



ANALISIS SENTIMEN VAKSIN SINOVAC PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES

Sri Lestari¹⁾, Sudin Saepudin²⁾

^{1, 2)} Program Studi Sistem Informasi, Universitas Nusa Putra
Jl. Raya Cibolang No. 21 Cibolang Kaler, Cisaat, Sukabumi, Jawa Barat 43152
e-mail: sri.lestari_si18@nusaputra.ac.id¹⁾, sudin.saepudin@nusaputra.ac.id²⁾

* Korespondensi: e-mail: sri.lestari_si18@nusaputra.ac.id

ABSTRAK

Saat ini angka orang yang terinfeksi virus Covid-19 terus meningkat terutama di Indonesia. Untuk menekan angka yang terinfeksi semakin tinggi, Indonesia melaksanakan vaksinasi Covid-19 secara gratis kepada penduduk Indonesia dimulai dari Presiden, kemudian tenaga kesehatan sampai dengan lansia dan anak-anak. Salah satu Vaksin Covid-19 yang dipakai di Indonesia yaitu vaksin Sinovac. Dengan adanya vaksinasi ini banyak sekali tanggapan atau opini dari masyarakat Indonesia, yang menimbulkan pro dan kontra. Pada media sosial twitter, Sinovac menjadi kata populer yang banyak dibahas oleh masyarakat Indonesia saat ini. Opini pada twitter ini yang kemudian akan menjadi data untuk penelitian analisis sentimen opini masyarakat terhadap vaksin sinovac. Pada penelitian ini menggunakan metode Naive bayes dengan Tools RapidMiner untuk pengklasifikasian sentimen opini positif atau negatif terhadap vaksin Sinovac pada tweet berbahasa Inggris. Hasil dari penelitian ini menunjukkan bahwa tweets dengan sentimen positif sebanyak 86%, sedangkan tweets dengan sentimen negatif sebanyak 14%. Dengan nilai akurasi dari perhitungan menggunakan algoritma Naive Bayes adalah 92,96%.

Kata Kunci: analisis sentimen, naïve bayes, sinovac, twitter

ABSTRACT

Currently, the number of people infected with the Covid-19 virus continues to increase, especially in Indonesia. To reduce the number of infected people getting higher, Indonesia carried out free Covid-19 vaccinations for the Indonesian population starting from the President, then health workers to the elderly and children. One of the Covid-19 vaccines used in Indonesia is the Sinovac vaccine. With this vaccination, there are many responses or opinions from the Indonesian people, which raise pros and cons. On social media twitter, Sinovac is a popular word that is widely discussed by Indonesian people today. This opinion on Twitter will then become data for research on sentiment analysis of public opinion on the Sinovac vaccine. In this study, using the Naive Bayes method with RapidMiner Tools for classifying positive or negative opinion sentiments on the Sinovac vaccine on English-language tweets. The results of this study show that tweets with positive sentiments are as much as 86%, while tweets with negative sentiments are 14%. With the value of accuracy of calculations using the Naive Bayes algorithm is 92.96%.

Keywords: sentiment analysis, nave bayes, sinovac, twitter

I. PENDAHULUAN

Saat ini dunia sedang dilanda permasalahan yang sama mengenai adanya wabah virus covid-19. Virus covid-19 sendiri ditemukan pertama kali di daerah Wuhan, China. Wabah ini dapat tertular melalui interaksi langsung dengan orang yang terinfeksi. Karena adanya wabah baru ini, peneliti di seluruh dunia membuat penawar atau obat yang bisa menyembuhkan wabah menular ini agar bisa menghilangkan wabah covid-19 dari dunia, salah satu obat yang dibuat adalah vaksin Sinovac.

Vaksin Sinovac adalah vaksin untuk mencegah infeksi virus SARS-CoV-2 atau COVID-19. Vaksin Sinovac yang dikenal juga dengan nama CoronaVac sudah mendapat izin penggunaan darurat dari Badan Pengawas Obat dan Makanan (BPOM) RI. CoronaVac merupakan vaksin yang mengandung virus SARS-CoV-2 yang sudah tidak aktif. Penyuntikan vaksin Sinovac akan memicu sistem kekebalan tubuh untuk mengenali virus yang sudah tidak aktif ini dan memproduksi antibodi untuk melawannya sehingga tidak terjadi infeksi COVID-19. Vaksin Sinovac dikembangkan oleh Sinovac Biotech Ltd. Vaksin ini sudah melewati uji klinis fase ketiga yang dilakukan di Brazil, Turki, dan Indonesia [1].

