

Rizkina, Hasan - cek.02

by Nisa Qonita Rizkina, Firman Noor Hasan

Submission date: 12-Jul-2023 10:05AM (UTC+0700)

Submission ID: 2129912250

File name: Rizqina,_Hasan_-_cek.02.docx (394.52K)

Word count: 4646

Character count: 29172



Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes

³ Nisa Qonita Rizkina¹, Firman Noor Hasan²

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr.Hamka, Jakarta, Indonesia

Jl. Tanah Merdeka No.6, RT.10, RW.5, Rambutan, Kec. Ciracas, Kota Jakarta Timur, DKI Jakarta, 13830, Indonesia

Alamat Lengkap (Nama Jalan, Nomor, Kode Pos, Kota), Indor²ia

Email: ¹nisaqonitar@email.com, ²firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Submitted: 99/99/999; Accepted: 99/99/999; Published: 99/99/999

Abstrak—Perkembangan teknologi saat ini sangat pesat membuat penyebaran informasi semakin cepat, penyebaran informasi sangat mudah didapat di media sosial seperti Twitter. Media sosial Twitter sendiri memberikan fitur untuk user-nya agar mereka dapat mengirim dan membaca informasi dalam bentuk teks ataupun video. Kejadian pembubaran konser NCT 127 menuai komentar pro dan kontra. Sehingga banyak pengguna Twitter berkomentar tentang kejadian itu. Data Twitter yang dikumpulkan dengan kata kunci “nct” dan rentang waktu 4-6 November 2022 mendapatkan sebanyak 2541 dataset, dilanjutkan proses cleansing yang menghasilkan data sebanyak 2541 lalu proses pelabelan dan setelah itu pengimplementasian algoritma Naive Bayes. Tujuannya, penelitian ini berguna mengetahui berapa banyak komentar tweets yang positif dan yang negatif dan ³³ mengindikasikan akurasi dari implementasi metode Naive Bayes. Berdasarkan eksperimen yang dilakukan bahwa jumlah komentar positif sebanyak 559 data dan komentar negatif sebanyak 1.892 data. Kemudian hasil pengujian evaluasi dimana ¹⁹ menunjukkan accuracy 82.01%. Setelah itu hasil sentimen negatif mendapatkan hasil recall 68,52%, precision 93,84%, dan f1-score 79,21%. Sedangkan untuk sentimen positif mendapatkan hasil recall 95,50%, precision 75,21%, dan f1-score 84,15%. Dapat disimpulkan bahwa algoritma naive bayes dapat mengklasifikasi dan memproses akurasi lebih stabil hingga mendekati hasil yang sempurna

Kata Kunci: Analisis Sentiment; NCT 127; Twitter; Rapidminer; Naive Bayes

Abstract—The ²⁸ rate of technological advancement has resulted in the rapid spread of information, which is easily available through social media platforms such as Twitter. Users of Twitter can send and read content in the form of text or videos using the facilities that Twitter itself offers. Numerous Twitter users have commented on the NCT 127 concert's recent dissolution, which has drawn both supportive and critical remarks. A dataset of 2541 tweets was created by gathering information from Twitter using the keyword "nct" between November 4 and November 6, 2022. The data was subsequently cleaned, yielding a total of 2541 useable data points. Labeling and the Naive Bayes algorithm were then applied to the data. The goal of this study was to count the number of favorable and unfavorable tweets and evaluate how well the Naive Bayes algorithm was applied. According to the trials done, there were 559 favorable remarks and 1,892 negative ones. The accuracy of the evaluation tests was 82.01%. Additionally, the analysis of negative sentiment produced a f1-score of 79.21%, a recall of ²³ 68.52%, and precision of 93.84%. Contrarily, the evaluation of positive attitude produced a f1-score of 84.15%, a recall of 95.50%, and a precision of 75.21%. The Naive Bayes method, it may be inferred, can categorize and process with a very consistent accuracy that approaches near-perfect outcomes.

Keywords: Sentiment Analysis; NCT 127; Twitter; Rapidminer; Naive Bayes

1. PENDAHULUAN

Dalam era modernisasi teknologi yang pesat, penyebaran informasi ke seluruh dunia semakin cepat, terutama melalui internet. Informasi mengenai perkembangan negara, budaya, dan kultur dapat dengan mudah diakses dan dirasakan oleh banyak orang, termasuk masyarakat di Indonesia [1]. Perkembangan ini dapat dilihat dari pertumbuhan media sosial seperti Twitter. Media sosial dapat menciptakan informasi berbentuk teks sehingga pengguna mendapatkan kemudahan mencari informasi yang akurat dan terbaru yang diperlukan hanya aplikasi yang ²⁴ menyajikan data.

Twitter adalah satu dari media sosial yang mempunyai fitur untuk ^{penggunanya} mengirim dan menerima informasi dalam bentuk teks dan video [2]. Awalnya, Twitter pada saat itu user hanya bisa mengirim dan membaca informasi berbasis teks sampai 140 karakter lalu di tahun 2017 di upgrade menjadi 280 karakter yang sekarang disebut sebagai “tweet” [3]. Tweet mengandung berupa opini dari penggunaannya terhadap suatu kejadian yang telah dialami. Pada masa sekarang Twitter dapat digunakan untuk mengemukakan suatu pendapat tentang suatu hal seperti kegiatan sosial, memberikan kabar tentang kondisi lalu lintas ataupun berita cuaca dan bencana alam serta dapat memperingati tentang suatu fenomena yang dihadapi [4].

