

# Literatur Review: Metode Klasifikasi Pada Sentimen Analisis

Mona Cindo, Dian Palupi Rini\*, Ermatita

Fakultas Ilmu Komputer, Magister Teknik Informatika, Universitas Sriwijaya, Palembang, Indonesia

Email: monacindo@gmail.com, \*)dprini@unsri.ac.id, ermatitaz@yahoo.com

## Abstrak

Sentimen analisis merupakan suatu ilmu yang mencari emosi ataupun opini masyarakat baik berupa teks, maupun dari *vocal*. Sentimen analisis merupakan salah satu cabang dari *text mining* yang sedang marak dilakukan saat ini. Pada dasarnya sentimen analisis memiliki dua tipe pendekatan yaitu *machine learning* dan *lexicon based*. Beberapa metode klasifikasi *machine learning* seperti *naïve bayes*, *support vector machine*, *logistic regression* dan *lexicon-based* sering digunakan untuk mendapatkan hasil yang terbaik. Selain itu pemilihan fitur ekstraksi yang tepat dapat meningkatkan tingkat akurasi data dalam menentukan sentimen. Fitur ekstraksi ini memanfaatkan seluruh sumber *resource* dan karakteristik data yang telah dikumpulkan sehingga membantu metode klasifikasi dalam menentukan sentimen.

**Kata Kunci:** Sentiment Analisis, *Text Mining*, Klasifikasi, *Machine Learning*, *Lexical-Based*, Fitur Ekstraksi

## 1. PENDAHULUAN

Perkembangan teknologi informasi yang sangat pesat mengakibatkan berubahnya cara masyarakat berkomunikasi, pada saat menggunakan teknologi telepon selular. Salah satu teknologi yang populer dikalangan masyarakat adalah media sosial. Media sosial seperti Facebook, Twitter, Instagram, Youtube dan Google+ sering sekali dimanfaatkan untuk sarana bertukar informasi, menyampaikan pendapat atau pandangan terhadap sesuatu dan mengekspresikan opini, pengalaman maupun hal lain yang menjadi *trend* perhatian pada masyarakat. Hal tersebut sering disebut dengan sentimen.

Analisis sentimen merupakan suatu proses yang digunakan untuk menentukan opini, emosi dan sikap yang dicerminkan melalui teks, biasanya diklasifikasikan menjadi opini negatif dan positif. Analisis sentimen pada suatu kalimat menggambarkan bagian pertimbangan penilaian terhadap entitas atau kejadian tertentu [1]. Analisis sentimen mengolah bahasa alami untuk mengetahui pandangan masyarakat tentang produk atau topik tertentu. Analisis sentimen merupakan sistem untuk mengumpulkan dan memeriksa opini tentang produk atau layanan yang dibuat dalam postingan web, blog, ataupun komentar di media sosial. Penggunaan analisis sentimen dapat diterapkan pada berbagai opini seperti opini kebencian, opini umum, film, politik, merek produk dan lainnya pada media sosial.

Hal tersebut menandakan bahwa analisis sentimen dapat digunakan untuk mendapatkan informasi yang ada di dalam pesan para pengguna jejaring sosial terhadap topik yang dibahas. Informasi akan diklasifikasikan ke dalam kelas positif apabila informasi yang disampaikan bernilai baik atau setuju terhadap sesuatu [2]. Ada beberapa pendekatan yang digunakan peneliti untuk melakukan sentimen analisis pada sosial media. [3] Seperti menggunakan pendekatan *supervised learning* untuk melakukan sentimen analisis. Peneliti menggunakan metode *Support Vector Machine* untuk memprediksi pergerakan stok pada saham suatu perusahaan. Beberapa penelitian lain menggunakan pendekatan *lexicon based* untuk melakukan sentimen analisis dengan *emoticon*. Selain itu yang paling sering digunakan sekarang adalah penggabungan dari pendekatan tersebut. [4] menggabungkan pendekatan *lexicon-based* yang kemudian diolah lagi dengan menggunakan pendekatan *Naïve bayes* yang merupakan pendekatan *supervised learning*.

## 2. METODOLOGI PENELITIAN

### 2.1 Analisis sentimen

Analisis sentimen merupakan salah satu penelitian yang cukup kompleks. Adapun karakteristik dari analisis sentimen adalah sebagai berikut ini[5].

