

Analisis Data Twitter : Ekstraksi dan Analisis Data Geospasial

Twitter Data Analytics: Geospatial Data Extraction and Analysis

Edi Surya Negara, Ria Andryani, Prihambodo Hendro Saksono
Data Science Interdisciplinary Research Center, Universitas Bina Darma
Jl. A. Yani No. 3, Palembang 30624, Indonesia
Email: e.s.negara@binadarma.ac.id

Abstract

Geospatial data on the media social like Twitter can be used to determine the spatial information (location), which is the location of the source of the emergence of public perception of an issue in social media. The amount of data production geospatial generated by Twitter provides a great opportunity to be used by various parties so as to produce more valuable information through Twitter Data Analytics. The process of data utilization Twitter geospatial process begins with the extraction of the spatial information such as the coordinates of Twitter users. Point coordinates obtained from the Twitter users location sharing is done by Twitter users. For extracting and analyzing geospatial data on Twitter necessary knowledge and frameworks of social media analytics (SMA). In this research, the extraction and analysis of geospatial data Twitter to an emerging public issues and develop the prototype software used to acquire geospatial data that exist on Twitter. Extraction and analysis process carried out through four stages, namely: crawling, storing, analyzing, and visualization. This study is exploratory focused on the development of extraction techniques and analysis of geospatial data twitter.

Keywords: social media, data analytics, social media analytics, twitter data analytics, data mining, machine learning

Abstrak

Data geospasial pada media sosial Twitter dapat dimanfaatkan untuk mengetahui informasi spasial (lokasi) yang merupakan lokasi sumber munculnya persepsi publik terhadap sebuah isu di media sosial. Besarnya produksi data geospasial yang dihasilkan oleh Twitter memberikan peluang besar untuk dapat dimanfaatkan oleh berbagai pihak sehingga menghasilkan informasi yang lebih bernilai melalui proses *Twitter Data Analytics*. Proses pemanfaatan data geospasial Twitter dimulai dengan melakukan proses ekstraksi terhadap informasi spasial berupa titik koordinat pengguna Twitter. Titik koordinat pengguna Twitter didapatkan dari *sharing location* yang dilakukan oleh pengguna Twitter. Untuk mengekstrak dan menganalisis data geospasial pada Twitter dibutuhkan pengetahuan dan kerangka kerja tentang *social media analytics* (SMA). Pada penelitian ini dilakukan ekstraksi dan analisis data geospasial Twitter terhadap suatu isu publik yang sedang berkembang dan mengembangkan prototipe perangkat lunak yang digunakan untuk mendapatkan data geospasial yang ada pada Twitter. Proses ekstraksi dan analisis dilakukan melalui empat tahapan yaitu: proses penarikan data (*crawling*), penyimpanan (*storing*), analisis (*analyzing*), dan visualisasi (*visualizing*). Penelitian ini bersifat *exploratory* yang terfokus pada pengembangan teknik ekstraksi dan analisis terhadap data geospasial twitter.

Kata kunci: media sosial, data analytics, social media analytics, twitter data analytics, data mining, machine learning

1. Pendahuluan

Perkembangan media sosial yang semakin pesat,

memberikan kemudahan bagi penggunanya untuk melakukan komunikasi. Selain untuk media komunikasi, media sosial juga telah banyak dimanfaatkan pada berbagai bidang seperti sosial, politik, ekonomi, pertahanan, keamanan dan lain-lain melalui *social media analytics* (SMA) [1] [2] [3][4][5][6]. Berdasarkan survei yang dilakukan oleh Global Web Index pada Januari 2014, pengguna internet di Indonesia mencapai

Received: 30 Nov 2015; Revised: 22 Juni 2016; Accepted: 17 Juni 2016; Published online: 21 Nov 2016 ©2016 INKOM 2016/16-NO433 DOI: <http://dx.doi.org/10.14203/j.inkom.433>

72.700.000 pengguna dari total jumlah penduduk Indonesia sebanyak 251.160.124 jiwa. Survei tersebut juga menunjukkan pengguna aktif media sosial mencapai 79,7% dari total pengguna internet di Indonesia [7].

Perkembangan teknologi Web 2.0 menjadi titik awal bermunculannya jenis media sosial. Kemampuan kolaborasi, interaksi, komunikasi dua arah dan jejaring pertemanan dengan sesama pengguna media sosial serta kegiatan - kegiatan seperti, percakapan (*chat*), memberi komentar (*comment/retweet*) dan respon suka atau tidak suka terhadap sebuah *posting*, dll. menjadi salah satu penyebab lonjakan yang tinggi terhadap jumlah pengguna media sosial dan popularitas situs media sosial [8]. Rilis hasil survei yang dilakukan oleh Pew Research Center dari tahun 2012 sampai tahun 2014 tentang jumlah pengguna media sosial, hasil tersebut menunjukkan bahwa media sosial yang paling yang paling populer digunakan oleh pengguna dengan umur diatas 18 tahun adalah Facebook (71%), LinkedIn (28%), Pinterest (28%), Instagram (26%), dan Twitter (23%) [8].

