoticon* yang pada teks.
4. **Tokenisasi.** Tokenisasi merupakan proses penguraian kata-kata dari sebuah kalimat. Kata-kata tersebut pada proses analisis teks diperlakukan sebagai sebuah entitas tunggal.
5. ***Case folding*.** Proses ini berfungsi untuk menyamaratakan karakter huruf yaitu mengkonversi huruf kapital atau huruf besar menjadi huruf kecil.
6. **Penghapusan angka.** Proses ini berfungsi untuk menghapus karakter angka pada teks. Angka dihapus karena tidak termasuk ke dalam komponen analisis teks.
7. ***Slang Word*.** Dilakukan perubahan *slang word* yaitu proses merubah atau mencari padanan kata.

8. **Menghapus *stopword*.** *Stopword* atau kata penghubung dalam analisis teks tidak diperlukan oleh sebab itu perlu dihapus. Pada proses ini kata-kata yang tidak digunakan dalam analisis dihapus. Tidak hanya kata penghubung, kata-kata yang yang tidak perlu dianalisis dapat dihapus.

2.3 Feature Selection

Feature selection merupakan tahapan penting dalam text mining. Salah satu fungsi penting yang disediakan oleh proses ini adalah untuk dapat memilih term atau kata apa saja yang dapat dijadikan sebagai wakil penting untuk kumpulan dokumen yang akan kita analisis (Ma'arif, 2016). Dalam penelitian ini, *feature* yang digunakan adalah *term frequency*.

Term frequency adalah gagasan standar frekuensi dalam *corpus-based* pengolahan bahasa alami (Yamamoto dan Church, 2001). *Term frequency* menghitung kemunculan suatu *term* dalam suatu *corpus* berdasarkan bobot suatu *term* pada dokumen. Berikut merupakan contoh *tweet* bersih, seperti yang terlihat pada Tabel 1.

Tabel 1. Contoh *tweet*

Dokumen	Tweet
D1	pilih pasti melayani warga jakarta @basuki_btp
D2	ini dia gubernur wakil gubernur baru jakarta @aniesbaswedan @sandiuno
D3	satu doa untuk jakarta bersama habaib kyai ulama @agusyudhoyono #agussylvimenang

Berdasarkan data *tweet* pada Tabel 1 didapatkan *corpus* kata unik sebanyak 22 kata unik. Misalkan untuk menghitung kemunculan *term* 'jakarta', maka dilihat berapa kali kemunculan *term* tersebut pada semua dokumen. *Term* 'jakarta' muncul pada dokumen D1, D2 dan D3. Maka total kemunculan untuk *term* 'jakarta' adalah 3. Daftar *term* frekuensi dapat dilihat pada Tabel 2.

Tabel 2. *Term frequency*

NO	Term	Term Frequency			Total Kemunculan
		D1	D2	D3	
1	#agussylvimenang	0	0	1	1
2	@agusyudhoyono	0	0	1	1
3	@aniesbaswedan	0	1	0	1

NO	Term	Term Frequency			Total Kemunculan
		D1	D2	D3	
4	baru	0	1	0	1
5	@basuki_btp	1	0	0	1
6	bersama	0	0	1	1
7	dia	0	1	0	1
8	doa	0	0	1	1
9	gubernur	0	2	0	1
10	habaib	0	0	1	1
11	ini	0	1	0	1
12	jakarta	1	1	1	3
13	kyai	0	0	1	1
14	melayani	1	0	0	1
15	pasti	1	0	0	1
16	pilih	1	0	0	1
17	@sandiuono	0	1	0	1
18	satu	0	0	1	1
19	ulama	0	0	1	1
20	untuk	0	0	1	1
21	wakil	0	1	0	1
22	warga	1	0	0	1

2.4 Content Analysis

Langkah berikutnya adalah melakukan analisis *cocurrence*. Analisis *co-cocurrence* dilakukan untuk mengetahui pasangan kata yang sering muncul dalam satu *tweet*. Serta mencari keterkaitan antara satu dengan kata yang lain. Yaitu kata-kata apa saja yang sering digunakan secara bersama-sama dalam satu *tweet*.

3. Hasil dan Pembahasan

Pada bagian ini dibagi menjadi tiga bagian pembahasan. Bagian pertama membahas mengenai hasil dari frekuensi kemunculan kata pada *tweet* yang diunggah. Bagian kedua membahas mengenai analisis *co-cocurrence* dilakukan untuk mengetahui pasangan kata yang sering muncul dalam satu *tweet*. Serta mencari keterkaitan antara satu dengan kata yang lain. Yaitu kata-kata apa saja yang sering digunakan secara bersama-sama dalam satu *tweet*.