- Pengkategorian sentiment yang akan membedakan antara kalimat subjektif dan objektif.
- Tingkatan analisis dibagi menjadi 3 bagian yaitu *message level*, *sentence level* dan *aspect level*.
- Pendapat yang memberikan perbandingan terhadap sesuatu serta pendapat yang hanya sekedar pendapat. Ini memiliki maksud setiap orang dapat memberikan pendapat dengan membandingkan suatu hal dengan hal yang lain atau hanya sekedar memberikan pendapat.
- Pembagian pendapat menjadi eksplisit dan implisit. Pendapat yang diungkapkan secara jujur, tegas serta lugas dan jelas atau pendapat yang diungkapkan secara tidak jelas.

### 2.2 Pendekatan Pada Sentimen Analisis

Pada dasarnya pendekatan pada sentiment analisis terbagi menjadi dua, yaitu : pendekatan *supervised learning* dan pendekatan berbasis *lexicon* [6].

#### a. Pendekatan *Supervised learning*

Beberapa algoritma yang biasa digunakan terhadap pendekatan *supervised learning* [6], diantaranya *naïve bayes*, *maximum entropy*, dan *support vector machine*. *Supervised learning* bergantung pada data pelatihan. Model klasifikasi berdasarkan data latih yang telah diberi label dalam satu *domain*, sering berkinerjanya buruk dengan *domain* yang berbeda. Meskipun adaptasi *domain* telah dipelajari oleh para peneliti, namun teknologi ini masih jauh dari sempurna.

#### b. Pendekatan *Lexicon-based*

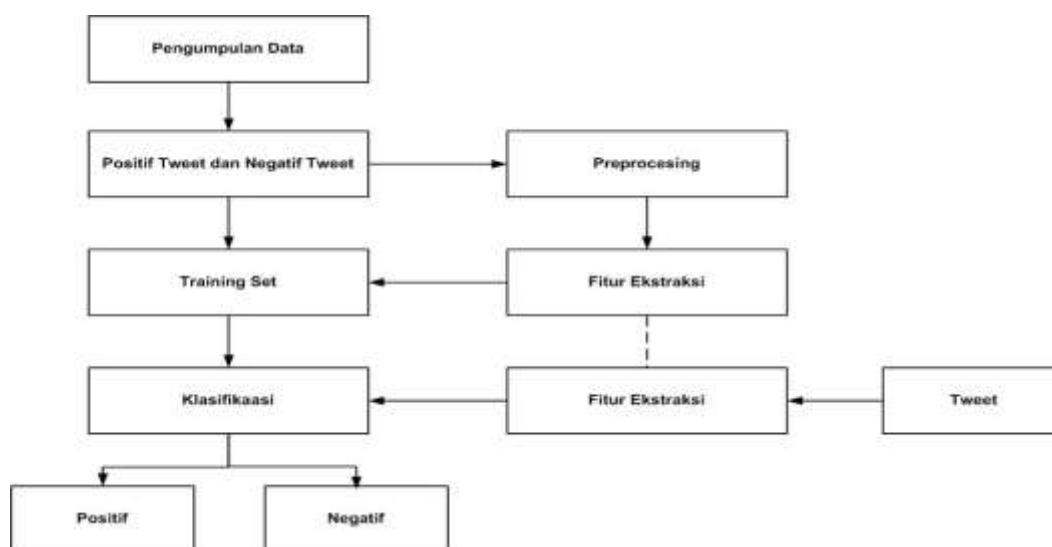
Pendekatan *lexicon* merupakan pendekatan dengan menggunakan kamus sentimen yang berisi kata-kata opini dan membandingkannya dengan data untuk mengidentifikasi nilai suatu kata. Pada kamus *lexicon*, kata-kata dibandingkan dengan kata polaritasnya. Untuk melakukan pendekatan *lexicon* yang pertama dilakukan adalah dengan menentukan kata-kata yang akan dianalisis dari *corpus*. Untuk melakukan pemilihan kata tersebut dapat menggunakan *part-of-speech* yang dapat mencari kata-kata dengan tipe tertentu yang dapat diambil, seperti *noun*, *adjective*, dan *adverb* [6].