Cepatnya penyebaran informasi melalui sosial media Twitter, salah satunya dapat menyebabkan pertukaran budaya seperti budaya Korea atau bisa disebut Korean Wave dan ^{Hal} ³⁵ yang sudah menyebar ke seluruh dunia seperti drama korea dan Kpop (Korean Music Pop)[5]. Budaya Korea ini sudah muncul di Indonesia sejak tahun 2002 dan berkembang hingga saat ini [6]. Munculnya budaya Korea ini disukai oleh para remaja dan orang dewasa.



Salah satu contoh budaya Korea yang paling banyak disukai adalah Kpop, karena ini Kpop menjadi kunci kesuksesan dari *Korean Wave* [7].

NCT 127 merupakan grup idol asal Korea dan termasuk dalam budaya Korea yaitu Kpop. NCT sendiri memiliki banyak penggemar. Di Indonesia sendiri NCT memiliki penggemar yang sangat banyak. Dikarenakan hal itu, pada tanggal 4 dan 5 November 2022 NCT 127 mennggelar konser mereka di Indonesia. Namun, saat konser hari pertama digelar terjadi kerusuhan sehingga konser dibatalkan. Pada media sosial Twitter banyak netizen terutama penggemar yang berkomentar setelah informasi pembubaran konser itu tersebar. Banyak netizen yang berkomentar positif dan negatif mengenai topik pembubaran konser tersebut. Komentar tidak hanya dari penggemar NCT tetapi juga banyak komentar dari masyarakat umum. Kejadian ini juga menjadi daya tarik penerbit berita. Sehingga kejadian pembubaran konser NCT 127 ini banyak menuai pro dan kontra.

Banyaknya berbagai r²⁰am komentar yang dicurahkan oleh pengguna Twitter, maka komentar ini dapat di analisis menggunakan cara analisis sentimen. Analisis sentimen adalah proses memahami, mengesktrak, dan memproses data yang di dapat dari pengumpulan data yang berbentuk teks secara otomatis. Dengan menggunakan analisis sentimen ini peneliti dapat menilai komentar dari sudut pandang orang lain. Proses *text mining* sendiri adalah bidang ilmu yang baru berkembang untuk menemukan atau mengesktrak informasi baru dari koleksi teks alami [8].

Sebelumnya telah dilakukan penelitian a⁷nlisis sentimen menggunakan naïve bayes di teliti oleh Amelia di tahun 2022. Hasil penelitian ini mendapatkan *accuracy* 69%, *precision* 73%, *recall* 69% dan *f1-score* 69% [9]. Setelah itu penelitian yang pernah dilakukan oleh Nurwati, Supriyati, L²⁹istyorini pada tahun 2020. Dalam penelitian itu mendapatkan hasil *accuracy* sebesar 75% [10]. Lalu penelitian yang dilakukan ol²²Astari, Divayana, dan Indrawan pada tahun 2020 mendapatkan hasil *accuracy* sebesar 67% [11]. Kemudian penelitian yang dilakukan oleh Andika, Azizah, dan Respatiwalan pada tahun 2019 mendapatkan *accuracy* 82,90%[12]. Selanjutnya pada penelitian sentimen analisis yang dilakukan oleh Wandani, Fauziah, dan Andrianingsih mendapatkan *accuracy* sebesar 83,3% untuk *naïve bayes*, 80,59% random forest, dan 82,94% untuk KNN akurasi tersebut r¹⁵gunakan kata kunci “flash sale”. Sedangkan untuk ka⁴ kunci “flash sale shopee” mendapatkan *accuracy* 81,48% *naïve bayes*, 77,78% KNN, 74,07% random forest. Dari hasil penelitian itu disimpulkan bahwa algoritma naïve bayes menjadi rekomendasi untuk pengklasifikasikan analisis sentimen dikarena akurasi lebih besar dan lebih stabil [13].

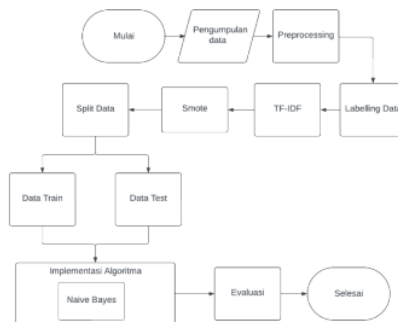
Dari beberapa penelitian tersebut, peneliti gunakan sebagai acuan dan perbandingan pada penelitian ini. Adapun yang membuat beda dari penelitian ini dengan penelitian sebelumnya yang pertama topik yang dibahas berbeda, kemudian cara pengumpulan data dan jumlah data berbeda, dan yang ketiga pada pada penelitian ini saat proses *preprocessing* peneliti menggunakan tahapan *filter token* berbeda dari penelitian diatas mereka tidak meng¹⁷akan tahap filter token. Pada penelitian ini hanya metode saja yang sama dengan penelitian sebelumnya yaitu menggunakan metode *Naïve Bayes*. M¹⁰ode *Naïve Bayes* ini banyak digunakan karena metode ini dapat menghasilkan akurasi yang baik dan stabil. Oleh karena itu, pada penelitian ini peneliti menggunakan metode *naïve bayes* sebagai m¹⁴ode untuk analisis sentimen komentar netizen di Twitter dari topik pembubaran konser NCT 127. Capaian dari penelitian ini untuk mengetahui berapa banyak sentimen positi dan negatif komentar netizen di Twitter dan mengindikasi tingkat akurasi pada analisis sentimen menggunakan metode *naïve bayes*.

