

ESKPLORASI TRENDING TOPIK TWITTER MENGGUNAKAN TEXT MINING

Ebi Mesak¹, Yesi Novaria Kunang², Ria Andryani³

^{1,2,3}Universitas Bina Darma Palembang

^{1,2,3} Jalan Jendral Ahmad Yani No.3 Kota Palembang, Sumatera Selatan 30264

¹ebimesak95@gmail.com, ²yesinovariakunang@binadarma.ac.id, ³riaandryani@binadarma.ac.id

ABSTRACT

Twitter is a website owned and operated by twitter.inc which offers social networking microblogs allowing users to send and read messages called tweets or tweets in text form with a capacity of up to 140 characters. The purpose of this study is to explore the trending topic of twitter using text mining, Text mining, is a process to examine a large collection of written information, and transform unstructured text into structured data. By doing this research, to know an information trending topic in the form of graph table and histogram. By utilizing Application Programming Interface (API) provided by twitter, the process of data withdrawal (Collecting), analysis (Analyzing), visualization (Vizualizing). Visualization of twitter topic trending data is made in the form of histogram.

Keywords: Twitter, Histogram, Application Programming Interface (API), Collecting, Analyzing, Vizualizing.

Twitter merupakan sebuah situs web yang dimiliki dan dioperasikan oleh twitter.inc yang menawarkan jejaring sosial berupa mikroblog sehingga memungkinkan penggunaannya untuk mengirim dan membaca pesan yang disebut kicauan atau tweets dalam bentuk teks dengan kapasitas hingga 140 karakter. Tujuan penelitian ini melakukan eksplorasi trending topik twitter menggunakan *text mining*, *Text mining*, adalah proses untuk memeriksa kumpulan informasi tertulis yang besar, dan mentransformasikan teks tidak terstruktur ke dalam data terstruktur. Dengan dilakukannya penelitian ini, untuk mengetahui sebuah informasi trending topik berupa tabel graf dan histogram. Dengan memanfaatkan Application Programming Interface (API) yang disediakan oleh twitter, dilakukan proses penarikan data (Collecting), analisis (Analyzing), visualisasi (Vizualizing). Visualisasi terhadap data trending topik twitter dibuat dalam bentuk histogram.

Kata kunci : Twitter, Histogram, Application Programming Interface (API), Collecting, Analyzing, Vizualizing.

1. PENDAHULUAN

Perkembangan media sosial saat ini berkembang pesat, ini dapat dilihat dari banyaknya jumlah pengguna media sosial. Pesatnya perkembangan media sosial tersebut dikarenakan hampir setiap orang memiliki mediana sendiri. Seseorang pengguna media sosial bisa mengakses menggunakan media sosial dengan jaringan internet. Internet bisa dikatakan hanya sebagai media yang bisa digunakan untuk diskusi atau debat politik, pertukaran ide maupun gagasan, hingga membangun wacana sebagai jawaban terhadap realitas politik. Internet juga merupakan integrasi beberapa peralatan teknologi komunikasi Bahasa dan jaringan komputer yang dapat menghubungkan peralatan komunikasi yang tersebar diseluruh pelosok dunia interaktif. Kementerian Komunikasi dan Informatika (KomenKominfo) mengatakan bahwa pengguna internet di Indonesia sampai saat ini mencapai 63 juta orang. Angka tersebut 95 persennya mengakses media sosial. (Kominfo.go.id 07/11/2013). Adapun jenis media sosial yang paling digemari saat ini adalah *facebook*, *twitter*, *google plus*, *Instagram*, *pinterest*, *tumblr* dan *sosial chat*. (Triastuti 2017).

Direktur Pelayanan Informasi International Ditjen Informasi dan Komunikasi Publik (IKP) Selamatta Sembiring mengungkapkan bahwa situs jejaring sosial yang paling banyak digunakan atau diakses adalah *Facebook* dan *Twitter*. Indonesia sendiri menempati peringkat keempat pengguna *facebook* terbesar di dunia setelah USA, Brazil, dan India. Sedangkan untuk *twitter* Indonesia menduduki peringkat kelima pengguna *twitter* di dunia, posisi Indonesia hanya kalah dari negara USA, Brazil, Jepang dan Inggris. (Kominfo.go.id 07/11/2013).

