

MODEL ANALISIS SENTIMENT DI PILKADA JAWA BARAT 2018

Munawar, Abdul Rahim
Teknik Informatika, Universitas Esa Unggul
Jl. Arjuna Utara No. 9 Kebon Jeruk, Jakarta 11510
munawar@esaunggul.ac.id

Abstrak

Nowadays, social media is an integral part of personal life. Business, social life and political preferences are now commonly expressed through social media. Therefore, it can be said that social media are representation of netizen's opinion. One of the famous social media in Indonesia is Twitter. Sentiment analysis is a part of text mining with its focus on document / text analysis. This study maps the sentiment analysis in the 2018 West Java Regional Election using the Naive Bayes Classifier method. The first step in this research is collecting Twitter data via API using the keyword around the 2018 West Java Regional Election. Furthermore, data pre-processing and data analysis are carried out. Data collected from Twitter from June 20 2016 to July 20, 2018. There are 11,527 tweets. Based on the test results, it was found that the average accuracy rate for Rindu was 95.00%, Dedy-Dedi 90.00%, Hasanah 84.17% and Asyik 74.08%. Based on data analysis, it can be concluded that Rindu get the highest positive response. However, there is no correlation between the number of positive responses from Twitter and the results of the KPU, because there are many Twitter opinion data whose contents are the same, RT (retweet) results from other users, almost the same time for tweets, and unusual user names.

Keywords : Sentiment analysis, twitter, naive bayes classifier

Media sosial saat ini sudah menjadi bagian yang tidak terpisahkan dari kehidupan pribadi. Bisnis, kehidupan sosial dan pandangan politik kini umum diekspresikan di media sosial. Oleh karenanya tidak berlebihan kiranya jika media sosial bisa merepresentasikan opini pemiliknnya. *Twitter* adalah salah satu media sosial yang cukup terkenal di Indonesia. Analisis sentimen merupakan cabang dari *text mining*, dengan fokus utamanya adalah analisis dokumen / teks. Penelitian ini memetakan analisis sentimen di Pilkada Jawa Barat 2018 menggunakan metode *Naive Bayes Classifier*. Langkah pertama pada penelitian ini adalah pengumpulan data *Twitter* via API dengan keyword seputar Pilkada Jawa Barat 2018. Selanjutnya dilakukan *pre-processing* data dan analisis data. Data yang berhasil dikumpulkan dari *Twitter* dari 20 Juni 2018 sampai dengan 20 Juli 2018 ada 11.527 *tweet*. Dari hasil pengujian didapatkan tingkat akurasi rata – rata untuk pasangan Rindu sebesar 95.00%, Pasangan Dedy-Dedi 90.00%, Pasangan Hasanah 84.17% dan Pasangan Asyik 74.08%. Dari hasil pengolahan data dapat disimpulkan bahwa pasangan Rindu mendapatkan respon positif tertinggi. Meski demikian tidak ada korelasi antara jumlah respon positif dari *Twitter* dengan hasil KPU, karena data opini *Twitter* banyak yang isinya sama, hasil RT (retweet) dari user lain, waktu *tweet* yang hampir sama, dan nama user yang tidak lazim.

Kata kunci: Analisis sentimen, twitter, naive bayes classifier .

Pendahuluan

Sentiment Analysis (analisis sentimen) adalah statistik, atau metode pembelajaran mesin untuk mengekstrak, mengidentifikasi, atau sebaliknya mencirikan isi sentimen dari sebuah teks. *Sentiment analysis* atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisis pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu (Liu, 2012). Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan

pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Dehaff, 2010). Kita dapat mencari pendapat tentang produk-produk, merek atau orang-orang dan menentukan apakah mereka dilihat positif atau negatif di web.

Penelitian ini dimaksudkan untuk menerapkan metode *naive bayes classifier* untuk melihat sentimen masyarakat di media Twitter terhadap hasil Pilkada Jawa Barat 2018. Sengaja digunakan metode *naive bayes classifier* karena metode ini menunjukkan akurasi yang cukup tinggi (Fahrurrozi, 2012).

Dengan menggunakan data – data twitter dan data dari KPU serta menggunakan metode *text mining* dan algoritma *naive bayes classifier* diharapkan bisa melihat sentimen masyarakat di media sosial Twitter terhadap hasil Pilkada Jawa Barat 2018. Diharapkan dari hasil penelitian ini akan bisa diketahui korelasi antara respon (positif atau negatif) dari sentimen masyarakat di media sosial terhadap hasil perolehan suara pilkada Jawa Barat 2018. Penelitian ini meliputi beberapa tahapan sebagaimana dijelaskan pada Tabel

Tabel 1
Tahapan Penelitian

Tahapan	Tujuan
1. Studi Literatur	1. Sentimen analisis 2. Pilkada 3. Opini 4. Text mining 5. Praprosesing 6. Naive Bayes
2. Pengumpulan Data	Data <i>Twitter</i> dari 20 Juni 2018 sampai dengan 20 Juli 2018
3. Pre-Processing	1. Remove duplicate 2. Replace/ Filtering 3. Transform Cases 4. Tokenizing
4. Analisis Data	1. Klasifikasi data 2. Naive bayes classifier 3. Penarikan kesimpulan