### 2.3 Metode Klasifikasi pada Sentimen Analisis

Dalam pengklasifikasian pada sentimen analisis metode [7] *Naive Bayes Classifier* membuat asumsi yang sangat kuat (naif) akan independensi dari masing-masing kelas kejadian yang diberikan label. *Naive Bayes Classifier* digunakan untuk klasifikasi sentimen dari data yang telah diperoleh. Metode klasifikasi selanjutnya yaitu *Support Vector Machine (SVM)* menggunakan *machine learning (supervised learning)* yang memprediksi kelas berdasarkan model atau pola dari hasil *proses training*. Nilai atau pola yang dihasilkan dari Metode *Support Vector Machine* sebenarnya adalah sebuah garis pemisah yang disebut dengan *Hyperplane* [8]. Metode yang juga sering digunakan pada klasifikasi sentimen analisis adalah [8] *maximum entropy*, berbasiskan probabilitas yang termasuk dalam kelas model eksponensial. Prinsip dari *maximum entropy* mencari distribusi  $p(a|b)$  yang akan memberikan nilai *entropy* maksimum.

## 3. ANALISA DAN PEMBAHASAN

Sentimen analisis *text* pada dasarnya melalui beberapa tahap sebelum mendapatkan *output* hasil yang diinginkan. *Input* suatu *text* diubah menjadi beberapa *vector* dari suatu fitur ekstraksi yang kemudian diproses lebih lanjut untuk dijadikan *training* dataset. Secara umum proses sentimen analisis ini terdiri dari beberapa tahap diantaranya pengumpulan dataset, pelebelan data secara manual, pembersihan dataset (*preprocessing*), ekstraksi fitur, klasifikasi, dan evaluasi hasil *output*. Berikut diagram alir proses sentimen analisis secara umum.



Gambar 1. Proses Sentimen Analisis Secara Umum

Pada proses sentimen analisis ini terbagi menjadi 5 tahapan dalam menentukan sentimen, yaitu:

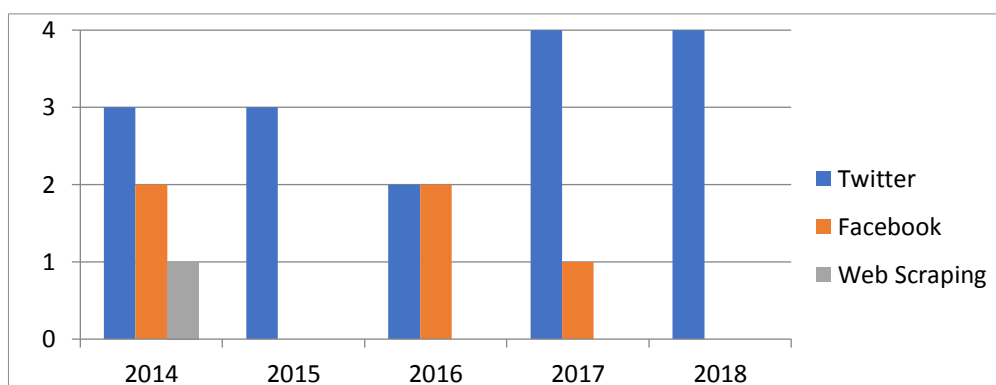
a. Tahap Pengumpulan Data

Pada tahapan ini proses pengambilan sentimen teks dilakukan dengan berbagai macam cara dan berbagai macam sumber seperti facebook, twitter, website, dan *movie review*. Berikut beberapa penelitian terdahulu yang menggunakan sentimen analisis ditunjukkan pada tabel 1 dan gambar grafik 1.

Tabel 1. Dataset yang digunakan peneliti sebelumnya

Penulis	Tahun	Dataset
[9] Alvaro Ortigosa, José M. Martín, Rosa M. Carro Department	2014	Facebook (Opini umum)
[10] Ahmad Fathan Hidayatullah, Azhari SN	2014	Twitter (Tokoh Politik)
[11] N. Azmina M. Zamani, siti z. Z. Abidin, nasiroh omar, m. Z. Z.abiden	2014	Facebook (Opini Umum)
[12] Noviah Dwi Putranti, Edi Winarko	2014	Twitter (Opini publik)
[13] Georgios S. Solakidis, Konstantinos N. Vavliakis and Pericles A. Mitkas	2014	web Scraping
[14] Geetika Gautam, Divakar yadav	2014	Twitter (Opini Pelanggan)
[15] Waghode Poonam B, Prof. Mayura Kinikar	2015	Twitter (Opini Publik)
[2] M. Fachrurrozi, M.T., Novi Yusliani, M.T	2015	Twitter (Opini Umum)
[16] Akshi Kumar, Prakhar Dogra, Vikrant Dabas	2015	Twitter (Opini Umum)
[17] Mondher Bouazizi, Tomoaki Ohtsuki	2016	Twitter (Opini Pelanggan)
[18] Antonius Rachmat C, Yuan Lukito	2016	Facebook (Opini Politik)
[19] A. Ortigosa, J. M. Martín, and R. M. Carro	2016	Twitter (Kebencian)

