VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Visualisasi Informasi untuk Memetakan Berita yang Berkorelasi dengan Tweet Menggunakan Analisis Leksikal dan Naive Bayes

Oleh

Yodi Dila Padilla

10/300710/PA/13343

Internet mengubah setiap bagian yang ada pada media masa juga melahirkan media baru yang bernama media sosial, salah satu diantaranya Twitter, dimana Twitter memiliki sejumlah potensi dan potensi itu mampu dideteksi oleh media massa. Kemudian, potensi tersebut dimanfaatkan oleh media massa dengan cara masuk ke dalam media sosial. Hal ini justru menyebabkan rusaknya alasan media sosial lahir yaitu sebagai media penyeimbang. Media sosial menjadi tempat media massa untuk membuat dan menyebarkan opini publik. Sejauh ini, opini publik yang muncul tentang suatu hal baru sebatas gambaran umum sementara alasan yang menyebabkan opini publik tersebut muncul tidak pernah diketahui. Oleh karena itu, dibutuhkan sebuah sistem yang mampu melihat penyebab opini publik muncul dan pola-pola yang terbentuk yang disebabkan oleh opini publik yang berkorelasi dengan berita yang telah dibuat oleh media massa.

Penelitian ini mencoba untuk membuat sistem yang mampu melihat opini publik dalam ukuran yang lebih kecil berdasarkan asumsi bahwa opini publik disebabkan oleh sejumlah berita. Berita yang dikumpulkan merupakan berita yang memicu masyarakat untuk membuat opini di media sosial, dalam kasus ini adalah Twitter. Tweet, satuan data di Twitter, akan diklasifikasikan berdasarkan sentimennya. Kemudian, sistem juga akan memprediksi lokasi berita yang kemudian akan dipetakan berdasarkan lokasinya. Kombinasi sentimen dan lokasi digunakan untuk melihat apakah ada pola yang muncul.

Metode yang digunakan dalam penelitian ini untuk mengklasifikasikan Tweet berdasarkan sentimennya menggunakan Multinomial Naive Bayes (MNB), sedangkan metode untuk memprediksi lokasi berita adalah analisis leksikal. Penelitian ini menghasilkan akurasi klasifikasi sentimen sebsar 69,43%. Kemudian, akurasi untuk memprediksi lokasi adalah sebesar 80,1% untuk *relaxed accuracy*, 90% untuk *regular accuracy*, dan 71% untuk *hard accuracy*. Penelitian ini juga melakukan percobaan menggunakan sejumlah kata kunci untuk menunjukan bahwa sistem dapat bekerja dengan cara menunjukan sejumlah hasil secara statistik dan visualisasi.

Kata kunci : Informasi Visualisasi, Naive Bayes, NER, Twitter, media massa.

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Information Visualization to Mapping News Correlated with Tweets Using Lexical Analysis and Naive Bayes

By

Yodi Dila Padilla

10/300710/PA/13343

Internet shifted every part of mass media also triggered the newly born media called social media, such as Twitter, which has a number potential within itself that could be detected by mass media. Then, the potential of Twitter exploited by mass media by infiltrating into social media. This broke one of the reasons why social media born was that people wished that it could become a balancer tool against mass media. Social media turned out to be the place for mass media to create and to spread public opinion. As far as it goes, public opinion that had been emerged was still in general picture while the reasons why it emerged were never been discovered. Therefore, a system is needed to see the causes of the public opinions and patterns that were made based on public opinion which correlated with news that were conducted by mass media.

The research is trying to make a system to see public opinion in smaller scope based on assumption as public opinion was triggered by news. News that were collected are news that provoked community to make opinion,especially in social media, which in this case is in Twitter. In regards of case Tweet, denumeration data post of Twitter, is going to be classified based on sentiment. Afterwards, the system is going to predict the location of the news and be spatially mapped. The combination of sentiment and location used for seeing whether any pattern exist.

The method used by this reasearch to classify sentiment of Tweet is Multinomial Naive Bayes (MNB), while the method used to predict location of the news is lexical analyisis. This reasearch produces accuracy of sentiment classification for about 69,43%. Then, the accuracy of location prediction are 80,1% , 90%, and 71% for relaxed, regular, and hard accuracy respectively. This research also tests the general system using several keyword in order to show that system can work well by produce several statistic summary and visualization.