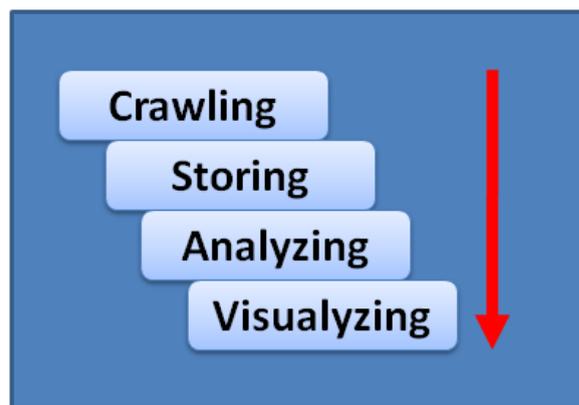
Dari beberapa media sosial yang ada, Twitter merupakan salah satu media sosial yang populer digunakan. Dalam waktu yang tergolong singkat, Twitter sanggup memikat hati banyak pengguna internet. Saat ini, tercatat lebih dari 500 juta pengguna Twitter dari seluruh dunia, dan nama Indonesia masuk dalam peringkat kelima dengan jumlah user sebanyak 29 juta akun [9], sedangkan Amerika Serikat berada di peringkat pertama dengan jumlah pengguna mencapai 140 juta orang [10]. Besarnya pertumbuhan pengguna twitter dari tahun ke tahun berdampak terhadap semakin banyaknya data yang dihasilkan, fenomena ini disebut dengan *Big Data*.

Fenomena *Big Data* yang dihasilkan oleh media sosial dapat berupa persepsi publik, perilaku sosial masyarakat, titik geospasial dari pengguna media sosial (*location*), dll. E.S. Negara dan P.H. Saksono (2015) melakukan analisis media sosial Twitter mengenai peristiwa jatuhnya pesawat AirAsia QZ8501 yang terjadi pada tanggal 28 Desember 2014 di selat Karimata, Indonesia. Penelitian tersebut memperlihatkan hasil sumber negara (*location*) munculnya *hashtag* dan komentar pada Twitter terkait dengan peristiwa tersebut. Tidak hanya negara, dalam penelitian ini juga memperlihatkan rentang usia para pengguna Twitter yang aktif dalam mengamati peristiwa ini [11]. Selain itu, jenis kelamin dan tipe pengguna Twitter, yaitu apakah mereka (pengguna Twitter) berasal dari organisasi atau merupakan personal ditunjukkan pada hasil penelitian ini [11].

Besarnya manfaat yang didapatkan melalui *social media analytics* memberikan peluang untuk meneliti sumber geospasial dari setiap data media sosial. Data Geospasial merupakan data

tentang lokasi geografis, dimensi atau ukuran, dan karakteristik objek alam atau buatan manusia yang berada di bawah, pada, atau di atas permukaan bumi [12]. Analisis dan Ekstraksi data geospasial pada Twitter dapat dimanfaatkan untuk melihat sumber lokasi persepsi publik dan perilaku sosial masyarakat berasal terhadap sebuah isu, sehingga informasi ini dapat dimanfaatkan oleh berbagai pihak yang berkepentingan.

Artikel ini membahas tentang ekstraksi dan analisis data geospasial Twitter terhadap suatu isu publik yang sedang berkembang dan mengembangkan prototipe perangkat lunak yang digunakan untuk mendapatkan data geospasial yang ada pada Twitter. Proses ekstraksi dan analisis dilakukan melalui empat tahapan yaitu: proses penarikan data (*crawling*), penyimpanan (*storing*), analisis (*analyzing*), dan visualisasi (*visualizing*) [13], lihat Gambar 1. Dalam penelitian ini penulis menggunakan metode penelitian deskriptif untuk mendeskripsikan lokasi dari fenomena yang terjadi diseluruh dunia melalui media sosial twitter. Dengan menganalisa data *tweet* yang di *posting* oleh pengguna Twitter seputar informasi terhangat yang terjadi di seluruh dunia.



Gambar 1. Proses Ekstraksi dan Analisis. [13]

2. Tinjauan pustaka

2.1. Media sosial

Media sosial sebagai sebuah kelompok aplikasi berbasis internet yang dibangun di atas dasar ideologi dan teknologi Web 2.0, dan memungkinkan penciptaan dan pertukaran *user-generated content*. Web 2.0 menjadi *platform* dasar media sosial. Media sosial ada dalam berbagai bentuk yang berbeda, termasuk *social network*, *forum internet*, *weblogs*, *social blogs*, *micro blogging*, *wikis*, *podcasts*, gambar, video, *rating* dan *bookmark social* [14].

Internet dan web 2.0 menyediakan suatu *platform* yang digunakan untuk meningkatkan pelayanan yang dapat digunakan untuk: membuat dan berbagi pemikiran dan cerita (*Blogger* dan *Twitter*); berbagi informasi dan

links (*Delicious*, *Digg* dan *Twine*); berbagi multimedia (*Youtube* dan *Flickr*); membuat dan berbagi pengetahuan (*Wikipedia*, *Yahoo Answer* dan *SlideShare*) dan membuat dan berbagi relasi (*Facebook*, *MySpace* dan *LinkedIn*) oleh grup-grup yang besar. Layanan inilah yang secara bersama-sama dikenal sebagai sosial media [15].

Sosial media merupakan suatu *platform* yang memberikan pelayanan dua arah yaitu membuat dan berbagi yang digunakan sebagai alat komunikasi baru di dalam era digital yang dapat membentuk jaringan pada komunitas yang memungkinkan untuk berkomunikasi secara *online* untuk membuat, mengatur, mengedit, mengomentari, menandai, mendiskusikan menggabungkan, mengkoneksikan dan bertukar informasi apapun di dalamnya. Salah satu jenis sosial media yang populer saat ini adalah *Twitter*. *Twitter* merupakan sebuah *microblogging* yang dapat mengirim pesan hingga 140 karakter secara instan melalui berbagai *platform*. 90% interaksi *Twitter* bukan berasal dari *website Twitter* melainkan dari *sms mobile*, pesan instan atau aplikasi desktop [16].

