



Analisis Sentimen Netizen Twitter Terhadap Pelayanan *Provider* Telkomsel: Komparasi Naive-Bayes dan K-Nearest Neighbors

Reza Ariftiarno¹, Monica Dwijayanti²

^{1,2}*Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260, DKI Jakarta, Indonesia*
*Email Penulis Koresponden: 2111601296@student.budiluhur.ac.id

Abstrak:

Penggunaan media sosial secara masif saat ini menghadirkan berbagai macam respon masyarakat terhadap suatu hal. Misalnya pada informasi-informasi terkait utang Badan Usaha Milik Negara (BUMN) di Indonesia. Salah satu BUMN yang terbesar adalah PT Telkom Indonesia (Persero) yang memiliki produk berupa provider jaringan bernama Telkomsel. Penelitian bertujuan untuk mengetahui bagaimana sentimen netizen terhadap provider Telkomsel di Twitter dan mengetahui algoritma dengan akurasi terbaik yang dapat digunakan untuk memprediksi sentimen netizen terkait provider Telkomsel. Hasil dari penelitian menunjukkan terdapat sentimen kurang puas sebesar 89,3%. Selain itu, algoritma K-NN memiliki akurasi terbaik dibandingkan algoritma Naïve Bayes dalam memprediksi sentimen netizen terhadap pelayanan *provider* Telkomsel.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license



Kata Kunci:

analisis sentimen,
naive bayes,
k-nearest neighbor

Riwayat Artikel:

Diserahkan 22 Agustus, 2023
Direvisi 13 Januari, 2024
Diterima 23 Januari, 2024

DOI:

10.22441/incomtech.v14i1.22597

1. PENDAHULUAN

Teknologi informasi yang semakin berkembang setiap waktu membuat manusia semakin mudah untuk berinteraksi satu sama lain secara cepat tanpa perlu berinteraksi secara langsung. Kebanyakan dari produk teknologi informasi tersebut dapat diakses secara gratis oleh masyarakat sehingga dapat menjangkau seluruh kalangan masyarakat, salah satu contohnya adalah media sosial. Dengan adanya media sosial membantu seseorang untuk memperoleh informasi secara langsung dari orang lain tanpa perlu menunggu waktu yang lama. Setiap orang juga bebas berekspresi di media sosial. Mereka dapat menuangkan emosi perasaannya pada media sosial, baik itu perasaan gembira, sedih, maupun marah. Salah satu platform dunia maya yang sering digunakan oleh publik untuk mengutarakan opini dengan bebas yaitu pada media sosial Twitter. Mengirim dan membaca tweet berbasis teks sampai 140 karakter bisa dilakukan pada media sosial Twitter [1].

Penggunaan media sosial secara masif saat ini menghadirkan berbagai macam respon masyarakat terhadap suatu hal. Misalnya pada informasi-informasi terkait utang Badan Usaha Milik Negara (BUMN) di Indonesia. Salah satu BUMN yang terbesar adalah PT Telkom Indonesia (Persero) yang memiliki produk berupa provider jaringan bernama Telkomsel. Dalam perkembangan teknologi saat ini jaringan internet sangat dibutuhkan setiap waktunya. Hal ini dapat memberikan gambaran bagaimana Telkomsel sebagai penyedia layanan jaringan terutama internet harus selalu dapat memberikan pelayanan terbaik agar kebutuhan akan teknologi internet para penggunanya tidak terganggu. Berbagai macam tanggapan masyarakat dilontarkan kepada Telkomsel di media sosial Twitter atas pelayanannya baik yang bersifat mendukung maupun kecewa. Respon-respon yang dikeluarkan tersebut dapat dikumpulkan menjadi satu yang disebut sebagai sentimen [2].

