



## Analisis Sentimen *Complain* dan Bukan *Complain* pada Twitter Telkomsel dengan *SMOTE* dan *Naïve Bayes*

Budi Kurniawan Rachmat <sup>1\*</sup>, Achmad Suwarisman <sup>2</sup>, Iis Afriyanti <sup>3</sup>, Aditya Wahyudi <sup>4</sup>, Dedi Dwi Saputra <sup>5</sup>

<sup>1,2,3,4,5</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusamandiri.

### article info

#### Article history:

Received 28 June 2022

Received in revised form

14 October 2022

Accepted 16 November 2022

Available online January 2023

#### DOI:

<https://doi.org/10.35870/jti.k.v7i1.691>

#### Keywords:

Public Sentiment; Naïve Bayes; SMOTE; Twitter; Text Mining.

#### Kata Kunci:

Sentimen Publik; Naïve Bayes; SMOTE; Twitter; Text Mining.

### abstract

This analysis aims to find out the public sentiment towards Telkomsel posted on Indonesian twitter, which makes market research on public opinion very useful. The dataset was taken from Twitter social media in a query Indonesian by crawling method using the RapidMiner application and the result of crawling the data set there were 1000 tweets with sentiment complaints and not complaints. Therefore, from 1000 tweets, preprocessing will be carried out with the SMOTE Upsampling and Naivebayes methods as well as several filtering such as transform case, tokenize, tokenize (by length) stemming filters and stopwords so that the data can stay in words and there is a balance in the sentiment on the dataset. It can be concluded that in the classification of sentiment there is a balance between complaints and non-complaints as many as 581. Where the accuracy rating level is 81.58%, the precision assessment is 86.82% and the recall assessment is 74.87 and the resulting AUC is 0.803.

### abstrak

Analisis ini bertujuan untuk mengetahui sentimen publik terhadap Telkomsel yang diposting di twitter Indonesia, yang menjadikan riset pasar tentang opini publik sangat berguna. Dataset diambil dari media sosial Twitter dalam query bahasa Indonesia dengan metode crawling menggunakan aplikasi RapidMiner dan hasil dari crawling data set tersebut terdapat 1000 tweet dengan sentiment komplain dan bukan komplain. Maka dari 1000 tweet akan dilakukan preprocessing dengan metode SMOTE Upsampling dan Naivebayes serta beberapa filtering seperti transform case, tokenize, filter tokenize (by length) stemming dan stopword sehingga data yang di dapat tinggal kata-kata dan terjadi keseimbangan pada sentimen pada dataset. Dapat disimpulkan bahwa pada klasifikasi sentimen terdapat keseimbangan antara complain dan bukan komplain sebanyak 581. Dimana tingkat accuracy penilaian 81,58%, penilaian precision 86,82% dan penilaian recall sebesar 74,87 serta AUC yang dihasilkan sebesar 0.803.

\*Corresponding author. Email: 11213080@nusamandiri.ac.id <sup>1</sup>.

## 1. Latar Belakang

Analisis sentimen merupakan salah satu cara untuk mengumpulkan pendapat orang banyak terhadap sesuatu, contohnya seperti layanan *public*, isu atau permasalahan tertentu, kinerja pemerintahan atau hal lain yang berkaitan. Pada dasarnya, analisis sentimen digunakan untuk melakukan evaluasi terhadap layanan yang telah diberikan oleh suatu perusahaan atau kelompok tertentu. Analisis ini dapat dilakukan melalui berbagai cara, salah satunya adalah dengan mengumpulkan atau mengambil pendapat orang banyak melalui laman media sosial [1]. Mengumpulkan opini dan sentimen dari bahasa alami memberikan tantangan tersendiri karena memerlukan tingkat pemahaman mendalam terkait aturan bahasa yang eksplisit dan implisit, reguler dan tidak teratur, serta sintaksis dan semantik. Berbagai penelitian terkait analisis sentimen berusaha menghadapi masalah NLP yang belum terselesaikan, seperti resolusi koreferensi, penanganan negasi, resolusi anafora, pengenalan entitas bernama, dan disambiguasi kata-kata yang terdapat dalam data yang dikumpulkan [2]. Analisis sentimen dan penggalian opini merupakan salah satu bidang studi yang menganalisis pendapat, sentimen, evaluasi, sikap, dan emosi orang dari bahasa tertulis. Hal ini merupakan salah satu area penelitian paling aktif dalam pemrosesan bahasa alami dan dipelajari secara meluas dalam ruang lingkup penambangan teks, penambangan Web, dan penambangan data. Semakin meningkatnya penelitian terkait analisis sentimen berbanding lurus dengan perkembangan media sosial seperti ulasan, diskusi forum, blog, mikro-blog, Twitter, dan jejaring sosial lainnya. Untuk pertama kalinya dalam sejarah manusia, kita sekarang memiliki sejumlah besar data opini yang direkam dalam bentuk digital untuk dianalisis [3].