Dengan banyaknya media sosial yang ada *twitter* merupakan salah satu yang paling populer. Kegiatan yang sering dilakukan di sosial media adalah dengan melakukan *update* status, melakukan *tweet*, *comment* ataupun *re-tweet* yang merupakan kegiatan masyarakat dalam kehidupan sehari-hari yang dicerminkan melalui sosial media. (Mayfield 2008). Hal itu merupakan persepsi atau pandangan dari

para pengguna sosial media terhadap isu-isu atau peristiwa yang sedang terjadi di tengah-tengah masyarakat saat ini.

Twitter merupakan situs *microblogging* yakni *blog* yang memungkinkan pengguna menuliskan teks singkat yang disebut *tweet*, *tweet* dibatasi hanya 140 karakter dan dapat dilihat oleh pengikut pengguna. (kwak 2010). Oleh karena itu saat ini *twitter inc* mengalami peningkatan saham yang signifikan, setelah layanan *microblogging* ini mengumumkan pertumbuhan pengguna yang lebih dari perkiraan kuartal pertama. Ini disebabkan meningkatnya minat pengguna terhadap berita dan politik. (techno.okezone.com (2017)

Pesatnya pertumbuhan pengguna *twitter* tiap tahunnya menyebabkan semakin banyaknya data yang dihasilkan, fenomena ini disebut juga dengan fenomena *big data*. Berdasarkan data-data tersebut banyak peneliti melakukan penelitian terhadap data yang dihasilkan oleh *twitter*. *Twitter* dijadikan sebagai objek sebuah penelitian, Seperti yang pernah dilakukan (Negara, E.S., Andryani, R., Saksono, P.H., 2016) "Analisis Data Twitter : Ekstraksi dan Analisis Data Geospasial" membahas masalah informasi spasial berupa titik koordinat pengguna *twitter*. pada penelitian ini dilakukan ekstraksi dan analisis pada data geospasial *twitter* terhadap suatu isu publik yang sedang berkembang. Pemanfaatan data *twitter* dan Facebook juga digunakan untuk berbagai keperluan analisis data media sosial seperti: melihat jaringan pertemanan, mengukur kekuatan penelitian suatu Negara, dan melakukan analisis sentiment public dengan menggunakan berbagai tools analytics data media sosial yang telah dikembangkan (Retnawiyati, E. and Fatoni, Negara E.S, Negara, E.S. and Andryani, R., 2018, Negara, E.S., Kerami, D., Wiryana, M., Kusuma, TM. 2017, Negara, E.S., 2018, Ria Andryani, and Negara, E.S. 2017).

II METODOLOGI PENELITIAN

A. Data Penelitian

Data penelitian yang digunakan dalam penelitian ini adalah data dari *twitter* yang merupakan hasil dari proses *collecting* data dengan menggunakan *python* melalui API yang disediakan oleh *twitter*.

B. Metode Penelitian

Dalam penelitian ini menggunakan metode penelitian deskriptif kualitatif untuk mendeskripsikan fenomena yang terjadi di masyarakat melalui sosial media *twitter* dan pada pengumpulan data berbentuk kata. Dengan memanfaatkan Application Programming Interface yang disediakan oleh *twitter*, pada penelitian ini akan dijelaskan bagaimana data akan diambil, dianalisis hingga menghasilkan informasi yang dapat digunakan untuk penelitian. Menurut Sugiyono (2013) metode kualitatif adalah metode penelitian yang berlandaskan yang berlandaskan paspositivisme yaitu memandang realita, fenomena yang diklarifikasikan tetap dan terukur.

Penelitian ini secara umum dapat diklasifikasikan pada *twitter data analytic* yaitu :

A. *Collecting Data Twitter* (Penarikan Data *twitter*)

Collecting data ialah proses pengumpulan data dari *twitter* dengan menggunakan *python* melalui API *twitter*. Penarikan data *twitter* (*collecting twitter data*) merupakan suatu cara yang dilakukan untuk mengumpulkan informasi mengenai pengguna *twitter*, *tweets* yang dihasilkan oleh pengguna, relasi pengguna *twitter* baik itu *follower twitter* maupun *retweet* yang dilakukannya dengan memanfaatkan API *public* yang disediakan oleh *twitter*.

