

BAB 2

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian ini tidak lepas dari hasil penelitian yang dilakukan terdahulu yang digunakan untuk perbandingan dan kajian. Berdasarkan dari hasil-hasil penelitian yang ada dijadikan perbandingan yang tidak lepas dari topik yang diangkat yaitu tentang analisis sentimen Twitter. Penelitian menurut Rofiqoh, dkk pada 2017 yang sudah dilakukan adalah menganalisis sentimen tingkat kepuasan pengguna pada penyedia layanan telekomunikasi seluler pada Twitter dengan metode Support Vector Machine dan Lexicon Based Features (Rofiqoh, Perdana, & Fauzi3, 2017).

Metode Support Vector Machine diimplementasikan pada analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler dengan hasil klasifikasi berupa sentimen positif atau negatif. Data yang digunakan untuk penelitian ini diambil dari Twitter berupa data teks yang merupakan *tweets* dari pengguna tiga penyedia layanan telekomunikasi seluler paling populer. Data tersebut berjumlah 300 yang dibagi menjadi 300 yang dibagi menjadi dua jenis dengan perbandingan 70%:30% yaitu data latih sebanyak 210 dan

data uji sebanyak 90 (Rofiqoh, Perdana, & Fauzi3, 2017).

Data latih dan uji diproses dalam analisis sentimen melalui tahapan praproses, pembobotan kata *tf-idf*, pembobotan *Lexicon Based Features* berupa perhitungan frekuensi kemunculan kata yang mengandung sentimen positif atau negatif pada tiap dokumen kemudian dinormalisasi dengan metode *Min-max*, dan klasifikasi *Support Vector Machine* untuk mendapatkan hasil analisis dokumen uji berupa sentimen positif atau negatif. Metode *Support Vector Machine* dan *Lexicon Based Features*, didapatkan nilai-nilai parameter *Support Vector Machine* yang optimal yaitu nilai *degree* untuk perhitungan *kernel polynomial* adalah 2 serta nilai konstanta *learning rate* dan jumlah iterasi maksimum untuk perhitungan *sequential training SVM* adalah 0,0001 dan 50. Apabila nilai *degree* diperbesar maka hasil akurasi akan mengalami kemunduran. Sementara nilai konstanta *learning rate* dan jumlah iterasi maksimum semakin diperbesar akan merubah *support vector* yang ditentukan sebagai titik data yang paling dekat dengan garis fungsi pemisah antar kelas data dan menyebabkan tingkat akurasi menurun (Rofiqoh, Perdana, & Fauzi3, 2017).

Penambahan *Lexicon Based Features* memiliki pengaruh untuk analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada Twitter, namun tingkat akurasinya lebih rendah dibandingkan dengan analisis sentimen yang tidak menggunakan *Lexicon Based Features*. Hal ini dikarenakan terdapat kalimat sindiran atau muncul kata-kata

bersentimen positif pada data uji yang seharusnya bersentimen negatif dan sebaliknya (Rofiqoh, Perdana, & Fauzi3, 2017).

Pada penelitian analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada Twitter dengan metode *Support Vector Machine* dan *Lexicon Based Features* menghasilkan nilai *accuracy* sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97%, dan *f-measure* sebesar 78%. Nilai *recall* yang sangat besar tersebut dipengaruhi oleh banyaknya data uji yang sebenarnya positif dan terdeteksi sebagai positif oleh sistem. Hal ini membuktikan bahwa analisis sentimen dengan metode *Support Vector Machine* dan *Lexicon Based Features* baik digunakan untuk klasifikasi analisis sentimen (Rofiqoh, Perdana, & Fauzi3, 2017).

Penelitian juga dilakukan oleh Setiawan dan Setyohadi (2017) dengan melakukan analisis komunikasi sosial media *Twitter* sebagai layanan pelanggan pada salah satu *provider* internet dan seluler di Indonesia. Tujuan utama dari penelitian ini adalah untuk mengetahui bagaimana interaksi yang terjadi antara perusahaan dan pelanggan, serta perilaku perusahaan dalam berinteraksi dengan pelanggan. Analisis fokus pada interaksi yang terjadi melalui layanan pelanggan di sosial media *Twitter* (Setiawan & Setyohadi, 2017).