Pada era digital saat ini, media sosial merupakan salah satu tempat dimana berbagai topik atau isu terkini di ungkapkan dan di diskusikan. Salah satu media sosial yang paling banyak digunakan oleh masyarakat saat ini adalah twitter. Sebagian besar pengguna twitter yang mengutarakan pendapatnya terkait penggunaan salah satu provider jaringan komunikasi terbesar di Indonesia yaitu telkomsel. Berbagai pendapat tersebut dapat digunakan untuk mengembangkan Telkomsel untuk dapat mengevaluasi produknya agar tetap bertahan menghadapi para pesaingnya. Komunikasi

perangkat menjadi semakin mudah karena semakin berkembangnya beberapa teknologi baru seiring dengan perbaikan yang dilakukan pada sistem yang tersedia untuk komputasi dan protokol di Internet. Terdapat kebutuhan data yang akan diwakili untuk mendapatkan layanan pada tingkat yang lebih baik lagi dan kerangka kerja perusahaan untuk mencapai kinerja yang lebih baik pula. Pada akhirnya data mentah yang telah dikumpulkan dari berbagai sumber dapat diakses oleh sistem kemudian dianalisis agar dapat melakukan olah data untuk ekstraksi pengetahuan lebih lanjut [4].

Pelanggan merupakan bagian penting dari sebuah perusahaan, sehingga penting bagi perusahaan untuk membangun hubungan yang baik dengan para pelanggannya. Media Twitter saat ini merupakan salah satu jejaring yang cukup populer di masyarakat. Berbagai informasi yang beredar, berita, opini, komentar, kritik dan saran, baik positif maupun negatif dan netral beredar sangat cepat. Hal ini yang menjadi salah satu latar belakang dilakukannya analisis sentimen. Analisis sentimen adalah teknik untuk mengevaluasi dan mengidentifikasi emosi dan opini positif dan juga negatif [5]. Salah satu metode yang digunakan pada analisis ini adalah Naïve Bayes. Metode ini digunakan untuk mengklasifikasikan opini menjadi beberapa bagian seperti opini negatif, opini netral dan opini positif. Penelitian ini mencoba untuk menganalisis sentimen publik yang diambil dari media twitter berbahasa Indonesia terkait salah satu perusahaan penyedia jaringan selular. Hasil dari penelitian ini adalah tingkat akurasi dari metode yang digunakan serta persentase untuk masing-masing sentiment berdasarkan tweet-tweet yang telah didapat sebelumnya.

Kehadiran jaringan internet sebagai salah satu media komunikasi modern saat ini telah membuat dunia menjadi semakin mudah dijangkau. Hampir semua orang memiliki perangkat komunikasi pintar yang digunakan untuk berkomunikasi dengan semua orang diberbagai belahan dunia melalui media sosial. Menurut Boyd, media sosial merupakan kumpulan dari perangkat lunak yang memungkinkan setiap individu atau komunitas untuk berkumpul, berbagi, dan berkomunikasi serta saling berkolaborasi atau bermain dalam konteks tertentu [6]. Media sosial memiliki kekuatan pada *usergenerated content* (UGC), dimana konten yang dihasilkan oleh para pengguna.

Selain itu, Van Dijk juga memberikan pengertian terkait media sosial, beliau mengatakan bahwa media sosial adalah suatu platform media atau tempat yang fokus pada eksistensi pengguna dan memfasilitasi mereka dalam beraktifitas maupun berkolaborasi. Dapat disimpulkan bahwa media sosial merupakan suatu tempat atau medium atau fasilitator daring yang mampu menguatkan hubungan antar pengguna sekaligus menjadi sebuah ikatan sosial [6].

