

SENTIMEN ANALISIS DI TWITTER (STUDI KASUS : RESEPSI PERNIKAHAN PUTRI PRESIDEN RI KE-7)

Lasmedi Afuan

Fakultas Teknik, Jurusan/Prodi Teknik Informatika, Universitas Jenderal Soedirman
Jl Mayjen Sungkono KM 5 Blater Purbalingga
Universitas Jenderal Soedirman
Email : Lasmedi.afuan@unsoed.ac.id

Abstrak

Penelitian ini membahas mengenai analisis sentimen di Twitter terkait dengan resepsi pernikahan putri Presiden Joko Widodo. Penelitian ini diharapkan dapat digunakan untuk membantu melakukan riset atas sentimen masyarakat yang mengandung sentimen positif, netral atau negatif. Data yang digunakan adalah data tweet dari sosial media twitter dengan menggunakan hastag 'JokowiMantu' yang diambil dari twitter menggunakan Tweepy pada tanggal 08 November 2017. Metode yang digunakan dalam penelitian ini, untuk preprocessing data menggunakan normalisasi kalimat dan tokenisasi. Untuk proses analisis sentimennya menggunakan pendekatan Lexicon Based. Hasil dari penelitian ini adalah 43 % netizen memberikan opini yang positif, 48% netral dan 9 % negatif.

Kata Kunci : sentimen analisis, twiter, tweepy, lexicon based

Pendahuluan

Pada 08 November 2017 lalu merupakan hari yang berbahagia bagi Presiden RI Bapak Jokowi Widodo, tepat pada tanggal tersebut Presiden ke-7 RI yang akrab disapa dengan Jokowi melaksanakan acara resepsi pernikahan putrinya Kahiyang Ayu dengan M. Bobby Nasution. Sebagai orang nomor satu di Republik ini pastinya banyak sekali rakyat Indonesia yang memberikan komentarnya mengenai acara tersebut, tidak terkecuali para netizen. Para netizen di media sosial (salah satunya Twitter) turut berkomentar

Sebagai salah satu media sosial yang memiliki banyak pengguna, Twitter tidak hanya digunakan sebagai media untuk aktualisasi diri dan bersosialisasi, tetapi juga sebagai sarana mengekspresikan opini, pengalaman mengenai kejadian yang sedang hangat dibicarakan (*trending topic*) yang sering kali disebut

dengan sentimen. Pernikahan Kahiyang Ayu dan M. Bobby Nasution sempat menjadi *trending topic* di Twitter tidak luput dari sentimen netizen.

Pada makalah ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen. Dimulai dari tahapan pengumpulan data, preprocessing, kemudian tahap analisis sentimen dengan menggunakan metode *Naïve Bayes Classifier*.

Tinjauan Pustaka

Penelitian terkait sentimen analisis telah dilakukan oleh para peneliti diantaranya penelitian yang dilakukan oleh [1] mengklasifikasikan status mahasiswa di Twitter menggunakan metode SVM dan *Maximum Entropy*, dengan nilai akurasi 93% untuk SVM dan 70% untuk *Maximum Entropy*. Penelitian[2] melakukan sentimen analisis terhadap harga bahan pokok, dengan mengklasifikasikan sentimen menjadi 2 kelompok yaitu positif dan negatif, sentimen diukur menggunakan *Chi Square* dan Prosedur Marascuilo. Analisis sentimen juga dilakukan oleh[3][4] dengan menggunakan metode SVM dan *Backpropagation*. Sedangkan [5] menggunakan pendekatan *Lexicon Based*. Penelitian [6] membandingkan *Naïve Bayes Classifier* (NBC) dengan SVM untuk sentimen analisis pada *tweet* terkait calon Gubernur DKI Jakarta.

Metodologi Penelitian

Tahapan – tahapan pada penelitian adalah sebagai berikut:

1. Pengumpulan data

Data yang digunakan adalah data *tweet* dari media sosial twitter. Data diambil dengan melakukan *streaming* dengan Tweepy pada Twitter yang dilakukan pada tanggal 08 November 2017, dimana pada tanggal tersebut Presiden Jokowi melaksanakan resepsi pernikahan putrinya yaitu Kahiyang Ayu dan M.

