

KAJIAN TERHADAP TOOLS DAN FRAMEWORK SOCIAL MEDIA ANALYTICS UNTUK PEMANFAATAN DATA SOCIAL MEDIA DALAM PENELITIAN ILMU SOSIAL

Edi Surya Negara¹

¹Fakultas Ilmu Komputer, Universitas Bina Darma Palembang

Masuk : 11 Oktober 2016, revisi masuk : 18 Desember 2016, diterima : 19 Januari 2017

ABSTRACT

The increasing of using social media in Indonesia has the impact to the available of data about society perception toward many issues of life, even though it comes as member/personality of live and citizen in the society. These phenomena have given an opportunity for stakeholders, such as government and private that need information about the society perception or social attitudes in their social life to improve quality and quantity of information through the using of the social media. The well understanding about technology of internet, social media, databases, data structure, theories of information, data mining, machine learning, and the technique for visualization data and information are needed to get and analyze the data to determine certain information which will be used by the users (stakeholders of the information). The target of this research is to master technology of social media analytic and develop prototype of software that will be used as tools of collecting data in social media, therefore, the users of the data can focus to get comprehension/understanding about social phenomena and take decision without looking for techniques for collecting and data analysis issues. This research is designed into explanatory with focus to the understanding technology through basis of social media, moreover, it is used to review advantages and disadvantages of techniques currently used in social media and media analytic researches. The result of this research is the tools and framework that benefit for media social analysis.

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

INTISARI

Tingginya tingkat penggunaan media sosial di Indonesia berdampak pada tersedianya data tentang persepsi masyarakat terhadap berbagai isu, baik itu isu kehidupan sebagai anggota kelompok sosial ataupun kehidupan berbangsa dan bernegara, dalam jumlah yang sangat besar. Fenomena ini membuka peluang, baik pemerintah maupun swasta, yang memerlukan informasi tentang persepsi masyarakat atau perilaku sosial masyarakat dalam kegiatan rutinnnya untuk meningkatkan kualitas dan kuantitas informasi dengan memanfaatkan data media sosial. Pemahaman yang mendalam tentang teknologi internet, media sosial, basis data, struktur data, teori informasi, *data mining*, *machine learning*, sampai teknik visualisasi data dan informasi diperlukan dalam pengumpulan dan pengolahan data sebelum data tersebut dianalisis dan sintesis menjadi informasi yang akan digunakan oleh pihak-pihak yang memerlukan. Sasaran penelitian ini adalah menguasai teknologi analisis media sosial dan mengembangkan prototipe perangkat lunak sebagai alat bantu dalam pengumpulan data media sosial sehingga pihak-pihak yang menggunakan data tentang persepsi masyarakat dan perilaku sosial masyarakat bisa memfokuskan kegiatannya pada usaha memahami fenomena sosial masyarakat dan pengambilan keputusan, tanpa harus direpotkan dengan isu-isu teknis pengumpulan dan pengolahan data. Penelitian ini bersifat *exploratory* terfokus pada pemahaman teknologi yang menjadi basis media sosial dan melakukan kajian tentang kelebihan dan kekurangan dari berbagai teknik yang ada dan sudah digunakan dalam penelitian *social media analytic* sebelumnya.

Penelitian ini menghasilkan *tools* dan *framework* yang dapat digunakan untuk melakukan analisis sosial media.

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

PENDAHULUAN

Berdasarkan survei yang dilakukan oleh Global Web Index pada Januari 2014 (Global Web Index, 2014), pengguna internet di Indonesia mencapai 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 (Global Web Index, 2014), Mengingat jumlah pengguna media sosial di Indonesia yang sangat besar dan terus bertambah menjadikan pemanfaatan data media sosial sangat ideal untuk melengkapi data yang dihasilkan dari metode pengumpulan data konvensional seperti survei dan sensus yang memerlukan biaya besar dan memakan waktu yang lama terutama untuk kegiatan penelitian dalam bidang ilmu sosial yang memerlukan data tentang persepsi dan/atau perilaku sosial masyarakat.

Di Indonesia sendiri sejak Oktober 2012 pemerintah, melalui badan Badan Perencanaan dan Pembangunan Nasional (BAPPENAS), telah memanfaatkan data media sosial untuk mengetahui besarnya permasalahan, sentimen dan suasana hati (*mood*) masyarakat berkaitan dengan harga dan ketersediaan pangan dan bahan bakar (Global Web Index, 2014).

