



Online Lecture Sentiment Analisis On Twitter Social Media During The Covid-19 Pandemic Using Sentistrength Algorithm

Analisis Sentiment Kuliah Daring di Media Sosial Twitter Selama Pandemi Covid-19 Menggunakan Algoritma *Sentistrength*

Reza Wahyu Hardian^{1*}, Pradita Eko Prasetyo², Ulfa Khaira³, Tri Suratno⁴

^{1,2,3,4}Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia

E-mail : ¹reza74204@gmail.com, ²pradita.eko@unja.ac.id, ³ulfa.ilkom@gmail.com, ⁴triel123@gmail.com

Received Dec 31th 2020; Revised Feb 21th 2021; Accepted Aug 20th 2021
Corresponding Author: Reza Wahyu Hardian

Abstract

Sentiment analysis is a form of activity used to analyze public opinion about an event such as online lectures during the Covid-19 pandemic through one of Twitter's social media. Online lectures are a learning system that is done by requiring the internet in the process of teaching and learning. While Twitter is a microblogging platform used to write an opinion or quiet opinion about an event. Sentistrength method is one of the methods that can be used to perform sentiment analysis on online lectures policies used to classify a tweet based on its emotional level. The results of the sentiment analysis process using sentistrength from public opinion in this case Twitter users against the online lecture policy resulted in a neutral score of percentage 54,3% and an accuracy rate of 85% and error rate of 15%.

Keyword : Classification, Online Lecture, Sentistrength Sentiment Analisis, Twitter

Abstrak

Analisis *sentiment* adalah suatu bentuk kegiatan yang digunakan untuk menganalisis opini masyarakat tentang suatu kejadian seperti kuliah daring selama pandemi Covid-19 melalui salah satu media sosial *Twitter*. Kuliah daring adalah suatu sistem pembelajaran yang dilakukan dengan membutuhkan internet dalam proses belajar- mengajar. Sedangkan *Twitter* adalah suatu platform microblogging yang digunakan untuk menulis suatu opini atau pendapat tentang suatu peristiwa. Metode *SentiStrenth* merupakan salah satu metode yang dapat digunakan untuk melakukan analisis *sentiment* terhadap kebijakan Kuliah Daring yang digunakan untuk mengklasifikasi suatu tweet berdasarkan tingkat emosinya. Hasil proses analisis *sentiment* menggunakan *sentistrength* berdasarkan opini masyarakat dalam hal ini pengguna *Twitter* terhadap kebijakan kuliah daring menghasilkan nilai netral dengan persentase 54%, dan tingkat akurasi 85%, serta tingkat kesalahan 15%.

Kata Kunci: Analisis Sentiment, Klasifikasi, Kuliah Daring, Sentistrength, Twitter

1. PENDAHULUAN

Di era pandemi saat ini mengakibatkan semua kegiatan manusia harus di transformasi ke dunia digital. Salah satunya dunia pendidikan, pada saat ini dunia sedang berupaya dalam mengatasi permasalahan yang ada agar semua kegiatan manusia dapat berjalan sesuai dengan tatanan masyarakat yang ada. Tidak terkecuali Indonesia, pada saat ini pemerintah dalam hal ini Kementerian Pendidikan dan Kebudayaan berupaya mengatasi masalah sistem pembelajaran di Indonesia di masa pandemic Covid-19 saat ini. Marqus & Hilgard menyatakan bahwa belajar adalah sistem pencarian jati diri agar terjadi perubahan ke arah yang lebih baik dengan cara pelatihan, pembelajaran, dan sebagainya [1].

Salah satu bentuk upaya dilakukan Kementerian Pendidikan dan Kebudayaan dalam mengatasi pembelajaran di Indonesia ialah dengan menerapkan sistem pembelajaran daring. Sistem Pembelajaran Daring adalah sistem belajar yang memanfaatkan jaringan internet dengan aksesibilitas, konektivitas, fleksibilitas dan kemampuan untuk memunculkan berbagai jenis interaksi pembelajaran [2]. Hal itu merupakan bentuk salah satu pemanfaatan kemajuan Teknologi Informasi yang berkembang sangat pesat saat ini.