Keyword : Information Visualization, Naive Bayes, NER, Twitter, Mass Media.

VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

BAB I
PENDAHULUAN

1.1 Latar Belakang

Internet menggiring perubahan terhadap berbagai sektor. Salah satu sektor yang berubah adalah media massa. Internet secara tidak langsung memaksa media massa berubah dari media cetak ke media dijital karena semenjak hadirnya internet media cetak mengalami kemunduran. Perubahan yang terjadi tidak sekedar perubahan medium yang digunakan tapi segala aspek yang ada di media massa. Walaupun media massa dari segi cetak mengalami kemunduran dan harus mengubah berbagai aspek, perubahan media massa ini ternyata tetap disambut baik oleh masyarakat. Menurut survey APJII (2012), halaman web yang sering diakses oleh masyarakat Indonesia adalah mengunjungi halaman web media massa dengan total 38%. Jumlah pengguna internet yang mengunjungi halaman web media massa relatif tinggi sehingga terus memacu media massa untuk terus memproduksi berita.

Selain mendorong media massa untuk berubah, internet juga melahirkan kelompok media yang lain bernama media sosial salah satu diantaranya adalah Twitter yang perkembangannya relatif sangat pesat. Berdasarkan data dari Statista (2015), jumlah pengguna Twitter adalah sekitar 30 juta pengguna aktif pada awal tahun 2010, dan pada akhir tahun 2014 menjadi 288 juta pengguna aktif. Kemudian, menurut Internet Live Stats (n.d.) jumlah Tweet yang dihasilkan Twitter adalah sekitar 500 juta Tweet per hari di tahun 2013 dengan jumlah pengguna aktif sebesar 241 juta pada akhir tahun tersebut.

Potensi besar yang dimiliki Twitter mampu dideteksi oleh media massa. Media masa melakukan inflitrasi ke media sosial untuk menyebarkan berita dengan cara membuat akun dan menjadi pengguna aktif di dalamnya. Media massa yang bergabung ke media sosial ternyata disambut baik oleh masyarakat dengan banyaknya akun pribadi yang menjadi pengikut atau pendengar bagi akun media massa. Sebagai contoh, satu dari lima akun Twitter Indonesia dengan pengikut terbanyak adalah akun media massa, yaitu detik.com (Dior, 2014). Hal ini tentu hal yang wajar karena sekarang adalah era masyarakat haus akan informasi, dan media

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

massa memenuhi kebutuhan tersebut karena memang tujuan utama dari media massa untuk menyebarkan informasi atau berita.

Walaupun demikian, berita yang dihasilkan ternyata tidak sekedar informasi biasa. Berita yang dihasilkan tidak sedikit yang justru membentuk opini publik terhadap suatu hal. Menurut Zhang et al. (2014) opini yang terbentuk di media sosial sebagian besar disebabkan oleh pemimpin opini dimana dalam penelitiannya mendefinisikan pemimpin opini sebagai akun Twitter dengan jumlah pengikut yang relatif banyak. Sejalan dengan penelitian Zhang et al. (2014), penelitian Asur et al. (2011) menunjukan bahwa tren dalam media sosial sangat didominasi oleh media massa, dan media sosial jauh dari yang namanya alternatif sumber berita padahal masyarakat berharap sebaliknya. Hal ini menunjukan bahwa ada hegemoni media massa walaupun di media sosial. Sehingga, jika diamati lebih dalam sebenarnya opini masyarakat yang terbentuk disebabkan oleh pemberitaan dari media massa dan yang menjadi pertanyaan adalah berita mana saja yang memicu masyarakat untuk beropini dan apakah opini masyarakat muncul dalam rangkaian dan pola tertentu dari berita-berita yang dihasilkan.

Proses untuk melihat pola opini masyarakat sangat bergantung terhadap perspektif yang digunakan. Berbagai macam perspektif telah digunakan untuk melihat pola-pola tersebut salah satunya perspektif pengamatan berdasarkan lokasi. Menurut Amitay et al. (2004) dengan mengetahui suatu lokasi dari sebuah teks, masyarakat bisa tahu area mana yang sedang populer tentang suatu produk atau topik tertentu. Sehingga, melakukan pengamatan berita berdasarkan lokasi, masyarakat atau peneliti bisa melihat pola apa saja yang terjadi di suatu lokasi.