Tantangan yang dihadapi dalam memanfaatkan data media sosial antara lain: masing-masing situs media sosial menggunakan platform yang berbeda, volume, kompleksitas dari informasi, dan data yang tidak terstruktur (Stieglitz and Linh, 2013). SMA menghadapi tantangan ini dengan menyediakan perangkat (*tools*) dan kerangka kerja (*framework*) untuk mengumpulkan, memantau, menganalisis, menyimpulkan, dan memvisualisasi data sosial media (Zeng, 2010)

Untuk dapat memanfaatkan data media sosial dalam mendukung kegiatan

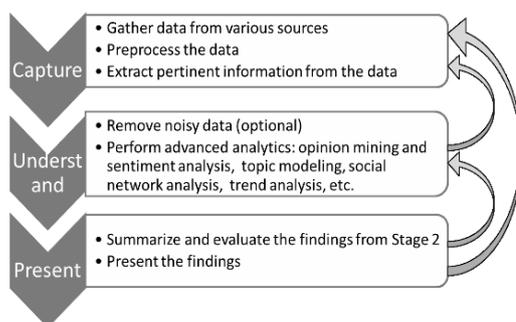
penelitian yang berhubungan dengan persepsi masyarakat dan perilaku sosial masyarakat diperlukan perangkat dan kerangka kerja yang dikembangkan secara khusus untuk mengumpulkan, memantau, menganalisis, menyimpulkan, dan memvisualisasikan data tersebut. Masalah yang perlu dijawab oleh penelitian ini adalah perangkat dan kerangka kerja seperti apa yang diperlukan untuk mendukung kegiatan penelitian masalah-masalah sosial menggunakan data media sosial. Mengingat kompleksitas dan berbagai jenis platform yang digunakan oleh masing-masing media sosial, maka penelitian ini hanya akan membahas SMA pada situs *microblogger Twitter* atau sering juga disebut dengan *Twitter Data Analytics*.

Keberadaan SMA diharapkan memperkaya informasi yang bersumber pada teknik pengumpulan data tradisional seperti survei dan sensus. Ketersediaan SMA *tools* yang mudah digunakan dan handal akan sangat membantu bagi pihak-pihak yang dalam kegiatan sehari-harinya menggunakan data tentang persepsi dan perilaku sosial masyarakat. Dengan demikian pihak-pihak tersebut bisa memfokuskan sumber daya mereka dalam memahami fenomena sosial masyarakat dan melakukan pengambilan keputusan tanpa harus direpotkan dengan isu-isu teknis pengumpulan dan pengolahan data.

Saat ini beragam jenis media sosial seperti: *Social networks, Blogs, Wikis, Podcast, Forums, Content communities, Microblogging* dan lain-lain dapat digunakan untuk berbagai tujuan tertentu ((Antony (2008), Sterne and Scott (2010)). Dengan cara-cara mengimplementasikan teori-teori *social presence, media richness* dan *social processes*, Kaplan dan Haenlein mengklasifikasikan media sosial menjadi enam jenis yaitu (Kaplan and Haenlein,

2010): 1) *Collaborative projects*, 2) *Blogs and microblogs*, 3) *Content communities*, 4) *Social networking sites*, 5) *Virtual game worlds*, dan 6) *Virtual communities*. 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* (Bradley, 2010).

Social Media Analytics (SMA) merupakan kegiatan yang berkaitan dengan pengembangan dan evaluasi *tools* informatika dan *framework* untuk mengumpulkan, proses memantau, proses menganalisis, merangkum dan memvisualisasi data media sosial (Zeng, 2010). Gartner Reasearch juga mendefenisikan SMA merupakan proses pemantauan, analisis, mengukur dan memprediksi interaksi-interaksi digital, relationships, topik, ide atau konten pada media sosial (Gartner Reasearch). 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* (Zeng, 2010). Tahapan SMA dapat dilihat pada Gambar 1.



Gambar 1. Social Media Analytics Process (Fan and Gordon (2014))

METODE

Objek dari kegiatan penelitian ini adalah data *tweet* dan *retweet* pengguna *Twitter* di Indonesia serta teknik-teknik yang digunakan dalam pengambilan, pemantauan, analisis, inferensi, dan visualisasi data tersebut. *Twitter* merupakan sebuah media sosial online

yang diluncurkan pada tahun 2006 oleh Jack Dorsey. Berbeda dengan beberapa pendahulunya seperti *Frienster*, *MySpace*, *YouTube*, *Bebo*, *Facebook*, dan lain-lain, *Twitter* memungkinkan penggunanya menyampaikan pesan pendek dengan jumlah karakter maksimum 140 karakter yang dikenal dengan istilah *tweet*. *Tweet* dapat dipost melalui berbagai cara seperti website *Twitter*, mobile application, atau aplikasi pihak ketiga setelah melakukan otentikasi.

