

Pengembangan Tools dan Framework Social Media Analytics Untuk Pemanfaatan Data Sosial Media Dalam Penelitian Ilmu Sosial

Edi Surya Negara¹, Suryayusra²

¹ Ilmu Komputer, Universitas Bina Darma.
e.s.negara@binadarma.ac.id, suryayusra@binadarma.ac.id

Abstrak. Tingginya tingkat penggunaan media sosial di Indonesia berdampak pada tersedianya data tentang persepsi masyarakat terhadap berbagai isu, baik itu isu tentang kehidupan sebagai anggota kelompok sosial ataupun kehidupan berbangsa dan bernegara, dalam jumlah yang sangat besar. Fenomena ini membuka peluang bagi berbagai pihak, baik itu pemerintah maupun swasta, yang memerlukan informasi tentang persepsi masyarakat atau perilaku sosial masyarakat dalam kegiatan rutusnya untuk meningkatkan kualitas dan kuantitas informasi yang digunakan dengan memanfaatkan data media sosial. Pemahaman yang mendalam tentang teknologi internet, media sosial, basis data, struktur data, teori informasi, data mining, machine learning, sampai dengan teknik visualisasi data dan informasi sangat diperlukan dalam pengumpulan dan pengolahan data sebelum data tersebut dianalisis dan sintesis menjadi informasi yang akan digunakan oleh pihak-pihak yang memerlukan. Sasaran dari penelitian ini adalah menguasai teknologi analisis media sosial (social media analytic) dan mengembangkan prototipe perangkat lunak yang nantinya akan dikembangkan 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 telah 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.

1 PENDAHULUAN

Berdasarkan survei yang dilakukan oleh Global Web Index pada Januari 2014^[1], 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 Indo-

Program Pascasarjana
Universitas Bina Darma
Palembang, INDONESIA

nesia^[1], 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 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^[1].

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^[2]. SMA menghadapi tantangan ini dengan menyediakan perangkat (tools) dan kerangka kerja (framework) untuk mengumpulkan, memantau, menganalisis, menyimpulkan, dan memvisualisasi data sosial media^[3].

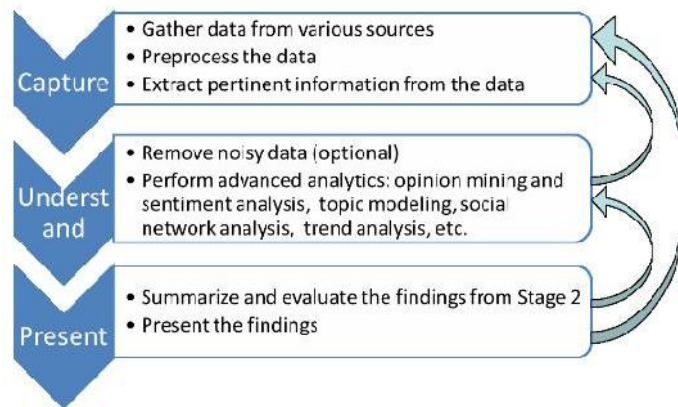
Untuk dapat memanfaatkan data media sosial dalam mendukung kegiatan penelitian yang berhubungan dengan persepsi masyarakat dan perilaku sosial masyarakat diperlukan perangkat dan kerangka kerja yang dikembangkan secara khusus untuk mengumpulkan, memantau, menganalisis, menyimpulkan, dan memvisualisasikan 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, Microbloging* dan lain-lain dapat digunakan untuk berbagai tujuan tertentu^{[4][5]}. Dengan mengimplementasikan teori social presence, media richness dan social processes Kaplan dan Haenlein mengklasifikasikan media sosial menjadi enam jenis yaitu^[6] : 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 Collaboration, 3) Social Publishing, dan 4) Social Feedbacks^[7].

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 social^{[8][3]}. Gartner Research juga mendefinisikan SMA merupakan proses pemantauan, analisis, mengukur dan memprediksi interaksi digital, relationships, topik, ide atau konten pada media sosial (Gartner Research). 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^[3]. Tahapan SMA dapat dilihat pada Gambar 1.



Gambar 1. Social Media Analytics Process^[3].

2 METODE PENELITIAN

Obyek 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 pendahulunya seperti Friendster, MySpace, YouTube, Bebo, Facebook, dan lain-lain, Twitter memungkinkan penggunanya mem-posting pesan pendek dengan jumlah karakter maksimum 140 karakter yang dikenal dengan istilah tweet. Tweet dapat di-post melalui berbagai cara seperti website Twitter, mobile application, atau aplikasi/website pihak ketiga setelah melakukan autentikasi.

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^{[9][3][1]},

Program Pascasarjana
Universitas Bina Darma
Palembang, INDONESIA

secara umum penelitian-penelitian tersebut dapat di- kelompokkan ke dalam penelitian yang menggunakan data content dan data jejaring (network) dengan menggunakan teknik-teknik yang berbeda.