Sejauh ini, untuk melihat opini masyarakat, metode yang dirancang baru sebatas gambaran umum. Berdasarkan asumsi-asumsi di atas, maka gambaran umum dapat dipecah berdasarkan opini yang dipicu dan berkorelasi dengan sebuah berita dengan harapan dapat ditemukan pola pemberitaan. Selain itu, penggunaan lokasi sebagai perspektif pengamatan ternyata juga mampu melihat tren. Sehingga, dibutuhkan suatu penelitian untuk melihat opini masyarakat yang berkorelasi

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

dengan berita-berita yang terjadi di suatu lokasi agar gambaran umum yang terjadi bisa dianalisis lebih lanjut.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang dijelaskan di dalam latar belakang, maka dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana memetakan berita berdasarkan lokasinya.
2. Bagaimana mengetahui sentimen pada media sosial terhadap berita yang dipetakan berdasarkan lokasinya.

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Sumber data yang dipilih menggunakan Bahasa Indonesia.
2. Berita yang dipilih untuk digunakan sebagai penelitian adalah data berupa teks dari media berita online yaitu kompas.com, tempo.co, republika.co.id, okezonenews.com, liputan6.com, dan metronews.com.
3. Data yang digunakan untuk analisis sentimen adalah dari salah satu media sosial yaitu Twitter.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah merancang sistem yang mampu memetakan dan mengkelompokan berita-berita berdasarkan lokasinya serta mengetahui sentimen pengguna media sosial terhadap berita-berita tersebut.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat berkontribusi untuk penelitian ilmu komputer terutama pada ranah *text mining* atau yang bersifat kontemporer dengan manfaat sebagai berikut :

1. Arsitektur sistem menjadi dasar pengembangan untuk melakukan pencarian Tweet yang berkorelasi dengan berita.

VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

2. Menjadi salah satu rujukan dalam proses menggabungkan klasifikasi sentimen pada Tweet dan lokasi yang terdapat pada berita sebagai fungsi agregasi yang saling berhubungan.

1.6 Metode Penelitian

Metode yang digunakan dalam penelitian ini terdiri atas :

1. Studi Literatur

   Proses pengumpulan data dan informasi yang berkaitan dan mendukung penelitian khususnya yang meliputi visualisasi informasi, analisis sentimen, proses ekstraksi lokasi, dan proses pengenalan entitas-bernama. Sumber yang digunakan sebagai rujukan sebagian besar adalah buku, jurnal, artikel ilmiah, dan tulisan-tulisan dari berbagai halaman web.

2. Analisis

   Proses analisis dilakukan untuk mengetahui kebutuhan sistem berdasarkan permasalahan dan tujuan penelitian. Proses analisis juga bertujuan untuk mengidentifikasi struktur dan sifat yang terdapat pada berbagai data dan metode. Data dan metode yang telah diidentifikasi kemudian dibandingkan lalu dipilih data dan metode yang akan digunakan.

3. Rancangan

   Rancangan sistem terbagi atas tiga bagian utama yang terdiri dari pengumpulan, praproses, dan pemetaan. Perancangan sistem dibuat ke dalam diagram alur untuk masing-masing proses.

4. Implementasi

   Implementasi sistem dilakukan dengan cara membuat perangkat lunak dalam *platform* web menggunakan bahasa pemrograman PHP dan basis data MySQL.

VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

5. Pengujian

   Pengujian sistem meliputi pengukuran secara kuantitatif terhadap sistem yang telah diimplementasikan meliputi pengujian klasifikasi sentimen, pengujian pengenalan entitas-bernama geospasial, dan pengujian antar muka.

6. Kesimpulan dan Saran

   Hasil pengujian akan memperlihatkan kemampuan sistem dan dari kemampuan yang telah diukur maka akan dapat disimpulkan keunggulan dan kelemahan sistem sehingga keunggulan dan kelemahan yang diperoleh bisa digunakan sebagai pembelajaran untuk menuju penelitian yang berkesinambungan.