11

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Penelitian ini menggunakan data yang dikumpulkan berkaitan pada komentar pembubaran konser k³⁰T 127 di Indonesia. Penelitian ini memakai metode *Naïve Bayes*, berikut alur penelitian yang peneliti gunakan dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

30



Berikut penjelasan dari Gambar 1. yang ada diatas:

- Pengumpulan data, memakai bahasa pemrograman *python* dan *library python snsrape* lalu menembak website Twitter.
- Preprocessing*, data yang telah dikumpulkan dilakukan pembersihan menggunakan beberapa tahapan *preprocessing*.
- Pelabelan data, data yang telah dibersihkan kemudian diberi label.
- Implementasi algoritma, pada langkah ini dataset sudah dilabel kemudian dibagi menjadi dua pembagian data yaitu, data latih dan data uji. Setelah itu implementasi algoritma *naive bayes*.
- Evaluasi, hasil pengimplementasi algoritma *naive bayes* berupa akurasi, *recall.f1-score*, dan *precision*.

2.2 Pengumpulan Data

Pada proses mengumpulkan data atau *crawl* data merupakan tahap awal dalam penelitian ini. Data ditarik dengan *keyword* yang relevan topik penelitian dari media sosial Twitter. Data yang diambil merupakan *tweet* yang berasal dari pengguna Twitter. Pengambilan data ini menggunakan bantuan dari *Auth* yang disediakan oleh Twitter [14].

2.3 Preprocessing

Dalam penelitian analisis sentimen, *preprocessing* merupakan tahap yang perlu dilakukan. Tahapan ini juga merupakan tahap pertama dalam pemrosesan kalimat dalam data[15]. Kalimat di dokumen akan di normalisasikan sehingga dapat memudahkan pada tahap pembobotan [16]. Terdapat beberapa langkah pada *preprocessing* yaitu, sebagai berikut:

2.3.1 Cleansing

Langkah ini merupakan tahapan pembersihan dokumen atau data untuk menghilangkan *username(@)*, RT atau *retweet*, URL, simbol, tanda baca, dan angka yang tidak diperlukan. Bertujuan agar data mengandung informasi relevan dan sesuai pada kebutuhan di penelitian [17].

2.3.2 Transform Case

Transform Case adalah salah satu tahapan umum dalam preprocessing. Tahapan ini sering digunakan, terutama parameter TF-IDF yang ada di Rapidminer. Pada tahapan ini kalimat dalam dokumen diubah menjadi huruf kecil (*lowercase*)[18].

2.3.2 Tokenizing

Tokenisasi merupakan tahap memotong kata-kata dalam kalimat menjadi token, yaitu unit terkecil yang memiliki makna [17]. Proses ini dilakukan dengan memberikan jarak menggunakan spasi antar kata [19].

2.3.3 Stopwords Removal

Stopwords adalah tahap menghapus kata terlihat umum dipakai namun tidak bermakna penting. Dapat dilihat contoh dari kata stopword menggunakan Bahasa Indonesia yaitu “yang”, “dan”, “di”, “dari” lalu masih banyak lagi. Tahap ini bertujuan untuk menghapus kata-kata yang tidak relevan dan memfokus pada kata-kata yang penting [20].

2.3.4 Stemming

Proses *stemming* berfungsi mengubah kata dalam kalimat menjadi bentuk kata dasar. Proses ini dilakukan dengan menggunakan library yang menjadi acuan sebagai referensi [21]. Tujuan dari proses ini untuk menyamakan kata-kata sehingga mempermudah analisis dan pengolahan data.

2.4 Algoritma Naïve Bayes

Metode *Naïve bayes* adalah satu dari banyak algoritma yang digunakan dalam proses *text mining* [2]. *Naïve bayes* sering digunakan untuk mengklasifikasi dataset yang berbentuk tulisan. Hal yang khas dari *Naïve bayes* yaitu ini terkuat (naif) terhadap tingkat kebesaran dari suatu kondisi atau kejadian. Rumus dari metode *naïve bayes*, sebagai berikut [21].

$$P(C|A) = \frac{p(A|C).p(C)}{P(A)}$$

Keterangan:

- A = Data yang *class* tidak diketahui
- C = Hipotesis A yaitu *class spesifik*
- $P(C|A)$ = Probabilitas C berkaitan dengan A (*posteriori probability*)
- $P(C)$ = Probabilitas Hipotesis (*prior probability*)
- $P(A|C)$ = Probabilitas A berkaitan dengan hipotesis C
- $P(A)$ = Probabilitas A



3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pada proses pengumpulan data penelitian, data dikumpulkan melalui media sosial Twitter lalu menembak tweets pengguna dengan menggunakan bahasa pemrograman Python menggunakan library *Sns scrape* dengan teknik *web scapping*. Data dikumpulkan berdasarkan keyword “NCT” dan “Konser NCT rusuh”. Tujuan pengumpulan data yaitu untuk mendapatkan informasi dari suatu situs web[22]. Pada Gambar 2 dapat dilihat bagaimana ilustrasi proses pengumpulan data.



Gambar 2. Ilustrasi Pengambilan Data

Pengambilan data dari media sosial Twitter menggunakan rentang waktu 4-6 November 2022, dengan total data yang terkumpul sebanyak 2451. Dataset tersebut kemudian disimpan ke file format CSV dan diberi nama file *Konser_NCT.csv*.