Saat ini beragam jenis media sosial seperti : *social networks*, *blogs*, *wikis*, *podcast*, *forums*, *content communities*, *microbloging*, dll. dapat digunakan untuk berbagai tujuan tertentu [16], [17]. Dengan mengimplementasikan teori *social presence*, *media richness* dan *social processes*, Kaplan dan Haenlein mengklasifikasikan media sosial menjadi enam jenis yaitu : 1) *Collaborative projects*, 2) *Blogs and microblogs*, 3) *Content communities*, 4) *Social networking sites*, 5) *Virtual game worlds*, dan 6) *Virtual communities* (Kaplan and Haenlein,2010). Sedangkan jika dilihat berdasarkan kategorinya, media sosial dibagi menjadi empat kategori yaitu : 1) *Social Networking*, 2) *Social Colaboration*, 3) *Social Publishing*, dan 4) *Social Feedbacks* [18].

2.2. Social media analytics

Social Media Analytics (SMA) merupakan kegiatan yang berkaitan dengan pengembang dan evaluasi *tools* informatika dan *framework* untuk mengumpulkan, memantau, menganalisis, merangkum dan memvisualisasi data media sosial [19], [20]. Gartner Research juga mendefinisikan SMA merupakan proses pemantauan, analisis, mengukur dan memprediksi interaksi digital, relationships, topik, ide atau konten pada media sosial [21]. SMA bertujuan untuk melakukan proses analisis dan sintesis data media sosial sehingga menghasilkan informasi yang dapat dipergunakan oleh pihak-pihak yang memerlukan. Proses SMA dilakukan melalui tiga tahapan, yaitu : *capture*, *understand* dan *present* [20]. Tahapan SMA dapat dilihat pada Gambar 1.

Tahapan *Capture* pada proses SMA merupakan proses mengumpulkan data media sosial yang relevan dengan kebutuhan dengan cara *collecting*

data menggunakan *crawler tools* yang terkoneksi melalui *Application Programming Interface* (API) ke media sosial seperti: Facebook, Twitter, LinkedIn, YouTube, Pinterest, Google+, Tumblr, Foursquare, *Internet forums*, *blogs* dan *microblogs*, Wikis, *news sites*, *picture sharing sites*, *podcasts*, and *social bookmarking sites*, dll. Data yang dihasilkan dari proses *Capture* disimpan kedalam basis data dan dipersiapkan untuk proses berikutnya yaitu *Understand*. Pada tahapan ini data juga diproses untuk menghasilkan informasinya yang sesuai dengan kebutuhan, termasuk memodelkan bentuk datanya [19].

Setelah menyelesaikan proses *Capture*, tahapan berikutnya adalah proses *Understand*. Proses *Understand* pada SMA merupakan proses pemilihan data yang relevan untuk melakukan pemodelan data, menghilangkan *noise* yang terdapat pada data, menseleksi data yang berkualitas dan melakukan proses analisis untuk memperoleh informasi yang lebih berkualitas [19]. Proses analisis data pada tahapan ini menggunakan metode statistik, *text mining*, *data mining*, *natural language processing* (NLP), *machine translation*, *machine learning* dan *network analysis* [22]. Beberapa teknik analisis data media sosial yang dapat digunakan untuk menghasilkan informasi spesifik antara lain : *Opinion mining (or sentiment analysis)*, *Topic modeling*, *Social network analysis*, *Trend analysis*, dan *Visual analytics*[19].

Tahapan terakhir dari proses SMA adalah *Present*. Proses *Present* merupakan proses untuk menampilkan atau memvisualisasikan informasi yang dihasilkan dari tahap *Understand* [19]. Berbagai teknik visualisasi dapat digunakan untuk menampilkan informasi yang didapatkan dari proses analisis.

2.3. Twitter data analytics

Shamanth Kumar, Fred Morstatter dan Huan Liu menyebutkan ada beberapa proses dalam melakukan *Twitter Data Analytics* yaitu [13]:

- Penarikan data Twitter (*Crawling Twitter data*).
- Penyimpanan data Twitter (*Storing Twitter data*).
- Analisis data Twitter (*Analyzing Twitter data*).
- Visualisasi data Twitter (*Visualizing Twitter data*).

Berdasarkan *tweet* yang dihasilkan setiap harinya oleh pengguna Twitter, dapat menjadi suatu sumber informasi sehingga dapat dilakukan proses *crawling* data Twitter dengan menggunakan *API Public* yang telah disediakan oleh Twitter. *Application Program Interface* (API) untuk mengakses data twitter dapat diklasifikasikan menjadi dua tipe berdasarkan desain dan metode akses, yaitu :

- a. REST API berdasarkan pada arsitektur REST yang sekarang digunakan mendesain web API. Tipe API ini menggunakan *pull strategy* untuk mendapatkan kembali informasi data. Untuk mengumpulkan informasi seorang pengguna harus secara eksplisit memintanya.
- b. Streaming API menyediakan aliran informasi publik yang berkelanjutan dari Twitter. Tipe API ini menggunakan *push strategy* untuk mendapatkan kembali informasi data. Satu kali melakukan permintaan informasi, *Streaming* API menyediakan aliran data yang selalu *update* tanpa *input* lebih lanjut dari pengguna.