Twitter merupakan salah satu layanan jejaring sosial dan *microblog* daring yang memungkinkan pengguna untuk dapat mengirim dan membaca pesan berbasis teks atau lebih dikenal dengan sebutan tweet atau kicauan. Media twitter ini memungkinkan pengguna untuk dapat mengirim teks hingga 140 karakter [5] Sebagai salah satu platform komunikasi, Twitter semakin masuk ke dalam kehidupan sehari-hari terlepas dari lokasi geografis seseorang. Berdasarkan data yang dikumpulkan oleh Statistic Brain tahun 2013 di perkiraan 554,7 juta orang di seluruh dunia yang secara aktif menggunakan layanan ini dan memposting 58 juta tweet kolektif setiap hari [7], selain itu didapati pula bahwa terdapat 135.000 pengguna baru yang bergabung dengan jaringan Twitter setiap harinya [8].

Text mining merupakan salah satu bidang ilmu baru yang mempelajari cara untuk mendapatkan informasi yang bermanfaat dari dokumen atau teks asli yang belum dimanipulasi [3]. Dapat disimpulkan bahwa text mining merupakan proses menganalisa suatu teks agar mendapatkan informasi berguna untuk mencapai tujuan tertentu. Text mining sendiri adalah varian kreasi baru hasil perkembangan data mining. Dalam proses text mining terdapat istilah *pre-processing data*, yaitu suatu proses pendahuluan yang diterapkan terhadap data pada teks. Proses text mining memiliki beberapa tahapan, antara lain melakukan tahapan *tokenizing*, *filtering*, dan *stemming* [9].

*The synthetic minority oversampling technique* (SMOTE) merupakan algoritma yang dianggap sebagai algoritma *benchmark* untuk mengatasi masalah *class imbalance learning* (CIL)[10]. Teknik ini merupakan Pendekatan untuk menyeimbangkan data sampel pada kelas yang memiliki ketidak seimbangan berlebihan (mayoritas) dengan focus terhadap kelas minoritas, dengan tujuan meningkatkan kinerja dari klasifikasi

dataset [11]. Sebagian besar algoritma merupakan pembelajaran yang dapat dilihat pada teori minimalisasi risiko empiris, menjadikan pembelajaran ketidakseimbangan kelas (CIL) sebagai salah satu tugas yang cukup menantang dalam proses pembelajaran. Oleh karena itu, model pembelajaran tradisional cenderung dilihat berpihak pada kelas mayoritas dan mengabaikan kinerja kelas minoritas. Teknik pengambilan sampel minoritas sintesis [10]. SMOTE adalah algoritme pengambilan sampel ulang pada tingkat data yang populer, hal ini secara efektif mengatasi masalah *overfitting* yang disebabkan oleh *oversampling* acak dan kehilangan informasi yang disebabkan oleh *undersampling* acak [12]. Pada tahun 2002 SMOTE telah menjadi tolok ukur standar untuk mempelajari data yang tidak seimbang, SMOTE menghasilkan contoh sintesis, salah satunya adalah minoritas lingkungannya [11].

*Naive Bayes* merupakan salah satu teknik klasifikasi dokumen dengan melakukan penyaringan spam (*spam filtering*) [13]. Algoritma *Naive Bayes* merupakan sebuah klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes yang menjadi dasar dari pengklasifikasi Naive Bayes [14]. *Classifier* atau pengklasifikasi adalah kumpulan dari algoritma klasifikasi yang bekerja berdasarkan prinsip bahwa semua fitur yang diklasifikasikan independen satu sama lain [15]. Ciri utama dr *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi atau kejadian [9].