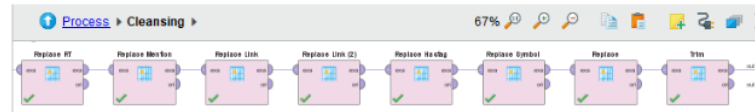
A	C	J	O	P
1	Tweet		Language	Username
2	@NCT_Indonesia Kasian deh	in		Helgaaaaaav
3	Pengen liat konser sebelum bi	in		elsaauriga
4	@lovemyselff @mecimapro v	in		abcdefghyunjin_
5	@MQMOAYASE ituu konser nc	in		naspadayangulal
6	NCT kapan konser ke indo aku	in		noviaapr
7	Abis konser nct niatnya	in		aleenaskyy
8	KEPADA MUTUALKU MAAF AKI	in		Syrenael1
9	GUE GAK BERHENTI RT TENTAN	in		coordiesh

Gambar 3. Tampilan file CSV

Terlihat pada Gambar 3 diatas, merupakan isi file.csv, beberapa hasil data yang telah dikumpulkan berkaitan dengan komentar netizen mengenai pembubaran konser nct 127. Menggunakan bahasa pemrograman *python* dengan library *sns scrape*.

3.2 Cleansing

Dalam tahap ini, dilakukan pembersihan data dengan menghapus informasi yang tidak relevan, seperti *retweet* (RT), URL, *hashtag*(#), tanda baca, *username*(@), angka, dan karakter yang tidak diperlukan. Setelah proses pembersihan, dataset akan tersisa sesuai dengan kebutuhan yang diinginkan.



Gambar 4. Proses cleansing menggunakan operator *replace* dan *trim*

Dapat dilihat pada Gambar 4 yang merupakan proses *cleansing* data dalam penelitian ini dilakukan menggunakan Rapidminer, dengan menggunakan tujuh operator *replace* dan *trim*. Penjelasan dari masing-masing operator sebagai berikut:

- Operator *replace* RT: untuk menghapus kata-kata RT dalam kalimat.
- Operator *replace* *username*(@): untuk menghapus @*username*.
- Operator *replace* *hashtag*(#): untuk menghapus *hashtag*(#).
- Operator *replace* Link: untuk menghapus URL.
- Operator *replace* simbol: untuk menghapus symbol atau karakter tidak relevan.
- Operator *replace*: Untuk mengapus angka.
- Operator *trim*: Untuk menghapus baris kosong.

Setelah berhasil melakukan proses *cleansing* data, maka dataset menjadi bersih dari informasi yang tidak relevan untuk topik penelitian. Terdapat perbedaan antara *tweets* sebelum dan setelah proses *cleansing*, dapat dilihat di Tabel 1. Pada tabel tersebut terlihat jelas perbedaan setelah dilakukan *cleansing* yaitu tanda baca dan angka pada *tweets* terhapus.

Tabel 1. Ilustrasi *cleansing*



Sebelum	Sesudah
Buat yg nonton Hari 2, tolong yg tertib ya karna ketentuan konser nct selanjutnya di indo ada ditangan kalian	Buat yg nonton Hari tolong yg tertib ya karna ketentuan konser nct selanjutnya di indo ada ditangan kalian
Setiap konser nct bermasalah mulu, alias fandom indonya rusuh otak dangkal !	Setiap konser nct bermasalah mulu alias fandom indonya rusuh otak dangkal

3.3 Preprocessing



Gambar 5. Preprocessing menggunakan rapidminer

Terdapat pada Gambar 5 merupakan proses dari *preprocessing*. Tahapan ini melakukan preprocessing data komentar netizen pada kejadian konser nct 127 yang telah dilakukan cleansing menggunakan rapidminer. Tahapan preprocessing ini menggunakan rapidminer. Preprocessing sendiri terdiri dari beberapa Langkah yaitu, *cleansing*, *transform case*, *tokenizing*, *stopwords removal*, *filter tokens* dan *stemming* [23].

3.3.1 Transform Case Case Folding

Langkah pertama dari preprocessing adalah transform case. langkah ini dilakukan perubahan semua huruf pada data yang telah dibersihkan menjadi huruf kecil. Langkah ini bertujuan untuk memastikan bahwa kalimat dalam data memiliki format yang konsisten dan standar[11].

Tabel 2. Ilustrasi transform case

Sebelum	Sesudah
Buat yg nonton Hari tolong yg tertib ya karna ketentuan konser nct selanjutnya di indo ada ditangan kalian	buat yg nonton hari tolong yg tertib ya karna ketentuan konser nct selanjutnya di indo ada ditangan kalian
Setiap konser nct bermasalah mulu alias fandom indonya rusuh otak dangkal	setiap konser nct bermasalah mulu alias fandom indonya rusuh otak dangkal

Terlihat perbedaan pada Tabel 2 kalimat sebelum dan sesudah di transform case. Pada kolom sebelah kiri terlihat beberapa huruf kapital, kemudian pada kolom sebelah kanan huruf kapital tersebut berubah menjadi huruf kecil setelah proses *transform case* dijalankan.

3.3.2 Tokenizing

Langkah kedua yaitu *tokenizing*, tahap ini melakukan tokenisasi pada kalimat data bertujuan untuk memisahkan kalimat menjadi kata-kata terpisah. Pemisahan dilakukan menggunakan spasi sebagai pemisah kata[24].

Tabel 3. Ilustrasi tokenizing

Sebelum	Sesudah
buat yg nonton hari tolong yg tertib ya karna ketentuan konser nct selanjutnya di indo ada ditangan kalian	['buat', 'yg', 'nonton', 'hari', 'tolong', 'yg', 'tertib', 'ya', 'karna', 'ketentuan', 'konser', 'nct', 'selanjutnya', 'di', 'indo', 'ada', 'ditangan', 'kalian']
setiap konser nct bermasalah mulu alias fandom indonya rusuh otak dangkal	['setiap', 'konser', 'nct', 'bermasalah', 'mulu', 'alias', 'fandom', 'indonya', 'rusuh', 'otak', 'dangkal']

Terlihat pada Tabel 3 kolom sebelah kiri merakan teks sebelum dilakukan *tokenize* sehingga masih berbentuk kalimat. Sedangkan pada kolom sebelah kanan hasil dari setelah dilakukan *tokenize* sehingga kalimat juga berbentuk per unit kecil atau per kata.