2.1 Hasil Term Frequency

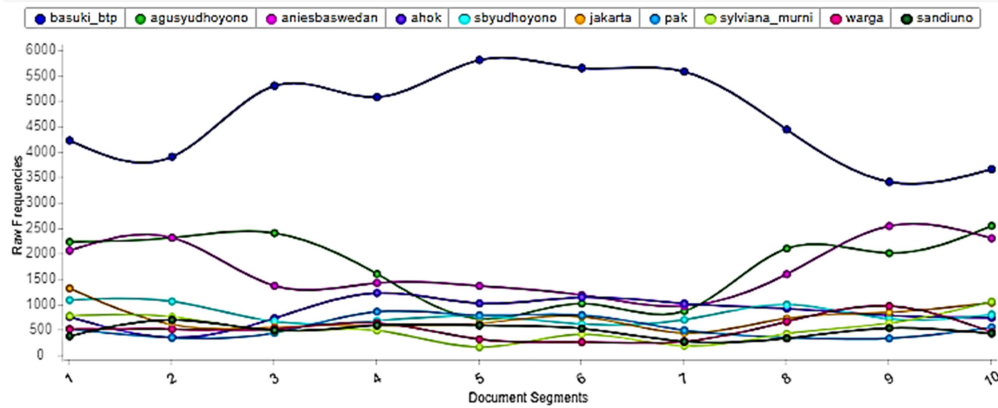
Hasil frekuensi kemunculan kata pada *tweet* yang diunggah oleh netizen terkait pilkada DKI Jakarta 2017, dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Term Frequency*

No	Term	Count
1	basuki_btp	46996
2	agusyudhoyono	17673
3	aniesbaswedan	16987
4	ahok	8526
5	sbyudhoyono	7933
6	jakarta	7331
7	pak	5305
8	sylviana_murni	5218
9	warga	4948
10	sandiuno	4671
11	konsergue	4449
12	ma	4439
13	amin	4240
14	gunromli	3643
15	jokowi	3541

Tabel 3. menampilkan 15 daftar kata yang sering muncul dalam *tweet*. Dari 79.856 *tweet* yang dikumpulkan, kata “basuki_btp” merupakan kata yang paling sering muncul. Yaitu dengan jumlah kemunculan sebanyak 46996 kali. Disusul dengan kata “agusyudhoyono”, “aniesbaswedan”, “ahok”, “sbyudhoyono” dan “jakarta”.

Pada data *tweet* yang berjumlah 79.856 *tweet*, dibagi ke dalam 10 *segment* dengan jumlah yang sama. Kemudian dilakukan proses perhitungan kemunculan kata, didapatkan bahwa trend kemunculan kata paling tinggi tiap *segment tweet* adalah kata “basuki_btp”. Hal ini dapat dilihat pada Gambar 2 di bawah ini.



Gambar 2. Grafik Trend kemunculan kata

2.2 Analisis Teks pada Tweet

Analisis teks *tweet* yang dilakukan adalah analisis *frequency* kemunculan kata. Dari setiap *tweet* yang diunggah oleh setiap akun Twitter yang mengandung kata-kata “@basuki_btp”, “@aniesbaswedan” atau “@agusyudhoyono”, ternyata kata “basuki_btp” menjadi kata yang paling sering digunakan oleh user Twitter dalam setiap *tweet*. Kata-kata yang sering muncul diilustrasikan pada Gambar 3.



Gambar 3. Visualisasi kata-kata yang paling sering muncul dalam Tweet

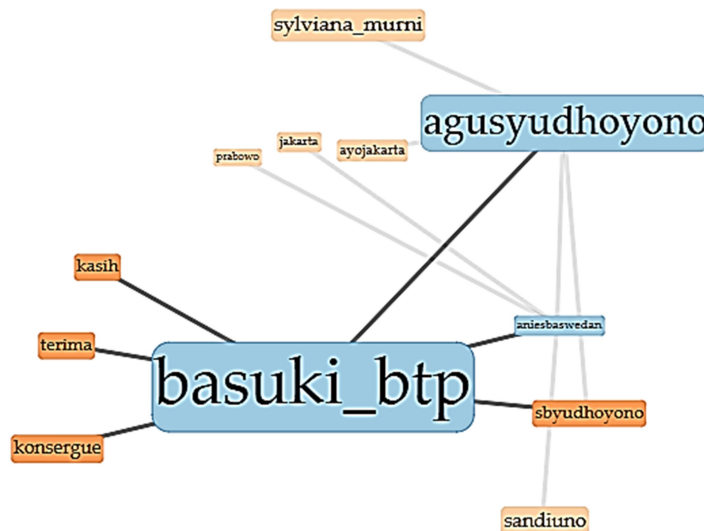
Pada Gambar 3, semakin besar ukuran huruf, maka berarti semakin sering kata-kata yang bersangkutan muncul pada *tweet* yang berhasil dikumpulkan. Dari Gambar 3, kata-kata “basuki_btp” menjadi kata yang sering muncul dalam *tweet* yang diunggah oleh netizen. Hal ini menggambarkan bahwa

akun Twitter “@basuki_btp” milik Basuki Tjahaja Purnama menjadi akun Twitter calon Gubernur DKI Jakarta yang sering di mention atau dipanggil oleh netizen dalam *tweet* menjelang Pemilihan Gubernur DKI Jakarta 2017. Sementara itu akun Twitter “@agusyudhoyono” milik calon gubernur DKI Jakarta Agus Harimurti Yudhoyono menjadi urutan yang kedua. Diurutan yang ketiga, akun Twitter “@aniesbaswedan” milik calon gubernur DKI Jakarta Anies Baswedan.