Pada era globalisasi saat ini, sentimen atau opini masyarakat semakin meluas dan bebas untuk diungkapkan di berbagai media. Sentimen dapat menjadi potensi besar bagi perusahaan yang ingin mengetahui umpan balik (*feedback*) dari masyarakat terhadap merek dagang atau produk mereka. Merek telah dianggap sebagai salah satu asset yang tidak berwujud (*intangible asset*) yang paling berharga dan manajemen merk adalah prioritas penting bagi pihak manajemen perusahaan atau organisasi [16]. Jumlah pengguna (*user*) aktif dalam komunikasi daring memiliki jumlah data yang sangat banyak sehingga memunculkan data besar (*big data*). Teknologi *big data* yang merupakan himpunan data dalam jumlah yang

sangat besar, rumit, dan tidak terstruktur menjadi salah satu sumber daya besar yang dapat diolah untuk memperoleh hasil sentimen yang lebih akurat [17]. Jumlah *review* dan opini dalam jumlah yang sangat banyak akan menyulitkan dan memakan waktu untuk membaca secara keseluruhan, sehingga dilakukan suatu perancangan sistem yang secara otomatis akan mengelompokkan opini dan *review* yang ada sesuai kelasnya. Kelas sentimen dibagi menjadi kelas sangat positif, positif, netral, negatif, dan sangat negatif sehingga pengguna dapat memilih ingin membaca opini sesuai dengan klasifikasi yang diinginkan [18]. Sistem analisis sentimen yang dibangun menggunakan algoritma klasifikasi Naive Bayes memiliki ciri utama yaitu asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian [13]. Kelebihan dari Naive Bayes adalah proses klasifikasi data dapat disesuaikan dengan sifat dan kebutuhan masing-masing [5].

## 2. Metode Penelitian

Metode penelitian yang dilakukan untuk menganalisis permasalahan menggunakan teknik data mining dengan algoritma naive bayes.



Gambar 1. Alur Diagram Penelitian

### Pengambilan Data

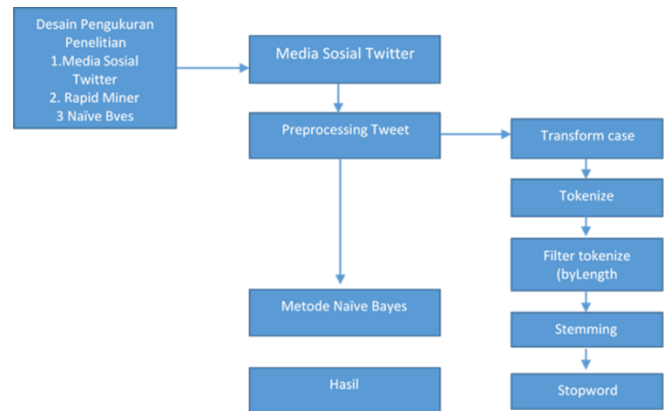
Dataset berupa file excel yang digunakan dalam penelitian ini adalah data mengenai tentang Telkomsel yang diambil dari media twitter dengan sentiment complain dan bukan complain dengan metode *crawling* [19] menggunakan *RapidMiner*.

### Preprocessing

Data pre-processing adalah teknik data mining yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Langkah data pre-processing diperlukan untuk menyelesaikan beberapa jenis masalah termasuk noisy data, data redundansi, nilai data yang hilang [20]. Tahap ini digunakan untuk mengolah data teks menjadi sebuah sentiment Terdapat beberapa *filtering* yang digunakan pada tahap ini di antaranya:

- 1) *Transform Case*: tahap ini merubah karakter huruf besar menjadi huruf kecil

- 2) *Tokenize*: merupakan proses untuk memisahkan data teks menjadi beberapa token
- 3) *Filter Token (Bylength)*: digunakan untuk pemilihan token dengan ukuran minimal yang dipakai penelitian ini adalah 3 karakter(huruf) meski diataranya adalah stopword
- 4) *Stopword*: digunakan untuk membuang kata yang tidak dipakai
- 5) *Stemming*: digunakan untuk memperbaiki dan menerjemahkan kata
- 6) *Naive Bayes*: digunakan untuk mengklasifikasi data pada kelas tertentu



Gambar 2. Kerangka Alur Pemikiran

### Naive Bayes

Metode algoritma yang dilakukan pada penelitian ini adalah dengan menggunakan metode Algoritma *Naive Bayes Classification* (Deepa, Sathya Priya and Devi, 2022), pada tahap ini data di analisis lalu di terapkan model-model yang sesuai dengan jenis data. Model yang diusulkan dalam analisis ini yaitu *naive bayes*. Kata-kata atau istilah asing ditulis dengan huruf miring (Italic). Namun, penggunaan istilah asing sebaiknya dihindari untuk artikel berbahasa Indonesia. Paragraf baru dimulai 10 mm dari batas kiri, sedangkan antar paragraf tidak diberi spasi antara. Semua bilangan ditulis dengan angka arab, kecuali pada awal kalimat.