1.7 Sistematika Penulisan

Sistematika dalam penulisan tugas akhir ini akan dijelaskan di bawah ini :

1. BAB I PENDAHULUAN: Berisi mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan.
2. BAB II TINJAUAN PUSTAKA: Bab ini memuat pembahasan mengenai penelitian terdahulu yang digunakan sebagai bahan referensi dalam penulisan penelitian ini. Referensi yang akan dibahas di dalam bab ini adalah berkenaan visualisasi informasi terutama yang menggunakan peta sebagai komponen untuk memvisualisasikannya.
3. BAB III LANDASAN TEORI: Bab ini berisi teori-teori yang menjadi landasan dalam penulisan penelitian ini yaitu mengenai visualisasi informasi, analisis sentimen, *random subsampling cross validation*, pengenalan-entitas bernama, *geocoding*, pengenalan entitas-bernama geospasial dan Twitter API.

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

4. BAB IV ANALISIS DAN PERANCANGAN SISTEM: Bab ini berisi analisis terhadap sistem yang akan dikembangkan, serta penjelasan mengenai perancangan sistem berdasarkan hasil analisis yang telah dilakukan.
5. BAB V IMPLEMENTASI: Pada bab ini akan dijelaskan hasil implementasi dari sistem yang telah dibangun berdasarkan perancangan yang telah dilakukan sebelumnya.
6. BAB VI PENGUJIAN: Pada bab ini akan dijelaskan hasil yang diperoleh setelah sistem selesai diimplementasikan. Disertai pula pembahasan hasil implementasi tersebut.
7. BAB VII KESIMPULAN DAN SARAN: Bab ini berisi mengenai kesimpulan dari penelitian yang telah dilakukan dan saran untuk pengembangan penelitian selanjutnya.

VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

BAB II
TINJAUAN PUSTAKA

Zong et al. (2005) merancang sebuah algoritma untuk mengetahui lokasi dalam teks Bahasa Inggris dan menandainya. Algoritma yang dirancang merupakan modifikasi dari GATE, perangkat lunak pengenalan entitas-bernama dipadukan dengan algoritma pembelajaran semi-terbimbing dengan menggunakan *Gazetteer* dan analisis leksikal. Algoritma yang disusun terdiri dari tiga subproses yaitu ekstraksi kandidat lokasi, deteksi ambiguitas kandidat lokasi, dan penyusunan lokasi. Algoritma yang dihasilkan menghasilkan *relaxed-accuracy* sebesar 91.3% dan *page-centric-accurracy* sebesar 86.8%.

Ahmed et al. (2010) mengadakan sebuah penelitian untuk menciptakan sebuah metode pencarian yang mampu mengumpulkan artikel-artikel yang berhubungan dengan suatu kejadian. Kejadian-kejadian yang dianalisis adalah kejadian tindak kekerasan. Artikel-artikel yang terkumpul direpresentasikan ke dalam peta disertai kronologi kejadian. Metode yang digunakan dalam penelitian ini adalah *incremental clustering* yang mampu bekerja dalam waktu nyata. Hasil penelitian menghasilkan rata-rata presisi sebesar 81%.

Mena dan Ornelas (2011) merancang sistem untuk mengekstrak lokasi dari berbagai halaman web berita dan memproyeksikannya ke dalam peta. Mena dan Ornelas menggunakan Geoparsing Web Service (GWS) untuk mengekstrak koordinat geografis dari suatu berita. Sistem yang dirancang dilengkapi dengan dengan fungsi semantik agar sistemnya mampu melakukan pencarian berita yang saling berelasi dan melakukan penyaringan sehingga pengguna hanya membaca bagian-bagian yang penting saja dari berita-berita tersebut. Penelitian ini menyimpulkan bahwa kesulitan utama bukan dalam proses perancangan GWS untuk data tidak terstruktur, tapi lebih kepada mengekstrak informasi yang utama dan berguna dalam dokumen tersebut.