Pengumpulan Data

Hasil dari proses *crawling* data Twitter dengan menggunakan R Studio dari tanggal 20 Juni – 20 Juli 2018 didapatkan 11.527 *tweet*. Secara keseluruhan, data yang berhasil dikumpulkan dari hasil *crawling* data twitter adalah sebagai berikut:

Tabel 2.
Data Terkumpul

No.	Keyword	Jumlah Data
1	Rindu (Ridwan Kamil-Ruzhanul Ulum)	1.158
2	Hasanah (Tubagus – Anton C)	541
3	Asyik (Sudrajad - Ahmad Syaikhu)	4.253
4	Deddy-Dedi (Dedi Mizwar – Dedi Mulyadi)	825
5	Pilkadajabar2018	1.063
6	Pilgubjabar2018	3.638
Total Data		11.527

Pra Pemrosesan Data

Sebelum dilakukan klasifikasi berdasarkan sentimen, data yang telah dikumpulkan perlu dilakukan pra-proses data terlebih dahulu. Tujuan dilakukannya prapemrosesan data ini adalah untuk membuat data mentah menjadi data yang berkualitas agar hasil data mining nantinya juga berkualitas. Pada *text mining*, informasi yang akan dianalisis merupakan informasi – informasi yang mempunyai struktur sembarang. Oleh karena itu, diperlukan proses pengubahan bentuk menjadi data yang terstruktur sesuai kebutuhan untuk proses *text mining*. (Feldman dan Sanger,2007).

Adapun tahapan dalam prapemrosesan data hasil dari *crawling* data twitter secara ringkas bisa disarikan sebagai berikut:

- a) *Remove Duplicate*, yaitu menghapus data yang duplikasi / berulang.
- b) *Replace / Filtering*, berupa penghapusan semua karakter selain string serta penghapusan beberapa karakteristik dari data, misalnya *@username*, *#hashtag*, *http://*, dan “RT”. Dalam *filtering* juga dilakukan penghapusan terhadap *stopward*. Hal ini bermanfaat untuk mengurangi *load* dan *performance* saat melakukan training maupun testing data.
- c) *Transform Cases*, yaitu mengubah semua huruf kapital menjadi huruf kecil atau *lowercase*.
- d) *Tokenizing*, yaitu pemecahan berdasarkan perkata. Hal ini bisa dilakukan dengan menandai karakter sebagai pembatas. Tahapan yang dilakukan adalah menghapus nomor dan tanda baca.

Dengan memanfaatkan aplikasi RapidMiner, bisa didapatkan data hasil prapemrosesan data sebagai berikut:

Name	Date modified	Type	Size
praproses_asyik_1	27/07/2018 15.02	Microsoft Excel C...	102 KB
praproses_asyik_2	27/07/2018 10.08	Microsoft Excel C...	102 KB
praproses_asyik_3	27/07/2018 10.09	Microsoft Excel C...	19 KB
praproses_coba	27/07/2018 13.16	Microsoft Excel C...	10 KB
praproses_deddydedi_1	27/07/2018 09.22	Microsoft Excel C...	8 KB
praproses_deddydedi_2	27/07/2018 09.27	Microsoft Excel C...	1 KB
praproses_deddydedi_3	27/07/2018 10.12	Microsoft Excel C...	82 KB
praproses_hasanah_1	27/07/2018 10.14	Microsoft Excel C...	9 KB
praproses_hasanah_2	27/07/2018 10.14	Microsoft Excel C...	5 KB
praproses_hasanah_3	27/07/2018 10.16	Microsoft Excel C...	51 KB
praproses_pilgub_1	27/07/2018 10.16	Microsoft Excel C...	7 KB
praproses_pilgub_2	27/07/2018 10.23	Microsoft Excel C...	388 KB
praproses_pilgub_3	27/07/2018 10.24	Microsoft Excel C...	4 KB
praproses_pilkada_1	27/07/2018 10.25	Microsoft Excel C...	5 KB
praproses_pilkada_2	27/07/2018 09.39	Microsoft Excel C...	3 KB
praproses_pilkada_3	27/07/2018 10.34	Microsoft Excel C...	82 KB
praproses_rindu_1	27/07/2018 12.56	Microsoft Excel C...	10 KB
praproses_rindu_2	27/07/2018 13.06	Microsoft Excel C...	64 KB
praproses_rindu_3	27/07/2018 13.09	Microsoft Excel C...	2 KB

Gambar 1
Hasil Prapemrosesan Data

Ternyata hasil prapemrosesan tahap ini belum sepenuhnya data bersih dan siap untuk diproses. Masih banyak RT (Retweet) dan karakter @. Oleh karena itu perlu dilakukan prapemrosesan tahap kedua guna menghilangkan semua hal tersebut. Berikut ini adalah hasil yang didapatkan dari prapemrosesan tahap kedua ini.