## 3. Hasil dan Pembahasan

### Analisis hasil pengujian

Sistem yang dibangun adalah sistem yang dapat digunakan untuk menganalisis sentimen pada complain dan bukan complain pada media Twitter Telkomsel. Sistem bekerja dengan melakukan *crawling review* pada laman twitter yang nantinya akan diproses

pada sistem. Perhitungan dilakukan dengan menerapkan langkah yang sudah di jelaskan di bagian metode penelitian, antara lain:

*Pengambilan data*

Data set yang digunakan dalam penelitian opini merupakan komentar berbahasa Indonesia tentang provider telkomsel. Data diambil dari twitter dengan proses crawling menggunakan aplikasi RapidMiner dengan query telkomsel dan diambil sebanyak 1000 data.

Tabel 1. Alur Proses Crawling

No	Proses	Deskripsi
1	Search Twitter	Mengambil tweet yang ada di twitter untuk dijadikan dataset analisis sentiment
2	Select Attributes	Memilih atribut dari contoh data set dan menghapus atribut lainnya
3	Remove Duplicate	Menghilangkan kata dari data set berarti ganda
4	Write csv	Mengubah dataset yang didapat menjadi file csv atau Ms. Excel



Gambar 3. Proses Crawling

*Pre-processing*

Data pre-processing adalah teknik data mining yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Langkah data pre-processing diperlukan untuk menyelesaikan beberapa jenis masalah termasuk noisy data, data redundansi, nilai data yang hilang (Kotsiantis, S. B., D.Kanellopoulos and Pintelas, 2006). Pada proses ini dilakukan pengolahan data teks menjadi analisis sentiment dengan tahapan *preprocessing* menggunakan *SMOTE Upsampling* untuk memproses ketidakseimbangan data, adapun prosesnya sebagai berikut:

Tabel 2. Alur Analisis Sentimen naïve bayes

No	Proses	Deskripsi
1	Read Excel	Memanggil File dataset yang bertipe csv dan untuk mengimport dataset yang telah di crawl
2	Nominal To Text	Mengubah jenis atribut nominal pada dataset menjadi teks
3	Set Role	Mengubah peran satu atribut atau lebih
4	Process Document	Memproses semua data menjadi satu dokumen dan juga memasukkan filter-filter seperti: <ul style="list-style-type: none"> <li>• Transform Case</li> <li>• Tokenize</li> <li>• Filter Tokenize (byLength)</li> <li>• Stemming</li> <li>• Stopword</li> </ul>
5	Smote Upsampling	Memproses ketidakseimbangan klasifikasi data
6	Cross Validation	Metode yang digunakan untuk menganalisis Dokumen dan disini menggunakan model algoritma <i>Naïve bayes</i>



Gambar 4. Proses Filtering



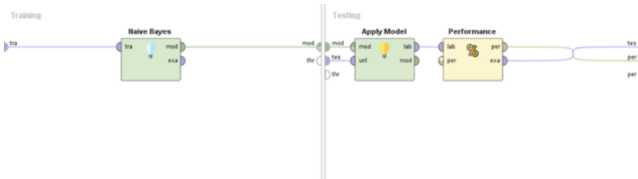
Gambar 5. Proses Preprocessing menggunakan SMOTE

Pada proses ini, data yang sudah dikumpulkan dilakukan proses filtering dan hanya menyisakan kata-kata saja tanpa ada tanda baca dalam data.