Penulis	Tahun	Dataset
[7] Syahmia Gusriani, Kartina Diah Kusuma Wardhani, Muhammad Ihsan Zu	2016	Facebook (Layanan publik)
[20] Mohd Suhairi Md Suhaimin, Mohd Hanafi Ahmad Hijazi, Rayner Alfred, Frans Coenen	2017	Facebook (Sarkasme)
[21] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. S. Harish	2017	Twitter (Sarkasme)
[22]Prananda Antinasari, Rizal Setya Perdana, M. Ali Fauzi	2017	Twitter (Komentar film)
[23] Agnes Rossi Trisna Lestari1, Rizal Setya Perdana, M. Ali Fauzi	2017	Twitter (Pilkada)
[24] Avinash Chandra Pandey, Dharmveer Singh Rajpoot, Mukesh Saraswa	2017	Twitter (Opini Umum)
[5] Sigit Suryono, Ema Utami, Emha Taufiq Luthfi	2018	Twitter (Opini Umum)
[25] Hajime Watanabe, Mondher Bouazizi, and Tomoaki Ohtsuki	2018	Twitter (Hatespeech)
[4] K.Kaviya and D. M. S. , K.K.Shanthini	2018	Twitter (Opini Publik)
[26] Haga Simada Ginting, Kemas Muslim Lhaksana, Danang Triantoro Murdiansyah	2018	Twitter (Opini Pilkada)



Gambar 2 Grafik Dataset yang digunakan peneliti sebelumnya

Dari tabel dan gambar diatas dapat disimpulkan bahwa sebagian besar peneliti menggunakan dataset dari Twitter. Twitter merupakan *microbloging* yang paling sederhana. Twitter memiliki batasan yaitu 140 kata untuk satu postingan. Selain itu, Twitter memiliki fasilitas twitter API yang memudahkan pengguna dan peneliti agar dapat mengumpulkan informasi-informasi yang ada pada twitter.

b. Tahap *Pre-processing*

*Pre-processing* merupakan teknik *data mining* yang melibatkan perubahan data mentah menjadi sebuah format yang terstruktur dan dimengerti. Data mentah seringkali tidak lengkap, tidak konsisten, dan mungkin mengandung banyak kesalahan. Teknik *pre-processing* terbukti dapat menyelesaikan masalah tersebut. Ada beberapa tahapan dalam teknik *pre-processing* seperti *tokenization*, *case folding*, *filtering* dan *stemming*. [18] Dalam studi khusus penelitian status dan komentar terhadap Facebook *Page* calon presiden Republik Indonesia pada Pemilu tahun 2014 tahapan *preprocessing* menggunakan tahap tokenisasi, *stemming* dan pembobotan token. [1] Menggunakan Proses *Case folding* untuk merubah dokumen menjadi huruf kecil dan menggunakan *tokenization* sebagai pemisah rangkaian kata. [21] Dalam *pre-processing* tahapan *filtering* menggunakan pemetaan emoji dan kamus *slang* untuk meningkatkan akurasi.

c. Tahap Fitur Ekstraksi

Setelah mendapatkan dataset baru tahapan selanjutnya yaitu tahap fitur ekstraksi. Tahapan ini digunakan untuk mendapatkan fitur yang ada dalam suatu teks sehingga akan digunakan sebagai klasifikasi sentimen. Pada tahapan ini yang pertama dilakukan adalah melakukan pengambilan ciri suatu objek yang dapat menggambarkan objek tersebut. Beberapa peneliti sebelumnya menggunakan *unigram* sebagai salah satu yang paling populer dalam model fitur klasifikasi untuk sentimen analisis [27].selanjutnya *bigram*, dan *n-gram* juga digunakan dalam model fitur ekstraksi pada sentiment analisis [13] [14]. Dan beberapa menggunakan *lexical-based* sebagai fitur ekstraksi yang berfungsi untuk mendapatkan nilai sentimen suatu kata [9]. Selain itu, ada fitur tambahan yang sangat penting seperti yang dilakukan oleh [17] dengan menambahkan deteksi sarkasme dalam fitur sentimen analisis.