Di Indonesia sendiri vaksin sinovac telah diberikan kepada masyarakat dalam jumlah banyak ke



seluruh Indonesia dimulai dari petugas kesehatan hingga lansia. Pemberian vaksin ini diharapkan dapat meminimalisir penyebaran wabah covid-19 di Indonesia yang saat ini sudah ada klaster-klaster baru. Berikut ini merupakan grafik dari jumlah penerima vaksin dosis 1 dan dosis 2 di Indonesia sampai dengan tanggal 16 Juni 2021.



Sumber : <https://vaksin.kemkes.go.id/>

Gambar.1. Grafik penerima vaksin dosis 1 dan 2 di Indonesia

Respon atau tanggapan masyarakat terhadap adanya vaksin Sinovac ini beragam. Ada yang setuju dengan pengadaan vaksin ini dan ada pula yang menentang sehingga tidak ingin di vaksin. Dengan kemajuan teknologi saat ini, masyarakat memberikan opini dan tanggapan nya melalui media sosial salah satunya twitter. Pada media sosial twitter, Sinovac menjadi kata populer yang banyak dibahas oleh masyarakat saat ini, terutama di Indonesia. Opini pada twitter ini yang kemudian akan menjadi data untuk penelitian analisis sentimen opini pengguna twitter terhadap vaksin sinovac.

Analisis sentimen adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat dalam tulisan menggunakan teknik analisis teks [2]. *Sentiment analysis* adalah proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam *platform* media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada *platform* tersebut. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber [3].

Metode yang digunakan dalam penelitian ini yaitu metode Naive bayes dengan *Tools RapidMiner* untuk pengklasifikasian sentimen opini positif atau negatif terhadap vaksin Sinovac pada tweet berbahasa Inggris serta nilai akurasi dari implementasi algoritma Naive Bayes.

Fajar Ratnawati, (2018) dalam penelitiannya yang berjudul Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter menyimpulkan bahwa banyaknya opini yang dituliskan di twitter membutuhkan pengklasifikasian sesuai sentimen yang dimiliki agar mudah untuk mendapatkan kecenderungan opini tersebut terhadap film apakah cenderung beropini positif atau negatif. Algoritma yang digunakan pada penelitian ini adalah Algoritma Naive Bayes. Berdasarkan hasil eksperimen, hasil akurasi akan semakin tinggi dan itu menandakan sistem berhasil melakukan klasifikasi dengan baik. Akurasi tertinggi didapat pada fold kedua yaitu 90%, precision 92%, Recall 90% dan f-



measure 90%. [4]

Penelitian ini memberikan solusi untuk mengetahui tingkat persentase komentar dan opini positif atau negatif dari pengguna akun sosial media Twitter terhadap vaksin sinovac. Sehingga berdasarkan penelitian ini kita dapat mengetahui bagaimana tanggapan pengguna twitter terhadap pengadaan vaksin Sinovac., dimana dengan persentase dan respon tersebut akan menjadi masukan dan evaluasi bagi perusahaan dan pemerintah tentang adanya pengadaan vaksin Sinovac. Penelitian ini memberikan manfaat selain bisa memberikan informasi pada pihak pengembang yaitu Sinovac Biotech Ltd tentang bagaimana respon masyarakat terhadap vaksin sinovac tersebut.

II. TINJAUAN PUSTAKA

A. Penelitian terkait

Penelitian sebelumnya yang dilakukan oleh (Samsir et al. 2021) dengan judul Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes menghasilkan kesimpulan bahwa pembelajaran daring memiliki 30% sentimen positif, 69% sentimen negatif, dan 1% netral pada periode tersebut [5].

Adapun penelitian yang dilakukan oleh (Rustiana & Rahayu, 2017) dengan judul Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes menghasilkan nilai akurasi sebesar 93%, dan tingkat precision positif sebesar 90%, precision negatif sebesar 90% dan sentimen yang netral 100% sehingga naïve bayes mampu menganalisis dengan akurasi yang tinggi [6].

Penelitian selanjutnya yang berkaitan dengan analisis sentimen yaitu penelitian yang dilakukan oleh (Nugraha, 2014). Pada penelitian ini dilakukan pengklasifikasian terhadap sentiment positif dan negatif mengenai review film berbahasa Inggris dengan menggunakan metode K-Nearest Neighbor (K-NN). Pada penelitian ini membandingkan metode K-NN dengan metode Naive Bayes. Hasil dari akurasi yang didapat menunjukkan bahwa metode Naive Bayes lebih baik dibandingkan dengan metode K-NN dimana metode Naive Bayes memiliki nilai akurasi tertinggi sebesar 81% dibandingkan dengan menggunakan metode K-NN yang bernilai akurasi sebesar 71%. Dan pada penelitian ini juga masih belum dilakukannya perbaikan kata tidak baku, sehingga tingkat akurasi tertinggi yang dicapai untuk metode Naive Bayes hanya sebesar 81% [7].

