

Metode Klasifikasi Analisis Sentimen pada Media Sosial

Bagja Nugraha¹

¹Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Email: *Bagja.nugraha@staff.unsika.ac.id

Abstrak. Di era industri 4.0 dimana banyaknya kegiatan dengan menggunakan internet, persaingan dalam dunia bisnis sangat meningkat dan mendorong setiap perusahaan atau suatu bisnis untuk terus melakukan perkembangan dan mengikuti setiap perubahan yang terjadi. Untuk meningkatkan kualitas suatu perusahaan dan inovasi dari suatu bisnis terdapat beberapa cara salah satunya dengan mengetahui opini apa yang berkembang di masyarakat mengenai perusahaan tersebut. Dari banyaknya opini yang diberikan oleh masyarakat terhadap perusahaan tersebut, selanjutnya akan diproses dengan menggunakan analisis sentimen. Dari banyaknya penelitian, analisis sentimen memiliki banyak metode dalam penyelesaian masalahnya. Namun belum ada penelitian mengenai metode klasifikasi yang paling banyak digunakan dan metode mana yang memiliki tingkat akurasi paling tinggi. Oleh karena itu penelitian ini dilakukan untuk melihat metode klasifikasi yang paling banyak digunakan dan metode klasifikasi yang memiliki tingkat akurasi paling tinggi untuk analisis sentimen pada Sosial Media menggunakan metode *Systematic Literature Review (SLR)*. Hasil dari penelitian ini disimpulkan bahwa metode klasifikasi yang paling banyak digunakan adalah metode *Naïve Bayes Classifier* dan tingkat akurasi dari suatu metode dipengaruhi oleh beberapa faktor seperti jumlah data yang digunakan.

Kata kunci: Analisis Sentimen, metode klasifikasi, *Naïve Bayes Classifier*, Bisnis, *Systematic Literature Review (SLR)*

1 Pendahuluan

Perkembangan teknologi mempengaruhi berbagai aspek atau bidang dalam kehidupan, salah satunya dalam bidang bisnis dan per-industrian. Per Juli 2019 *World Economic Forum* mencatat terdapat 26 perusahaan di dunia yang telah menerapkan industri 4.0, dimana Indonesia termasuk dalam daftar tersebut [1]. Di era industri 4.0 dimana banyaknya kegiatan dengan menggunakan internet, persaingan dalam dunia bisnis sangat meningkat dan mendorong setiap perusahaan atau suatu bisnis untuk terus melakukan perkembangan dan mengikuti setiap perubahan yang terjadi.

Untuk meningkatkan kualitas suatu perusahaan dan inovasi dari suatu bisnis terdapat beberapa cara, salah satunya dengan mengetahui opini apa yang berkembang di masyarakat mengenai perusahaan tersebut. Opini atau komentar tersebut dapat dilihat dari aplikasi atau website perusahaan tersebut, atau dari

sosial media. Dari banyaknya opini yang diberikan oleh masyarakat terhadap perusahaan tersebut, akan membutuhkan waktu yang lama untuk memisahkan secara manual beberapa macam komentar seperti komentar positif, negatif dan netral. Komentar yang diberikan masyarakat dalam bentuk teks selanjutnya akan diproses dengan menggunakan analisis sentimen yang merupakan cabang ilmu dari *text mining*. *Text mining* merupakan salah satu bidang dalam *Natural Language Processing* (NLP) yang biasa digunakan untuk memproses data khusus yaitu data dalam bentuk teks [2].

Di bidang pendidikan, banyak penelitian dengan menggunakan analisis sentimen yang telah dilakukan oleh mahasiswa, dosen, guru, dan peneliti. Dari banyaknya penelitian, analisis sentimen memiliki banyak metode dalam penyelesaian masalahnya. Namun belum ada penelitian mengenai metode apa yang paling banyak digunakan dan metode mana yang memiliki tingkat akurasi paling tinggi untuk penelitian di bidang bisnis.

Karena jumlah penelitian mengenai analisis sentimen di bidang bisnis cukup banyak, oleh karena itu penulis akan menggunakan metode *Systematic Literature Review* (SLR). *Systematic Literature Review* atau yang biasa disingkat dengan SLR merupakan suatu metode penelitian dimana isinya berupa ulasan kembali mengenai topik tertentu yang menekankan rumusan masalah yang sudah ditetapkan sebelumnya berdasarkan penelitian sebelumnya yang sesuai dengan rumusan masalah penelitian [3]. Tujuan dari penelitian ini adalah untuk melihat metode apa yang paling banyak digunakan dan metode apa yang memiliki tingkat akurasi paling tinggi untuk analisis sentimen pada bidang bisnis dengan menggunakan metode *Systematic Literature Review* (SLR).