3.3.3 Stopwords Removal

Langkah ketiga yaitu stopwords removal yang bertujuan mengeliminasi kata-kata yang sekiranya kurang memiliki makna yang jelas, seperti kata sambung atau kata penghubung, untuk meningkatkan kejelasan pada kalimat. Ilustrasi dari *stopword* seperti Tabel 4. Pada tabel dibawah ini terlihat pada kolom sebelah kiri merupakan teks sebelum di *stopwords*, sedangkan kolom sebelah kanan merupakan hasil setelah dilakukan *stopwords*

Tabel 4. Ilustrasi stopwords



Sebelum	Sesudah
['buat', 'yg', 'nonton', 'hari', 'tolong', 'yg', 'tertib', 'ya', 'karna', 'ketentuan', 'konser', 'nct', 'selanjutnya', 'di', 'indo', 'ada', 'ditangan', 'kalian']	['buat', 'nonton', 'hari', 'tolong', 'tertib', 'ya', 'karna', 'ketentuan', 'konser', 'nct', 'selanjut', 'indo', 'tangan', 'kalian']
['setiap', 'konser', 'nct', 'bermasalah', 'mulu', 'alias', 'fandom', 'indonya', 'rusuh', 'otak', 'dangkal']	['setiap', 'konser', 'nct', 'bermasalah', 'mulu', 'fandom', 'indo', 'rusuh', 'otak', 'dangkal']

3.3.4 Filter Token

Selanjutnya langkah filter token digunakan untuk menghapus kata-kata yang memiliki jumlah karakter di antara 4 sampai 25, yang memiliki tujuan untuk memperbaiki kejelasan kalimat [25]. Dengan menggunakan filter token maka kata yang memiliki kurang dari 4 dan lebih dari 25 di eliminasi, sehingga informasi dapat lebih jelas karena ketidakadaan kata yang tidak informatif.

Tabel 5. Ilustrasi filter token

Sebelum	Sesudah
['buat', 'nonton', 'hari', 'tolong', 'tertib', 'ya', 'karna', 'ketentuan', 'konser', 'nct', 'selanjut', 'indo', 'tangan', 'kalian']	['buat', 'nonton', 'hari', 'tolong', 'tertib', 'karna', 'ketentuan', 'konser', 'selanjut', 'indo', 'tangan', 'kalian']
['setiap', 'konser', 'nct', 'bermasalah', 'mulu', 'fandom', 'indo', 'rusuh', 'otak', 'dangkal']	['setiap', 'konser', 'bermasalah', 'mulu', 'fandom', 'indo', 'rusuh', 'otak', 'dangkal']

Hasil dari filter token dapat dilihat pada Tabel 5, Terdapat beberapa kata seperti 'ya' dan 'nct' terhapus dikarenakan kata tersebut memiliki length kurang dari 4 sesuai dengan ketentuan dari parameter filter token.

3.3.5 Stemming

Langkah selanjutnya yaitu stemming berfungsi untuk mengubah kata yang berbentuk imbuhan diubah ke bentuk dasar yang sesuai dengan kamus atau library yang digunakan [26]. Untuk mengetahui mana saja kata imbuhan maka langkah ini membutuhkan dictionary.

Tabel 6. Ilustrasi stemming

Sebelum	Sesudah
['buat', 'nonton', 'hari', 'tolong', 'tertib', 'karna', 'ketentuan', 'konser', 'selanjut', 'indo', 'tangan', 'kalian']	['buat', 'nonton', 'hari', 'tolong', 'tertib', 'karna', 'tentu', 'konser', 'selanjut', 'indo', 'tangan', 'kalian']
['setiap', 'konser', 'bermasalah', 'mulu', 'fandom', 'indo', 'rusuh', 'otak', 'dangkal']	['setiap', 'konser', 'masalah', 'mulu', 'fandom', 'indo', 'rusuh', 'otak', 'dangkal']

Dapat dilihat hasil stemming pada Tabel 6 sebelumnya terdapat kata "ketentuan" dan "bermasalah" setelah dilakukan stemming maka kata tersebut berubah menjadi kata dasar yaitu "tentu" dan "masalah". Seperti itulah ilustrasi proses dari stemming.

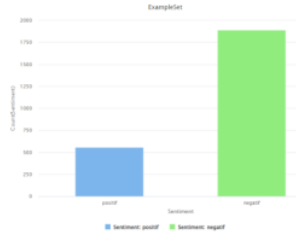
3.4 Pelabelan Data/Labelling

Tahap setelah preprocessing adalah pelabelan data. Pelabelan data Twitter di penelitian ini dikerjakan dengan manual. Dataset telah terkumpul kemudian dibagi menjadi dua kelas sentimen, yaitu positif dan negatif. Ilustrasi perbandingan antara sentimen positif dan negatif dapat dilihat pada Tabel 7. Disana terlihat teks yang telah di preprocessing dilakukan pelabelan menjadi sentimen positif dan negatif.