Bobby Nasution. Data tersebut diambil secara *random* dari pengguna biasa atau media online di twitter. Hasil dari proses ini adalah daftar opini pembaca serta metadata seperti: nama user dan waktu.

2. Mengambil Kamus sentimen

Pada proses ini dilakukan pengambilan kata kunci sentimen dari kamus. Kamus yang digunakan dalam sistem ini adalah kata kunci sentimen positif (*positif keywords*), kata kunci sentimen negatif (*negatif keywords*).

3. Preprocessing

Preprocessing bertujuan untuk menyiapkan kalimat sebelum dilakukan ekstraksi kata kunci dan penentuan sentimen. Proses yang dilakukan antara lain:

1. Normalisasi kalimat Bertujuan untuk menormalkan kalimat sehingga kalimat gaul menjadi normal.
2. Tokenisasi Bertujuan untuk memecah kalimat menjadi token-token yang merupakan kandidat *keyword*.

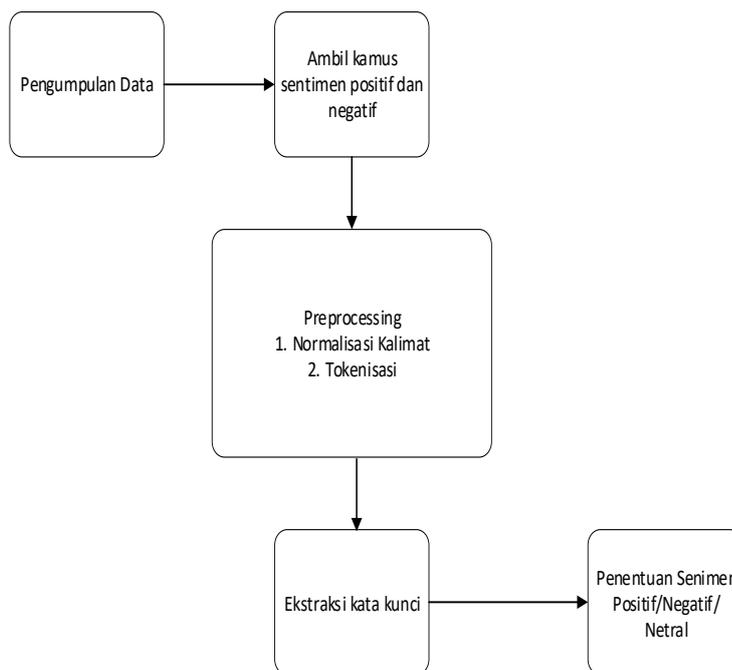
4. Ekstraksi Kata Kunci

Proses ini bertujuan untuk mengekstraksi kata kunci penentu sentimen positif dan negatif.

5. Penentuan sentimen

Proses ini bertujuan untuk menentukan sentimen suatu kalimat opini, penentuan sentimen dilakukan dengan menghitung probabilitas kemunculan kata kunci positif dan kata kunci negatif.

Secara umum, metode penelitian disajikan pada gambar 1.



Gambar 1. Metodologi penelitian

Pembahasan

Dataset

Dataset yang digunakan adalah opini *netizen* terkait dengan resepsi pernikahan Kahiyang Ayu dan M. Bobby Nasution dengan menggunakan *hashtag* #JokowiMantu. Dataset yang berhasil dikumpulkan sebanyak 100 *tweet*. Contoh dataset *tweet* disajikan pada Tabel 1.