Endarnoto et al. (2011) melakukan sebuah penelitian untuk memperoleh informasi tentang lalu lintas di Jakarta berdasarkan informasi yang diperoleh dari Twitter tepatnya dari akun Twitter bernama @TMCPoldaMetro. Informasi tentang

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

lalu lintas tersebut direpresentasikan ke dalam peta yang berjalan di perangkat bergerak platform Android. Informasi lalu lintas tersebut juga diklasifikan menjadi tiga tingkatan dengan cara memberi warna ke dalam peta menggunakan 3 warna yang merepresentasikan tingkat kepadatan lalu-lintas yang berbeda-beda. Adapun proses yang dilakukan adalah melakukan Natural Language Processing terhadap Tweet dari akun Twitter tersebut. Tweet yang diperoleh dilakukan serangkaian proses, yaitu tokenisasi, POS-*tagging*, analisis struktur kalimat, *template filter* dan terakhir proses visualisasi. Penelitian ini diklaim bahwa sistem dapat bekerja dengan baik untuk mengekstrak informasi tentang lalu lintas.

Ma (2012) mengajukan sebuah metode untuk mengubah teks berita dari Sina News, sebuah halaman web berita berbahasa China ke dalam grafis dengan cara memetakan setiap berita berdasarkan lokasinya. Penelitian ini terdiri dari serangkaian proses. Proses yang pertama adalah melakukan segmentasi pada teks untuk menghasilkan informasi utama pada sebuah halaman web semisal teks utama pada berita, waktu, dan lokasi berita tersebut. Kemudian, setelah mendapat lokasi, proses selanjutnya adalah menerjemahkan lokasi yang masih dalam Bahasa China ke dalam Bahasa Inggris. Selanjutnya, lokasi yang telah dalam Bahasa Inggris bisa digunakan untuk menghasilkan *latitude* dan *longitude* menggunakan Google Map API. Setelah data yang dibutuhkan terkumpul semua, data tersebut kemudian disimpan ke dalam basis data dimana basis data tersebut nanti digunakan untuk memetakan berita berdasarkan lokasi dan memvisualisasikannnya menggunakan Google Map. Penelitian yang dilakukan Ma menghasilkan sebuah layanan untuk memetakan berita yang bisa dimanfaatkan oleh perusahaan-perusahaan yang bergerak dalam bidang media.

Nakaji dan Yanai (2012) menciptakan sebuah metode untuk mengumpulkan foto-foto yang mengandung lokasi dari Twitter. Foto-foto yang memiliki informasi lokasi tersebut dilakukan pemetaan sedemikian rupa sehingga foto-foto tersebut dapat dikumpulkan berdasarkan suatu kejadian atau kegiatan. Metode yang dirancang menggunakan Twitter Streaming API untuk mengumpulkan Tweet selama kurang lebih satu tahun yang jumlahnya mencapai 18 juta Tweet yang

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

mengandung foto dan lokasi. Tweet yang terkumpul tersebut dilakukan praproses terlebih dahulu kemudian dilakukan klastering berdasarkan foto dan lokasi yang terkandung pada Tweet menggunakan algoritma k-means. Kemudian, Nakaji dan Yanai menggunakan metode GeoVisualRank untuk memilih foto yang benar-benar representatif terhadap suatu kejadian atau acara, disamping menggunakan lokasi dan waktu saja sebagai fitur. Foto-foto yang representatif tersebut direpresentasikan ke dalam peta berdasarkan lokasi dan waktu dari kejadian. Hasil penelitian ini mengklaim bahwa metode mampu menghasilkan foto-foto yang representatif terhadap suatu kejadian walaupun belum mampu berjalan dalam waktu nyata.

Ji et al. (2013) melakukan klasifikasi sentimen terhadap data di Twitter untuk pengawasan terhadap masalah kesehatan masyarakat dengan merancang sebuah sistem yang bernama Epidemic Sentiment Monitoring System (ESMOS). ESMOS dalam proses klasifikasi Tweet melalui 2 proses. Pertama, Tweet diklasifikasi ke dalam Tweet personal dan Tweet berita. Selanjutnya, kumpulan Tweet personal tersebut kemudian diklasifikasikan ke dalam Tweet negatif dan non-negatif (netral). Adapun algoritma yang digunakan adalah Naive Bayes, Multinomial Naive Bayes, dan Support Vector Machine. Multinomial Naive Bayes menghasilkan F-Measure terbaik. Sedangkan untuk kebutuhan visualisasi agar mampu dimunculkan dalam peta, penelitian ini menggunakan Google Map API. Hasil akhir dari penelitian ini adalah ESMOS yang memiliki kemampuan untuk melakukan visualisasi terhadap Tweet yang berhubungan dengan berbagai macam penyakit dan perkembangannnya ke dalam peta, serta memberikan peringatan sejak dini jika ada penyakit yang mewabah karena ESMOS berjalan dalam waktu nyata.