B. Analisis Data *Twitter* (*Analyzing Data Twitter*)

Analisis data *twitter* ialah proses ekstraksi data hasil *crawling* data dengan menggunakan *text mining*. *Text mining* adalah proses untuk memeriksa kumpulan informasi tertulis yang besar, dan mentransformasikan teks tidak terstruktur ke dalam data terstruktur. (Aggarwal 2012). Data yang di dapat dibersihkan dari *noise* dengan menggunakan ekstraksi dan *stop word* atau *filtering* untuk mengambil kata-kata penting dan membuang kata-kata yang kurang penting.

C. Visualisasi Data (*Visualizing*)

Visualisasi adalah konversi data kedalam format visual atau tabel sehingga karakteristik dari data dan relasi di antara item data atau atribut dapat di analisis dan dilaporkan . Teknik visualisasi memudahkan

manusia untuk menangkap konsep dari data yang ditampilkan karena pada dasarnya manusia lebih mudah mengartikan sebuah gambar daripada teks yang menjelaskan tentang makna gambar tersebut.

III Hasil dan Pembahasan

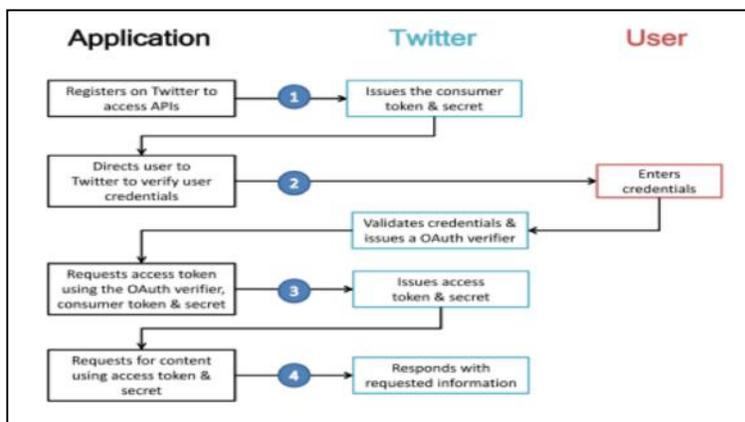
A. Collecting Data Twitter

Twitter sebagai media sosial menyimpan data yang besar yang bersumber dari tweets dari pemilik akun tersebut. Untuk mendapatkan data dari twitter, maka kita harus mempunyai PIN atau user id khusus yang dapat mengakses ke API twitter. Dalam hal ini disebut OAuth (Open Authentication) yang merupakan standard terbuka untuk otentikasi yang diadopsi oleh twitter untuk menyediakan akses melindungi informasi.

Otentikasi permintaan API pada twitter dilakukan menggunakan OAuth. Twitter API hanya dapat diakses menggunakan aplikasi. Diperlukan beberapa langkah-langkah untuk dapat melakukan API twitter dari aplikasi menggunakan OAuth, yaitu :

1. Aplikasi juga disebut consumers dan semua aplikasi diwajibkan mendaftarkan dirinya dengan twitter. Dari proses ini akan didapatkan consumer key dan consumer secret yang nantinya akan digunakan untuk proses otentikasi pada twitter
2. Aplikasi ini menggunakan consumer key dan consumer secret untuk membuat link unik twitter yang digunakan pengguna untuk otentikasi. Pengguna memberikan izin pada aplikasi melalui otentikasi dirinya di twitter. Twitter memverifikasi identitas pengguna dan memberikan verifikasi OAuth yang disebut dengan PIN
3. Pengguna menyediakan PIN ini untuk aplikasi. Aplikasi menggunakan PIN ini untuk melakukan permintaan access token dan access token secret yang dapat digunakan user
4. Dengan menggunakan access token dan access token secret, aplikasi mengotentikasi pengguna di twitter dan menggunakan key API untuk kepentingan pengguna. (Shamant Kumar dkk, 2013).