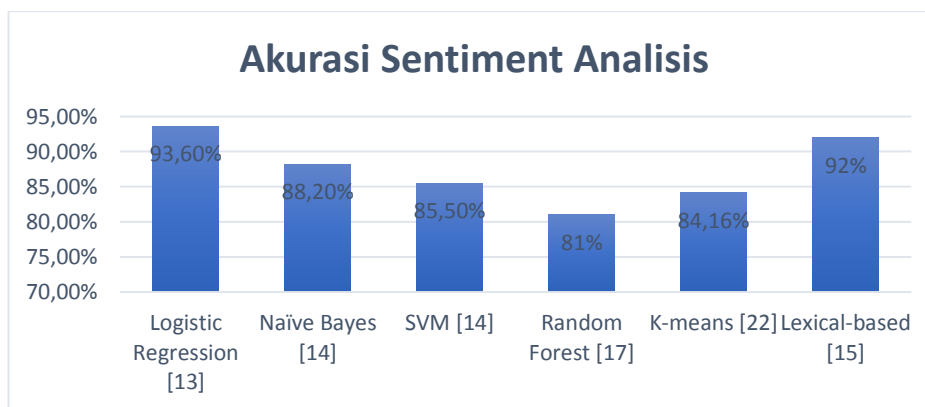
d. Tahap Klasifikasi

Setelah mendapatkan fitur untuk data uji dan latih, maka klasifikasi dilakukan agar dapat mendapatkan kelas pengenalan yang sesuai. Pada dasarnya metode klasifikasi pada sentimen analisis terbagi menjadi dua yaitu: pendekatan dengan menggunakan *machine learning* dan *lexical-based*. Pada pendekatan *machine learning* beberapa peneliti menggunakan *Support Vector Machine* [9] [13] [14] [28], *Naive bayes* [9] [13] [14] [28] [24], *Random Forest* [17], dan *logistic regression* [13]. Selain itu, peneliti [21] memanfaatkan metode *hybrid* baru yaitu *clustering k-means* yang digabung dengan metode *cuckoo search* dalam melakukan klasifikasi. Sedangkan pada penelitian [14] [15] [29] menggunakan *lexical-based* sebagai metode klasifikasi.

e. Tahap Evaluasi

Pada tahapan ini pengujian dari hasil klasifikasi dilakukan. Beberapa penelitian sebelumnya menggunakan matrik akurasi sebagai tahapan pengujian.

Berikut merupakan hasil dari tahapan pengujian dari beberapa peneliti pendahulu :



**Gambar 3.** Grafik Akurasi yang digunakan peneliti sebelumnya

Dari gambar di atas merupakan hasil terbaik dari beberapa metode yang diusulkan dari penelitian tahun 2014 hingga 2018. Hasil terbaik pada peneliti [13] dengan menggunakan metode *logistic regression*. Penelitian tersebut memanfaatkan deteksi *emoticon* sebagai suatu pelengkap fitur lain untuk membantu mendeteksi emosi penulis. Selain itu hasil yang baik juga di dapat dengan menggunakan pendekatan *lexicon-based* yang memanfaatkan kamus WordNet [15].

#### 4. KESIMPULAN

Kajian ini membahas tentang metode dan fitur ekstraksi yang memungkinkan untuk dapat melakukan sentimen analisis. Beberapa penelitian sebelumnya lebih banyak menggunakan Twitter untuk mengumpulkan dataset. Metode klasifikasi yang paling sering digunakan adalah *Naïve Bayes* dan *Support Vector Machine*. Namun, hasil yang terbaik ada pada metode *logistic regression* dan *lexical-based*. Hasil tersebut mungkin berdampak dari pemilihan fitur ekstraksi yang tepat dan penyusunan dataset yang baik.