2 Metode Penelitian

Penelitian ini merupakan sebuah *Systematic Literature Review* (SLR). Sumber data yang dipakai berasal dari beberapa jurnal yang dicari melalui *Google Scholar*, berupa penelitian sebelumnya mengenai analisis sentimen Opini Masyarakat pada Sosial Media baik berupa laporan penelitian, skripsi, jurnal maupun prosiding. Pengumpulan data dilakukan dengan menentukan *variable* apa yang diperlukan dengan cara studi literatur dan paper yang digunakan merupakan paper dari tahun 2015 – 2019. Tujuannya adalah untuk mengetahui metode yang paling banyak dan paling baik untuk penelitian terhadap metode analisis sentimen untuk kepentingan bisnis.

3 Hasil & Pembahasan

Pada penelitian yang dilakukan oleh Hilda Rachmi pada tahun 2017 mengenai analisis sentimen terhadap *review* barang pada suatu website yang menyediakan *review* untuk berbagai kategori, metode yang digunakan adalah *Support Vector Machine* (SVM), *Principal Component Analysis* dan *Genetic*

Algorithm. Penelitian ini dilakukan untuk mengklasifikasikan *review* pengiriman barang menjadi 2 kelas yaitu kelas positif dan kelas negatif. Untuk hasil tingkat akurasi dari metode SVM yaitu 86.00% sedangkan untuk metode SVM yang telah dioptimalkan menggunakan *Principal Component Analysis* dan *Genetic Algorithm* memiliki tingkat akurasi yang lebih tinggi yaitu 97%. [4]

Pada penelitian [5] metode yang digunakan adalah *Fuzzy Logic* dimana data yang dikumpulkan berasal dari social media yaitu twitter dan facebook. Parameter sentiment pada penelitian ini telah ditentukan yaitu untuk parameter positif 211 kata dan untuk parameter entimen negative 389 kata. Pada penelitian ini penulis membandingkan antara hasil dari sistem analisis sentimen dengan cara manual yaitu pengecekan satu persatu percakapan. Hasil dari penelitian ini keakuratan antara perbandingan 2 metode yang telah dijelaskan adalah 63.15%, dimana media twitter dinilai lebih baik dibandingkan media facebook dikarenakan pada media twitter pembuatan status dibatasi hanya 10 karakter sedangkan media facebook tidak dibatasi sehingga lebih banyak *noise* dalam data.

Penelitian yang dilakukan oleh Faishal Nuruz Zahri dan Andry Alamsyah menggunakan metode *Naïve Bayes Classifier*. Data yang digunakan adalah opini berbahasa Indonesia mengenai merk Smartfren di forum Kaskus, dimana jumlah datanya sebanyak 7386 kalimat opini. Untuk pengujianya peneliti membuat 2 skenario yaitu, skenario pertama komposisi data latih tidak seimbang atau acak sedangkan scenario kedua komposisi data latih seimbang antara sentimen positif dan sentiment negatif. Hasil dari penelitian ini adalah tingkat akurasi sebesar 98.40% dengan menggunakan 1000 data latih dan data uji sebanyak 6388 dimana hasil klasifikasi menunjukkan 4049 adalah sentimen positif dan 2339 adalah sentiment negatif [6].

Pada penelitian [7] data yang digunakan adalah data berupa opini yang didapatkan dari akun twitter perusahaan jasa ekspedisi JNE. Jumlah data uji yang digunakan adalah 1500 yang kemudia diklasifikasikan ke dalam 3 kelas yaitu positif, negative dan netral. Seluruh data uji dan data latih dilakukan tanpa tahapan *preprocessing* menggunakan metode *Naïve Bayes Classifier* dan menghasilkan tingkat akurasi sebesar 83%.

Pada penelitian [2] metode yang digunakan adalah *K-Nearest Neighbor*, dengan jumlah data sebedar 335 data berupa komentar yang didapatkan dari akun resmi J&T Indonesia di facebook yang sudah dikelompokkan secara manual oleh pihak J&T. Hasil dari penelitian ini adalah tingkat akurasi yang didapatkan sebesar 79.21% dengan jumlah data latih yaitu 101. Hasil dari pengklasifikasian dipengaruhi oleh jumlah data latih dan nilai k yang digunakan.

Penelitian yang dilakukan oleh Muhammad Apriandito Arya Saputra dan Andry Alamysah menggunakan data yang didapat dari Paperlust.Co dan menggunakan metode *Naïve Bayes* sebagai metode dalam pengklasifikasian. Pada tahap pengujian perbandingan antara data latih dan data uji adalah 30:70.

Penelitian ini menghasilkan tingkat akurasi sebesar 81.18% dimana persentase antara sentimen netral, negative dan positif yaitu 67% untuk sentimen netral, 8% untuk sentimen negatif, dan 25% untuk sentimen positif [8].

Pada penelitian [9] sumber data berasal dari *mention* akun JNECare di sosial media twitter dengan rentang data dari 14 Februari 2018 – 19 Februari 2018 yang menghasilkan data sebanyak 2349 data. Metode yang digunakan adalah *Naïve Bayes* yang menghasilkan 97% dimana sebagian besar data berisi komentar netral dan negatif dikarenakan untuk pelanggan yang tidak mengalami kendala tidak memberikan komentar.