Untuk dapat memanfaatkan data *tweet* yang dimiliki *Twitter* sebagai sumber data dalam penelitian ilmu-ilmu sosial, peneliti perlu terlebih dahulu mengetahui jenis dan karakteristik dari data tersebut dan bagaimana cara memperolehnya. Dari berbagai publikasi penelitian ilmu sosial yang menggunakan data *Twitter* (Zeng, 2010), secara umum penelitian-penelitian tersebut dapat dikelompokkan ke dalam penelitian yang menggunakan data *content* dan data jejaring (*network*) dengan menggunakan teknik-teknik yang berbeda.

Penelitian ini merupakan proses penelitian *exploratory* yang dapat diimplementasikan dengan melakukan proses penelitian sekunder (*secondary research*), dimana peneliti akan melakukan kajian ulang atas hasil penelitian-penelitian sebelumnya dalam bidang ilmu sosial yang menggunakan *tweet*, baik sebagai data primer maupun data sekunder. Untuk memastikan penelitian ini mencapai tujuan yang diharapkan maka penelitian ini dibagi ke dalam empat tahapan yang sekaligus merupakan milestone dari penelitian ini. Tahap pertama yang akan dilakukan oleh peneliti adalah proses mengumpulkan dan mengevaluasi berbagai *framework* dan rancangan hal sistem yang telah digunakan untuk mengumpulkan dan mengolah data pada penelitian terdahulu berdasarkan hasil yang telah dipublikasikan. Tahap kedua adalah memilih *framework* dan desain sistem yang terbaik berdasarkan hasil evaluasi untuk digunakan dalam penelitian ilmu sosial. Tahap ketiga difokuskan pada usaha untuk melakukan klasifikasi dan karakterisasi dari data

tweet yang diperoleh dengan menggunakan *framework* dan sistem yang telah dipilih. Tahap terakhir mencari teknik visualisasi yang sesuai untuk digunakan dalam merepresentasikan data *tweet*.

PEMBAHASAN

Social Media Analytics (SMA) bertujuan untuk melakukan proses analisis dan sintesis data media sosial sehingga menghasilkan informasi yang dapat dipergunakan oleh pihak-pihak yang memerlukan. Salah satu *framework* yang dapat digunakan untuk melakukan *social media analytics* (SMA) dilakukan melalui tiga tahapan, yaitu: *Capture*, *Understand* dan *Present* (Zeng, 2010).

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 proses *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*, dan *social bookmarking sites* dan lain-lain. Data yang dihasilkan dari proses *Capture* disimpan ke dalam *database* dan dipersiapkan untuk proses berikutnya yaitu *Undetstad*. Pada tahapan ini data juga diproses untuk menghasilkan informasi yang sesuai dengan kebutuhan, termasuk memodelkan bentuk datanya (Fan and Gordon, 2014).

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 (Fan and Gordon, 2014). Proses analisis data pada tahapan ini menggunakan metode statistik, *text mining*, *data mining*, *natural language processing* (NLP), *machine tranlation*, *machine learning*, dan *network analysis* (Fan and Gordon, 2014).

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* (Fan and Gordon, 2014).

Tahapan terakhir dari proses SMA adalah *Present*. Proses *Present* merupakan proses untuk menampilkan atau memvisualisasikan informasi yang dihasilkan dari tahap *Understand* (Fan and Gordon, 2014). Berbagai teknik visualisasi dapat digunakan untuk menampilkan informasi yang didapatkan dari proses analisis.