Tabel 7. Contoh komentar positif dan negatif di Twitter

Tweets	Sentimen
['buat', 'nonton', 'hari', 'tolong', 'tertib', 'karna', 'tentu', 'konser', 'selanjut', 'indo', 'tangan', 'kalian']	positif
['setiap', 'konser', 'masalah', 'mulu', 'fandom', 'indo', 'rusuh', 'otak', 'dangkal']	negatif

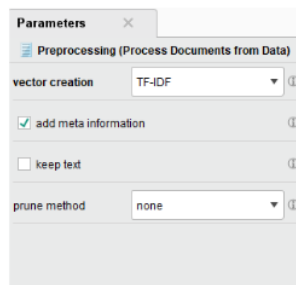
Tabel di atas merupakan contoh komentar yang diberi label sentimen positif dan negatif. Kemudian setelah dataset di label, maka dapat dilihat visualisasi jumlah perbandingan sentimen positif dan negatif pada Gambar 6.



Gambar 6. Visualisasi perbandingan sentimen positif dan negatif

3.5 TF-IDF

Tahap ini merupakan metode yang berguna untuk mengakurasi nilai pada setiap kata yang sering muncul dalam dokumen [27]. Untuk melakukan pembobotan dengan TF-IDF ini dengan menggunakan operator process dokumen yang menampung operator dari tahapan preprocessing yaitu, *transform case*, *tokenize*, *stopwords*, *filter token*, dan *stem*. Setelah itu pilih parameter dan atur bagian vector creation ke TF-IDF, seperti yang terlihat di Gambar 7.



Gambar 7. TF-IDF

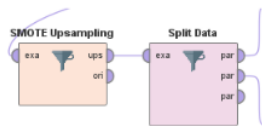
Sesudah parameter telah diatur, selanjutnya operator *process document* dihubungkan ke operator *wordlist to data*. Operator ini untuk memvisualisasi nilai kata yang lebih banyak muncul. Jika banyak kata yang muncul maka pada visualisasi lebih terlihat jelas. Bentuk visualisasi yang dimaksud dapat dilihat pada Gambar 8.



Gambar 8. Visualisasi kata menggunakan wordcloud

3.6 Split Data

Sebelum melakukan tahap implementasi algoritma, dataset yang telah diberi label dimasukkan ke dalam operator smote upsampling. Kegunaan dari operator ini digunakan untuk mengatasi ketidakseimbangan label dalam dataset. Tujuan dari operator ini adalah memproses data yang tidak seimbang sentimennya kemudian diolah sehingga menjadi seimbang.

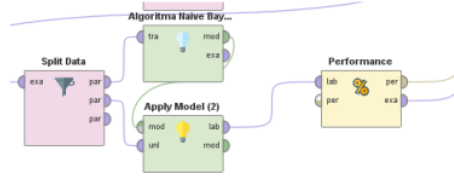


Gambar 9. Operator smote dan split data

Setelah proses *smote*, selanjutnya dataset bedakan menjadi data latih dan data uji. *Train data* yaitu data yang dijadikan pola pelabelan dan setelah itu *test data* digunakan untuk mengevaluasi pola *train data* [28]. Pada

pembagian data latih dan data uji per⁶ti menggunakan operator *split data* seperti pada Gambar 9. Dengan perbandingan rasio 80:20. Sebesar 80% data latih dan sebesar 20% data uji [29]. Kemudian data yang dibagi masuk ke tahap implementasi.

3.7 Implementasi Algoritma



Gambar 10. Implementasi algoritma naive bayes

Tahap selanjutnya adalah implementasi algoritma naïve bayes. Setelah pembagian data latih dan data uji menggunakan operator *split data* selanjutnya dihubungkan ke operator naïve bayes dan *apply model*. Kemudian untuk mengetahui evaluasi pada penerapan algoritma naïve bayes maka digunakan operator *performance*. Proses implementasi algoritma menggunakan rapidminer seperti pada Gambar 10.

3.8 Evaluasi

Tahapan akhir⁶ yang dilakukan adalah evaluasi, evaluasi ini sebuah tahapan yang penting, bertujuan agar mendapatkan nilai *accuracy*, *recall*, *precision*, dan *f1-score*.

accuracy: 82.01%

	true positif	true negatif	class precision
pred. positif	361	119	75.21%
pred. negatif	17	259	93.84%
class recall	95.50%	68.52%	

Gambar 11. Confusion matrix

Penjelasan dari Gambar 11:

- Hasil pengolahan pada 2451 data yang di uji, mendapatkan 361 merupakan data True Positif (TP).
- Hasil pengolahan pada 2541 data yang di uji, mendapatkan 119 data False Positif (FP)
- Hasil pengolahan pada 2541 data yang di uji, mendapatkan 17 data False Negatif (FN)
- Hasil pengolahan pada 2541 data yang di uji, mendapatkan 259 data yang menjadi True Negatif (TN)

Berikut adalah hasil akhir⁹ implementasi algoritma *Naive Bayes confusion matrix* pada Gambar 11, hasil dari data yang di bagi menjadi data latih dan data uji klasifikasi menggunakan algoritma naïve bayes pada rapidminer. Mendapatkan nilai akurasi sebesar 82,01%. Kemudian di dapat hasil precision sebesar 75,21%, recall sebesar 95,505, dan f1-score sebesar 84,15%. Perbedaan dari hasil penelitian ini dengan penelitian terdahulu ialah penelitian ini mendapatkan nilai *accuracy* yang stabil. Hasil *recall*, *f1-score*, dan *precision* lebih besar dari pada sentimen negatif.