Tabel 3.
Data Hasil Prapemrosesan

No	Keyword	Data Mentah	Data Praproses
1	Pilkadajabar2018	1063	24
2	Pilgubjabar2018	3638	12
3	Rindu	1158	20
4	Hasanah	541	42
5	Asyik	4253	175
6	Deddy-Dedi	825	23
Total Data		11527	296

Hasil dan Pembahasan

Data hasil prapemrosesan yang didapatkan ini selanjutnya diklasifikasikan sehingga bisa didapatkan sentimen yang bisa dikategorikan menjadi positif, negatif atau netral. Dengan RapidMiner, klasifikasi sentimen bisa dilakukan dengan operator *Analyze Sentiment* dari Aylie. Hasil dari pengolahan analisis sentimen ini bisa dilihat pada tabel 4.

Tabel 4.
Perbandingan Data Sentimen

No	Keyword	Positif	Negatif	Total Data
1	Rindu	14 77.77%	4 22.23%	18
2	Hasanah	35 92.10%	3 7.90%	38
3	Asyik	132 85.71%	22 14.29%	154
4	Deddy-Dedi	19 95.00%	1 5.00%	20

Merujuk pada penelitian Ginting, et all (2018), klasifikasi sentimen menggunakan *naive bayes classifier*, tingkat akurasinya dapat digunakan untuk mengukur *preference value* pada kasus pemilihan kepala daerah. Atas dasar ini selanjutnya dilakukan perbandingan *performance* dengan mengeluarkan sentimen netral dalam perhitungan. Selanjutnya hasil ini dibandingkan dengan perolehan data dari KPU, sebagaimana terlihat pada Tabel 5.

Tabel 5.
Tingkat akurasi data sentimen

No	Keyword	Data Twitter		Perolehan
		Akurasi	Positif	Suara KPU
1	Rindu	95.00 %	32.88 %	77.77%
2	Hasanah	84.17 %	12.62 %	92.10%
3	Asyik	74.08 %	28.74 %	85.71%
4	Deddy-Dedi	90.00 %	25.77 %	95.00%

Berdasarkan Tabel 5, dapat disimpulkan bahwa pasangan Rindu mendapatkan respon positif tertinggi, berdasarkan data *tweet* yang diambil dari tanggal 20 Juni sampai dengan 20 Juli 2018. Meski demikian tidak ada korelasi antara jumlah respon positif dari Twitter dengan hasil KPU, karena data opini twitter banyak yang isinya sama, hasil RT (retweet) dari user lain, waktu *tweet* yang hampir sama, dan nama user yang tidak lazim. Akibatnya hasil akurasi dan tahap prapemrosesan banyak data yang dihilangkan dan hanya mendapatkan beberapa data saja yang bisa digunakan untuk penarikan kesimpulan.

Kesimpulan

Dari hasil pengumpulan data awal dari twitter berhasil didapatkan 11527 *tweet* dari 6 *keyword*. Hasil pra-proses didapatkan data 296 data yang sudah bersih dan siap diklasifikasikan. Hasil klasifikasi dengan *naive bayes classifier* didapatkan hasil dengan akurasi sebesar 95.00% untuk pasangan Rindu, kemudian disusul oleh pasangan Deddy-Dedi sebesar 90.00%, pasangan Hasanah sebesar 84.17%, dan terakhir pasangan Asyik sebesar 74.08%.

Hasil di atas tidak berkorelasi dengan data pemenang pemilu versi KPU, mengingat banyak data opini di twitter yang isinya sama, hasil retweet dari pengguna yang lain, aktu *tweet* yang hampir sama serta nama user yang tidak lazim. Oleh karena itu dirasa perlu untuk mengintegrasikan hasil penelitian ini dengan mengambil data dari media sosial yang lain seperti Facebook, instagram, Youtube dan lain-lain agar bisa didapatkan gambaran yang lebih riil tentang opini masyarakat yang ada di media sosial.

Daftar Pustaka

- Dehaff, M. (2010). *Sentiment Analysis, Hard But Worth It!*. Diakses dari http://www.customerthink.com/blog/sentiment_analysis_hard_but_worth_it pada Oktober 2017
- Fahrurozzi, I. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi.
- Feldman, R., & Sanger. J. (2007). *Text Mining Handbook: Advanced Approaches in. Analyzing Unstructured Data*. New York: Cambridge University Press
- Ginting , S.H., Muslim, K dan Murdiansyah, DT. (2018). Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa 2018 di Twitter Menggunakan Naive Bayes. E-proceeding of engineering : Vol.5, No 1. 1793.
- Liu, Bing. (2012). *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- <http://www.dcc.ufrj.br/~valeriab/DTMSentiment-AnalysisAndOpinionMining-BingLiu.pdf>. Diakses November 2017