

Analisis Konten Interaksi Pengguna Twitter pada Masa 100 Hari Pertama Pemerintahan Baru DKI Jakarta Menggunakan *Text Mining*

Content Analysis on Twitter Users Interaction within First 100 Days of Jakarta's New Government by Using Text Mining

Muhammad Rifqi Ma'arif

Fakultas Teknik dan Teknologi Informasi, Universitas Jenderal Achmad Yani Yogyakarta
Jl. Siliwangi, Jl. Ring Road Bar., Kabupaten Sleman, Daerah Istimewa Yogyakarta 55293, Telp./Fax : (0274) 552489

rifqi@unjaya.ac.id

Diterima : 30 Agustus 2017 | Revisi : 1 Februari 2018 | Disetujui: 16 Oktober 2018

Abstrak - Sebagai salah satu negara demokrasi terbesar di dunia, Indonesia memiliki masyarakat yang sangat responsif terhadap berbagai macam kebijakan yang dibuat oleh pemerintah, baik pemerintah pusat maupun daerah. Di Indonesia, masa 100 hari pemerintahan baru merupakan saat-saat yang cukup krusial, yakni masyarakat akan cukup intens dalam memantau kinerja mereka. Berkembangnya platform media sosial seperti Twitter, perbincangan ataupun interaksi antara masyarakat dengan akun resmi dari gubernur/wakil gubernur terpilih akan mengalami kenaikan yang cukup signifikan. Momentum Pilkada DKI Jakarta tahun 2017 yang menghasilkan gubernur baru, Anies Baswedan, dan wakil gubernur, Sandiaga Uno, merupakan salah satu hajatan politik yang sangat banyak mendapat perhatian masyarakat. Implikasinya adalah tingginya atensi masyarakat terhadap masa 100 hari pertama kinerja mereka. Twitter Indonesia, dengan basis pengguna yang sangat besar, cukup reaktif dengan berbagai langkah yang dilakukan Anies-Sandi, sehingga platform ini dapat digunakan untuk secara akurat memonitor opini publik terhadap kinerja 100 hari pertama ini. Dalam penelitian ini, konten interaksi masyarakat dengan akun Twitter resmi dari Anies Baswedan (@aniesbaswedan), Sandiaga Uno (@sandiuno) dan Pemerintah DKI Jakarta (@DKIJakarta) dalam masa 100 hari pertama akan dianalisis dengan menggunakan *text mining*.

Kata kunci: 100 hari Anies Sandi, *text mining*, twitter

Abstract - As one of the biggest democratic country in the world, Indonesia have very responsive society towards the various political move in either the central or local governments. In Indonesia, the first 100 days of new government era usually became a crucial timelines in which the society give their intents attention to government's performance. By the massive penetration of online social networking platform likes Twitter, the interaction between netizens and government figure official account will significantly increase. The provincial election for of Jakarta momentum was one of the most attractive political events in which Anies Baswedan and Sandiaga Uno were elected the lead of Jakarta government. The implication is the society would look intensively at their performance within their first 100 days. With the huge amount of Twitter users in Indonesia, Indonesian Twitter was very reactive towards any movement of Anies Sandi. Thus, this platform can be utilized to monitor the public opinion towards first 100 days of Anies Sandi government. In this study, we analyzed the content of conversations in interaction among the netizen and official Twitter Account of Anies Baswedan (@aniesbaswedan), Sandiaga Uno (@sandiuno) and the Government of Jakarta (@DKIJakarta). A set of text mining techniques was employed in this study.

Keywords: first 100 days of Anies Sandi, *text mining*, twitter

PENDAHULUAN

Pemilihan Kepala Daerah (Pilkada) DKI Jakarta merupakan salah satu hajatan demokrasi yang menyita begitu banyak perhatian publik di tahun 2017 yang lalu. Tidak hanya warga DKI Jakarta saja yang turut meramaikan gelaran tersebut, namun juga mencakup masyarakat Indonesia secara luas, mengingat DKI Jakarta adalah ibukota Indonesia. Hal tersebut tercermin dalam banyaknya unggahan maupun interaksi yang terjadi di berbagai platform media

sosial. Imbas dari fenomena tersebut adalah pada tingginya tingkat perhatian masyarakat terhadap kinerja pemenang kontestasi Pilkada DKI Jakarta khususnya pada 100 hari pertama masa pemerintahan. Pemenang Pilkada DKI yakni gubernur Anies Baswedan dan wakil gubernur Sandiaga Salahuddin Uno menjadi properti yang sangat banyak diperbincangkan di media sosial dalam kurun waktu Oktober hingga Desember 2017, khususnya mengenai

sepak terjang mereka di 100 hari pertama pemerintahan DKI Jakarta periode 2017 – 2022.