Kata-kata lain yang sering muncul adalah “jakarta” dan “warga” yang mana hal ini erat kaitannya dengan lokasi kota pemilihan gubernur yaitu di Jakarta, serta warga masyarakat jakarta yang terlibat dalam pemilihan gubernur. Kata-kata “konsergue” menjadi kata yang cukup sering muncul juga dalam *tweet*. Konser “Gue 2” merupakan konser para seniman dan selebriti pendukung pasangan Ahok dan calon wakilnya, Djarot Saiful Hidayat, pada Pemilihan Kepala Daerah (Pilkada) DKI Jakarta 2017.

2.3 Analiss Keterkaitan Kata

Melakukan analisis konten hanya berdasarkan frekuensi kemunculan kata tentulah tidak akurat karena bisa memunculkan keambiguan. Untuk melengkapi analisis konten teks *tweet*, maka dilakukan analisis yang kedua yakni analisis kata-kata yang muncul bersama dalam satu *tweet*. Hasil dari analisis ini direpresentasikan dalam bentuk graf, seperti yang diilustrasikan pada Gambar 4.



Gambar 4. Relasi *co-occurrence* antar kata yang sering muncul dalam *tweet*

Seperti yang diilustrasikan pada Gambar 4, kata “basuki_btp” yang merupakan kata yang paling sering muncul memiliki *co-occurrence* dengan beberapa kata yang juga memiliki frekuensi kemunculan yang tinggi yakni “konsergue”, “terima”, “kasih”, dan “jakarta”. Hal tersebut berarti dalam satu *tweet*, kata-kata “basuki_btp” acap kali berdampingan dengan kata-kata tersebut. Pada Tabel 4 menunjukkan beberapa contoh *tweet* yang mengandung kata-kata “basuki_btp” dengan kata-kata yang muncul bersamaan dengan kata “basuki_btp” tersebut.

Tabel 4. Contoh *tweet* yang mengandung pasangan kata tertentu

Pasangan Kata	Contoh <i>Tweet</i>	Jumlah <i>Tweet</i>
basuki_btp - jakarta	bisabisa jakarta tidak karuan tidak dipegang @basuki_btp #badjamelaju	4422
basuki_btp - konsergue	#konsergue menjadi salah satu bukti @basuki_btp masih layak memimpin #dkijakarta #periode	4301
basuki_btp - terima	@basuki_btp terima kasih atas dukungan besar sukarela keberanian bersuara para musisi pekerja seni bagi pesta demokrasi	3262

4. Penutup

Dalam penelitian ini, telah berhasil dianalisis dan dirangkum sejumlah besar pesan dalam bentuk *tweet* terkait dengan calon Gubernur pada kasus Pilkada DKI Jakarta 2017. Analisis teks *tweet* yang dilakukan berhasil mengidentifikasi bahwa akun Twitter yang paling sering di-*mention* dalam *tweet* terkait dengan dengan calon gubernur pada kasus pilkada DKI Jakarta 2017 adalah “@basuki_btp”. Disusul dengan akun “@agusyudhoyono” di urutan kedua, dan akun Twitter “@aniesbaswedan” di urutan ketiga.

Analisis keterkaitan kata yang dilakukan berhasil menunjukkan bahwa kata “basuki_btp” yang merupakan kata yang paling sering muncul memiliki *co-occurrence* dengan beberapa kata yang juga memiliki frekuensi kemunculan yang tinggi yakni “konsergue”, “terima”, “kasih”, dan “jakarta”.

Daftar Pustaka

- Haddi, E., Liu, X., dan Shi, Y., 2013. The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science* 17. (2013). 26 – 32.
- Hemalatha, I., Varma, P.G., dan Govardhan, A., 2012. Preprocessing the Informal Text for Efficient Sentiment Analysis. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*. Vol. 1. July – August 2012. ISSN 2278-6856.
- Ma'arif, M.R., 2016. Perbandingan Naïve Bayes Classifier dan Support Vector Machine untuk Klasifikasi Judul Artikel. *Jurnal Informatika Sunan Kalijaga (JISKA)*. Vol 1. No.2. September 2016. 90 – 93. ISSN 2527 - 5836
- Malhotra, A., Malhotra, C. K., & See, A. 2012. How to get your messages retweeted. *MIT Sloan Management Review*. 53(2). 61.
- Takhteyev, Y., Gruzd, A., & Wellman, B. 2012. Geography of Twitter networks. *Social networks*. 34(1). 73-81.
- Yamamoto, M., dan Church, K.W., 2001. Using Suffix Arrays to Compute Term frequency and Document Frequency for All Substrings in A Corpus. *Computational Linguistics*. 27(1). 1-30.