Kemajuan teknologi informasi memberikan ruang lebih bagi masyarakat untuk berpendapat tentang semua hal yang terjadi, salah satunya media sosial *Twitter*. Media Sosial *Twitter* adalah salah satu situs

microblogging yang memungkinkan penggunaannya untuk menulis, membaca tentang fenomena terkini dan membahas isu – isu yang sedang terjadi . Pengguna *Twitter* terdiri dari berbagai kalangan, seperti pejabat, penggiat seni, hingga masyarakat biasa pada umumnya. Namun disini *Twitter* membatasi layanan kepada pengguna untuk mengirim atau membaca *tweets* yang telah dibagikan dengan membatasi karakter maksimal 140. Adanya layanan tersebut menyebabkan masyarakat lebih memilih untuk menyampaikan pendapatnya lewat *Twitter* daripada menyampaikan secara langsung. Dari pendapat – pendapat tersebut dapat dimanfaatkan untuk mencari sebuah informasi. Namun, dalam pemanfaatannya membutuhkan analisis dalam bidang ilmu *text mining* yang tepat sehingga informasi yang dihasilkan dapat membantu dalam pengambilan keputusan yang lebih tepat dan efektif.

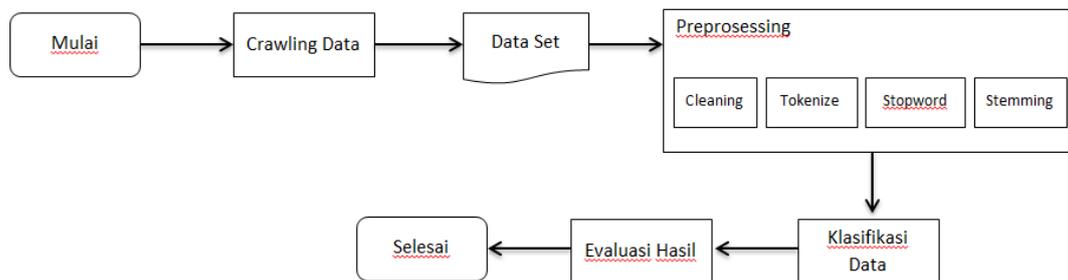
Salah satu bentuk penunjang pengambilan keputusan bagi pemerintah melalui pemanfaatan data teks *tweet* pada *Twitter* yaitu melalui analisis *sentiment* pada bidang ilmu *text mining*. *Text Mining* adalah bidang ilmu baru yang berkembang untuk mengekstrak informasi yang bermakna dari teks alami, itu dapat dicirikan sebagai proses analisis teks ke ekstrak informasi yang berguna untuk tujuan tertentu[3]. Analisis *sentiment* adalah riset komputasional dari opini, *sentiment* dan emosi yang diekspresikan secara tekstual [4]. Analisis *sentiment* sangat berperan penting dalam menyaring komentar – komentar yang terdapat di media sosial *Twitter*. Analisis *sentiment* pada komentar dilakukan untuk mengetahui tingkat pandangan masyarakat dalam hal ini pengguna *Twitter* akan kebijakan kuliah daring yang diterapkan selama pandemi.

Dari latar belakang tersebut, maka pada analisis ini peneliti ingin melihat bagaimana respons masyarakat terhadap kebijakan pemerintah mengenai kuliah daring dengan menggunakan metode atau algoritma *SentiStrength*. Dengan banyak *tweet* yang diberikan pengguna *Twitter* menjadi faktor pendukung bagi peneliti untuk mengetahui partisipasi masyarakat akan kuliah daring selama pandemic covid-19.

2. METODOLOGI PENELITIAN

2.1. Prosedur Penelitian

Pada proses analisis *sentiment* terkait partisipasi pengguna *Twitter* terhadap kuliah daring, memiliki beberapa tahapan. Pertama, *Crawling Data*, selanjutnya tahap praproses data klasifikasi *tweet*, dan evaluasi hasil klasifikasi, seperti Gambar 1 dibawah ini:



Gambar 1. Prosedur Penelitian

2.2. Crawling Data

Tahapan yang pertama kali dilakukan ialah crawling data (pengumpulan data). *Crawling* data adalah kegiatan pengumpulan data yang diambil dari suatu website atau platform media sosial dengan memanfaatkan API yang diberikan oleh website atau platform media sosial. *Crawling* data adalah proses mengidentifikasi dan mengambil data pada suatu halaman Web, biasanya digunakan untuk pengindeksan dan hal tertentu berdasarkan grafik link Web [5]. *Crawling* data digunakan untuk memperoleh dataset yang digunakan untuk proses analisis menggunakan API *Twitter*. Pada penelitian ini data yang di *crawling* dari *Twitter* sebanyak 854 tweet. Kata kunci yang digunakan pada proses pengambilan data yaitu kata atau kalimat yang mengacu tentang “Kuliah daring”. *Tools - tools* yang digunakan untuk menyokong penelitian ini ialah *plugin Twitter-python*, *csv*, *string*, *re*, *pandas*, *nlk*, *numpy*, *sastrawi*, *textblob*, dan *matplotlib*. Ditulis dengan menggunakan bahasa pemrograman python melalui IDE *Jupyter Notebook*.

2.3. Data Preprocessing

Preprocessing data adalah tahapan perbersihan data dari suatu kalimat atau kata, link, atau hal – hal yang tidak diperlukan untuk melakukan analisis *sentimentt*. Dataset yang diperoleh dari proses *crawling* masih berbentuk kata atau kalimat yang tidak beraturan atau tidak terstruktur, sehingga membutuhkan proses yang disebut *preprocessing* data agar menghasilkan data yang bersih supaya hasil analisis *sentimentt* menjadi akurat. Adapun tahapan *preprocessing* data terdiri dari *Cleaning* data, *Case Folding*, *Tokenizing*, *Stopword*, dan *Stemming*.

2.3.1. Proses *Cleaning*

Cleaning data adalah proses menghilangkan atau menghapus data dari suatu kalimat yang memiliki unsur, seperti username (@username), *hashtag* (#), URL, angka, kalimat yang *redundance*, *emoticon* (:@, :D), dan tanda baca yang tidak diperlukan pada saat proses analisis *sentiment*. Pada proses *cleaning* ini pun juga terjadi proses perubahan data ke *lowercase*.

2.3.2. Proses *Tokenizing*

Tokenizing adalah proses dimana memecahkan kalimat menjadi kata – perkata atau disebut dengan proses pemenggalan kata. *Tokenizing* adalah tahap pemotong *string input* berdasarkan kata – kata yang tersusun. Token yang dihasilkan berupa kata tunggal yang akan memudahkan dalam proses selanjutnya [6].

2.3.3. Proses *Stopword* atau *Stop Removal*

Stopword adalah proses menghilangkan atau menghapus kata – kata hasil *tokenizing* yang tidak memiliki makna seperti “ke”,”di”, “mau”,”itu” dan sebagainya. *Stopword* adalah daftar kata – kata yang tidak menampilkan isi dari suatu dokumen teks *stopword* dilakukan untuk menghapus kata atau *term* yang tidak memiliki arti[7].

2.3.4. Proses *Stemming*

Stemming adalah proses mengubah kata – kata yang berimbuhan menjadi kata dasar sesuai kaidah Bahasa Indonesia yang baku. *Stopword* adalah kata umum yang muncul dalam jumlah yang besar di suatu data dan dianggap tidak bermakna.

2.4. Klasifikasi Data

Klasifikasi merupakan suatu teknik pada text mining yang berguna untuk mengelompokkan suatu kalimat atau data berdasarkan kemiripan [8]. Klasifikasi pada analisis *sentiment* sering dimodelkan sebagai dua arah (positif/negatif) atau tiga arah (positif/negatif/netral) klasifikasi suatu data, meskipun beberapa metode menghasilkan klasifikasi yang lebih halus atau skor intensitas berkelanjutan [9]. Pada analisis sebelumnya ada yang menggunakan metode *textblob* dan ada yang menggunakan *Sentistrength*, namun pada analisis ini, klasifikasi data *tweet* dilakukan dengan mengaplikasikan algoritma *Sentistrength*. *SentiStrength* adalah algoritma yang digunakan untuk mengklasifikasi suatu index yang berfokus pada kekuatan *sentiment* di sebuah kalimat. Senti Strength akan menghasilkan nilai positif dan negatif, dimana jangkauan *value* dari angka 1 sampai 5. *Value* 1 menunjukkan *tweet* tersebut tidak memiliki sentimen positif maupun negatif, sedangkan *value* 5 menunjukkan *tweet* tersebut memiliki sentimen positif atau sangat negatif[10]. Keputusan akhir *sentiment* didasarkan pada aturan dibawah ini:

If positive value > negative value then Positive Sentiment
If positive value < negative value then Negative Sentiment
If value = negative value then Neutral Sentiment

2.5. Uji Akurasi *Tweet*

Pengujian hasil klasifikasi *sentiment* pada data *tweet* dilakukan dengan membandingkan klasifikasi prediksi dengan klasifikasi aktual. Klasifikasi prediksi merupakan hasil dari proses klasifikasi dengan menggunakan algoritma, sedangkan klasifikasi aktual, sedangkan klasifikasi aktual merupakan hasil klasifikasi yang dihitung secara manual. Algoritma sentimen *Sentistrength* bertujuan untuk mengukur kekuatan sentimen negatif dalam sebuah teks, dengan anggapan bahwa *sentiment* positif dan *sentiment* negatif bisa terdapat bersamaan dalam teks[11].

Untuk menghitung tingkat akurasi klasifikasi digunakan metode *confusion matrix*. *Confusion matrix* pada model $n \times n$ yang terkait dengan pengklasifikasian menghasilkan klasifikasi prediksi dan aktual, dimana n berada pada jumlah kelas yang berbeda [12]. Pada analisis ini peneliti menggunakan model *confusion matrix* 3×3 atau disebut dengan *extended confusion matrix*, yang ditunjukkan pada Tabel 1.

Table 1. Tabel *Confusion Matrix*

		Data Prediksi		
		Positif	Negatif	Netral
Data Aktual	Positif	AA	AB	AC
	Negatif	BA	BB	BC
	Netral	CA	CB	CC

Untuk menguji keakuratan klasifikasi menggunakan persamaan 1,

$$\text{Akurasi} = \frac{AA+BB+CC}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \quad (1)$$

$$\text{Error} = 1 - \text{Akurasi} \quad (2)$$

3. HASIL DAN PEMBAHASAN

3.1. Hasil Preprocessing

3.1.1. *Cleaning Data*

Proses *cleaning* ini bisa dikatakan suatu hal yang harus dilakukan dalam suatu kegiatan *text mining*, karena suatu data dapat menghasilkan yang akurat apabila data tersebut bersifat bersih dari hal yang mengganggu hasil dari *text mining*. Pada analisis ini peneliti melakukan proses *cleaning* data dengan hasil pada tabel 2.

Table 2. Tabel Hasil *Cleaning Data*

Sebelum	Sesudah
@skincareforfun Mau eye serum sama eye creamnyaðŸ˜-. Semenjak kuliah online matakmu jadi kayak mata pandaðŸ•¼. https://t.co/eZRmILXVGw	mau eye serum eye creamnya. semenjak kuliah online matakmu jadi kayak panda

3.1.2. *Tokenizing*

Setelah melakukan *cleaning* data peneliti melakukan proses *tokenizing* dengan tujuan untuk memecahkan kalimat pada data tweet menjadi perkata dengan hasil pada tabel 3.

Table 3. Tabel Hasil *Tokenizing*

Sebelum	Sesudah
mau eye serum eye creamnya. semenjak kuliah online matakmu jadi kayak panda	mau, eye, serum, eye, creamnya, semenjak, kuliah, online, matakmu, jadi, kayak, panda,

3.1.3. *Stopword* atau *Stop Removal*

Pada tahapan ini, peneliti menggunakan hasil *tokenizing* untuk menghilangkan kata yang tidak memiliki makna pada data hasil *tokenizing* dengan hasil pada tabel 4.

Table 4. Tabel Hasil *Stopword*

Sebelum	Sesudah
mau, eye, serum, eye, creamnya, semenjak, kuliah, online, matakmu, jadi, kayak, panda	eye, serum, eye, creamnya, semenjak, kuliah, online, matakmu, kayak, panda

3.1.4. *Stemming*

Setelah melakukan data diatas ternyata data masih belum bersih secara total, hal itu dibuktikan masih terdapat kata yang berimbuhan. Sehingga peneliti melakukan proses *stemming* untuk mengembalikan kata yang berimbuhan ke kata baku dengan hasil pada tabel 5.