Penelitian ini akan mengeksplorasi konten-konten apa saja yang dibicarakan oleh masyarakat terkait dengan Anies Baswedan dan Sandiaga Uno dalam kurun waktu Oktober sampai Desember 2017. Penelitian akan berfokus pada pembicaraan yang terjadi di platform media sosial Twitter. Hal yang mendasari dipilihnya Twitter adalah basis pengguna Twitter yang sangat besar dan aktif di Indonesia. Indonesia, bersama Amerika Serikat, Brazil, Inggris Raya, dan Jepang merupakan Negara dengan basis pengguna Twitter terbesar di dunia (Takhyetev et.al, 2012). Dengan basis pengguna yang sangat besar, Twitter Indonesia cukup reaktif dengan berbagai macam berita yang beredar di Internet. Hal tersebut membuat opini masyarakat yang diunggah ke Twitter atau lebih dikenal dengan sebutan *tweet*, memiliki akurasi yang cukup tinggi apabila digunakan untuk memonitor atau mengetahui pendapat masyarakat mengenai hal-hal tertentu, termasuk reaksi masyarakat terhadap kinerja 100 hari pertama pemerintahan DKI Jakarta di bawah Gubernur dan Wakil Gubernur Baru, Anies Baswedan dan Sandiaga Uno

Media sosial telah menjadi bagian hidup dari masyarakat kontemporer saat ini. Media sosial telah dimanfaatkan dalam berbagai domain kehidupan mulai dari ekonomi/bisnis, pendidikan, budaya, religiusitas, politik, dsb. Masifnya pemanfaatan media sosial memberikan pengaruh yang kuat terhadap pola berpikir dan gaya hidup masyarakat, serta pola berinteraksi diantara satu sama lain. Media sosial telah menjadi sebuah ruang untuk memunculkan ide dan gagasan, perdebatan, dan bahkan opini publik. Hal ini telah menarik perhatian banyak kalangan, mulai dari pelaku bisnis, akademisi, hingga para politis dan pengambil kebijakan untuk memanfaatkan *insight* yang didapatkan dengan mengamati media sosial untuk berbagai macam keperluan.

Pada penelitian-penelitian sebelumnya, data yang berasal dari Twitter sudah digunakan di berbagai macam domain. Pada domain bisnis contohnya, Alamsyah (2016) menggunakan Twitter untuk mengamati unggahan *netizen* yang terkait dengan pelayanan industri telekomunikasi di Indonesia untuk menganalisis pola pemasaran yang efektif. Penelitian lain yang dilakukan oleh Ray & Chakrabarti (2017) menggunakan data Twitter untuk mengetahui tanggapan konsumen terkait penggunaan produk

smartphone. Sementara itu, Urolagin (2017) menggunakan data Twitter untuk mengetahui sentimen masyarakat terkait produk dan pelayanan suatu perusahaan. hasil sentimen tersebut digunakan untuk memprediksi satu harga saham perusahaan di pasar saham.

Twitter juga banyak digunakan sebagai subjek penelitian untuk bidang politik dan kebijakan publik. Imtiyazi et.al (2016), dengan mengamati penggunaan kata dan asosiasi kata pada unggahan *netizen* di Twitter (*tweet*), melihat trend yang terjadi terkait dengan pemilihan kepala daerah (Pilkada) serentak pada tahun 2016. Penelitian yang dilakukan oleh Rumata (2017) menggunakan data Twitter untuk melihat trend opini masyarakat terkait dengan peluncuran program pengampunan pajak (amnesti pajak). Pada masa kampanye pemilihan presiden Amerika Serikat, Dominguez, et al, (2017) melakukan Penelitian tentang agenda politik terkait Presiden di Negara Amerika Serikat. Data Twitter yang digunakan adalah data *tweet* dari akun pribadi @realDonaldTrump dan akun Institusi @POTUS (*the President of United States*). Penelitian ini membandingkan isi dan wacana dari dua akun tersebut. Hal yang serupa juga dilakukan oleh Korakakis, et al, (2017) Menggunakan data Twitter untuk mengetahui sentimen masyarakat Amerika Serikat terkait dengan pemilihan presiden 2016.

