

BAB II

TINJAUAN PUSTAKA

Penelitian ini menggunakan beberapa tinjauan pustaka sebagai dasar teori dan/atau metodologi penelitian. Tinjauan pustaka tersebut berupa penelitian-penelitian terdahulu yang relevan dengan topik dan metode penelitian yang diusulkan. Peneliti dapat mempelajari sebagian metode yang disampaikan pada penelitian sebelumnya untuk merencanakan pengembangan metode yang telah dibuat maupun mengembangkan metode baru. Berikut ini adalah deskripsi singkat tentang penelitian-penelitian terdahulu yang digunakan sebagai tinjauan pustaka pada penelitian ini.

Malinský & Jelínek (2012) pernah melakukan penelitian tentang penggunaan analisis sentimen untuk menilai popularitas kandidat presiden AS. Penelitian ini menggunakan Barack Obama dan Mitt Romney pada pemilihan presiden AS tahun 2012 sebagai objek. Latar belakang yang mendasari penelitian ini adalah karena pemilihan presiden merupakan peristiwa yang sangat penting dan diperhatikan diseluruh dunia. Berita-berita tentang pemilu dapat dengan mudah ditemukan di internet, yaitu dengan memanfaatkan sosial media berbasis Web 2.0 (blogger, blog, google news, twitter, myspace, youtube, dan facebook). Peneliti mengumpulkan data linguistik dengan melakukan *crawling* dari sosial media berbasis Web 2.0. Masalah yang ditemukan dalam proses *crawling* adalah konten web yang bervariasi, perbedaan teknologi, dan struktur web yang berbeda. Implementasi dilakukan dengan diawali pra-proses dan dilanjutkan dengan analisis sentimen. Pada tahap analisis sentimen, data akan diberi nilai sentimen yang terdiri

dari nilai positif, nilai negatif, dan nilai obyektif dengan menggunakan SentiWordNet. Peneliti menghitung popularitas kandidat presiden dengan menghitung agregat dari dua kelompok kandidat (Obama dan Romney). Data yang telah diproses kemudian akan ditampilkan pada tahap visualisasi.

DiGrazia dan kawan-kawan (2013) melakukan sebuah penelitian untuk membuktikan bahwa banyaknya jumlah *tweet* yang didapatkan dari Twitter tentang kandidat kongres Amerika Serikat berbanding dengan suara yang didapatkan saat pemilihan presiden. Hal ini dipicu oleh anggapan bahwa komunikasi secara *online* dapat menjadi indikator yang valid terhadap tingkah laku atau kebiasaan *offline*. Metode pengumpulan data yang dilakukan adalah menyusun data set setingkat distrik dari Twitter, komisi pemilihan federal, dan biro sensus Amerika Serikat. Setelah itu, dilakukan perhitungan untuk menghitung perhatian yang diberikan di Twitter tentang kandidat yang berkaitan dengan lawannya dalam kasus atau topik tertentu. Peneliti lalu mencoba menganalisa dengan cara mengestimasi efek aktivitas di Twitter dengan hasil pemilihan menggunakan tiga model regresi *ordinary least squares* (OLS). Batasan yang menjadi pertimbangan pada penelitian ini yaitu analisa tersebut tidak menggunakan informasi tentang lokasi pengguna Twitter, sehingga kurang sesuai dengan situasi aktualnya yang menerapkan sistem *electoral college*.

Granka (2013) melakukan penelitian untuk memprediksi pemilihan presiden AS dengan menganalisa trafik pencarian *online*. Granka berpendapat bahwa pencarian *online* merupakan bentuk informasi yang aktif yang dapat menunjukkan pengetahuan dan motivasi dari pencariinya. Metode yang digunakan

dalam penelitian ini adalah dengan menghitung agregat dari data pencarian dengan menerapkan *query* kandidat presiden dan isu politik. Peneliti menggunakan Google Trends sebagai sumber data. Model yang digunakan untuk melakukan prediksi adalah *baseline model*. Melalui penelitian ini dapat ditarik beberapa keterbatasan, yaitu penggunaan agregat dalam *query* pencarian tidak dapat mengetahui ideologi politik, umur, gender, dan tingkah laku penggunanya. Suatu *query* pencarian dapat mengacu pada banyak topik yang mungkin tidak berhubungan dengan kandidat presiden atau isu politik (contoh: *query* “ekonomi” dapat mengacu pada topik yang berhubungan dengan mobil ekonomi nasional, bahkan pelajaran ekonomi di sekolah).