Penelitian yang dilakukan oleh Dudih Gustian, Indah Suciati, dan Sudin Saepudin mengenai Sistem Pakar Dengan Algoritma Naive Bayes Untuk Prediksi Hasil Produksi Ayam Broiler Plasma (Studi Kasus: Pt.Sekawan Sinar Surya) menyimpulkan bahwa algoritma Naive Bayes mampu menghasilkan prediksi klasifikasi keuntungan dan kerugian sebelum perhitungan realisasi pada bagian produksi. Untuk mengetahui nilai keakuratan pengujian dilakukan dengan menggunakan Software Weka. Akurasi rata-rata dari algoritma yang dihasilkan, Naive Bayes dari 96.36% mendekati angka 100% menunjukkan hasil yang maksimal. dan dengan Jurnal Rekayasa Teknologi Nusa Putra. Vol. 6, No. 1, September 2019: Hal 54 - 62 57 nilai Receiver Operating Curve (ROC) 0,9995 mendekati angka 1 maka model yang dihasilkan lebih baik [8].

B. Text Mining

Text mining (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda [9].

C. Text Preprocessing

Tahap *text preprocessing* adalah tahap awal dari *text mining*. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery sistem text mining* [9]. Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil dan *Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut [10].

D. Feature Selection

Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat [9]. Pada tahap ini tindakan yang dilakukan adalah menghilangkan *stopword* (*stopword removal*) dan *stemming* terhadap kata yang berimbuhan [11].

E. Sentiment Analysis

Sentiment analysis atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisis pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu [12].

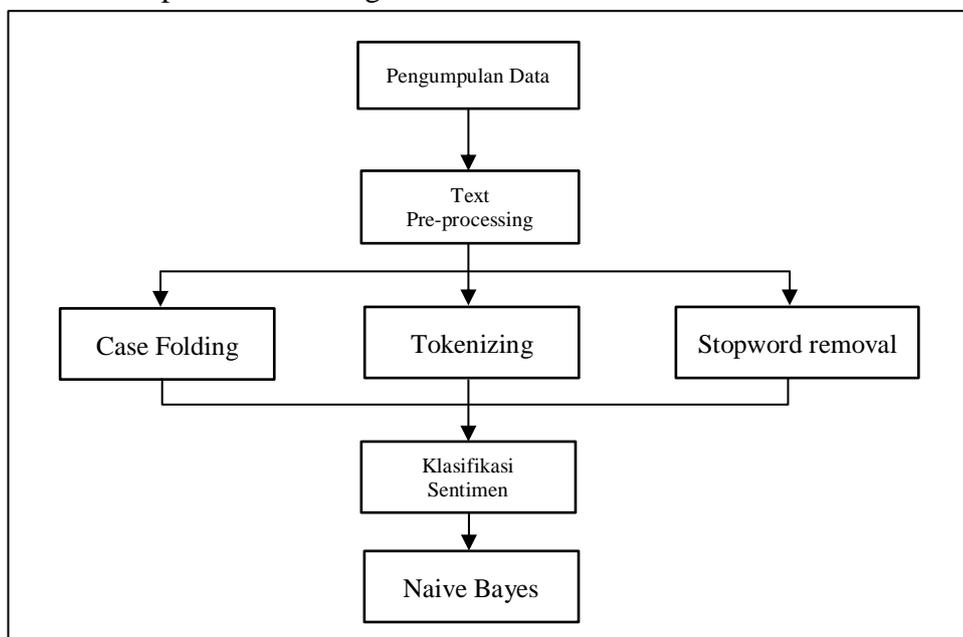
Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral [13]. *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

F. Naïve Bayes Classifier

Algoritma *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [9]. Dalam penelitian ini yang menjadi data uji adalah dokumen tweets. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

III. METODOLOGI PENELITIAN

Dalam penelitian ini objek yang diteliti oleh penulis adalah opini masyarakat terhadap vaksin Sinovac di media sosial twitter. Data yang digunakan berupa *tweet* (*mention, reply, like, retweet*) berbahasa Inggris. Pada penelitian ini menggunakan Tool Rapidminer dalam melakukan beberapa langkah metode untuk mendapatkan dan menganalisa data.