Aliran API publik yang paling baik untuk digunakan dalam berbagai hal adalah *streaming* API sehingga dalam penarikan data Twitter akan digunakan tipe API ini. API Twitter hanya bisa diakses melalui permintaan otentikasi. Twitter menggunakan *Open Authentication* (OAuth) dan setiap permintaan harus dilakukan oleh pengguna Twitter yang sah. Akses ke dalam API Twitter dibatasi pada jumlah tertentu yang disebut *rate limit*. Batasan ini diterapkan pada tingkatan pengguna dan tingkatan aplikasi [13].

Pada tahapan sebelumnya telah dilakukan penarikan data Twitter dengan jumlah yang sangat besar. Setelah dilakukan penarikan, hal selanjutnya yang perlu dilakukan adalah bagaimana menyimpan data yang sangat besar tersebut ke dalam suatu database yang nantinya data tersebut dapat digunakan kembali pada saat proses analisis data. Diperlukan suatu *database* yang dapat menyimpan data besar tersebut. Pada proses penyimpanan ini digunakan *database* yang menggunakan konsep NoSQL (*Not Only SQL*) yang dapat menyimpan data dalam jumlah yang besar dengan cara akses yang lebih mudah dari cara tradisional, model relasi. Ada beberapa implementasi NoSQL. Dalam hal ini digunakan MongoDB, yang memberikan beberapa kelebihan sebagai berikut [13]:

- a. Penyimpanan berorientasi dokumen. MongoDB menyimpan data dalam bentuk JSON. Hal ini membuat sangat mudah untuk menyimpan dokumen atau data mentah dari API Twitter.
- b. Mendukung index. MongoDB juga mengizinkan untuk melakukan *index* dalam berbagai *field*, yang dapat membuat lebih mudah untuk menciptakan optimasi *index* pada aplikasi.
- c. *Straightforward Queries*. Query MongoDB, secara sintaks berbeda dari SQL, namun hampir sama secara semantik. Sebagai tambahan, MongoDB mendukung MapReduce, yang lebih memudahkan dalam pencarian data.

Setelah proses *crawling data* dan *storing data* dilakukan kemudian selanjutnya melakukan *analyzing* data untuk mengetahui informasi tentang pengguna berdasarkan data yang telah dikumpulkan. Ketika pengguna berinteraksi di Twitter kemudian menghasilkan informasi mengenai jaringan, ketika mereka mempublikasikan *tweets*, kemudian menghasilkan informasi tekstual. *Tweet* sendiri memiliki informasi seperti lokasi pengguna tersebut, Sebagai tambahan, pada profil pengguna Twitter menjelaskan diri mereka sendiri, seperti nama dan website. Teknik visualisasi dapat membantu kita efisien menganalisis dan memahami bagaimana dan mengapa pengguna berinteraksi di Twitter [13].

2.4. *Microblogger twitter terminology*

Twitter merupakan salah satu media sosial yang sangat populer dan menempati ranking ke 8 pada *Alexa rank* [23]. Lahirnya Twitter berasal dari ide Jack Dorsey pada tahun 2006 yang melihat kebiasaan orang-orang ingin berbagi kegiatan mereka saat bersama orang lain [24]. Dalam pengembangan Twitter, Jack Dorsey mengkombinasikan pola komunikasi dari satu ke banyak untuk menjadi pola dasar komunikasi yang diimplementasikan pada Twitter. Hal ini memungkinkan pengguna Twitter untuk berbagi informasi kepada banyak orang.

Neppelenbroek *et al.* menggambarkan arsitektur pengembangan Twitter dengan menggunakan model "4+1" yang dikembangkan oleh Kruchten [24]. Model ini digunakan untuk mendeskripsikan arsitektur perangkat lunak yang terfokus pada *logical, process, physical, development* dan *scenario view*. Dengan *Kruchten view model*, Neppelenbroek *et al.* menggambarkan arsitektur Twitter dengan *Logical view, Process view, Physical view, Development view* dan *Scenario view* [24].

Arsitektur pengembangan Twitter dapat dilihat pada Gambar 2 [24]. Lapisan *Back-end Service* dari Twitter menyimpan semua *tweets* yang diunggah oleh anggota dengan menggunakan MySQL sebagai database penyimpanan data. Pada lapisan *Search Engine*, Twitter menggunakan Apaches Lucene. *Search Engine* pada Twitter menggunakan metode *inverted indexing*, metode ini memisahkan *tweets* menjadi kata-kata (*words of a sentence*). Lapisan *Middle Layer* pada arsitektur Twitter pada dasarnya digunakan sebagai sistem antrian, sehingga tidak membebani *Back-end Service*. Lapisan *Middle Layer* pertama kali diimplementasikan oleh Starling dengan menggunakan bahasa pemrograman *Ruby on Rails* [24].