Penelitian ini merupakan penelitian exploratory yang diimplementasikan dengan melakukan penelitian sekunder (*secondary research*), dimana peneliti akan melakukan kajian ulang atas hasil penelitian-penelitian sebelumnya dalam bidang ilmu sosial yang menggunakan tweet, baik sebagai 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 mengumpulkan dan mengevaluasi berbagai framework dan rancangan sistem yang 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.

2.2 Metode Pengumpulan Data

Dalam melakukan penelitian ini, metode pengumpulan data yang digunakan oleh penulis adalah kajian kepustakaan, dimana sumber data utama dalam penelitian ini berasal dari jurnal-jurnal ilmiah, buku-buku, artikel ilmiah, internet, dan laporan-laporan yang berkaitan dengan pemanfaatan data tweet atau media sosial lain.

3 HASIL DAN PEMBAHASAN

3.1. Tools dan Framework Social Media Analytics

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^[3].

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 dan lain-lain. Data yang dihasilkan dari proses Capture disimpan kedalam database dan dipersiapkan untuk proses berikutnya yaitu Undetstad. Pada tahapan ini data juga diproses untuk meng-

hasilkan informasinya yang sesuai dengan kebutuhan, termasuk memodelkan bentuk datanya^[8].

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, menyeleksi data yang berkualitas dan melakukan proses analisis untuk memperoleh informasi yang lebih berkualitas^[8]. Proses analisis data pada tahapan ini menggunakan metode statistik, text mining, data mining, natural language processing (NLP), machine translation, machine learning, dan network analysis^[8]. 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^[8].

Tahapan terakhir dari proses SMA adalah Present. Proses Present merupakan proses untuk menampilkan atau memvisualisasikan informasi yang dihasilkan dari tahap Understand^[8]. Berbagai teknik visualisasi dapat digunakan untuk menampilkan informasi yang didapatkan dari proses analisis.

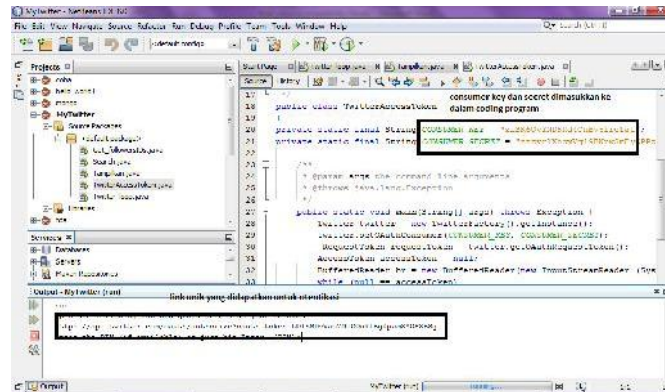
Capture Data

Dari penelitian yang telah dilakukan menghasilkan prototipe aplikasi data retrieval dan kerangka kerja untuk social media analytics. Pada penelitian ini, penulis menggunakan twitter sebagai objek penelitian. Untuk menganalisa data twitter ini 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 3. Setelah proses penarikan data twitter 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 diretweet 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.

Program Pascasarjana
Universitas Bina Darma
Palembang, INDONESIA

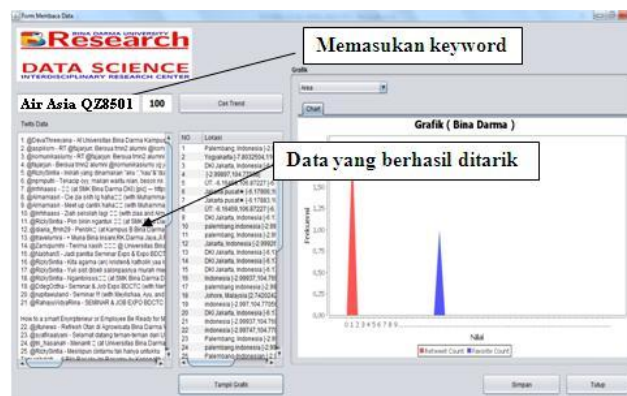


Gambar 2. Code PIN Authentication

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 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 bubble graph atau graph lainnya 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.

Program Pascasarjana
Universitas Bina Darma
Palembang, INDONESIA

DAFTAR PUSTAKA

1. Global Web Index. Survei data global web index, 2014. URL <https://www.globalwebindex.net/>.
2. S. Stieglitz and D. Linh. Social media analytics and political communication; a social media analytics framework. *Social Network Analysis and Mining*, 3(4):1277–1291, 2013.
3. D. Zeng, H. Chen, R. Lusch, and S. Li. Social media analytics and intelligence. 25 (6):13–16, 2010a.
4. Antony. What is social media ? Icrossing, 2008.
5. Jim Sterne and David Meerman Scott. *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*. John Wiley, March 2010.
6. Andreas M. Kaplan and Michael Haenlein. The early bird catches the news: Nine things you should know about micro-blogging. Kelley School of Business, Indiana University, 2010.
7. Anthony J. Bradley. *Becoming a social organization: Taking a strategic approach to social media*, 2010.
8. Weiguo Fan and Michael D Gordon. The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014.
9. 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.