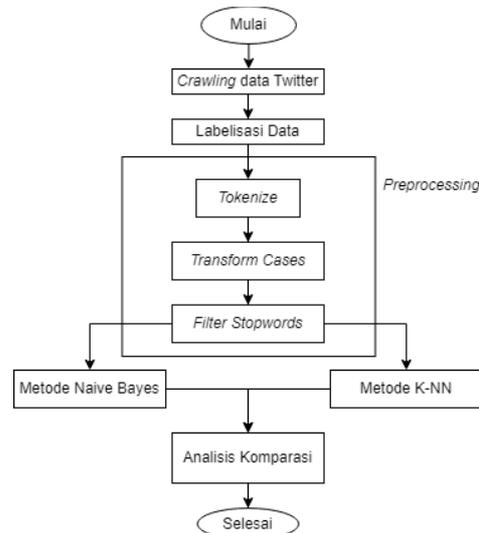
Analisis sentimen merupakan studi perbandingan antara suatu pandangan, sentimen, penilaian, sikap dan emosi orang terhadap objek dan aspeknya yang diekspresikan ke dalam teks [3]. Text mining merupakan metode yang dapat digunakan untuk melakukan analisis sentimen. Metode tersebut merupakan metode mengekstrak data teks sehingga diperoleh informasi yang berguna dari data teks tersebut. [4] dan [5] telah menerapkan metode teks mining untuk melakukan analisis sentimen. Dalam metode tersebut terdapat beberapa tahapan, seperti proses pengambilan data, preprocessing, dan pengujian dengan algoritma klasifikasi.

Penelitian [6] meneliti mengenai analisis sentimen user Twitter terhadap kinerja dan pelayanan perusahaan BUMN (PT PAL Indonesia) pada tahun 2022 dengan menggunakan algoritma K-NN, Naive Bayes, dan Decision Tree. Penelitian tersebut memperoleh hasil akurasi terbaik pada algoritma Naive Bayes dengan nilai akurasi sebesar 84,08%. Selain itu penelitian terkait analisis sentimen di Twitter kepada BUMN juga dilakukan oleh [7] dengan mengkomparasi lima algoritma klasifikasi Naive Bayes, SVM, Decision Tree, K-NN, dan Random Forest. Hasil dari penelitian tersebut K-NN memperoleh akurasi tertinggi sebesar 99,80%. Berdasarkan penelitian tersebut maka penelitian ini akan mencoba mengetahui sentimen netizen di media sosial Twitter terhadap pelayanan provider Telkomsel sebagai salah satu BUMN di Indonesia dan juga untuk mengetahui algoritma dengan akurasi terbaik di antara Naive Bayes dan K-NN untuk melakukan prediksi terhadap sentimen netizen terkait pelayanan provider Telkomsel sehingga tema pada penelitian ini adalah menganalisis sentimen netizen Twitter terhadap provider Telkomsel dengan menggunakan algoritma Naive Bayes dan K-Nearest Neighbor.

2. METODE

Penelitian ini memiliki tujuan untuk mengetahui bagaimana sentimen netizen terhadap provider Telkomsel di Twitter dan mengetahui algoritma dengan akurasi terbaik yang dapat digunakan untuk memprediksi sentimen netizen terkait provider Telkomsel. Penelitian ini menggunakan data tweet terkait provider Telkomsel yang diperoleh menggunakan Twitter API pada perangkat lunak Rapidminer pada bulan Juli tahun 2023. Tweet yang digunakan merupakan tweet yang menggunakan bahasa Indonesia. Kemudian tweet akan diberikan label sesuai dengan tanggapan netizen. Pelabelan terdiri dari label sentimen puas dan kurang puas. Setelah setiap tweet diberi label sesuai dengan sentimennya masing-masing maka perlu dilakukan

preprocessing dengan metode TF-IDF, yaitu dengan memberikan bobot pada setiap kata yang telah di tokenisasi [8]. Tahap selanjutnya adalah uji klasifikasi menggunakan algoritma Naive Bayes dan K-NN hingga tahap akhir pada penelitian ini melakukan analisis keakuratan kedua algoritma dan menentukan hasil akurasi terbaik dalam memprediksi sentimen netizen terhadap provider Telkomsel. Lebih lengkap tahapan penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Langkah Penelitian

Naive Bayes merupakan salah satu dari banyak jenis algoritma klasifikasi. Penggunaan Naive Bayes dalam prediksi sentimen bekerja dengan cara menghitung frekuensi setiap term dalam sebuah dokumen [9]. Secara umum teori Naive Bayes sesuai dengan teorema Bayes dapat digambarkan pada persamaan 1 berikut ini [10]

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \dots \dots \dots (1)$$

Keterangan:

$P(A)$ = Probabilitas kejadian A

$P(B)$ = Probabilitas kejadian B

$P(A|B)$ = Probabilitas A diberikan B

$P(B|A)$ = Probabilitas B diberikan A

K-Nearest Neighbor (K-NN) digunakan sebagai pembanding terhadap algoritma Naive Bayes sehingga dapat ditentukan algoritma terbaik dalam melakukan prediksi sentimen terhadap provider Telkomsel. K-NN merupakan algoritma yang memiliki fungsi untuk klasifikasi data berdasarkan kedekatan jarak pada objek yang diprediksi. K pada algoritma ini mendefinisikan mengenai jumlah tetangga yang diperiksa untuk mengklasifikasi titik pada query tertentu. Pemilihan tinggi rendahnya nilai K yang digunakan bergantung pada kondisi data, jika noise pada data cenderung tinggi maka pemilihan K yang tinggi sangat disarankan. [11], [12]. Cara kerja K-NN dapat digambarkan pada persamaan 2 berikut ini dengan menggunakan konsep *Euclidean distance*. [13]

$$d(x, y) = \sqrt{\sum_{i=0}^k (x_{1i} - y_{1i})^2 + (x_{2i} - y_{2i})^2 + \dots \dots \dots} \quad (2)$$

Keterangan:

$d(x,y)$ = Euclidean/jarak antara variabel x dan y

x = data yang akan diklasifikasikan

y = data sekitar

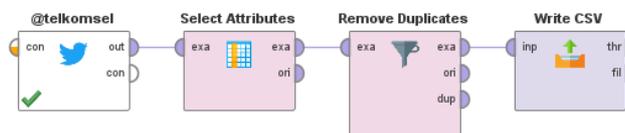
i = jumlah fitur

k= jumlah tetangga

3. HASIL DAN PEMBAHASAN

3.1 *Crawling data Twitter*

Pengambilan data atau crawling data yang dilakukan menggunakan software Rapidminer. Data yang diperoleh dari hasil crawling didapatkan sebanyak 500 tweet pada bulan Juli 2023. Proses crawling dilakukan dengan menggunakan dua operator search twitter untuk mengakomodir dua kata query pencarian, yaitu kata @telkomsel sebagai official username provider Telkomsel di Twitter seperti terlihat pada gambar 2. Selanjutnya dilakukan penyaringan terhadap tweet yang memiliki duplikat menggunakan operator remove duplicate sehingga berhasil diperoleh 439 tweet. proses akhir dari tahap crawling, yaitu menyimpan file ke dalam format csv. Pada tabel 1 menunjukan hasil crawling data yang diambil dari Twitter menggunakan software Rapidminer.



Gambar 2. Proses Crawling Data

Tabel 1. Hasil *Crawling Data* Twitter

<i>Text</i>
@Telkomsel Mana ini masih belum nambah saldo pulsa maupun uang saya. Tolonglah. Seenggaknya dijelasin kek ini kenapa bisa kayak gini, inimah cuma disuruh sabar nunggu doang tp ga ada penjelasan apa2!
@Telkomsel dh betmut gk ush ???
...
@Telkomsel tolong gabisa login di app, sdh versi terbaru & sdh coba re-install lagi https://t.co/qeKqCgJfPb

3.2 **Labelisasi Data**

Pelabelan data dilakukan untuk mengetahui jumlah masing-masing kategori sentimen yang diperoleh dari hasil crawling data sehingga dapat diketahui sentimen netizen terhadap provider Telkomsel di media sosial Twitter. Tahap ini menggunakan dua kategori sentimen, yaitu puas dan kurang puas. Seperti terlihat pada tabel 2 sebanyak 445 tweet pada dataset diberi label sesuai dengan jenis kalimatnya masing-masing. Dari hasil pelabelan manual diperoleh tweet yang

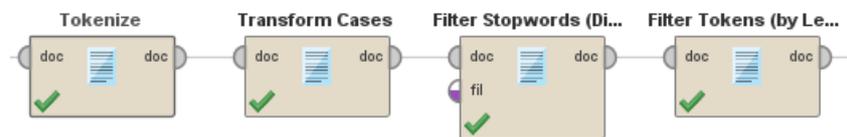
memiliki sentimen puas sebanyak 10,7%, dan sentimen kurang puas sebanyak 89,3%.