Penelitian lain yang menggunakan analisis sentimen adalah penelitian untuk lokalisasi atau mengelompokkan pendapat dari Twitter menggunakan analisis sentimen (Hridoy, et al., 2015). Peneliti menganalisa informasi yang disediakan melalui Twitter karena dapat memberi wawasan pendapat publik tentang hampir semua produk, jasa, atau sifat. Selain daripada itu, data dari media sosial dapat mewakili sentimen publik yang efektif dan akurat. Metodologi yang digunakan dalam penelitian ini antara lain: ekstraksi data, pra-proses , dan implementasi menggunakan SentiWordNet. Batasan yang disampaikan pada penelitian ini yaitu kualitas *tweet* yang sangat rendah karena masih ditemukan skor nol dengan SentiWordNet, walaupun sebelumnya sudah difilter menggunakan *tool natural language processing* (NLP).

Chifu dan kawan-kawan (2015) mengembangkan penelitian untuk menganalisa masukan atau *feedback* yang diberikan oleh pengguna dengan metode

web harvesting dan analisis sentimen. Peneliti memanfaatkan data yang tersedia di internet yang berupa opini pengguna tentang produk atau jasa, lalu mengklasifikasikannya dengan kelas sentimen *fuzzy* yang terdiri dari *very negative*, *strongly negative*, *moderately negative*, *moderately positive*, *strongly positive*, *very positive*). Algoritma *web harvesting* yang dilakukan yaitu dengan mengimplementasikan komunikasi melalui HTTP, *post-processing* dan sanitasi HTML, mengidentifikasi dan mengekstraksi opini atau ulasan pengguna, *post-processing* ulasan pengguna, mendapatkan hasil, dan menyimpulkan hasil analisa. Metode yang digunakan untuk analisis sentimen yaitu *part of speech* (POS) *tagging*, membuat *tuple* (pola sintetis atau n-gram atau frase), mengevaluasi *tuple* yang terdeteksi, dan klasifikasi dengan *fuzzy*. Kesimpulan yang didapatkan dari penelitian tersebut adalah diketahui bahwa penggunaan ulasan film dan buku kurang akurat dengan data yang digunakan (masukan atau *feedback* pengguna).

Adapun penggunaan analisis sentimen pada topik lainnya adalah penelitian untuk mengulas kepuasan pasien dengan menggunakan data yang ada di internet (Hopper & Uriyo, 2015). Hal ini didasari oleh kesusahan yang dialami oleh manajer di bidang kesehatan untuk mengolah data ulasan pasien menjadi informasi yang penting. Metode yang digunakan adalah dengan *sample* dari www.ratedms.com, prosedur kolektif, analisa data untuk meyakinkan ulasan pengguna dapat mencerminkan pandangan mereka, dan menampilkan informasi yang didapatkan. Dalam penelitian ini ditemukan beberapa masalah, antara lain: target yang tidak sesuai, tidak ada informasi yang lengkap (umur, gender, dsb.), dan identitas penulis ulasan.

Selain dalam bidang kesehatan, analisis sentimen juga sering digunakan untuk membuat ulasan produk (Aarti & Patil, 2014). Penelitian ini dilatar belakangi oleh susahnya pengunjung web membaca ulasan produk yang terlalu panjang. Penelitian ini mengimplementasikan analisis sentimen dengan menggunakan algoritma Naïve Bayes tingkat lanjut. Data yang digunakan adalah 400 ulasan dari tiga orang berbeda. Hasilnya, dari 400 ulasan yang dikumpulkan, 52% diberi label positif dan 48% diberi label negatif. Peneliti mengukur tingkat ketepatan analisis sentimen dalam melakukan klasifikasi ulasan dengan menggunakan *precision* dan *recall*. Adapun yang menjadi batasan pada penelitian ini adalah uji coba yang hanya menggunakan *sample* dari 3 orang yang berbeda.