Adapun penelitian ini merancang sebuah sistem yang memiliki fungsi untuk memprediksi lokasi dari berita-berita lalu memvisualisasikannya ke dalam peta berdasarkan lokasi, kemudian melakukan analisis sentimen terhadap Tweet yang berkorelasi dengan masing-masing berita. Proses ekstraksi lokasi dari dokumen tersebut menggunakan metode pengenalan entitas-bernama geospasial yang terdiri dari indeks entitas, analisis leksikal Bahasa Indonesia dan Google Geocoder API sedangkan untuk klasifikasi sentimen menggunakan Multinomial Naive Bayes.

VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

BAB III
LANDASAN TEORI

3.1 Visualisasi Informasi

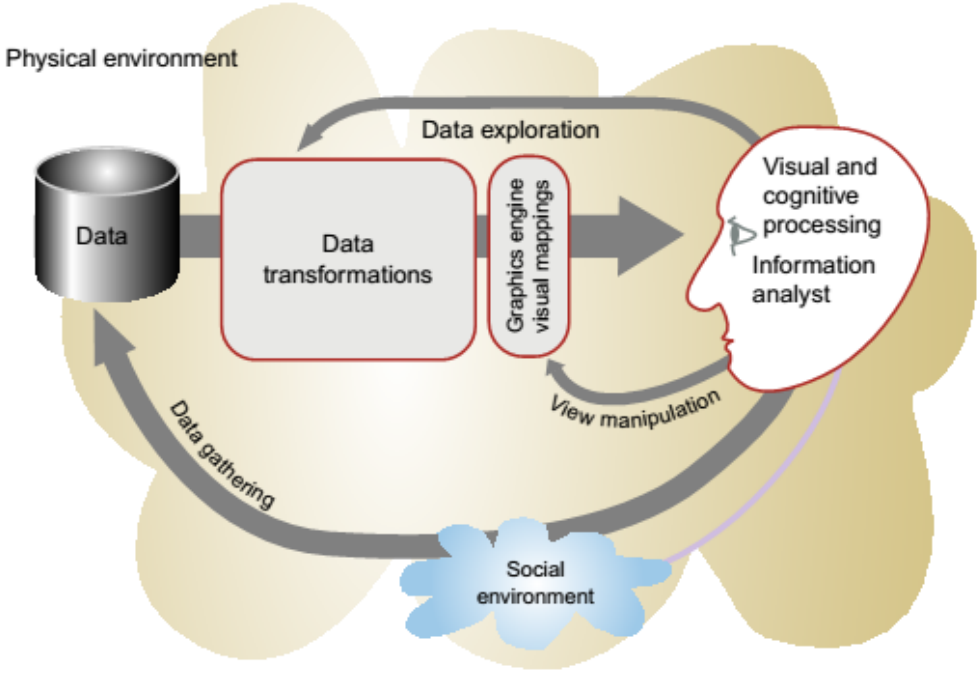
Menurut McCormick (1987) visualisasi adalah proses mentransformasikan data, informasi, dan pengetahuan ke dalam bentuk grafis untuk membantu dan mempermudah proses-proses seperti analisis data, eksplorasi informasi, prediksi tren atau kecenderungan, pengenalan pola, dan lain-lain. Informasi yang disajikan tanpa bantuan visualisasi akan menyulikan manusia untuk mengintrepretasikan informasi, data, atau pengetahuan. Hal ini bisa disebabkan karena keterbatasan kemampuan visual manusia, atau data,informasi, dan pengetahuan yang disajikan masih dalam bentuk yang mengandung banyak hal yang sulit dilihat dan abstrak. Berdasarkan kondisi-kondisi tersebut, visualisasi informasi berperan dalam mengolah data, informasi, dan pengetahuan menjadi sesuatu yang lebih berarti dan mudah dipahami.