3. Hasil dan pembahasan

3.1. Penarikan data (*Crawling data*)

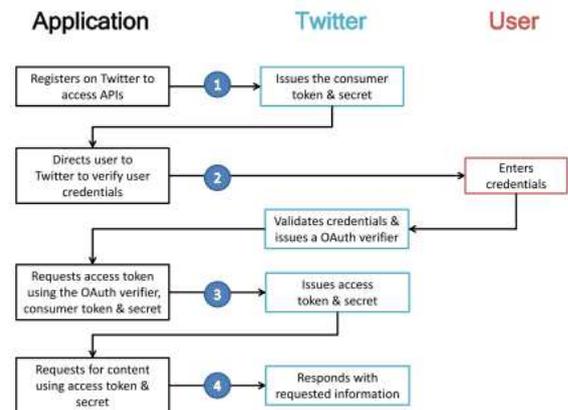
Proses penarikan (*crawling*) data Twitter dilakukan dengan memanfaatkan *Application Programming Interface* (API) yang telah disediakan oleh Twitter menghasilkan kumpulan data *text* berdasarkan update yang telah di unggah oleh pengguna Twitter. Selain itu proses penarikan ini juga menghasilkan data geospasial berupa titik koordinat yang bersumber dari lokasi pengguna Twitter pada saat melakukan proses *check in location* pada Twitter. Data vektor dan koordinat yang dihasilkan merupakan kebutuhan dari aplikasi yang dibangun dalam menentukan lokasi pengguna Twitter. Dari hasil data tersebut dapat dilihat sumber titik koordinat dari mana saja pengguna Twitter yang paling banyak membicarakan tentang fenomena atau peristiwa tersebut. Model data vektor menampilkan, menempatkan, dan menyimpan data spasial dengan menggunakan titik-titik, garis, atau polygon beserta atribut- atributnya. Bentuk- bentuk dasar representasi data spasial ini didalam sistem model data vektor didefinisikan oleh sistem koordinat kartesian dua dimensi (x, y). Pada model data vektor terdapat tiga entiti yaitu entiti titik, entiti garis, dan entiti polygon [12].

Dari penelitian yang telah dilakukan menghasilkan prototipe aplikasi *Crawler* dan kerangka kerja untuk *social media analytics* [11]. Penelitian ini tidak hanya dapat mengetahui negara pengguna Twitter saja, tetapi juga bisa mengetahui rentang usia pengguna Twitter, informasi jenis kelamin pengguna Twitter, dan penempatan data geospasial pengguna Twitter [11].

Dalam artikel ini, penulis melakukan penelitian berdasarkan berita yang menjadi *trending topic* pada bulan Juli 2015, yaitu diharamkannya Badan Penyelenggara Jaminan Sosial (BPJS) oleh Majelis Ulama Indonesia (MUI) pada tanggal 29 Juli 2015. Pada penelitian ini, penulis akan memperlihatkan sumber lokasi dari pengguna Twitter yang ikut mengamati tentang berita tersebut berdasarkan titik koordinatnya. Tidak hanya bahasa, penelitian ini juga memperlihatkan rentang usia para pengguna Twitter, informasi berdasarkan jenis kelamin pengguna Twitter, dan penempatan data geospasial pengguna Twitter yang akan ditunjukkan pada hasil penelitian ini. Proses analisis terhadap bahasa yang digunakan dan *user demografi* dilakukan dengan memanfaatkan *tools* *TexTalytics* [25].

Langkah yang pertama dilakukan dalam *Twitter data analytics* adalah *crawling* data twitter. Untuk melakukan *crawling* terlebih dahulu harus mendaftarkan aplikasi yang

dibuat ke Twitter untuk mendapatkan *user credential* yang nantinya digunakan dalam proses *crawling* data twitter. Proses ini bertujuan untuk mendapatkan *authentication* dari Twitter terhadap akses data yang dimiliki oleh Twitter. Proses *authentication* ini ditunjukkan pada Gambar 2. Setelah proses penarikan data Twitter berhasil dilakukan, langkah selanjutnya adalah menyimpan data tersebut ke dalam database MongoDB.



Gambar 2. Proses Authentication [13]

Pada saat *crawling* data ada beberapa faktor yang mempengaruhi yaitu koneksi jaringan internet, lamanya proses penarikan data dan *update* berita terbaru yang akan dilakukan *crawling*. Koneksi jaringan internet yang stabil akan memperlancar proses *crawling* data. Sebaliknya, apabila koneksi tersebut tidak stabil akan menyebabkan *error connection* atau proses penarikan data menjadi lambat dan terputus-putus. Faktor yang kedua adalah lamanya proses *crawling* data adalah semakin lama *crawling* data twitter dilakukan maka akan semakin banyak data yang didapatkan dari proses ini.

Dan faktor yang ketiga adalah *update* berita terbaru. Hal ini dikarenakan *crawling* data yang dilakukan twitter merupakan *realtime*. Untuk itu pada proses ini yang digunakan adalah *streaming API*. Dimana data yang diambil merupakan data *realtime*. Metode yang digunakan untuk menarik data tersebut adalah dengan menggunakan *REST-API*. Oleh karena itu, berita terbaru yang akan terjadi atau sedang terjadi, akan menjadi sangat mudah untuk diketahui perkembangannya.

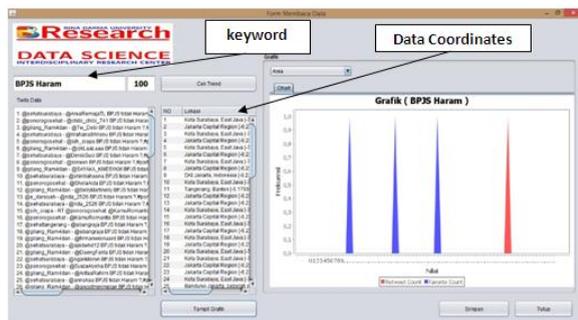
REST-API pada Twitter dapat digunakan untuk mengakses *status* atau *timelines* pengguna twitter. *REST-API* dapat mengambil 3.200 tweet terbaru dari pengguna, termasuk *re-tweet* [13].

- Parameter utama: Dalam setiap halaman, kita dapat mengambil 200 *tweet* dari pengguna.
- *Rate Limit*: Sebuah aplikasi diperbolehkan melakukan permintaan sebanyak 300 permintaan.