Penelitian lain yang terkait dengan data Twitter juga dilakukan oleh Liao & Tan (2016) menggunakan data Twitter untuk mengkaji pendapat konsumen terhadap maskapai penerbangan bertarif rendah di Malaysia. Di bidang industri, Chae (2014) meneliti manajemen rantai pasokan (*supply chain management*) dari suatu perusahaan dengan data Twitter. Bidang sosial juga menjadi sasaran penelitian oleh Dragiewicz & Burgess, (2016) yang melakukan penelitian mengenai tingkat kekerasan terhadap perempuan dalam rumah tangga oleh pria di Australia. Sementara itu Amirkhanyan & Meinel (2017) melakukan penelitian tentang kesadaran keamanan publik di Berlin. Penelitian ini menggunakan data Twitter pada akun Twitter kepolisian Berlin yang berisi informasi kepada penduduk tentang insiden yang terjadi di kota tersebut.

METODOLOGI PENELITIAN

Penelitian ini adalah penelitian deskriptif kuantitatif, yang bertujuan untuk menjelaskan

fenomena yang terjadi dari suatu peristiwa dengan menggunakan angka-angka ataupun rasio berdasarkan angka-angka tersebut. Data yang digunakan dalam penelitian ini merupakan unggahan (*tweet*) masyarakat pada platform Twitter terkait kinerja Anies Baswedan dan Sandiaga Uno di 100 hari pertama kepemimpinan mereka di DKI Jakarta. Sampel data adalah unggahan yang berinteraksi secara langsung dengan akun resmi Anies Baswedan (@aniesbaswedan), Sandiaga Uno (@sandiuno), dan Pemerintah Provinsi DKI Jakarta (@DKIJakarta), baik dalam bentuk *retweet* ataupun *mention*.

Penelitian dimulai dengan proses pengambilan data (*tweet*) yang sesuai dengan topik yang akan diteliti. Tahapan selanjutnya adalah mengolah data yang berhasil dikumpulkan menggunakan teknik *text mining*. Pemrosesan *text mining* terdiri dari dua tahapan. Tahapan yang pertama adalah melakukan *preprocessing* terhadap data yang berhasil dikumpulkan. Kemudian di tahapan yang kedua, dilakukan ekstraksi topik pembicaraan dari setiap *tweet* yang diunggah dengan menggunakan metode *Latent Dirichlet Allocation* (LDA). Bagian akhir dari penelitian ini adalah melakukan analisis konten pada data yang ada. Analisis konten yang dilakukan mencakup dua hal, yaitu analisis kemunculan topik dan analisis kemunculan tagar pendamping dari tagar #hutri72. Hasil analisis kemudian divisualisasikan dalam bentuk *wordcloud*.

Pengumpulan Data

Insight yang akurat dari media sosial seperti Twitter, dapat didapatkan jika pengguna terhadap *tweet-tweet* yang bermunculan di Twitter setiap hari. Akibat jumlah *tweet* yang sangat banyak serta pertambahannya yang cepat, hal tersebut mustahil dilakukan dengan cara manual tanpa bantuan alat bantu berbasis komputer. Oleh karena Twitter sudah menyediakan antar muka berupa *Application Programming Interface* (API) yang memungkinkan pengguna untuk secara simultan mengakuisisi *tweet* yang diunggah oleh pengguna Twitter, maka dalam penelitian ini, *Tweet* dikumpulkan dengan memanfaatkan fasilitas tersebut. Proses pengumpulan data dimulai dari tanggal 20 Oktober 2017 dan berakhir pada tanggal 20 Januari 2018. *Tweet* yang diambil adalah *tweet* yang di dalamnya terdapat interaksi (*mention* dan *retweet*) dengan akun @aniesbaswedan, @sandiuno, dan/atau @DKIJakarta.

Dari kegiatan pengumpulan data, diperoleh sejumlah 418.879 *tweet*.