Ware (2012) memaparkan bahwa visualisasi informasi terdiri atas 4 tahap :

1. Mengumpulkan data.
2. Praproses : proses untuk mengubah data menjadi lebih mudah untuk dimanipulasi. Proses ini mengalami dua subproses yang utama untuk kebutuhan visualisasi, yaitu reduksi data dan eksplorasi data. Reduksi data adalah proses memilih data yang memenuhi aspek sedangkan eksplorasi data proses mengubah data yang telah mengalami proses reduksi.
3. Pemetaan : proses untuk menyajikan data. Data yang telah dilakukan praproses direpresentasikan ke dalam format visual yang dituju, sebagian besar ditunjukan dengan menghasilkan sebuah citra. Proses ini juga memberikan kendali untuk pengguna untuk melakukan serangkaian aksi atau masukan.

VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

4. Analisis dan mengintrepretasikan hasil pemetaan. Proses ini melibatkan kemampuan kognitif manusia dalam mengintrepretasikan informasi yang disajikan.



Gambar 3.1 Tahapan visualisasi informasi (Ware, 2012)

Mazza (2009) menjelaskan bahwa pemetaan visual adalah proses menentukan struktur visual yang berkorespondensi dengan data yang ingin direpresentasikan ke dalam bentuk visual. Struktur visual tersebut terbagi ke dalam 3 bagian :

1. Substrat spasial : ruang untuk merepresentasikan hasil visual.
2. Elemen grafis : semua komponen yang tampak di substrat spasial. Elemen grafis secara umum terdiri atas 4 macam : titik, garis, permukaan, dan volume.
3. Properti grafis : properti pada elemen grafis yang paling sensitif bagi mata manusia. Properti grafis juga disebut sebagai variabel retina karena properti grafis paling berhubungan dengan kemampuan kognitif

UNIVERSITAS
GADJAH MADA
VISUALISASI INFORMASI UNTUK MEMETAKAN BERITA YANG BERKORELASI DENGAN TWEET MENGGUNAKAN ANALISIS
LEKSIKAL DAN NAIVE BAYES
YODI DILA PADILLA, Mardhani Riasetiawan, M.T.
Universitas Gadjah Mada, 2015 | Diunduh dari http://etd.repository.ugm.ac.id/

manusia. Properti grafis yang umum digunakan di antaranya ukuran, orientasi, warna, tekstur, dan bentuk.

3.2 Analisis Sentimen

Analisis sentimen adalah sebuah domain pengetahuan yang mempelajari tentang pendapat, sentimen, pujian, cacian, sikap atau ekspresi masyarakat menyikapi terhadap sebuah produk, layanan, organisasi, individu, isu, kejadian, atau topik. Padanan kata atau istilah untuk analisis sentimen akan banyak dijumpai seperti penggalian pendapat, ekstraksi pendapat, analisis subjektivitas dan lain-lain. Penggunaan istilah yang berbeda muncul karena penggunaan di sektor yang berbeda-beda, akan tetapi secara umum domain pengetahuan dan penelitian ini lebih banyak menggunakan istilah analisis sentimen dan penggalian pendapat (Liu, 2007).

Secara umum penelitian tentang analisis sentimen terbagai ke dalam tiga level yaitu level dokumen, kalimat, serta level aspek dan entitas.

1. Level dokumen. Level ini menganggap dokumen sebagai sebuah entitas. Tugas pada level ini adalah mengklasifikasikan dokumen berdasarkan sentimen salah satu diantaranya adalah mengklasifikasikan entitas ke dalam kelas negatif, netral, atau positif secara kontekstual.

2. Level kalimat. Analisis sentimen pada level ini disebut sebagai analisis subjektivitas. Level ini menganggap sebuah kalimat sebagai sebuah entitas. Tugas yang dilakukan pada level ini sama dengan tugas yang dilakukan pada level dokumen yaitu mengklasifikasikan dokumen ke dalam kelompok kalimat dengan kelas sentimen negatif, netral, atau positif.

3. Level aspek dan entitas. Analisis sentimen pada level dokumen dan kalimat dianggap belum mampu untuk menjawab klasifikasi sentimen sebuah entitas karena pada kedua level tersebut mencoba menganalisis sentimen secara kontekstual. Level ketiga melakukan analisis dengan