Penarikan data *coordinates* dan *location*

twitter dengan keyword BPJS Haram untuk menghasilkan data antara 10.000 - 11.000 tweets, dimana data yang memiliki informasi koordinat posisi atau lokasi pengguna Twitter sekitar 7500 tweets. Dari jumlah data yang dihasilkan proses penarikan lokasi (*location*) dilakukan pembatasan proses analisis lokasi *tweet* menggunakan aplikasi yang telah dibangun yakni sebanyak 200 *tweet* dalam setiap melakukan analisis.

Pada proses *crawling*, data *coordinates* dan *location twitter* data yang di tarik berupa *user_name*, *retweet_count*, *tweet_followers_count*, *source*, *tweet_mentioned_count*, *tweet_ID*, *tweet_text*, dan *coordinates* yang berisikan *longitude* dan *latitude* yang berguna untuk melakukan penempatan data geospasial. Hasil *crawling* data *coordinates* dapat dilihat pada Gambar 3.



Gambar 3. Proses *crawling* data *coordinates* dan *location twitter*

3.2. Penyimpanan data (*Storing data*)

Penyimpanan data merupakan tahap selanjutnya setelah melakukan proses penarikan data atau *crawling* data Twitter. Basis Data yang digunakan pada proses penyimpanan ini ialah MongoDB. Data yang berhasil disimpan kemudian dianalisis untuk mendapatkan data yang bersih yang bebas dari *noise*. Data yang telah bersih tersebut dapat dijadikan sebagai data untuk penelitian. Untuk mempermudah dalam melihat data yang telah dihasilkan sebagai informasi, maka data divisualisasikan ke dalam bentuk peta. Proses penyimpanan data harus dilakukan secara langsung atau *direct storing*. Hal ini dilakukan dikarenakan data yang ditarik merupakan *realtime data twitter*. Sehingga diperlukan suatu *database* yang memungkinkan untuk menyimpan data secara langsung.

3.3. Analisis Data (*Analyzing Data*)

Analisis data merupakan tahap selanjutnya setelah selesai melakukan proses penyimpanan data atau *storing data*. Analisis data merupakan bagian penting, terutama dalam pengolahan data. Pengolahan data ini dilakukan agar data yang didapat tidak mengandung *file* kosong atau *null*,

apabila terdapat *file* tersebut maka analisis data tidak mendapatkan hasil. Setelah data selesai dianalisis, selanjutnya dari hasil analisis tersebut bisa dilakukan tahapan visualisasi. Perangkat lunak yang digunakan pada tahapan analisis ini ialah *textalytics* untuk menganalisis data yang telah tersimpan ke dalam *database*. *Textalytics* berfungsi untuk menganalisis suatu data ke dalam beberapa bentuk kategori seperti *text classification*, *sentiment analysis*, *language identification*, *user demographics*, *topic extraction*. Dalam penelitian ini fungsi dari *textalytics* yang digunakan ialah *language identification* dan *user demographics*.

Selain analisis terhadap *language identification* dan *user demographics* dilakukan, proses ini juga melakukan pemetaan terhadap titik *logitude* dan *latitude* dari data Twitter yang telah tersimpan pada basis data. Titik *logitude* dan *latitude* menjadi informasi yang didapat dari proses *crawling* yang divisualisasikan dalam bentuk peta. Hasil data titik koordinat yang menunjukkan sumber persepsi publik tentang BPJS haram dapat dilihat pada Gambar 4.



Gambar 4. Hasil data geospasial tentang BPJS Haram

Gambar 4 menunjukkan hasil data geospasial yang paling banyak mengikuti perkembangan berita tentang diharamkannya Badan Penyelenggara Jaminan Sosial (BPJS) oleh Majelis Ulama Indonesia (MUI) pada tanggal 29 Juli 2015. Dengan keyword BPJS Haram adalah Kota Surabaya, diikuti kota Jakarta. Kota Surabaya yang paling banyak mengamati berita tersebut dikarenakan *tweet* yang diunggah oleh pengguna hampir semuanya di *re-tweet* berbeda dengan Kota Jakarta yang penggunaanya tidak hanya melakukan *re-tweet* tetapi ada juga yang memberikan pendapat dan saran.

3.4. Visualisasi Data (*Visualizing Data*)

Visualisasi merupakan suatu cara untuk mengkonversi data ke dalam format visual atau tabel sehingga karakteristik dari data dan relasi di antara item data atau atribut dapat dianalisis atau dilaporkan. 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.

Setelah proses analisis selesai maka tahapan selanjutnya adalah visualisasi data tersebut. Visualisasi data dimaksudkan agar data yang dihasilkan dari proses analisis terlihat lebih menarik serta dapat lebih mudah dipahami sebagai suatu informasi.

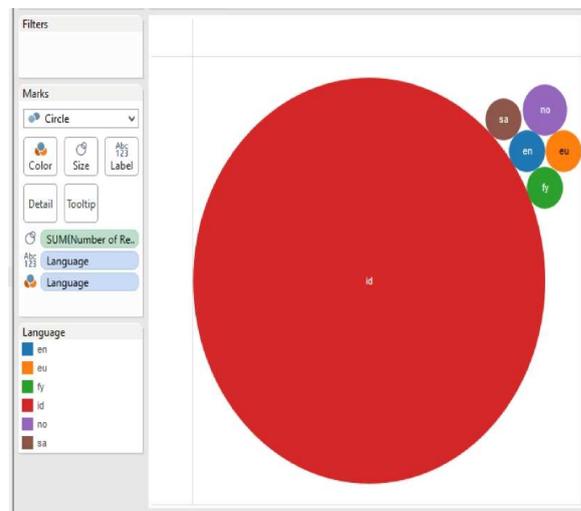
Untuk mempermudah dalam melihat data yang telah dihasilkan sebagai informasi, maka data divisualisasikan ke dalam bentuk *bubble graph* atau *graph* lainnya sesuai dengan kebutuhan. Seperti tujuan penelitian yang telah disebutkan sebelumnya, maka hasil dari penelitian ini adalah melakukan *crawling* data twitter dengan memanfaatkan *Application Programming Interface* (API) yang telah disediakan oleh Twitter. Data dari Twitter tersebut akan diolah menjadi suatu informasi yang dapat digunakan sebagai bahan penelitian.