KESIMPULAN

Berdasarkan hasil pengolahan data Twitter sebanyak 2451 data, dengan topik pembubaran konser NCT 127 dan diambil datanya pada rentang waktu 4-6 November 2022. Pengolahan data ini bertujuan untuk mengetahui bagaimana komentar netizen pengguna Twitter. Proses yang telah dilakukan pada dataset sebanyak 2451 tweets dari Twitter untuk mengetahui bagaimana komentar netizen terhadap pembubaran konser NCT 127 di hari pertama konser. Setelah melalui tahapan cleansing dan preprocessing data yang sudah bersih tetap sebanyak 2451 data. Tahap pelabelan secara manual seharusnya dapat melabeli secara otomatis, data sentimen³⁴ positif dari hasil pelabelan yaitu Setelah itu pembagian *train data* dan *testing data* menggunakan perbandingan dengan rasio 80:20, untuk data latih sebanyak 1961 dan data uji sebanyak 490. Setelah itu tahapan implementasi algoritma *naive bayes* dimana data latih dan data uji di klasifikasi. Hasil *accuracy* pada proses pengujian evaluasi menghasilkan 82,01%. Setelah itu hasil sentimen negatif mendapatkan hasil *recall* 68,52%, *precision* 93,84%, dan *f1-score* 79,21%. Dibandingkan sentimen positif¹ mendapatkan hasil *recall* 95,50%, *precision* 75,21%, dan *f1-score* 84,15%. Dapat disimpulkan bahwa algoritma *naive bayes* mendapatkan hasil akurasi lebih stabil dan mendekati hasil yang sempurna. Hasil akurasi pada penerapan metode ini berdasarkan dari komentar user Twitter yang mengganggu kejadian pasca pembubaran konser NCT 127, jika dilihat berdasarkan pelabelan dan akurasi. Sentimen negatif lebih banyak dikarenakan rasa kecewa dan malu akibat kejadian tersebut. Oleh karena itu saran untuk



penyelenggara konser berikutnya keamanan lebih di perketat dan untuk yang menonton harap menjaga sikap. Pengembangan penelitian ini dapat menggunakan metode klasifikasi lain untuk membandingkan hasil akurasi dengan berbagai metode klasifikasi lainnya.

REFERENCES

- [1] N. Magfirah Syahmar, I. I. Idrus, M. Ridwan, and S. Ahmad, "BUDAYA K-POP DAN KEHIDUPAN SOSIAL REMAJA (STUDI KOMUNITAS K-POP MAKASSAR: NCT-Zen MAKASSAR)," *Agustus*, vol. 3, no. 1, 2022.
- [2] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [3] M. A. Djamaludin, A. Triayudi, and E. Mardiani, "Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naive Bayes Classifier," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 2, p. 2022, 2022, doi: 10.35870/jti.
- [4] F. Rizali Rakhman, R. Wulan Ramadhani, and Y. Ari Kuncoroyakti, "ANALISIS SENTIMEN DAN OPINI DIGITAL KAMPANYE 3M DI MASA COVID-19 MELALUI MEDIA SOSIAL TWITTER," *Maret*, vol. 18, no. 8, 2021.
- [5] D. A. Hidayati, S. Dini, R. Fitriani, and S. Habibah, "Realitas Sosial Remaja Penggemar Budaya Korea (K-POP) di Bandar Lampung," *RESIPROKAL*, vol. 4, no. 2, pp. 212–232, 2022.
- [6] D. Guna Memenuhi Persyaratan *et al.*, "BUDAYA POPULER KOREA SELATAN (K-POP) DAN PERILAKU KONSUMTIF PENGGEMAR GRUP MUSIK KOREA SELATAN: STUDI KASUS EXO-L MARKAS LOTTO," Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2020.
- [7] M. Ananda, N. Hadi, and N. H. P. Meiji, "Di balik perilaku konsumtif NCTZEN dalam pembelian merchandise NCT (studi kasus komunitas NCTzen Malang)," *Jurnal Integrasi dan Harmoni Inovatif Ilmu-Ilmu Sosial*, vol. 1, no. 9, pp. 1011–1026, 2021, doi: 10.17977/um063v1i92021p1011-1026.
- [8] M. Kantardics, *Data Mining: Concept, models, methods, and algorithms*, Third Edition., vol. 03. Piscataway, NJ: John Wiley & Son, Inc, 2020.
- [9] R. Amelia, N. S. Prastiwi, and M. E. Purbaya, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *Jurnal Riset Komputer*, vol. 9, no. 2, pp. 338–343, 2022, doi: 10.30865/jurikom.v9i2.3895.
- [10] P. Nurmawati, E. Supriyati, and T. Listyorini, "ANALISIS SENTIMEN TERHADAP PENGGEMAR K-POP DI MEDIA SOSIAL TWITTER MENGGUNAKAN NAIVE BAYES (STUDI KASUS PENGGEMAR GRUP BTS)," *JIEET (Journal Information Engineering and Educational Technology)*, vol. 04, no. 02, 2020.
- [11] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.
- [12] L. Aji Andika and P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indonesian Journal of Applied Statistics*, vol. 2, no. 1, 2019.
- [13] A. Wandani, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes," 2021.
- [14] A. Wibowo, F. Noor Hasan, R. Nurhayati, and dan Arief Wibowo, "Jurnal Asimetrik: Jurnal Ilmiah Rekayasa Dan Inovasi Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naive Bayes Classifier," vol. 4, pp. 239–248, 2022.