Pada penelitian [10] metode yang digunakan adalah *Naïve Bayes Classifier* dan data yang digunakan merupakan opini konsumen pada suatu produk pada aplikasi Goodreads, Twitter, dan Tokopedia. Penelitian ini menghasilkan tingkat akurasi sebesar 70,43% - 88,03% dengan kategori dari hasil klasifikasi yaitu 3 kelas.

Tingkat akurasi dari setiap metode berbeda – beda, bahkan untuk metode yang sama juga memiliki tingkat akurasi yang berbeda. Hal ini disebabkan karena tingkat akurasi dari setiap metode dipengaruhi oleh jumlah data yang digunakan, dan jumlah kategori yang dibuat. Untuk data lebih detailnya dapat dilihat pada **Tabel 1**.

Tabel 1. Tingkat Akurasi

No	Metode	Tingkat Akurasi
1	<i>Support Vector Machine (SVM), Principal Component Analysis dan Genetic Algorithm</i>	SVM = 86% SVM yang telah dioptimalkan = 97%
2	<i>Fuzzy Logic</i>	63,15%
3	<i>Naïve Bayes Classifier</i>	70,43% - 98,40%
4	<i>K- Nearest Neighbor</i>	79,21%

Dapat dilihat pada **Tabel 1**. Metode *Naïve Bayes Classifier* memiliki tingkat akurasi paling tinggi pada tingkat akurasi maksimal yaitu 98,40% dan metode yang memiliki tingkat akurasi paling rendah adalah metode *Fuzzy Logic* sebesar 63,15%.

4 Kesimpulan

Berdasarkan hasil penelitian dengan menggunakan metode *Systematic Literature Review (SLR)* dapat disimpulkan bahwa metode yang paling tingkat akurasinya paling tinggi adalah metode *Naïve Bayes Classifier*. Untuk menentukan metode yang memiliki tingkat akurasi yang paling tinggi tidak dapat dilakukan begitu saja, karena tingkat akurasi dari suatu metode dipengaruhi oleh beberapa faktor seperti jumlah data yang digunakan. Analisis

sentimen pada Media Sosial dalam dunia bisnis digunakan untuk mengetahui bagaimana opini masyarakat terhadap suatu produk atau jasa sehingga perusahaan dapat meningkatkan kualitasnya.

5 Referensi

- [1] A. Lidwina and S. Fitra, “Inilah Negara-negara yang Sudah Menerapkan Industri 4.0,” *databoks*, 2019. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2019/10/14/inilah-negara-negara-yang-sudah-terapkan-industri-40>. [Accessed: 22-Dec-2019].
- [2] A. Salam, J. Zeniarja, R. Septiyan, U. Khasanah, P. Studi, T. Informatika, F. I. Komputer, U. D. Nuswantoro, and S. Analysis, “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa,” in *Prosiding SINTAK*, 2018, pp. 480–486.
- [3] W. Adisasmito, “Faktor Risiko Diare pada Bayi dan Balita di Indonesia : Systematic Review Penelitian Akademik,” *Makara, Kesehat.*, vol. 11, no. 1, pp. 1–10, 2007.
- [4] H. Rachmi, “Penerapan Principal Component Analysis dan Genetic Algorithm pada Analisis Sentimen Review Pengiriman Barang Menggunakan Algoritma Support Vector Machine,” *J. Evolusi*, vol. 5, no. 2, pp. 97–104, 2017.
- [5] B. Subaeki, F. Gunawan, and A. R. Atmadja, “Penggunaan Metode Fuzzy Logic untuk Pemantauan Sentimen Brand pada Media Sosial,” *QUERY J. Sist. Inf.*, vol. 01, no. 02, pp. 56–62, 2017.
- [6] F. N. Zuhri and A. Alamsyah, “Analisis Sentimen Masyarakat Terhadap Brand Smartfren Menggunakan Naive Bayes Classifier di Forum Kaskus,” *e-Proceeding of Management*, vol. 4, no. 1, pp. 242–251, 2017.
- [7] R. Raksanagara, Y. H. Chrisnanto, and A. I. Hadiana, “Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naive Bayes,” in *Prosiding SNST ke-7*, 2016, pp. 19–24.
- [8] M. A. A. Saputra and A. Alamsyah, “Analisa obrolan pelanggan menggunakan metode text mining untuk manajemen hubungan pelanggan (studi kasus: paperlust.co),” in *e-Proceeding of Management*, 2019, vol. 6, no. 1, pp. 781–785.
- [9] F. S. Jumeilah, “Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naive Bayes,” *J. Sist. Inf. Bisnis*, vol. 01, pp. 92–98, 2018.
- [10] S. Hanggara, T. M. Akhriza, and M. Husni, “Aplikasi Web untuk Analisis Sentimen pada Opini Produk dengan Metode Naive Bayes,” in *Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri 2017*, 2017, pp. 1–6.