#### REFERENCES

- [1] D. Rosdiansyah and S. Agustian, "Analisis Sentimen Twitter Menggunakan Metode K-Nearest Neighbor dan Pendekatan Lexicone," *Tek. Inform. UIN SUSKA Riau*, 2014.
- [2] M. Fachrurrozi and N. Yusliani, "ANALISIS SENTIMEN PENGGUNA JEJARING SOSIAL," no. June 2016, 2015.
- [3] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, 2015.
- [4] K.Kaviya and D. M. S. , K.K.Shanthini, "Micro-blogging Sentimental Analysis on Twitter Data Using Naïve Bayes Machine Learning Algorithm in Python Pre-processing Feature Selection Naïve Bayes classification Positive," no. April, pp. 46–51, 2018.
- [5] S. Suryono *et al.*, "Analisis Sentiment Pada Twitter Dengan Menggunakan," pp. 9–15, 2018.
- [6] V. A. Kharde and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 975–8887, 2016.
- [7] S. Gusriani, K. D. K. Wardhani, and M. I. Zul, "Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)," *4th Appl. Bus. Eng. Conf.*, vol. 1, no. 1, pp. 1–7, 2016.
- [8] "Analisis Sentimen Menggunakan Support Vector Machine dan Maximum Entropy Sentiment Analysis Using Support Vector Machine and Maximum Entropy Method," vol. 4, no. 2, pp. 2389–2395, 2017.
- [9] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Comput. Human Behav.*, vol. 31, no. 1, pp. 527–541, 2014.
- [10] A. Fathan Hidayatullah, A. Sn, J. Teknik, I. Fakultas, and T. Industri, "ANALISIS SENTIMEN DAN KLASIFIKASI KATEGORI TERHADAP TOKOH PUBLIK PADA TWITTER," *Semin. Nas. Inform.*, vol. 2014, no. semnasIF, pp. 115–122, 2014.
- [11] N. A. M. Zamani, S. Z. Z. Abidin, N. Omar, and M. Z. Z. Abiden, "Sentiment Analysis : Determining People ' s Emotions in Facebook 2 Related Work," *Appl. Comput. Sci.*, vol. ISBN: 978-, pp. 111–116, 2014.
- [12] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 8, no. 1, pp. 91–100, 2014.
- [13] G. S. Solakidis, K. N. Vavliakis, and P. A. Mitkas, "Multilingual sentiment analysis using emoticons and keywords," *Proc. - 2014 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. - Work. WI-IAT 2014*, vol. 2, pp. 299–305, 2014.
- [14] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," *2014 Seventh Int. Conf. Contemp. Comput.*, pp. 437–442, 2014.
- [15] B. W. Poonam and M. Kinikar, "Twitter Sentiment Analysis with Emoticons," *Ijecs.in*, vol. 4, no. 4, pp. 11315–11321, 2015.
- [16] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, "Opinion Mining and Sentiment Analysis on a Twitter Data Stream," *Int. Conf. Adv. ICT Emerg. Reg. (ICTer 2012)*, pp. 182–188, 2012.
- [17] M. Bouazizi and T. Ohtsuki, "Sentiment analysis in twitter: From classification to quantification of sentiments within tweets," *2016 IEEE Glob. Commun. Conf. GLOBECOM 2016 - Proc.*, 2016.
- [18] A. Rachmat and Y. Lukito, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016.
- [19] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," *Din. Inform.*, vol. 5, no. September, 2016.
- [20] M. S. M. Suhaimin, M. H. A. Hijazi, R. Alfred, and F. Coenen, "Natural language processing based features for sarcasm detection: An investigation using bilingual social media texts," *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, pp. 703–709, 2017.
- [21] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. S. Harish, "Sentiment analysis for sarcasm detection on streaming short text data," *2017 2nd Int. Conf. Knowl. Eng. Appl. ICKEA 2017*, vol. 2017–Janua, no. 2009, pp. 1–5, 2017.
- [22] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1733–1741, 2017.
- [23] A. R. T. Lestari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017.

- [24] A. Chandra Pandey, D. Singh Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," *Inf. Process. Manag.*, vol. 53, no. 4, pp. 764–779, 2017.
- [25] H. Watanabe, M. Bouazizi, and T. Ohtsuki, "Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection," *IEEE Access*, vol. 6, pp. 13825–13835, 2018.
- [26] H. S. Ginting, K. M. Lhaksana, and D. T. Murdiansyah, "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes," *e-Proceeding pf Eng.*, vol. 5, no. 1, pp. 1793–1802, 2018.
- [27] A. Z. Riyadh, N. Alvi, and K. H. Talukder, "Exploring Human Emotion Via Twitter," pp. 22–24, 2017.
- [28] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 11, no. 2, p. 187, 2017.