Tabel 2. Labelisasi Manual

<i>Text</i>	<i>Label</i>
@Telkomsel Mana ini masih belum nambah saldo pulsa maupun uang saya. Tolonglah. Seenggaknya dijelasin kek ini kenapa bisa kayak gini, inimah cuma disuruh sabar nunggu doang tp ga ada penjelasan apa2!	KURA NG PUAS
@Telkomsel dh betmut gk ush ???	KURA NG PUAS
...
@Telkomsel tolong gabisa login di app, sdh versi terbaru & sdh coba re-install lagi https://t.co/qeKqCgJfPb	KURA NG PUAS

3.3 Preprocessing Data

Tahap preprocessing data merupakan sebuah tahap pembersihan data tweet dari noise dan ketidak konsistenan sebelum dilakukan pengujian menggunakan algoritma [14]. Contoh preprocessing data adalah mengubah seluruh huruf pada tweet menjadi huruf kecil dan menghilangkan karakter yang tidak perlu. Pada penelitian ini tahap preprocessing yang digunakan terdiri dari proses tokenisasi, transform cases, filter stopwords, dan filterisasi token. Seluruh proses preprocessing dilakukan menggunakan software Rapidminer seperti terlihat pada gambar 3.



Gambar 3. Preprocessing Pada Rapidminer

Tabel 3 di bawah merupakan hasil konversi teks dari proses preprocessing data. Dapat dibandingkan dengan tabel 2 terlihat pada tabel 3 beberapa karakter sudah dihilangkan seperti @,?! dan lain sebagainya. Selain itu dengan menggunakan metode TF-IDF maka setiap kata tersebut akan dilakukan proses tokenisasi dan diberikan bobot pada setiap kata seperti terlihat pada tabel 4. Terdapat 1142 kata yang dapat di tokenisasi dari keseluruhan tweet.

Tabel 3. Hasil Preprocessing Data Twitter

<i>Text</i>
telkomsel nambah saldo pulsa uang tolonglah seenggaknya dijelasin kayak gini inimah disuruh sabar nunggu doang penjelasan
telkomsel betmut
...
telkomsel tolong gabisa login versi terbaru coba install https://t.co/qeKqCgJfPb

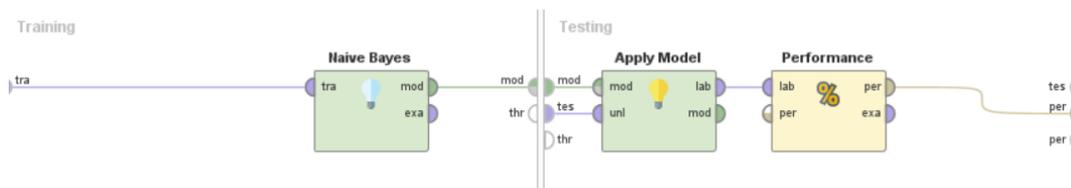
Tabel 4. Hasil Metode TF-IDF

<i>Text</i>	<i>abis</i>	<i>abissss</i>	<i>acaranya</i>	<i>... zphgnvaj</i>	<i>zrypbjnz</i>
-------------	-------------	----------------	-----------------	---------------------	-----------------

telkomsel nambah saldo pulsa uang tolonglah seenggaknya dijelasin kayak gini inimah disuruh sabar nunggu doang penjelasan	0.0	0.0	0.0	...	0.0	0.0
telkomsel betmut	0.0	0.0	0.0	...	0.0	0.0
...
telkomsel tolong gabisa login versi terbaru coba install https qekqcgjfpb	0.0	0.0	0.0	...	0.0	0.0

3.3 Metode Naive Bayes

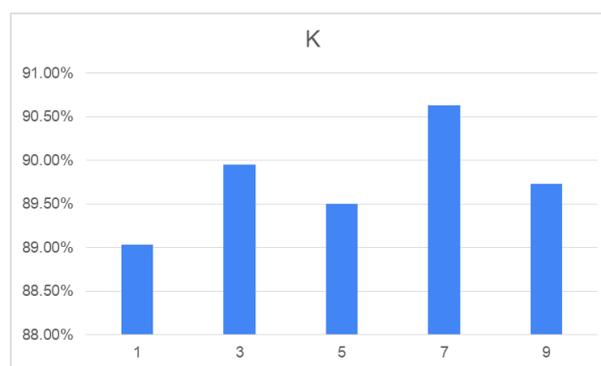
Metode Naive Bayes tersusun oleh operator algoritma Naive Bayes, Apply Model, dan Performance seperti terlihat pada gambar 4. Pengujian prediksi dilakukan menggunakan operator cross-validation yang terdiri dari bagian training dan testing. Operator Naive Bayes diletakan pada bagian training agar algoritma dapat mempelajari data tweet yang telah dilakukan tahap preprocessing. Selanjutnya, pada tahap testing algoritma yang telah dilatih akan diujikan menggunakan operator apply model.