Table 5. Tabel Hasil *Stemming*

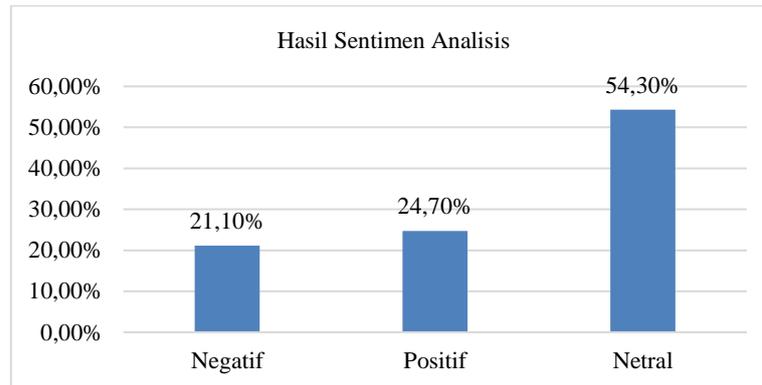
Sebelum	Sesudah
eye, serum, eye, creamnya, semenjak, kuliah, online, matakmu, kayak, panda	eye, serum, eye, cream, semenjak, kuliah, online, matakmu, kayak, panda

3.2. Hasil Klasifikasi Tweet

Setelah melakukan proses preprocessing diperoleh data yang bersih dan siap dilakukan proses analisis dengan jumlah data sebesar 304 data *tweet*. Pengklasifikasian data *tweet* dilakukan dengan menggunakan algoritma *SentiStrength* berbasis *leksikon*. *Sentistrength* akan menghasilkan *value* yang positif, negatif, dan netral. Keputusan dalam menentukan polarity dari teks dilihat dari tingkat emosi yang memiliki *value* terbesar.

Pada analisis ini data tweet yang digunakan dipresentasikan dalam bentuk bahasa Indonesia, sehingga membutuhkan sentimen *lexicon* yang di muat berbentuk berbahasa Indonesia seperti yang dilakukan oleh penelitian Ulfa Khaira et al[13], yang menggunakan kamus bahasa Indonesia yang telah disusun oleh Devid Haryalesmana dan Azhari SN. Sehingga, pada saat pengujian klasifikasi menjadi lebih tepat karena sesuai dengan kamus yang digunakan. Penilaian dilakukan dimana *value* tweet yang memiliki polarity *value* positif

lebih besar dari polarity value negatif maka tweet bernilai *positive*, begitu dengan tweet yang memiliki polarity value positif sama dengan polarity value negatif maka tweet bernilai *negative*, begitupun sebaliknya.



Gambar 2. Hasil Visualisasi Sentimen *Tweets*

Gambar 2 di atas merupakan hasil pengklasifikasian sentimen menggunakan *sentistrength* yang menghasilkan tweet dengan klasifikasi netral sebesar 54.3%, positif 24,7%, dan negatif 21,1%. Berdasarkan hasil presentasi tersebut dapat ditarik kesimpulan bahwa tanggapan masyarakat terhadap kebijakan kuliah daring bersifat netral. Menurut peneliti nilai netral diperoleh lebih banyak dikarenakan pengguna *Twitter* sering berpikir matang terlebih dahulu dan menggunakan pertimbangan logika dan emosional saat memberikan opininya tentang kuliah daring. Pernyataan ini didukung dari hasil penelitian Dodi dan In [14], mengatakan pengguna *Twitter* cenderung mempunyai emosional dan pikiran negatif namun cenderung bertindak secara positif sehingga klasifikasi *tweet* bernilai netral.

3.3. Pengujian Akurasi

Pada tahap pengujian akurasi dibutuhkan hasil klasifikasi *sentistrength* dan klasifikasi secara manual sebagai faktor pembandingan dalam pengujian akurasi hasil klasifikasi data menggunakan metode *sentistrength*. Pengujian ini dilakukan guna untuk mengetahui seberapa besar tingkat akurasi hasil klasifikasi dengan menggunakan *sentistrength* dibandingkan dilakukan secara manual, hal itu dikarenakan jika dilakukan klasifikasi secara manual dipengaruhi oleh keadaan emosional manusia yang berbeda dalam memberi ukuran sentimen dari suatu teks. Untuk mengetahui tingkat akurasi tersebut dilakukan perhitungan secara manual dengan menggunakan metode *Confusion Matrix*.