Tabel 1. Contoh Dataset *tweet* di Twitter

No	Tweet
1	RT @okezonenews: Soal Pembatasan Jumlah Undangan di Hajatan Pejabat Negara, Risma Bela Jokowi https://t.co/DprZoCbSJ8 #JokowiMantu
2	RT @RadioElshinta: Selamat ya mba Kahiyang dan mas Bobby "Semoga menjadi keluarga Sakinah Mawadah Warahmah." Warganet yg mau kasih ucaâ€!
3	RT @Metro_TV: Detik-detik ijab kabul. Selamat menempuh hidup baru Kahiyang Ayu-Bobby Nasution #JokowiMantu https://t.co/Z1yG1PUtwa
4	Selamat utk mbak Kahiyang dan mas Bobby, semoga langgeng sampe kakek nenek #JokowiMantu https://t.co/gaMiWqigko

5	RT @KicauSunyi: Kasihan warga Solo... mereka ga kebagian antri untuk memberi selamat ke kedua mempelai... karena "kuota" mereka "diâ€¦
---	--

Kamus

Kamus adalah komponen penting dalam sistem yang menggunakan pendekatan *lexicon-based*. Kamus digunakan dalam proses normalisasi kalimat dan ekstraksi kata kunci [6]. Contoh kata kunci sentimen positif dan negatif disajikan pada Tabel 2 dan Tabel 3.

Tabel 2. Kata kunci sentimen positif

No	Kata kunci
1	Selamat
2	Anggun
3	Bagus
4	Bahagia
5	Bergembira sekali
6	Bergembira

Tabel 3. Kata kunci sentimen negatif

No	Kata kunci
1	Aneh
2	Antipati
3	Bahaya
4	Belum dikonfirmasi
5	Belum pasti

Preprocessing

Proses yang dilakukan untuk menormalisasi kalimat, tahapan yang dilakukan pada *preprocessing* adalah:

1. Mengkonversi *tweet* menjadi huruf kecil semua
2. Menghapus tanda baca (*punctuation*) dan simbol selain alphabet. Tujuannya adalah agar tanda baca dan simbol selain alphabet tidak masuk menjadi token pada saat proses tokenisasi.
3. Normalisasi kata. Untuk proses normalisasi kata dilakukan dengan mentransformasi *tweet* yang mengandung kata-kata dengan bahasa gaul atau singkatan menjadi bentuk aslinya. Aturan normalisasi disajikan pada Tabel 4 [5].

Tabel 4. Aturan normalisasi

Tidak normal/gaul	Normal
Akhiran -ny	Akhiran -nya
Akhiran -nk	Akhiran -ng
Akhiran -x	Akhiran -nya
Akhiran -z	Akhiran -s
Akhiran -dh	Akhiran -t

Hasil dari *tweet* yang telah dilakukan preprocessing disajikan pada Tabel 5.

Tabel 5. *Tweet* yang sudah dilakukan preprocessing

No	Tweet
1	soal pembatasan jumlah undangan di hajatan pejabat negara risma bela jokowi
2	selamat ya mba kahiyang dan mas bobby semoga menjadi keluarga sakinah mawadah warahmah warganet yang mau kasih ucapan
3	Detik detik ijab kabul selamat menempuh hidup baru kahiyang ayu bobby nasution
4	selamat untuk mbak kahiyang dan mas bobby semoga langgeng sampai kakek nenek
5	kasihan warga solo mereka tidak kebagian antri untuk memberi selamat ke kedua mempelai karena kuota mereka

Tokenisasi

Data *tweet* yang telah dilakukan *preprocessing*, selanjutnya dilakukan tokenisasi dengan memisahkan token-token dalam kalimat menggunakan pemisah spasi. Penelitian ini menggunakan 3 jenis token antara lain:

1. Unigram, token yang terdiri dari 1 kata. Misal: Menempuh.
2. Bigram, token yang terdiri dari 2 kata. Misal: Menempuh hidup.
3. Trigram, token yang terdiri dari 3 kata. Misal: Menempuh hidup baru.