Gambar.2. Alur Penelitian

a. Pengumpulan Data

Pengumpulan Data: Pengumpulan data dilakukan dengan menggunakan python dengan *tweepy library*. Penulis mengekstraksi 1.975 data tweet pengguna dengan kata kunci “sinovac” yang berbahasa Inggris pada media sosial twitter. Hasil pencarian kemudian akan menampilkan semua tweet yang mengandung kata kunci “sinovac”. Semua tweet yang diperoleh akan diolah ke proses



selanjutnya menggunakan *tools Rapidminer*.

b. *Text Preprocessing*

Text Preprocessing: Tahap *preprocessing* adalah tahapan dimana dilakukan seleksi data agar data yang akan digunakan menjadi lebih terstruktur. Tahap *text preprocessing* pada penelitian ini menggunakan beberapa tahapan, diantaranya: *case folding*, *tokenizing*, dan *filtering*.

- *Letter casing (case folding)*: Proses mengkonversi keseluruhan teks dalam tweet menjadi bentuk standar (huruf kecil atau *lowercase*).
- *Tokenizing*: Proses Mengubah tweet menjadi token. Token adalah kata-kata yang dipisahkan oleh spasi dalam teks.
- *Stopword removal*: Proses mengambil kata-kata penting dan membuang kata-kata yang tidak penting pada tweet (*stopwords*) misalnya kata penghubung seperti "dan", "atau", "kemudian" dan seterusnya atau kata-kata yang tidak berpengaruh pada proses klasifikasi.

c. *Klasifikasi*

Klasifikasi: tweet yang telah melalui *preprocessing* akan diklasifikasi sesuai kelasnya (*sentiment class*) untuk menentukan polaritas dari text tersebut apakah text tersebut termasuk dalam opini positif, negatif atau netral.

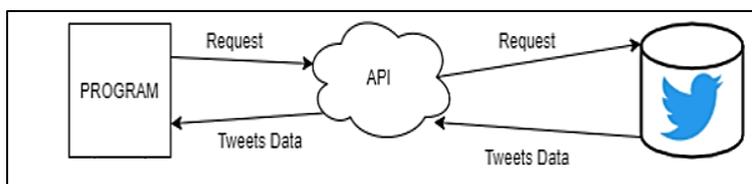
d. Penerapan metode Naive Bayes

Data yang sudah di klasifikasikan sentimen nya kemudian melalui tahap *cross validation* yaitu tahap untuk mendapatkan nilai akurasi dengan metode naive bayes.

gambar

IV. HASIL DAN PEMBAHASAN

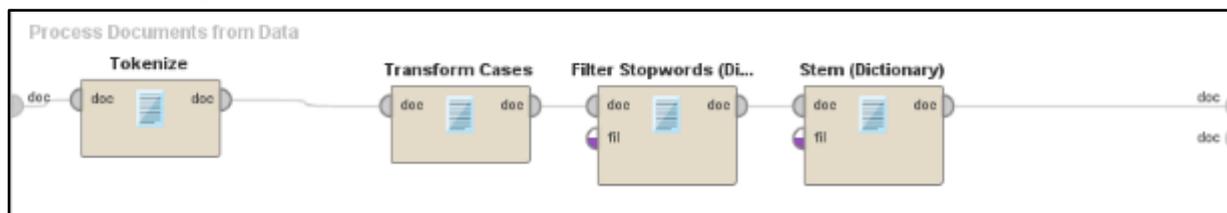
A. Pengumpulan data



Gambar.3. Pengumpulan data Tweets

Data *tweets* didapatkan menggunakan program python dengan *tweepy library* dari Twitter yang diakses langsung menggunakan Twitter *API key* dan *Access token*. Kata kunci yang digunakan untuk menangkap *tweets* yaitu "sinovac" yang diambil dari periode Januari - Juni 2021 dan dan berhasil mendapatkan 1.976 data *tweets* berbahasa Inggris. Setelah data terkumpul kemudian dikonversikan kedalam file excel (.xlsx).