Penelitian ini memberikan kontribusi membantu perusahaan mengetahui interaksi yang terjadi melalui layanan pelanggan di lingkungan sosial media *Twitter*. Selain itu penelitian ini mengkonfirmasi penggunaan *netnography* dalam penelitian komunitas *online*, khususnya *microblogging Twitter*. Hasil penelitian

mengungkapkan terdapat tiga kelompok utama interaksi pelanggan meliputi keluhan, komentar, dan pertanyaan. Interaksi keluhan memiliki presentase paling besar mencapai 42,20% diikuti komentar 35,55% dan pertanyaan 22,25%. Untuk dapat meningkatkan nilai kompetisi perusahaan harus segera memperbaiki layanan yang diberikan, terutama masalah jaringan dan kecepatan yang merupakan masalah yang paling banyak dikeluhkan. Selain itu, berkaitan dengan banyaknya keluhan yang disuarakan perusahaan perlu menciptakan manajemen keluhan yang baik untuk memaksimalkan kepuasan pelanggan (Setiawan & Setyohadi, 2017).

Secara umum sebagian besar pelanggan (53,13%) mengungkapkan kepuasan /ketidakpuasannya dalam bentuk ekspresi. Ekspresi positif didasarkan pada evaluasi rasional tentang kualitas layanan, sedangkan ekspresi negatif lebih didorong oleh ketidakpuasan terhadap layanan yang diterima pelanggan. Tercatat sedikitnya 13,53% pelanggan menggunakan ekspresi negatif. Temuan lainnya mengungkapkan perilaku perusahaan dalam berinteraksi dengan pelanggan. Secara konsisten perusahaan menggunakan pendekatan secara personal dan bersahabat ketika berkomunikasi dengan pelanggan. Selain itu dalam beberapa kasus perusahaan berusaha mengedepankan nilai-nilai perusahaan dalam berinteraksi dengan pelanggan. Perusahaan berusaha untuk senantiasa menghargai setiap masukan yang diberikan oleh pelanggan. Hal tersebut baik untuk pelanggan maupun perusahaan untuk meningkatkan kualitas hubungan yang lebih baik sehingga diharapkan dapat meningkatkan

kepuasan pelanggan dan citra baik perusahaan (Setiawan & Setyohadi, 2017).

Lebih lanjut, penelitian ini melakukan analisis terbatas hanya pada layanan pelanggan sosial media *Twitter* pada salah satu *provider* internet dan seluler di Indonesia. Ini berarti hasil temuan bersifat spesifik pada lingkungan tertentu dan hasil temuan mungkin tidak dapat digeneralisasikan untuk lingkungan lainnya. Oleh karena itu, studi dalam lingkungan berbeda dan konteks yang bervariasi direkomendasikan.

Penelitian Monarizqa, dkk (2014) tentang penerapan analisis sentimen pada *twitter* berbahasa Indonesia sebagai pemberi *Rating* pada proses ini didapatkan sejumlah 1.358.417 (14%) *twit* mengandung emotikon dengan 215.211 *tweet* mengandung *emosad* (negatif) dan 1.143.206 *tweet* mengandung *emolaugh* dan *emohappy* (positif). Sebagai model, hanya diambil 175.000 *tweet* saja dengan 8.750 *tweet* positif dan 8.750 *tweet* negatif yang kemudian tiap *tweet* nya diproses. Nilai akurasi dari konfigurasi algoritme SVM dengan preprocessing yang diterapkan penelitian ini untuk analisis sentimen pada teks berupa *tweet* berbahasa Indonesia adalah sebesar 73.43%. Konfigurasi ini kemudian menjadi core aplikasi analisis sentimen. Aplikasi analisis sentimen dibangun dengan menggunakan *Twitter API* untuk mengambil *data realtime*, pustaka *Lucene* sebagai alat stemming pada core aplikasi, *SVMlight*, dan *Java*. Ketika konfigurasi diterapkan dalam aplikasi analisis sentimen yang dibuat, nilai akurasi ketika menggunakan kata kunci "Jokowi" sebesar 68%, "Prabowo" sebesar 56%, "kalimilk"

sebesar 70% dan "sunmor" sebesar 74%. (Monarizqa, Nugroho, & Suna, 2014).