- [15] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes,” vol. 9, no. 2, pp. 785–795, 2022.
- [16] A. Muzaki and A. Witanti, “SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER,” *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [17] A. Bagus Sasmita, B. Rahayudi, and L. Mufflikhah, “Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes,” 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [18] S. H. Ramadhani and M. I. Wahyudin, “Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN,” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 4, p. 2022, 2022, doi: 10.35870/jti.
- [19] I. Mulya and C. M. Karyati, “Analisis Sentimen Terhadap Universitas Gunadarma Berdasarkan Opini Pengguna Twitter Menggunakan Metode Naive Bayes Classifier,” *Jurnal Ilmiah Komputasi*, vol. 19, no. 4, Dec. 2020, doi: 10.32409/jikstik.19.4.354.
- [20] R. Slamet, W. Gata, A. Novtariyany, K. Hilyati, and F. A. Jariyah, “Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal,” *INTECOMS: Journal of Information Technology and Computer Science*, vol. 5, no. 1, pp. 145–153, 2022, doi: 10.31539/intecom.s.v5i1.3933.
- [21] A. Kusuma and A. Nugroho, “Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 15, no. 2, 2021.
- [22] D. Anjas Ramadhan and E. Budi Setiawan SSi, “ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE,” vol. 06, p. 9736, Aug. 2019.
- [23] D. Rusdianan and D. Rosiyadi, “ANALISA SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE,” 2019.
- [24] A. Azis Adjie Sumanjaya and A. Ridok, “Analisis Sentimen Data Tweets terhadap Penanganan Covid-19 di Indonesia menggunakan Metode Naïve Bayes dan Pemilihan Kata Bersentimen menggunakan Lexicon Based,” 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [25] N. Hardi *et al.*, “SISTEMASI: Jurnal Sistem Informasi Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier.” [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [26] F. Noor, H. #1, S. #2, and P. Afikah, “Sentiment Analysis of Community Response on Cooking Oil Price Increase Policy with Naïve Bayes Classifier Algorithm.” [Online]. Available: <https://t.co/B6rkIjpkov>
- [27] I. R. Afandi, F. Noor, H. #2, A. A. Rizki, N. Pratiwi, and Z. Halim, “Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes.” [Online]. Available: <https://t.co/2HAdwg1drL>
- [28] G. Shmueli, P. C. Bruce, A. V Deokar, and N. R. Patel, *Machine Learning for Business Analytics: Concepts, Techniques and Applications in RapidMiner*, First Edition., vol. 1. Hoboken, USA: John Wiley & Son, Inc, 2023.AA
- [29] “Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression.” [Online]. Available: <https://t.co/23c4krbjp>

ORIGINALITY REPORT

11%

SIMILARITY INDEX

10%

INTERNET SOURCES

5%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1	123dok.com Internet Source	1%
2	ejurnal.seminar-id.com Internet Source	1%
3	Submitted to Yonkers High School Student Paper	1%
4	jutif.if.unsoed.ac.id Internet Source	<1%
5	digilib.unila.ac.id Internet Source	<1%
6	ejournal.uniramalang.ac.id Internet Source	<1%
7	www.researchgate.net Internet Source	<1%
8	Siti Lestari, Akmaludin Akmaludin, Mohammad Badrul. "IMPLEMENTASI KLASIFIKASI NAIVE BAYES UNTUK PREDIKSI KELAYAKAN PEMBERIAN PINJAMAN PADA KOPERASI ANUGERAH BINTANG	<1%

CEMERLANG", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2020

Publication

9	Submitted to Universitas Amikom Student Paper	<1 %
10	Submitted to Universitas Negeri Surabaya The State University of Surabaya Student Paper	<1 %
11	conference.binadarma.ac.id Internet Source	<1 %
12	djournals.com Internet Source	<1 %
13	media.neliti.com Internet Source	<1 %
14	Firman Noor Hasan. Jurnal Linguistik Komputasional (JLK), 2021 Publication	<1 %
15	tunasbangsa.ac.id Internet Source	<1 %
16	www.scielo.br Internet Source	<1 %
17	journal.maranatha.edu Internet Source	<1 %
18	journal.stekom.ac.id Internet Source	<1 %

<1 %

19

journal.univpancasila.ac.id

Internet Source

<1 %

20

jurnal.umk.ac.id

Internet Source

<1 %

21

library.umn.ac.id

Internet Source

<1 %

22

Akhmad Fadjeri, Atik Muhimatun Asroriyah, Atiq Rahmawati. "Analisis Teks Bahasa Indonesia Dan Inggris Dari Sebuah Citra Menggunakan Pengolahan Citra Digital", Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN), 2022

Publication

<1 %

23

George Azzopardi, Nicolai Petkov. "Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters", Pattern Recognition Letters, 2013

Publication

<1 %

24

Ikhwanul Hakim, Arifin Nugroho, Sulaeman Hadi Sukmana, Windu Gata. "Sentimen Analisis Stay Home menggunakan metode klasifikasi Naive Bayes, Support Vector Machine, dan k-Nearest Neighbor",

<1 %

Paradigma - Jurnal Komputer dan Informatika, 2020

Publication

25	ejournal.sisfokomtek.org Internet Source	<1 %
26	ejournal.undip.ac.id Internet Source	<1 %
27	eprints.ums.ac.id Internet Source	<1 %
28	id.123dok.com Internet Source	<1 %
29	journal.lembagakita.org Internet Source	<1 %
30	jurnal.umt.ac.id Internet Source	<1 %
31	repository.mercubuana.ac.id Internet Source	<1 %
32	repository.uin-suska.ac.id Internet Source	<1 %
33	sistemasi.ftik.unisi.ac.id Internet Source	<1 %
34	www.stmik-budidarma.ac.id Internet Source	<1 %
35	moneyduck.com Internet Source	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On