Gambar 4. Penggunaan Algoritma Naive Bayes Pada Operator Cross Validation

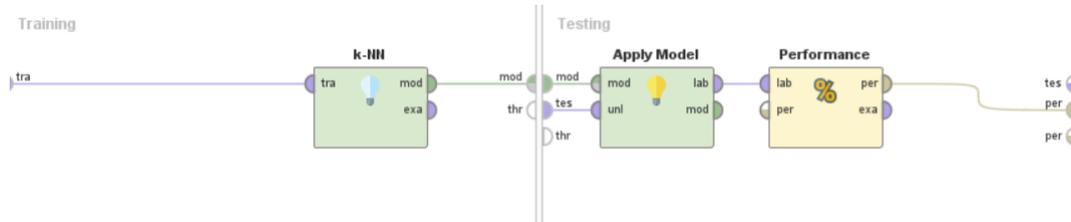
3.4 Metode K-Nearest Neighbor

Model prediksi metode K-Nearest Neighbor tersusun oleh operator algoritma K-NN, Apply Model, dan Performance seperti terlihat pada gambar 6. Operator K-NN diletakan pada bagian training di dalam operator cross-validation. Nilai K yang digunakan mengacu pada [8], [15], dan [16], yaitu nilai K berupa bilangan ganjil dan menghasilkan nilai akurasi tertinggi. Pada penelitian ini telah disediakan nilai k dari 1 sampai 9 dan berdasarkan gambar 5 terlihat bahwa $K = 7$ menghasilkan nilai akurasi paling tinggi sebesar 90.63% yang berarti nilai *confidence* sebelum dan setelah proses prediksi tidak banyak memiliki perbedaan. Selain itu, proses prediksi pada setiap data *testing* juga memiliki jarak *euclidean* yang lebih optimal untuk melakukan prediksi pada saat penggunaan $K=7$ dibandingkan dengan nilai K yang lainnya.



Gambar 5. Perbandingan Nilai Akurasi Pada Setiap Nilai K

Pada bagian training algoritma akan dilatih dengan menggunakan data dari hasil preprocessing. Kemudian hasil pengolahan operator K-NN akan masuk ke dalam bagian testing dari operator cross-validation. Pada tahap testing algoritma K-NN akan diuji untuk melakukan prediksi terhadap sentimen tweet.



Gambar 6. Penggunaan Algoritma K-NN Pada Operator Cross Validation

3.5 Analisa Komparasi

Tabel 5 merupakan penggambaran confusion matrix dari hasil pengujian model menggunakan Naive Bayes. Nilai akurasi prediksi dapat ditentukan dengan menghitung nilai yang ada pada tabel confusion matrix. Pengujian model prediksi menggunakan algoritma naive bayes menghasilkan nilai akurasi sebesar 77.86%.

Tabel 5. Hasil Pengujian Model dengan Naive Bayes

	<i>true KURANG PUAS</i>	<i>true PUAS</i>	<i>class precision</i>
pred KURANG PUAS	314	20	94.01%
pred. PUAS	77	27	25.96%
class recall	80.31%	57.45%	

Pengujian model prediksi menggunakan algoritma k-nn menghasilkan nilai akurasi sebesar 90.63% dengan nilai K = 7. Confusion matrix dari hasil pengujian algoritma K-NN dapat dilihat pada tabel 6 berikut.