Aturan yang digunakan untuk membentuk ketiga jenis token tersebut adalah dengan overlapping token. Berikut ilustrasi pembentukan token-token tersebut:

Opini: soal pembatasan jumlah undangan di hajatan pejabat negara risma bela Jokowi

Unigram	Soal, pembatasan, jumlah, undangan, di, hajatan, pejabat, negara, risma, bela, Jokowi
Bigram	Soal pembatasan, pembatasan jumlah, jumlah undangan, undangan di, di hajatan, hajatan pejabat, pejabat negara, negara risma, risma bela, bela Jokowi
Trigram	soal pembatasan jumlah, pembatasan jumlah undangan, jumlah undangan di, undangan di hajatan, di hajatan pejabat, hajatan pejabat negara, pejabat negara risma, negara risma bela, risma bela Jokowi

Tujuan tokenisasi dengan menggunakan 3 jenis token ini adalah karena pada bahasa Indonesia frasa tidak hanya terdiri dari satu kata saja, bisa dua atau tiga kata yang memiliki satu kesatuan arti.

Ekstraksi Kata kunci

Tahap selanjutnya setelah dilakukan tokenisasi (unigram, bigram, dan trigram) adalah ekstraksi kata kunci. Tiga jenis token yang dihasilkan dicocokkan dengan kamus sentimen positif dan negatif.

Penentuan Sentimen

Setelah dilakukan ekstraksi kata kunci pada ketiga jenis token yang memiliki nilai sentimen, langkah berikutnya menghitung probabilitas kemunculan sentimen positif dan negatif mana yang lebih dominan.

Jika nilai sentimen positif lebih dominan maka nilai sentimen untuk *tweet* tersebut adalah positif, namun jika nilai sentimen negatif lebih dominan maka nilai sentimen untuk *tweet* tersebut adalah negatif, namun jika nilainya sama antara sentimen negatif dan sentimen positif maka nilai sentimen untuk tersebut adalah netral. Persamaan yang digunakan untuk menentukan nilai sentimen seperti berikut ini.

$$\text{Nilai sentimen} = \begin{cases} 1, & \sum \text{sentimen} > 0 \\ 0, & \sum \text{sentimen} = 0 \\ -1, & \sum \text{sentimen} < 0 \end{cases}$$

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka didapatkan kesimpulan sebagai berikut:

1. Hasil dari penelitian ini adalah 43 % netizen memberikan opini yang positif, 48% netral dan 9 % negatif dari 100 *tweet* yang dianalisis.
2. Penentuan sentimen dengan *lexicon based* dipengaruhi oleh kelengkapan kamus yang digunakan, semakin lengkap kamus kata kunci positif dan negatif maka semakin baik dalam penentuan sentimen.

Daftar Pustaka

[1] T. H. Susilo and S. Rochimah, "Pengklasifikasian topik dan analisis sentimen dalam media sosial," in SNASTI, 2013.

[2] M. S. Naffisah and I. Surjandari, "Penggunaan Text Mining pada Analisis Sentimen Masyarakat terhadap Perubahan Harga Bahan Pokok melalui Twitter," pp. 1–20, 2014.

[3] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," e-Proceeding Eng., vol. 2, no. 1, pp. 1177–1183, 2015.

[4] R. Habibi and D. B. Setyohadi, "Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation," Informatika, vol. 12, no. 1, pp. 103–109, 2016.

[5] A. Nurfalah and A. A. Suryani, "Analisis Sentimen Berbahasa Indonesia dengan Pendekatan Lexicon-Based Pada Media Sosial," J. Masy. Inform. Indones., vol. 2, no. 1, pp. 1–8, 2017.

[6] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," Integer J., vol. 1, no. 1, pp. 32–41, 2017.

Biodata Penulis

Lasmedi Afuan, memperoleh gelar Sarjana Teknik Informatika (S.T), Jurusan Teknik Informatika Universitas Islam Indonesia, lulus tahun 2007. Memperoleh gelar Magister of Computer Science (M.Cs) Program Pasca Sarjana Ilmu Komputer

Universitas Gajah Mada Yogyakarta, lulus tahun 2011.
Saat ini menjadi staff pengajar di Jurusan Teknik
Informatika UNSOED dan sedang melanjutkan studi
S3 di Ilmu Komputer UGM tahun 2015 hingga
sekarang.