B. *Pre-processing Data*



Gambar.4. Tahap preprocessing data

Tahap preprocessing data meliputi tahap sebagai berikut :

- *Tokenize / Remove Punctuation and Clean Number* : yaitu menghapus semua karakter non alphabet, misalnya simbol, spasi, angka dan lain-lain menggunakan fungsi *tokenize* pada RapidMiner.



accuracy: 92.96% +/- 1.85% (micro average: 92.96%)

	true positive	true negative	class precision
pred. positive	1614	56	96.65%
pred. negative	83	222	72.79%
class recall	95.11%	79.86%	

Gambar.8. Nilai akurasi dengan algoritma Naïve Bayes

Nilai akurasi yang didapatkan dengan algoritma naïve bayes ini adalah 92,96%. Dengan nilai akurasi *recall* untuk hasil positif adalah 95,11%, sedangkan nilai akurasi *recall* untuk hasil negatif adalah 79,86%. Nilai akurasi *precision* untuk pred. positif adalah 96,65%, sedangkan nilai akurasi *precision* untuk pred. negatif adalah 72,79%. Dapat disimpulkan bahwa algoritma naïve bayes ini memiliki nilai akurasi yang tinggi sehingga hasil dari prediksinya akurat untuk dijadikan metode dalam penelitian Data Mining.

V. KESIMPULAN

Berdasarkan analisis dan pengujian yang telah dilakukan, maka kesimpulan yang didapat adalah sebagai berikut :

1. Hasil dari klasifikasi sentimen tweets pengguna twitter dengan kata kunci “sinovac’ menunjukkan bahwa tweets dengan sentimen positif sebanyak 86%, sedangkan tweets dengan sentimen negatif sebanyak 14%. Hal ini dapat diartikan bahwa ternyata respon pengguna twitter terhadap vaksin sinovac tidak terlalu buruk yang artinya mereka menerima dengan positif mengenai pengadaan vaksin sinovac ini.
2. Dalam pengolahan data menggunakan algoritma Naïve Bayes menghasilkan nilai akurasi 92,96%. Dengan nilai akurasi *class recall* untuk hasil positif adalah 95,11%, sedangkan nilai akurasi *class recall* untuk hasil negatif adalah 79,86%. Nilai akurasi *class precision* untuk pred. positif adalah 96,65%, sedangkan nilai akurasi *class precision* untuk pred. negatif adalah 72,79%. Dapat disimpulkan bahwa algoritma naïve bayes ini memiliki nilai akurasi yang tinggi sehingga hasil dari prediksinya akurat untuk dijadikan metode dalam penelitian selanjutnya mengenai Analisis Sentimen.

DAFTAR PUSTAKA

- [1] “Vaksin Sinovac - Manfaat, dosis dan efek samping - Alodokter.” <https://www.alodokter.com/vaksin-sinovac> (accessed Jun. 19, 2021).
- [2] “Sentiment Analysis: Pengertian, Teknik, dan Penggunaannya - Glints Blog.” <https://glints.com/id/lowongan/sentiment-analysis/#.YNHTL-gzBIU> (accessed Jun. 22, 2021).
- [3] “Mengenal peran sentiment analysis beserta cara kerjanya.” <https://www.ekrut.com/media/sentiment-analysis-adalah> (accessed Jun. 22, 2021).
- [4] U. Moh. Syafik, Daryanto, and A. M. Lutfi, “SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN SISWA BERPRESTASI MENGGUNAKAN METODE FUZZY MULTI ATTRIBUTE DECISION MAKING,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.
- [5] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [6] D. Rustiana and N. Rahayu, “Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 8, no. 1, pp. 113–120, 2017, doi: 10.24176/simet.v8i1.841.
- [7] M. W. A. Nugraha, “Sentiment Analysis Pada Review Film Dengan Menggunakan Metode K-Nearest Neighbor,” Universitas Widyatama, 2014.
- [8] D. Gustian, I. Suciati, and S. Saepudin, “SISTEM PAKAR DENGAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI HASIL PRODUKSI AYAM BROILER PLASMA (STUDI KASUS : PT . SEKAWAN SINAR SURYA),” no. 12.



- [9] R. Feldman, J. Sanger, and others, *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press, 2007.
- [10] S. M. Weiss, N. Indurkha, T. Zhang, and F. J. Damerau, "From textual information to numerical vectors," in *Text Mining*, Springer, 2005, pp. 15–46.
- [11] M. W. Berry and J. Kogan, "Text mining," *Appl. Theory. West Sussex, PO19 8SQ, UK John Wiley & Sons*, 2010.
- [12] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [13] M. Dehaff, "Sentiment Analysis, Hard But Worth It!," *Tersedia di [http://www. Cust.com/blog/sentiment_analysis_hard_but_worth_it](http://www.Cust.com/blog/sentiment_analysis_hard_but_worth_it)*, 2010.