- 1) *Transform Case*: RapidMiner akan mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf pada tahap ini.
- 2) *Tokenize*: tahap *Tokenize* untuk menghilangkan tanda baca, symbol dan karakter yang bukan huruf di setiap komentar.
- 3) *Stopword*: pada Tahap ini kita akan menyelesaikan filter token. Kata-kata yang dibuang adalah kata-kata yang tidak relevan atau tidak berarti (kata-kata kasar).
- 4) *Stemming*: pada tahap ini mengubah kata hubung menjadi kata dasar
- 5) *Filter token (By Length)*: pada tahap ini dilakukan pemilihan token dengan panjang minimal 3 karakter.

*Naïve Bayes*

Proses perhitungan penentuan akurasi menggunakan algoritma naïve bayes meliputi, *Training* dan *testing*. Prosesnya melalui pemberian skema data ke mesin. Nantinya mesin akan mempelajari pola-pola data tersebut agar dapat menggunakan hasilnya sebagai acuan untuk pola-pola baru yang memodelkan hasil secara akurat.



Gambar 6. Metode Naïve Bayes

*Hasil Analisis Sentimen*

Hasil dari pengujian performance accuracy naïve bayes untuk analisis sentiment pada media sosial twitter memperoleh hasil sebesar 81,59%.

accuracy: 81.58% +/- 2.30% (micro average: 81.58%)			
	true non compliant	true compliant	class precision
pred non compliant	513	146	77.85%
pred compliant	68	435	86.48%
class recall	88.30%	74.87%	

Gambar 7. Performance Analisis Akurasi

Hasil dari pengujian performance precision naïve bayes untuk analisis sentiment pada media sosial twitter memperoleh hasil sebesar 86,82%.

precision: 86.82% +/- 5.23% (micro average: 86.48%) (positive class: compliant)			
	true non compliant	true compliant	class precision
pred non compliant	513	146	77.85%
pred compliant	68	435	86.48%
class recall	88.30%	74.87%	

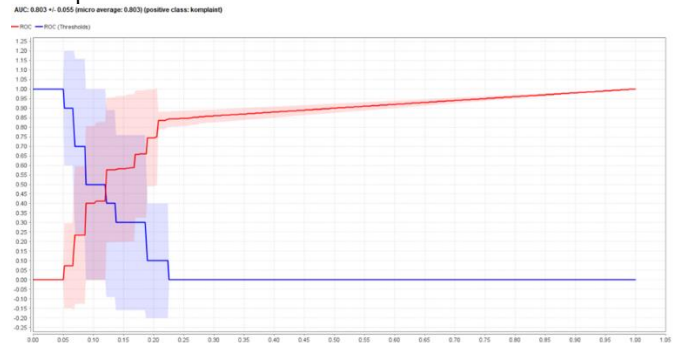
Gambar 8. Performance Analisis Presisi

Hasil dari pengujian performance recall naïve bayes untuk analisis sentiment pada media sosial twitter memperoleh hasil sebesar 74,87%.

recall: 74.87% +/- 4.17% (micro average: 74.87%) (positive class: compliant)			
	true non compliant	true compliant	class precision
pred non compliant	513	146	77.85%
pred compliant	68	435	86.48%
class recall	88.30%	74.87%	

Gambar 9. Performance Analisis Recall

Hasil dari pengujian performance AUC naïve bayes untuk analisis sentiment pada media sosial twitter memperoleh hasil sebesar 0.803.



Gambar 10. Performance Analisis AUC

**4. Kesimpulan**

Dapat disimpulkan bahwa analisis sentiment pada tweet #telkomsel atau @Telkomsel di twitter menunjukkan bahwa masih banyak orang yang memberikan sentiment negatif terhadap layanan Telkomsel. Hasil yang diperoleh dari 1000 dataset menyimpulkan bahwa hasil sentiment yang dihasilkan 86,48% complain dan 77,85% bukan complain dan dengan adanya *SMOTE Upsampling* maka total sentiment true complaint dan true non complain menjadi seimbang yaitu 581, dengan *Accuracy Micro average* 81,58, *precision micro average* 86,48 dan recall rate 74,87% serta AUC sebesar 0.803. Dengan adanya sistem analisis sentimen ini diharapkan dapat membantu perusahaan mengetahui umpan balik terhadap merek dagang atau produk yang mereka



hasilkan, dan masyarakat dalam menilai sebuah produk berdasarkan opini dan review yang ada.