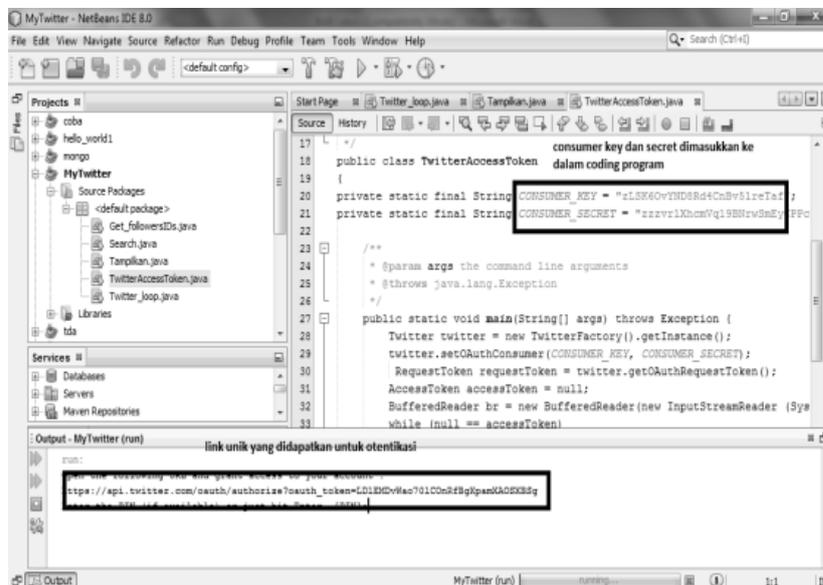
Dari penelitian yang telah dilakukan menghasilkan prototipe aplikasi *data retrieval* dan kerangka kerja untuk *social media analytics*. Pada penelitian ini, digunakan *Twitter* sebagai objek penelitian. Untuk menganalisa data *Twitter* diperlukan teknik-teknik khusus untuk menjadikan data yang berupa teks dari *Twitter* menjadi informasi yang selanjutnya dapat dijadikan bahan untuk penelitian.

Langkah yang pertama dilakukan dalam *social media analytics* adalah *retrieving* data *twitter*. Untuk melakukan *retrieving* terlebih dahulu harus mendaftarkan aplikasi yang dibuat ke *Twitter* untuk mendapatkan *user credential* yang nantinya digunakan dalam proses *retrieving* 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 berhasil dilakukan, langkah selanjutnya adalah menyimpan data tersebut ke dalam *database* MongoDB.

Pada proses *retrieving* ini, data yang diambil berupa *user_name*, *retweet_count*, *tweet_followers_count*, *source*, *tweet_mentioned_count*, *tweet_ID*, dan *tweet_text*. *User_name* merupakan nama pengguna *Twitter*, *retweet_count* yaitu berapa kali status tersebut di-*retweet* oleh pengguna lain, *tweet_followers_count* yaitu jumlah *follower* dari pengguna akun tersebut, *source* yaitu sumber media yang

digunakan untuk mengunggah tweets tersebut, *tweet_mentioned_count* yaitu berapa banyak tweet tersebut diikuti,

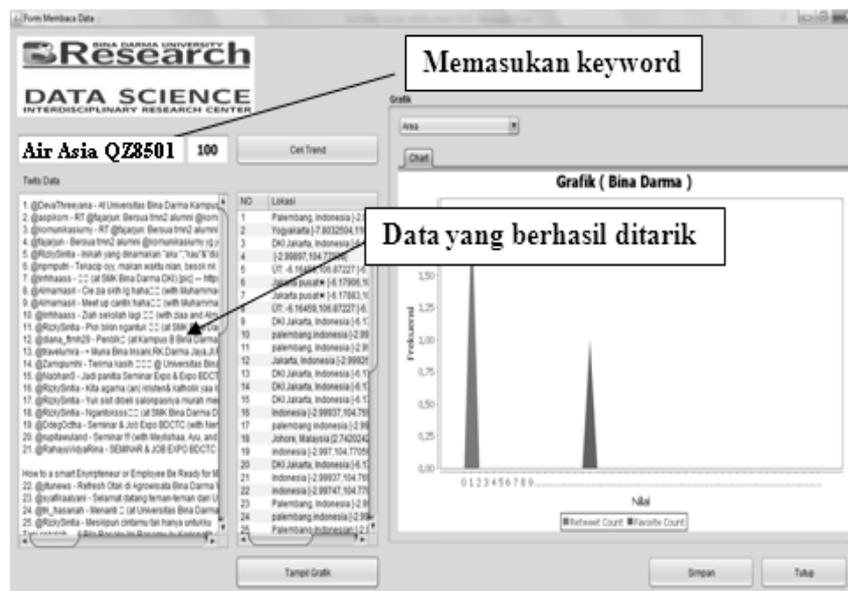
tweet_ID yaitu ID pengguna *Twitter* dan *tweet_text* yaitu isi dari *tweet* tersebut.



Gambar 2. Code PIN Authentication

Proses Understand Data, pada tahapan ini, data yang telah ditarik dari *twitter* dimasukkan ke dalam *database*. *Database* yang digunakan pada penelitian ini adalah MongoDB. Data yang berhasil disimpan kemudian dianalisis untuk mendapatkan data yang bebas dari *noise*.