Tabel 6. Hasil Pengujian Model dengan K-NN

	<i>true KURANG PUAS</i>	<i>true PUAS</i>	<i>class precision</i>
pred KURANG PUAS	388	38	91.08%
pred. PUAS	3	9	75.00%
class recall	99.23%	19.15%	

4. KESIMPULAN

Hasil dari komparasi akurasi yang dihasilkan oleh model prediksi Naive Bayes dan K-NN menghasilkan nilai masing-masing di atas 50%. Dari kedua model prediksi tersebut K-NN memiliki nilai akurasi yang lebih tinggi dibandingkan

dengan Naive Bayes, yaitu sebesar 90.63%. Naïve bayes memiliki nilai akurasi yang lebih kecil dibandingkan dengan K-NN karena naïve bayes memiliki kekurangan apabila algoritma tersebut melakukan *training* data terdapat kemungkinan ada kategori yang tidak masuk ke dalam tahap *training* sehingga saat pada saat tahap *testing* muncul kategori tersebut akan diubah nilainya menjadi 0 dan hal tersebut dapat mengurangi tingkat akurasi. Hal ini menunjukkan bahwa K-NN memiliki performa yang lebih baik dibandingkan Naive Bayes dalam memprediksi sentimen netizen di media sosial Twitter terkait dengan pelayanan provider Telkomsel. Selain itu dari 470 tweet yang diperoleh pada bulan Juli 2023 tanggapan netizen terhadap pelayanan provider Telkomsel di media sosial twitter cenderung memiliki sentimen kurang puas, yaitu sebesar 89,3%.

REFERENSI

- [1] F. Felicia and R. Loisa, "Peran Buzzer Politik dalam Aktivitas Kampanye di Media Sosial Twitter," *Koneksi*, vol. 2, no. 2, p. 352, May 2019, doi: 10.24912/kn.v2i2.3906.
- [2] A. R. Prananda and I. Thalib, "Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion," *Journal of Information Systems Engineering and Business Intelligence*, vol. 6, no. 1, p. 1, Apr. 2020, doi: 10.20473/jisebi.6.1.1-8.
- [3] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham: Springer International Publishing, 2012. doi: 10.1007/978-3-031-02145-9.
- [4] Oman Somantri and Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 2, 2019.
- [5] Anni Karimatul Fauziyyah and Deden Hardan Gautama, "Analisis Sentimen Pandemi Covid-19 Pada Streaming Twitter Dengan Text Mining Python," *Jurnal Ilmiah Sinus (JIS)*, vol. 18, no. 2, 2020.
- [6] F. S. Pattihha and H. Hendry, "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 506, Apr. 2022, doi: 10.30865/jurikom.v9i2.4016.
- [7] W. W. Cholil, F. Panjaitan, F. Ferdiansyah, A. Arista, R. Astriratma, and T. Rahayu, "Comparison of Machine Learning Methods in Sentiment Analysis PeduliLindungi Applications," in *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Nov. 2022, pp. 276–280. doi: 10.1109/ICIMCIS56303.2022.10017669.
- [8] Muhammad Rangga Aziz Nasution and Mardhiya Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *JURNAL INFORMATIKA*, vol. 6, no. 2, 2019.
- [9] Shima Fanissa, M. Ali Fauzi, and Sigit Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 8, 2018.
- [10] Fransiska Vina Sari and Arief Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [11] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.

- [12] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 2, p. 121, Apr. 2021, doi: 10.22146/ijccs.65176.
- [13] T. A.M and A. Yaqin, "Perbandingan Algoritma Naïve Bayes, K-Nearest Neighbors dan Random Forest untuk Klasifikasi Sentimen Terhadap BPJS Kesehatan pada Media Twitter," *InComTech : Jurnal Telekomunikasi dan Komputer*, vol. 12, no. 1, p. 01, Apr. 2022, doi: 10.22441/incomtech.v12i1.13642.
- [14] S. A. Dainamang, N. Hayatin, and D. R. Chandranegara, "Analisis Sentimen Media Sosial Twiiter terhadap RUU Omnibus Law dengan Metode Naive Bayes dan Particle Swarm Optimization," *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 2, pp. 211–218, Aug. 2022, doi: 10.34010/komputika.v11i2.6037.
- [15] K. Hulliyah, A. M. Almaisah, F. Mintarsih, S. U. Masrurah, D. Khairani, and S. Aripriyanto, "Analysis of Public Sentiment Using The K-Nearest Neighbor (k-NN) Algorithm and Lexicon Based on Indonesian Television Shows on Social Media Twitter," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, IEEE, Sep. 2022, pp. 1–5. doi: 10.1109/CITSM56380.2022.9936011.
- [16] N. Shahbazi, S. Bagheri, and G. B. Gharehpetian, "Identification and classification of cross-country faults in transformers using K-NN and tree-based classifiers," *Electric Power Systems Research*, vol. 204, p. 107690, Mar. 2022, doi: 10.1016/j.epsr.2021.107690.