Table 6. Tabel *Confusion Matrix*

		Data Prediksi		
		Positif	Negatif	Netral
Data Aktual	Positif	63	12	10
	Negatif	0	53	1
	Netral	6	15	144

$$\text{Akurasi} = \frac{63+53+144}{304} = 0,85 = 85\%$$

$$\text{Error} = 1 - 0,85 = 0,15 = 15\%$$

Berdasarkan perhitungan diatas didapatkan keakurasian klasifikasi dengan algoritma *sentistrength* dengan klasifikasi secara manual sebesar 85% lebih besar dibandingkan dengan tingkat kesalahan (*error*) sebesar 15%. Tingkat keakurasian klasifikasi tersebut menunjukkan seberapa besar akurat sistem dalam mengklasifikasi data secara signifikan. Akurasi hasil penelitian menunjukkan perbedaan dengan penelitian sebelumnya dengan akurasi 60,3% [13]. Jadi, terbukti bahwa analisis *sentiment* menggunakan algoritma *sentistrength* menunjukkan hasil yang signifikan dibuktikan dengan tingkat keakurasiannya sebesar 85%.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan disimpulkan, Algoritma SentiStrength mampu melakukan analisis *sentiment* di *Twitter*. Dari data tweet pengguna *Twitter* sebanyak 304 diperoleh hasil analisis *sentiment* dengan sikap netral sebesar 54,3%, positif 24,7%, dan negatif 21,1% serta diperoleh hasil pengujian keakurasian sebesar 85% dan tingkat error sebesar 15%.

REFERENSI

- [1] N. R. Rahmawati, F. E. Rosida, and F. I. Kholidin, "Analisis Pembelajaran Daring Saat Pandemi Di Madrasah Ibtidaiyah," *J. Prim. Educ.*, vol. 1, no. 2, pp. 139–148, 2020.
- [2] J. L. Moore, C. Dickson-Deane, and K. Galyen, "E-Learning, online learning, and distance learning environments: Are they the same?," *Internet High. Educ.*, vol. 14, no. 2, pp. 129–135, 2011, [Online]. Available: <http://dx.doi.org/10.1016/j.iheduc.2010.10.001>.
- [3] L. Kumar and P. K. Bhatia, "TEXT MINING : CONCEPTS , PROCESS AND APPLICATIONS," vol. 4, no. 3, pp. 36–39, 2013.
- [4] 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.
- [5] P. Bailey, N. Craswell, and D. Hawking, "Engineering a multi-purpose test collection for Web retrieval experiments," *Inf. Process. Manag.*, vol. 39, no. 6, pp. 853–871, 2003.
- [6] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018.
- [7] L. Ardiani and H. Sujaini, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak Implementation of Sentiment Analysis of Community Responses to Development in Pontianak City," vol. 8, no. 2, pp. 44–51, 2020.
- [8] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," *Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, no. January 2010, pp. 841–842, 2010.
- [9] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, "The State-of-the-Art in Twitter Sentiment Analysis," *ACM Trans. Manag. Inf. Syst.*, vol. 9, no. 2, pp. 1–29, 2018.
- [10] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 10, no. 2, p. 207, 2016.
- [11] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment Strength Detection for the Social Web," *J. Am. Soc. Inf. Sci. Technol.*, vol. 23, no. 5, pp. 739–742, 2017.
- [12] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection," *CEUR Workshop Proc.*, vol. 710, pp. 120–127, 2011.
- [13] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, "Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength," *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 1, p. 21, 2020.
- [14] D. Setian and I. Seprina, "ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET LAZADA MENGGUNAKAN TEXT MINING DAN ALGORITMA NAIVE BINA DARMA Conference on Computer Science $P(c|X) = P(x|C)P(c)P(x)$," *Bina Darma Conf. Comput. Sci.*, pp. 998–1004, 2019.