Informasi yang telah didapatkan tersebut juga merefleksikan bagaimana perilaku masyarakat terhadap suatu peristiwa yang sedang terjadi pada kehidupan nyata yang dituangkan pada sosial media. Hal ini akan menunjukkan apakah peristiwa tersebut mempunyai pengaruh pada masyarakat global. Berpengaruhnya peristiwa tersebut dapat terlihat dari *tweets* yang di-*update* oleh pengguna Twitter. Hal ini akan terlihat dari bahasa yang digunakan, bahasa tersebut akan menunjukkan kemungkinan dari negara mana pengguna Twitter tersebut berasal. Dari hasil analisis data twitter yang membicarakan tentang BPJS haram, terlihat bahwa pengguna Twitter yang banyak memberikan perhatian terhadap kejadian tersebut bersumber dari negara Indonesia.

Selain bahasa yang digunakan, perilaku masyarakat juga dapat dilihat dari usia pengguna Twitter. Dari suatu peristiwa yang terjadi dapat dilihat rentang usia pengguna Twitter yang aktif terhadap suatu peristiwa. Selain itu, apakah pengguna Twitter tersebut merupakan organisasi atau perseorangan serta jenis kelamin pengguna Twitter itu sendiri dapat diketahui melalui hasil penelitian ini.

Dalam penelitian ini, penulis melakukan penelitian sumber geospasial persepsi publik tentang BPJS haram. Gambar 5 menunjukkan negara mana yang paling banyak mengikuti perkembangan berita tentang diharamkannya Badan Penyelenggara Jaminan Sosial (BPJS) oleh Majelis Ulama Indonesia (MUI) pada tanggal 29 Juli 2015. Dengan *keyword* BPJS Haram ialah Indonesia (id) dengan lingkaran berwarna merah dengan jumlah 189 *user*, diikuti oleh negara Norway (no) dengan lingkaran berwarna ungu dengan jumlah 3 *user*. Indonesia sebagai negara yang paling banyak mengamati berita tersebut dikarenakan sumber berita tersebut memang berasal dari Indonesia. Sedangkan untuk negara

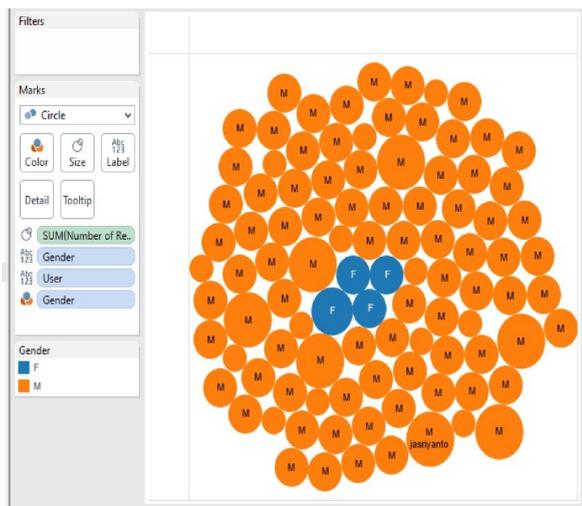
lainnya dapat mengetahui perkembangan berita tersebut dengan melihat dari *trending topic* pada halaman depan Twitter yang berkaitan dengan BPJS HARAM.



Gambar 5. Negara sumber data geospasial tentang BPJS Haram

Tidak hanya negara, dalam penelitian ini juga memperlihatkan rentang usia para pengguna Twitter yang aktif dalam mengamati peristiwa ini. Rentang usia yang paling banyak mengikuti perkembangan berita tentang diharamkannya Badan Penyelenggara Jaminan Sosial (BPJS) oleh Majelis Ulama Indonesia (MUI) pada tanggal 29 Juli 2015. Dengan *keyword* BPJS Haram adalah pengguna Twitter dengan kisaran usia 15-24 tahun dan kisaran usia 45-54 tahun, masing-masing berjumlah 48 *user*. Pengguna Twitter dengan kisaran usia 15-24 tahun dan kisaran usia 45-54 tahun yang paling banyak mengamati berita tersebut dikarenakan pada kedua kisaran usia tersebutlah yang lebih cenderung pada perkembangan dunia politik berbeda dengan kisaran usia 25-54 yang lebih sedikit dikarenakan pada usia ini lebih tertarik pada dunia olahraga, hiburan, dan perkantoran. Selain itu, jenis kelamin dan tipe pengguna Twitter, yaitu apakah mereka (pengguna Twitter) berasal dari organisasi atau merupakan personal ditunjukkan pada hasil penelitian ini. Mayoritas yang paling banyak mengikuti perkembangan berita tentang diharamkannya Badan Penyelenggara Jaminan Sosial (BPJS) oleh Majelis Ulama Indonesia (MUI) pada tanggal 29 Juli 2015. Dengan kata kunci BPJS Haram adalah laki-laki. Bisa dilihat pada Gambar 6, M = *Male* yaitu laki-laki dengan lingkaran berwarna jingga dengan jumlah 191 *user*. Dan F = *Female* yaitu perempuan dengan lingkaran berwarna biru dengan jumlah 9 *user*. M = *Male* yaitu laki-laki merupakan mayoritas yang paling banyak mengamati berita tersebut dikarenakan laki-laki pada umumnya lebih tertarik pada dunia politik. Pengguna Twitter berdasarkan

jenis kelamin yang paling banyak membicarakan tentang BPJS haram dapat dilihat pada Gambar 6.