Prediksi Pemilihan Presiden Amerika Serikat 2016 Dengan Menganalisis Sentimen Pada Sosial Media (Wicaksono, Suyoto, & Pranowo, 2016). Penelitian ini melakukan perbandingan empat metode klasifikasi sentimen atau analisis sentimen yaitu *Multinomial Naïve Bayes*, *Binarized Multinomial Naïve Bayes*, *SentiWordNet*, dan *AFINN-111*. *Multinomial Naïve Bayes* dan *Binarized Multinomial Naïve Bayes* merupakan model analisis sentimen berbasis pembelajaran (*learning-based*), model ini memerlukan pengetahuan dari data latih untuk dapat melakukan klasifikasi. Sedangkan *SentiWordNet* dan *AFINN-111* merupakan metode klasifikasi sentimen dengan menggunakan leksikon yang berisi kata atau *synset* yang telah diberi nilai sentimen (Monarizqa, Nugroho, & Suna, 2014).

Data uji yang digunakan pada penelitian ini berjumlah 359 yang terbagi menjadi 181 data positif dan 177 data negatif. Metodologi penelitian yang diterapkan pada penelitian ini mampu untuk memprediksi kandidat presiden atau partai politik yang akan memenangkan pemilu presiden Amerika Serikat tahun 2016. Hasil yang didapatkan dari penelitian adalah berdasarkan data yang telah dikumpulkan dari tanggal 7 Agustus 2016 hingga 25 September 2016, mayoritas prediksi pemilihan presiden AS tahun 2016 dimenangkan oleh Donald Trump atau partai Republican. Hasil perbandingan metode analisis sentimen telah disediakan dalam bentuk tabel dan grafik yang mencantumkan nilai indeks perbandingan yang digunakan serta waktu yang

diperlukan metode analisis sentimen untuk memproses data (Monarizqa, Nugroho, & Suna, 2014).

Penelitian Ridwan, dkk (2013) menggunakan algoritma *Naïve Bayes Classifier* (NBC). Penelitian ini secara garis besar meliputi beberapa kegiatan inti yaitu pembuatan proposal, pengumpulan data, pengolahan data, implementasi *NBC*, pengujian, dan analisis hasil. Pengujian ini digunakan data sampel mahasiswa angkatan 2005-2009 yang sudah dinyatakan lulus dari perguruan tinggi. Jumlah data yang digunakan dalam penelitian ini mencapai 100 data dengan menggunakan kelas "Tepat" dan "Tidak Tepat" dimana masing-masing berjumlah 33 dan 67. Proses pengujian ini data dibagi menjadi dua bagian yang terdiri dari data latih dan data uji. Proses percobaan dilakukan sebanyak lima kali percobaan dengan mendapatkan berbagai hasil yang berbeda dan data pengujian pada mahasiswa tingkat akhir tahun 2005-2009 dengan menggunakan NBC adalah menghasilkan nilai *precision*, *recall*, dan *accuracy* masing-masing 83%, 50%, dan 70%. Hasil *mining* NBC dapat digunakan untuk mengklasifikasikan kinerja akademik mahasiswa tahun ke-2 yang dalam penelitian yang dilakukan digunakan sebagai data target (Ridwan, Suyono, & Sarosa, 2013).

Tabel 2.1 Perbandingan dengan Penelitian Sebelumnya

Peneliti	Masalah	Metode	Batasan
Umi Rofiqoh, Rizal Setya Perdana & M. Ali Fauzi (2017)	Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter	Metode Support Vector Machine dan <i>Lexicon Based Features</i>	Terdapat kata bersentimen positif didalam data uji yang seharusnya bersentimen negatif dan sebaliknya.
Retno Agus Setiawan & Djoko Budiyanto Setyohadi (2017)	Analisis Komunikasi Sosial Media Twitter sebagai Saluran Layanan Pelanggan <i>Provider</i> Internet dan Seluler di Indonesia	Metode netnography	Penelitian ini melakukan analisis terbatas hanya pada layanan pelanggan social media <i>Twitter</i> pada salah satu <i>provider</i> internet dan seluler di Indonesia. Ini berarti hasil temuan bersifat spesifik pada lingkungan tertentu dan hasil temuan mungkin tidak dapat digeneralisasikan untuk lingkungan lainnya. Oleh karena itu, studi dalam lingkungan berbeda dan konteks yang bervariasi di rekomendasikan
Peneliti	Masalah	Metode	Batasan
Stefani	Analisis	Metode	Penelitian ini

Aprilia (2018)	Sentimen Masyarakat Tentang <i>Provider Telekomunikasi Indonesia</i> Pada <i>Twitter</i>	<i>Naïve Bayes Classifier</i>	melakukan analisis terbatas hanya pada layanan pelanggan social media <i>Twitter</i> dan kelas sentimen yang dapat diprediksi pada analisis sentimen hanya positif dan negatif
-------------------	--	-------------------------------	--