## 5. Daftar Pustaka

- [1] Suryono, S., Utami, E. and Luthfi, E.T., 2018. Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naïve Bayes Classifier. Seminar Nasional GEOTIK 2018.
- [2] Cambria, E., Schuller, B., Xia, Y. and Havasi, C., 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, 28(2), pp.15-21. DOI: 10.1109/MIS.2013.30.
- [3] Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), pp.1-167.
- [4] Deepa, N., Priya, J.S. and Devi, T., 2022. Towards applying internet of things and machine learning for the risk prediction of COVID-19 in pandemic situation using Naive Bayes classifier for improving accuracy. *Materials Today: Proceedings*. DOI: <https://doi.org/10.1016/j.matpr.2022.03.345>.
- [5] Naraswati, N.P.G., Nooraeni, R., Rosmilda, D.C., Desinta, D., Khairi, F. and Damaiyanti, R., 2021. Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *Sistemasi: Jurnal Sistem Informatika*, 10(1), pp.222-238. DOI: <https://doi.org/10.32520/stmsi.v10i1.1179>.
- [6] Sari, A. C. *et al.* 2019. Komunikasi dan media sosial. *Reserachgate.net*.
- [7] Statistic Brain. 2013. *Twitter statistics*. Available at: <http://www.statisticbrain.com/twitter-statistics>.
- [8] Olofinlua, T. 2019. Twitter: social communication in the twitter age, *Information, Communication & Society*, 22(13), pp. 2037–2038. DOI: 10.1080/1369118x.2019.1620824.
- [9] Vu, D.H., 2022. Privacy-preserving Naive Bayes classification in semi-fully distributed data model. *Computers & Security*, 115, p.102630. DOI: <https://doi.org/10.1016/j.cose.2022.102630>.
- [10] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, pp.321-357. DOI: <https://doi.org/10.1613/jair.953>.
- [11] Zhang, A., Yu, H., Zhou, S., Huan, Z. and Yang, X., 2022. Instance weighted SMOTE by indirectly exploring the data distribution. *Knowledge-Based Systems*, 249, p.108919. doi: <https://doi.org/10.1016/j.knosys.2022.108919>
- [12] Anandarajan, M., Hill, C. and Nolan, T. 2019 Text Preprocessing BT - Practical Text Analytics: Maximizing the Value of Text Data, in Anandarajan, M., Hill, C., and Nolan, T. (eds). Cham: Springer International Publishing, pp. 45–59. DOI: 10.1007/978-3-319-95663-3\_4.
- [13] Samuel, N. 2010. *Naive Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen*. Bandung: Institut Teknologi Bandung.
- [14] Bustami, B., 2013. Penerapan algoritma Naive Bayes untuk mengklasifikasi data nasabah asuransi. *TECHSI-Jurnal Teknik Informatika*, 5(2). DOI: 10.26555/jifo.v8i1.a2086.
- [15] Singh, M., Bhatt, M.W., Bedi, H.S. and Mishra, U., 2020. Performance of bernoulli's naive bayes classifier in the detection of fake news. *Materials Today: Proceedings*. DOI: <https://doi.org/10.1016/j.matpr.2020.10.896>.
- [16] Keller, K.L. and Lehmann, D.R., 2006. Brands and branding: Research findings and future priorities. *Marketing science*, 25(6), pp.740-759. DOI: <https://doi.org/10.1287/mksc.1050.0153>.

- [17] Tanesab, F.I., Sembiring, I. and Purnomo, H.D., 2017. Sentiment analysis model based on Youtube comment using support vector machine. *International Journal of Computer Science and Software Engineering*, 6(8), p.180.
- [18] Gunawan, B., Sastypratiwi, H. and Pratama, E.E., 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 4(2), pp.113-118. DOI: 10.26418/jp.v4i2.27526.
- [19] Sembodo, J.E., Setiawan, E.B. and Baizal, Z.A., 2016, October. Data Crawling Otomatis pada Twitter. In *Indonesian Symposium on Computing (Indo-SC)* (pp. 11-16). DOI: 10.21108/indosc.2016.111.
- [20] Kotsiantis, S.B., Kanellopoulos, D. and Pintelas, P.E., 2006. Data preprocessing for supervised learning. *International journal of computer science*, 1(2), pp.111-117.