Text Pre-Processing

Permasalahan yang muncul dari penggunaan API dalam proses pengumpulan/akuisisi data adalah data yang didapatkan dari antar muka tersebut masih dalam mentah yang belum siap untuk proses analisis lebih lanjut. Oleh karena itu, data dibersihkan terlebih dahulu melalui mekanisme *data preprocessing* dengan menggunakan teknik *text mining*. *Pre-processing* adalah tahapan pembersihan data untuk mengeliminasi data yang kurang sempurna, gangguan pada data (*noise*) serta data yang tidak konsisten (Hemalatha dkk, 2012).

Apabila data mentah yang masih mengandung beberapa properti tidak ditangani secara benar, akan membuat hasil analisis tidak akurat. Properti pada data teks mentah yang perlu dibersihkan diantaranya adalah angka, tanda baca, *emoticon*, *stopword*, dsb. Tahapan proses data, dari akuisisi hingga siap untuk dianalisis, setidaknya terdiri atas 8, sebagai berikut:

1. Penghapusan *Unifrom Resource Locator* (URL). URL (<http://www.situs.com>) dan email (nama@situs.com) dihapus karena tidak digunakan dalam analisis teks.
2. Penghapusan *username* dan tanda *retweet* pada Twitter. Proses ini dilakukan dengan menghapus karakter khusus Twitter yakni *username* (@username), dan tanda *retweet* (RT)
3. Penghapusan simbol, *emoticon* dan tanda baca. Proses ini berfungsi untuk menghapus tanda-tanda baca dan *wildchar* serta *emoticon* yang pada teks.
4. Tokenisasi. Tokenisasi merupakan proses penguraian kata-kata dari sebuah kalimat. Kata-kata tersebut diperlakukan sebagai sebuah entitas tunggal pada proses analisis teks.
5. *Case folding*. Proses ini berfungsi untuk menyamaratakan karakter huruf yaitu mengkonversi huruf kapital 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. *Stemming*. *Stemming* adalah tahap mencari akar kata dengan menghilangkan imbuhan pada sebuah kata.
8. Menghapus *stopword*. Kata penghubung dalam analisis teks tidak diperlukan oleh sebab itu perlu dihapus. Pada proses ini kata-kata yang tidak

digunakan dalam analisis dihapus (Hidayatullah & Ma'arif, 2016).

Analisis Topik dengan LDA

Blei, et al (2003) telah memperkenalkan *Latent Dirichlet Allocation* (LDA) sebagai model probabilistik generatif untuk kelompok data diskret seperti teks *korpora* (kumpulan teks). LDA merupakan *unsupervised machine learning technique*. Metode ini bertujuan untuk memodelkan dokumen sebagai kemunculan dari banyak topik, dimana topik didefinisikan sebagai distribusi melalui persyaratan kosa kata tetap.

Analisis Frekuensi Teks dan Hashtag

Analisis selanjutnya adalah analisis frekuensi kemunculan teks/kata-kata tertentu dan penggunaan *hashtag*. *Hashtag* yang ditulis dengan simbol # digunakan untuk mengindeks kata kunci atau topik di Twitter. Fungsi ini dibuat di Twitter dan memungkinkan pengguna untuk mengikuti topik yang mereka minati dengan mudah. Isi *hashtag* tidak selalu secara kasat mata sejalan dengan sesuatu yang disampaikan *netizen* dalam konten *tweet*-nya. *Hashtag* berfungsi untuk mengkaitkan *tweet* dengan topik tertentu yang sedang dibahas oleh para pengguna Twitter yang seringkali terkait dengan suatu peristiwa di dunia nyata

Frekuensi kemunculan kata dan *hashtag* tersebut divisualisasikan dengan menggunakan metode *wordcloud*. *Wordcloud* adalah teknik visualisasi data yang sudah umum digunakan. Pada visualisasi *wordcloud*, ukuran huruf (*font*) dari kata-kata akan berbanding lurus dengan frekuensi kemunculannya dalam sebuah set dokumen, sehingga semakin sering topik tersebut muncul, maka ukurannya akan tampak semakin besar.