Penelitian-penelitian tersebut digunakan peneliti sebagai referensi untuk membuat penelitian tesis. Peneliti menambahkan fitur-fitur baru serta mengembangkan dan membandingkan metode analisis sentimen dari metode yang telah digunakan pada tinjauan pustaka tersebut ditambah dengan metode lainnya. Penjelasan tentang perbandingan metodologi penelitian yang dirancang peneliti dengan penelitian-penelitian sebelumnya dituliskan dalam Tabel 2.1.

Tabel 2.1. Perbandingan dengan Penelitian Sebelumnya

Peneliti	Masalah	Metode	Batasan
Malinský & Jelínek (2012)	Analisis sentimen untuk menilai popularitas kandidat presiden Amerika Serikat	<i>Crawling</i> , pra-proses, <i>SentiWordNet</i>	Masalah saat <i>crawling</i> yaitu konten web yang bervariasi, perbedaan teknologi, struktur web yang berbeda

Tabel 2.1. Lanjutan

Peneliti	Masalah	Metode	Batasan
DiGrazia, et al. (2013)	Membuktikan bahwa banyaknya tweet yang didapatkan dari Twitter tentang kandidat kongres Amerika Serikat berbanding dengan <i>voteyang</i> didapatkan saat pemilihan presiden	Menyusun dataset setingkat distrik dari Twitter, menghitung perhatian yang diberikan di Twitter, dan mengestimasi efek aktivitas di Twitter dengan hasil pemilihan menggunakan 3 model regresi <i>ordinary least squares</i> (OLS)	Tidak menggunakan informasi tentang lokasi pengguna Twitter
Granka (2013)	Penggunaan pencarian <i>online</i> karena pencarian <i>online</i> merupakan bentuk informasi yang aktif yang dapat menunjukkan pengetahuan dan motivasi dari pencarinya	Agregat dari data pencarian dengan <i>query</i> kandidat presiden dan isu politik	Penggunaan agregat dalam <i>query</i> pencarian tidak dapat mengetahui ideologi politik, umur, gender, dan tingkah laku penggunanya. Dalam suatu <i>query</i> pencarian dapat mengacu banyak topik yang mungkin tidak berhubungan
Akib Anwar Hridoy, et al. (2015)	Menganalisa informasi yang disediakan melalui Twitter karena dapat memberi wawasan pendapat publik tentang hampir semua produk, jasa, atau sifat	<i>data extraction</i> , <i>data pre-processing</i> , dan implementasi menggunakan SentiWordNet	Kualitas tweet yang sangat rendah karena masih ditemukan skor nol di SentiWordNet

Tabel 2.1. Lanjutan

Peneliti	Masalah	Metode	Batasan
Şt CHIFU, et al. (2015)	Memanfaatkan data yang tersedia di internet yaitu opini pengguna tentang produk atau jasa, lalu mengklasifikasikan dengan fuzzy	<i>part of speech tagging</i> , membuat <i>tuple</i> , mengevaluasi <i>tuple</i> yang terdeteksi, dan klasifikasi dengan fuzzy	Penggunaan ulasan film dan buku kurang akurat dengan data yang digunakan
Hopper & Uriyo, (2015)	Susahnya manajer di bidang kesehatan untuk mengolah data ulasan pasien menjadi informasi yang penting	<i>Sample</i> , prosedur kolektif, analisa data, dan menampilkan informasi yang didapatkan	target yang tidak sesuai, tidak ada informasi yang lengkap tentang identitas penulis ulasan.
Aarti & Patil (2014)	Kesusahan pengunjung web membaca ulasan produk yang terlalu panjang untuk membaca ulasan produk yang terlalu panjang untuk membaca ulasan produk yang terlalu panjang untuk membaca ulasan produk yang terlalu panjang	Algoritma Naïve Bayes tingkat lanjut	Pengujian hanya dilakukan dengan sample pada 3 orang yang berbeda
Andy Januar W. (2016)	Banyaknya pilihan metode analisis sentimen. Tingginya perhatian masyarakat di dunia terhadap pemilihan presiden Amerika Serikat tahun 2016.	Pengumpulan data dengan format data untuk menyesuaikan penerapan <i>electoral college</i> (status, data pengguna, dan lokasi pengguna), Pra-proses, analisis sentimen, agregasi, dan implementasi <i>electoral college</i>	Data asli harus ditulis oleh pengguna yang tinggal di Amerika Serikat dan kelas sentimen yang dapat diprediksi pada analisis sentimen hanya positif dan negatif