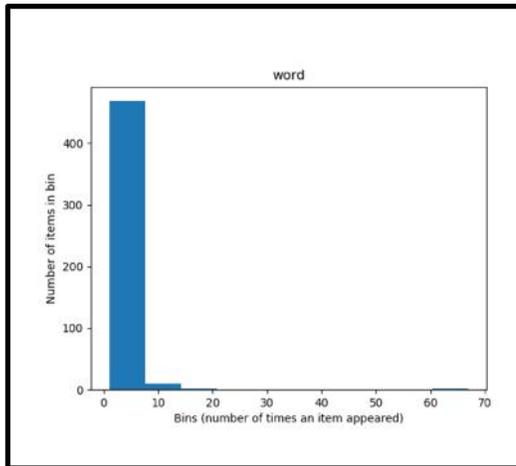
Seperti yang telah dijelaskan pada alur OAuth di atas bahwa hal yang pertama kali harus dilakukan untuk mendapatkan API twitter adalah dengan mendaftarkan aplikasi yang kita buat ke twitter. Setelah aplikasi tersebut terdaftar, maka kita akan mendapatkan key API yang dapat digunakan untuk mengakses ke twitter. Key API tersebut berupa kode unik khusus gabungan huruf dan angka yang disebut dengan consumer secret dan consumer key.



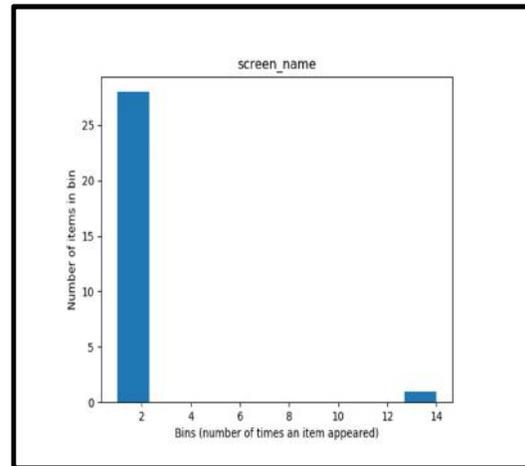
Gambar 2 Proses Aunthentication

Pada saat crawling data ada beberapa factor yang mempengaruhi yaitu koneksi jaringan internet, lamanya proses penarikan data dan update berita terbaru yang akan dilakukan crawling. Koneksi jaringan

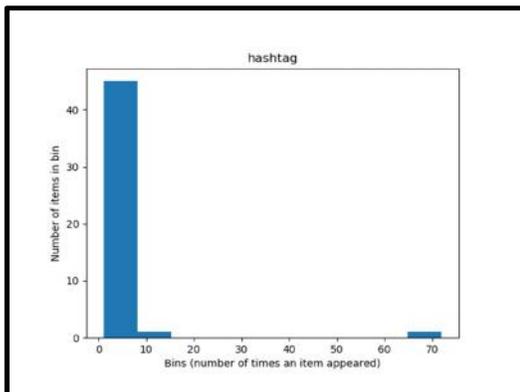
Visualisasi data merupakan salah satu dari teknik yang paling baik dan menarik di dalam hal *mengeksplorasi* data. *Visualisasi* juga dapat menggambarkan pola umum yang terjadi, trend yang sedang berkembang serta hal-hal yang tidak umum.



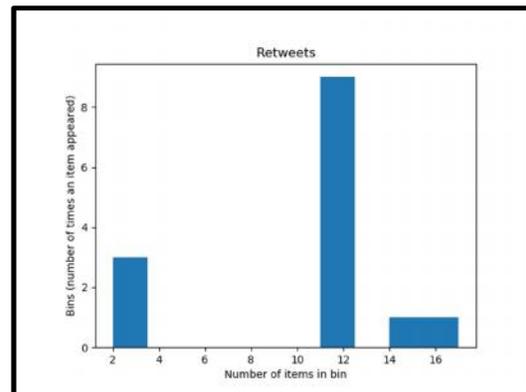
Gambar 8 histogram word



gambar 9 histogram screen name



Gambar 10 histogram hashtag



gambar 11 histogram retweet

Dari ke empat gambar histogram diatas menunjukkan hasil dari visualisasi pada tabel-tabel yang didapat dalam proses analisis. dimana proses penarikan data *word, hashtag, screen name, retweet* yang menghasilkan *histogram* diatas dengan perhitungan grafik menunjukkan bahwa *count* yang berulang kali muncul di hitung atau akumulasikan dalam bentuk grafik yang tertinggi, sedangkan *count* yang terbanyak dan hanya muncul 1 sekali (tidak berulang) di tunjukkan dengan grafik yang terendah.