HASIL DAN PEMBAHASAN

Hasil Analisis Topik dengan LDA

Hasil analisis topik menggunakan algoritma pemodelan topik LDA ditunjukkan dalam Gambar 1, dimana terlihat bahwa ada satu topik yang cukup dominan dibicarakan oleh pengguna Twitter, yakni topik pembicaraan yang berkaitan dengan pejabat gubernur sebelumnya yakni Basuki Tjahaja Purnama atau yang lebih dikenal dengan Ahok. Topik lain yang banyak dibicarakan adalah topik terkait dengan realisasi janji kampanye Anies dan Sandi, yakni

penghentian reklamasi teluk Jakarta, program rumah murah (uang muka 0%).



Gambar 1 Visualisasi *wordcloud* untuk topik yang sering dibicarakan

Hasil Analisis Teks dan Hashtag

Analisis teks *tweet* yang pertama adalah analisis frekuensi kemunculan kata. Dari hasil perhitungan, kata *Gaberner* adalah kata yang paling sering digunakan oleh pengguna Twitter dalam *tweet* yang terkait dengan 100 hari pertama masa pemerintahan Anies Baswedan dan Sandiaga Uno. Kata-kata yang sering muncul diilustrasikan pada Gambar 2, dimana semakin besar ukuran huruf, maka berarti semakin sering kata-kata yang bersangkutan muncul pada *tweet* yang berhasil dikumpulkan.



Gambar 2 Visualisasi *wordcloud* untuk kata-kata yang sering digunakan

Kata *Gaberner* yang merupakan *plesetan* dari kata Gubernur kebanyakan digunakan untuk mengekspresikan sentimen negatif terhadap gubernur baru, Anies Baswedan. Dalam visualisasi Gambar 2, dapat juga dilihat bahwa kata/nama Ahok cukup sering digunakan dalam unggahan pengguna Twitter. Hal ini sejalan dengan banyaknya topik yang berkaitan dengan Ahok di pembahasan sebelumnya.

DAFTAR PUSTAKA

- Alamsyah, A., Paryasto, M., Putra, F. J., & Himmawan, R. (2016, May). Network text analysis to summarize online conversations for marketing intelligence efforts in telecommunication industry. In *Information and Communication Technology (ICoICT), 2016 4th International Conference on* (pp. 1-5). IEEE.
- Amirkhanyan, A., & Meinel, C. (2017, April). Analysis of data from the Twitter account of the berlin police for public safety awareness. In *Computer Supported Cooperative Work in Design (CSCWD), 2017 IEEE 21st International Conference on* (pp. 209-214). IEEE.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Hidayatullah, A. F., & Ma'arif, M. R. (2017, January). Pre-processing Tasks in Indonesian Twitter Messages. In *Journal of Physics: Conference Series* (Vol. 801, No. 1, p. 012072). IOP Publishing.
- Korakakis, M., Spyrou, E., & Mylonas, P. (2017, July). A survey on political event analysis in Twitter. In *Semantic and Social Media Adaptation and Personalization (SMAP), 2017 12th International Workshop on* (pp. 14-19). IEEE.
- Hemalatha, I., Varma, G. S., & Govardhan, A. (2012). Preprocessing the informal text for efficient sentiment analysis. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 1(2), 58-61.
- Imtiyazi, M. A., Alamsyah, A., Junaedi, D., & Pradana, J. A. (2016, May). Word association network approach for summarizing Twitter conversation about public election. In *Information and Communication Technology (ICoICT), 2016 4th International Conference on* (pp. 1-4). IEEE.
- Márquez-Domínguez, C., López-López, P. C., & Arias, T. E. (2017, June). Social networking and political agenda: Donald trump's Twitter accounts. In *Information Systems and Technologies (CISTI), 2017 12th Iberian Conference on* (pp. 1-6). IEEE.
- Park, H., Rodgers, S., & Stemmler, J. (2013). Analyzing health organizations' use of Twitter for promoting health literacy. *Journal of health communication*, 18(4), 410-425.
- Ray, P., & Chakrabarti, A. (2017, February). Twitter sentiment analysis for product review using lexicon method. In *Data Management, Analytics and Innovation (ICDMAI), 2017 International Conference on* (pp. 211-216). IEEE.
- Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter networks. *Social networks*, 34(1), 73-81.
- Urolagin, S. (2017, September). Text Mining of Tweet for Sentiment Classification and Association with Stock Prices. In *Computer and Applications (ICCA), 2017 International Conference on* (pp. 384-388). IEEE.