Data yang telah bersih tersebut dapat dijadikan sebagai data penelitian. Untuk mempermudah dalam melihat data yang telah dihasilkan sebagai informasi, data divisualisasikan ke dalam bentuk *bubble graph* atau sesuai dengan kebutuhan.



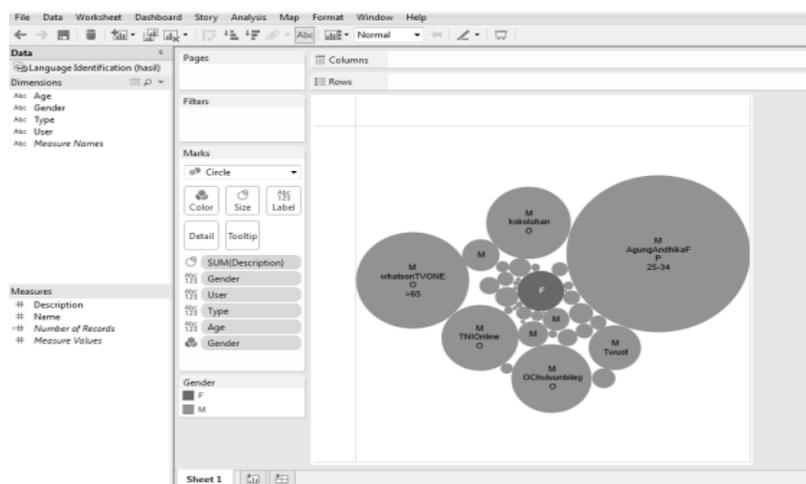
Gambar 3. Aplikasi Data Retrieval

Hasil *retrieving* data yang diperoleh dari proses penarikan data (*retrieving data*) merupakan data dalam bentuk teks atau dokumen. Lebih dari 100 data atau teks yang diambil pada saat melakukan *retrieving* data. Hal ini akan menyulitkan untuk membuat *primary key* secara manual karena pada saat *retrieving* data, data yang diterima harus langsung masuk ke dalam *database*.

Proses Present Data, pada proses 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 proses eksploitasi data.

Visualisasi dapat menggambarkan pola umum yang terjadi, trend yang sedang berkembang serta hal-hal yang tidak umum.

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 telah memanfaatkan *Application Programming Interface (API)* yang telah disediakan oleh *twitter*. Data tersebut diolah menjadi suatu informasi yang dapat digunakan sebagai bahan penelitian.



Gambar 4. Jenis Kelamin pengguna *Twitter*, Sumber Data *Twitter* (Air AsiaQZ88501)

KESIMPULAN

Pengembangan *tools* dan *framework social media analytics* pada penelitian ini telah mampu melakukan penarikan data dari media sosial *Twitter* dan mampu melakukan analisis dan visualisasi terhadap data informasi yang didapatkan. Kerangka kerja (*framework*) untuk melakukan *social media analytics* dilakukan melalui tahapan *capture data*, *understand data*, dan *present data*.

DAFTAR PUSTAKA

Kaplan, A.M. and Haenlein, M. (2010). *The early bird catches the news:*

Nine things you should know about micro-blogging. Kelley School of Business, Indiana University.

Bradley, A.J. (2010). *Becoming a social organization: Taking a strategic approach to social media.*

Antony (2008). *What is social media ? Icrossing.*

Holsapple, C., Hsiao, S., and Pakath, R. (2014). Business social media analytics: Definition, benefits, and challenges. *Proceedings of the 20th Americas conference on Information Systems* (AMCIS2014). Association for

- Information Systems, Association for Information Systems.
- (D. Zeng, H. Chen, R. Lusch, and S. Li(2010)) D. Zeng, H. Chen, R. Lusch, and S. Li. *Social media analytics and intelligence*. 25 (6):13–16, 2010a.
- Global Web Index (2014). Survei data global web index. URL: <https://www.globalwebindex.net/>
- Gartner Research. Social analytics. URL <http://www.gartner.com/it-glossary/social-analytics>
- Sterne, J. and Scott, D.M. (2010). *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*. John Wiley.
- Stieglitz, S and Linh, D. (2013). Social media analytics and political communication; a social media analytics framework. *Social Network Analysis and Mining*, 3(4):1277-1291.
- Fan, W. and Gordon, M.D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014.