IV Kesimpulan

Dari penelitian yang dilakukan terhadap *trending topik twitter*, dalam hal ini dapat disimpulkan bahwa :

1. Penarikan data atau *collecting data* dengan menggunakan media sosial *twitter* telah dilakukan, proses ini harus melakukan *Authentication* terlebih dahulu untuk bisa mengakses *Twitter API*.
2. Dari data yang didapat dan di proses melalui analisis *text mining* tersebut dapat diketahui jumlah *word, screen name, hashtag dan retweet* dengan frekuensi yang berbeda.
 - a) Pada *word* yaitu tingkat kemunculan tertinggi dengan *count* 10 sebanyak 25% dan *count* 5 30%.

- b) Pada *screen name* dengan *count* 1 muncul sebanyak 80 %.
- c) Pada *hashtag* dengan *count* 1 muncul sebanyak 35 %
- d) Pada *retweet* dengan *count* 12 muncul sebanyak 68,75%

V. Daftar Pustaka

- Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.
- Antoni, D., Negara, E.S. and Suweno, S., 2015. Ekstraksi Data Geo-Spatial Twitter (Studi Kasus: Badan Penyelenggara Jaminan Sosial Kesehatan). Seminar Nasional Teknologi Komputer, 22-24 Oktober 2015.
- Developer, Twitter, (2012). Twitter API. <https://dev.twitter.com/docs/api>
- kominfo.go.id (2017). Pengguna internet di Indonesia. URL
https://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+Indonesia+63+Juta+Orang/0/berita_satker
- Kumar, S., Morstatter, F., Liu, H., (2014). Twitter Data Analytics.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media? In Proceedings of the 19th international conference on World wide web (pp. 591-600). ACM.
- Mayfield Antony. (2008). What Is Social Media?. iCrossing-ebook.
- Negara, E.S. and Andryani, R., 2018. A Review on Overlapping and Non-Overlapping Community Detection Algorithms for Social Network Analytics. Far East Journal of Electronics and Communications, 18 (1), 1-27
- Negara, E.S., Kerami, D., Wiryana, M., Kusuma, TM. 2017. Researchgate data analysis to measure the strength of Indonesian research. Far East Journal of Electronics and Communications, 17 (5), 1177-1183
- Negara, E.S., 2018. Kajian Terhadap Tools dan Framework Social Media Analytics Untuk Pemanfaatan Data Social Media Dalam Penelitian Ilmu Sosial. Jurnal Teknologi Technoscintia, 9(2).
- Negara, E.S., Andryani, R. and Saksono, P.H., 2016. Analisis Data Twitter: Ekstraksi dan Analisis Data Geospasial. INKOM Journal, 10(1), pp.27-36.
- Sugiyono, (2013). Metode Penelitian. Kuantitatif Kualitatif dan R&D.
- Suweno, S, Antoni, D. dan Negara, E.S. 2012. Analisis Dan Ekstraksi Data GEO-SPITAL Twitter Untuk Penelitian. Jurnal Imiah xxxxxxxxxxxx Vol.X No.X.
- Ria Andryani, and Negara. E.S, 2017. Network of Friends to The Other Friends by Social Media on Facebook. The Turkish Online Journal of Design, Art and Communication, 12(12), pp.1363-1378.
- Retnawiyati, E. and Fatoni, M.M., M. Kom./Surya, Edi Negara, M. Kom. Analisis Sentimen Pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk.
- Techno.okezone.com (2017). Jumlah pengguna twitter.
<http://techno.okezone.com/read/2017/04/27/207/1678186/fitur-live-streaming-tingkatkan-jumlah-pengguna-twitter>
- Triastuti, E. (2017). Dimas Adrianto Akmal Nurul Pusat Kajian Komunikasi, FISIP Universitas Indonesia.