Gambar 6. Informasi jenis kelamin pengguna Twitter tentang BPJS Haram

4. Kesimpulan

Dari penelitian yang dilakukan terhadap *Twitter Data Analytics*, dapat disimpulkan bahwa :

- Proses *crawling* terhadap data twitter dengan memanfaatkan *Application Programming Interface* telah berhasil dilakukan dan menghasilkan data yang informatif melalui proses *Crawling*, *Storing*, *Analyzing* dan *Visualizing*.
- Berdasarkan data yang telah ditarik tersebut dapat diketahui negara asal pengguna twitter, informasi berdasarkan geospasial, jenis kelamin pengguna twitter, rentang usia pengguna twitter, dan penempatan data berdasarkan *tweet* yang telah di unggah.

Ucapan terima kasih

Ucapan terima kasih penulis sampaikan kepada *Data Science Interdisciplinary Research Center* atas dukungan sarana dan prasarana penelitian. Kepada Octavia Dwi Cahyanti dan Sutami Suweno yang telah membantu pada proses pengembangan prototipe aplikasi data retrieval.

Daftar pustaka

[1] UN Global Pulse, "Mining Indonesian tweets to understand food price crises," UN Global Pulse, Methods Paper, 2014.

[2] S. Stieglitz, T. Brockmann, and L. Dang-Xuan, "Usage of social media for political communication." in PACIS, 2012, p. 22.

[3] C. Holsapple, S. Hsiao, and R. Pakath, "Business social media analytics: Definition, benefits, and challenges," in Proceedings of the 20th Americas conference on Information Systems

(AMCIS2014), Association for Information Systems. Association for Information Systems, 2014.

[4] M. Rosemann, M. Eggert, M. Voigt, and D. Beverungen, "Leveraging social network data for analytical CRM strategies – the introduction of social BI," in Proceedings of the 20th European conference on information systems, Barcelona, Spain, 2012, p. 95.

[5] M. D. Sykora, T. W. Jackson, A. O'Brien, and S. Elayan, "National security and social media monitoring: A presentation of the emotive and related systems," in Intelligence and Security Informatics Conference (EISIC), 2013 European. IEEE, 2013, pp. 172–175.

[6] D. O'Callaghan, D. Greene, M. Conway, J. Carthy, and P. Cunningham, "An analysis of interactions within and between extreme right communities in social media," arXiv preprint arXiv:1206.7050, 2012.

[7] Global Web Index, "Survei data global web index," 2014. [Online]. Available: <https://www.globalwebindex.net/>

[8] M. Duggan, N. Ellison, C. Lampe, A. Lenhart, and M. Madden, "Social media update 2014," Pew Research Center, 2015.

[9] S. SemioCast 2nd, "Brazil becomes 2nd country on twitter, Japan 3rd, Netherlands most active country," 2013.

[10] E. Diaz-Aviles, A. Stewart, E. Velasco, K. Denecke, and W. Nejdl, "Epidemic intelligence for the crowd, by the crowd." in ICWSM, 2012.

[11] E. Negara and P. Saksono, "Social media analytics Data utilization of social media for research," MAKARA (Draft Usulan), 2015.

[12] R. Indonesia, "Undang undang no. 4 tahun 2011 tentang informasi geospasial," Lembaran Negara RI Tahun 2011 No. 49. Sekretariat Negara, 2011.

[13] S. Kumar, F. Morstatter, and H. Liu, Twitter Data Analytics, 2013. [Online]. Available: www.tweettracker.fulton.asu.edu

[14] A. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," Business Horizons, vol. 53, 2010, pp. 59–68.

[15] R. Brussee and E. t. Hekman. (2015) Social media are highly accessible media.

[16] Antony, What is social media? Icrossing, 2008.

[17] J. Sterne and D. M. Scott, Social Media Metrics: How to Measure and Optimize Your Marketing Investment. John Wiley, March 2010.

[18] A.J. Bradley., Becoming a social organization: Taking a strategic approach to social media. Gartner Inc., 2010.

[19] W. Fan and M.D. Gordon, "The Power of social media analytics", Communication of ACM, vol

57, no. 6, 2014, pp.74-81.

- [20] D.Zheng, H.Chen, R. Lusch, and S.H.Li, “Social media analytics and intelligence”, *Intelligent Systems, IEEE*, vol 25, no.6, 2010, pp.13-16.
- [21] Gartner Research, “Social analytics”. [Online]. Available: <http://www.gartner.com/it-glossary/social-analytics>.
- [22] W.Fan, L. Wallace, S.Rich, and Z.Zhang, “Tapping the power of text mining”, *Communication of the ACM*, vol 49, no. 9, 2006, pp.76-82.
- [23] Alexa.com, “Alexa internet.twitter.com”[Online] Available: <http://www.alexa.com/siteinfo/twitter.com>, 2015.
- [24] M.Neppelenbroek, M.Lossek, R.Janssen, and T.de Boer, “Twitter an architectural review”, 2011.
- [25] A.Abbasi, A.Hassan, and M.Dhar, “Benchmarking twitter sentiment analysis tools”. In *LREC*, 2014, pp.823-829.

