

Analisis Sentimen Terhadap Ulasan Penggunaan Shopee Melalui Tweet Pada Twitter Menggunakan Algoritma Naïve Bayes

Hilmy Zhafran Muflih¹, Hafizh Dhery Al Assyam¹, Faisal Akbar Pangestu¹ & Mia Kamayani¹

¹Teknik Informatika Fakultas Teknologi Industri dan Informatika Universitas Muhammadiyah Prof.Dr.Hamka, Jakarta
Email: hilmyzhfrnm@gmail.com, dheryhafidz@gmail.com, 2003015215@uhamka.ac.id

Article Info

Received: 11 Juli 2023

Accepted: 28 September 2023

Published: 30 September 2023

ABSTRAK

Meningkatnya penggunaan internet dikalangan masyarakat karena menjadi sarana untuk melakukan berbagai aktivitas, salah satunya yaitu jual beli online atau dikenal sebagai *e-commerce*. Salah satu *e-commerce* terbesar yang ada di Indonesia yaitu shopee. Shopee menawarkan berbagai berbagai fitur untuk para pengguna. Banyaknya pengguna shopee membuat banyaknya respon yang diberikan kepada shopee, sehingga peneliti ingin melakukan proses analisis sentimen mengenai respon pengguna terhadap shopee, apakah respon para pengguna shopee bernilai negatif atau positif. Respon atau pendapat para pengguna shopee diambil dari *tweets* yang ada di aplikasi twitter. *Tweet* yang diketik dan ditulis dan dipublish oleh para pengguna twitter mengenai shopee. Pada penelitian ini peneliti menggunakan aplikasi *RapidMiner* untuk mengumpulkan data *tweets* para pengguna twitter dan untuk menerapkan algoritma *Naïve Bayes*. Peneliti mengumpulkan data mengenai shopee sebanyak 200 data dari twitter. Hasil yang didapat dari analisis sentimen menggunakan algoritma *Naïve Bayes* mendapat 78% sentimen negatif dan 22% sentimen positif dari 200 dataset. Proses Pengujian algoritma *Naïve Bayes* menggunakan *confussion matrix* memperoleh nilai akurasi yaitu 77,50%

Kata kunci : analisis sentimen, shopee, twitter, naïve bayes

ABSTRACT

The increasing use of the internet among the public is because it is a means to carry out various activities, one of which is buying and selling online or known as e-commerce. One of the largest e-commerce in Indonesia is Shopee. Shopee offers various features for its users. The large number of shopee users results in the large number of responses given to shopee, so the researcher wants to carry out a sentiment analysis process regarding user responses to shopee, whether the response of shopee users is negative or positive. The responses or opinions of Shopee users are taken from tweets in the Twitter application. Tweets typed and written and published by Twitter users about shopee. In this study, researchers used the RapidMiner application to collect tweets data from Twitter users and to apply the Naïve Bayes algorithm. The researcher collected 200 data regarding shopee from Twitter. The results obtained from sentiment analysis using the Naïve Bayes algorithm get 78% negative sentiment and 22% positive sentiment from 200 datasets. The process of testing the Naïve Bayes algorithm using the confusion matrix obtains an accuracy value of 77.50%.

Keywords : *sentiment analysis*, shopee, twitter, naïve bayes

1. PENDAHULUAN

Jumlah pengguna media sosial terus meningkat seiring berjalannya waktu, hal ini dikarenakan adanya kemudahan yang ditawarkan oleh media sosial kepada masyarakat

(Rahma & Delliana, 2022). Pada saat ini Indonesia mengalami peningkatan pengguna internet apabila dilakukan perbandingan di tahun sebelumnya (Hamid, 2020). Internet berperan sebagai sarana untuk mendapatkan informasi dan berkomunikasi, namun juga berfungsi

sebagai platform kegiatan pemasaran, iklan, serta transaksi jual beli barang dan jasa, seperti saat ini sedang ramainya jual beli secara online atau biasa dikenal dengan *e-commerce* (Yanti et al., 2023). Kehadiran *e-commerce* telah memberikan kemudahan kepada masyarakat dalam melakukan pembelian produk, karena pelanggan tidak perlu mengunjungi toko secara langsung (Kurnia, 2023). Salah satu platform belanja secara online populer di Indonesia adalah shopee (Hasugian et al., 2023). Orang-orang pada saat ini banyak yang menggunakan shopee untuk berbelanja secara online karena mudahnya dalam membeli barang tanpa harus berpergian. Shopee juga memberikan keamanan dan kemudahan dalam melakukan proses transaksi.

Akan tetapi shopee mungkin bisa memiliki banyak kekurangan, seperti sebagian masyarakat masih ragu terhadap berbelanja secara online karena sedang ramainya barang yang dibeli merupakan barang palsu serta kasus barang yang diterima pembeli tidak sesuai dengan barang yang dibeli. Khususnya masyarakat masih ragu untuk berbelanja barang elektronik seperti produk smartphone. Jika mengalami ketidakpuasan yang dialami oleh para pengguna shopee dan kurangnya terhadap pelayanan dan keamanan khususnya untuk keaslian barang yang dijual. Hal yang dilakukan untuk meningkatkan kualitas layanan, maka perlu adanya analisis sentimen mengenai opini masyarakat terhadap penggunaan shopee. Twitter digunakan untuk mendapatkan respon masyarakat yang berada di sosial media (Alfandi Safira & Hasan, 2023). Twitter merupakan salah satu sosial media yang banyak dikenal dan berguna untuk menjadi wadah masyarakat dalam berkomunikasi (Fikri et al., 2020). Semua opini, kritik, dan saran yang diberikan pengguna media sosial twitter bisa dikumpulkan menjadi satu yaitu sentimen (Afandi et al., 2022). Kumpulan dari sebuah opini yang telah didapat tersebut akan digunakan untuk melakukan proses analisis sentimen.

Terdapat berbagai metode untuk menganalisis pendapat orang yang berada di media sosial, salah satunya adalah metode analisis sentimen yang mampu menganalisis sebuah pendapat atau opini, sehingga dapat memberikan penilaian atas pendapat yang telah dianalisis sebelumnya. Analisis sentimen merupakan metode yang digunakan oleh peneliti untuk melakukan pengolahan data opini dan melakukan pemahaman secara otomatis untuk mendapatkan gambaran terhadap sentimen yang ada pada opini (Sari & Wibowo, 2019). Analisis sentimen berfungsi sebagai metode untuk memperoleh pemahaman mengenai polaritas data yang telah diperoleh dari internet (Laurensz & Eko Sedyono, 2021). Analisis sentimen digunakan untuk menganalisis popularitas opini atau sentimen yang telah dikumpulkan, sehingga dapat dilakukannya proses prediksi apakah sentimen tersebut positif atau negatif. Peneliti melakukan analisis sentimen untuk mengetahui tingkat kepuasan pengguna. Dengan menggunakan metode analisis sentimen, kesimpulan dapat diambil dari data yang tidak terstruktur (Alghifari et al., 2022). Pemrosesan bahasa alami (NLP) adalah salah satu cabang analisis sentimen yang digunakan oleh para peneliti saat ini untuk mengenali dan mengklasifikasikan konten dataset yang berupa pendapat atau opini dalam bentuk teks terkait peristiwa atau kejadian

yang dapat memiliki positif, atau negatif (Fikri et al., 2020). Pada masa sekarang, distribusi dan akses informasi menjadi sangat mudah berkat kemajuan pesat internet. Hal ini juga dipengaruhi oleh munculnya media sosial seperti Twitter, yang memberikan pengguna fasilitas untuk dengan mudah membagikan dan mendapatkan informasi, serta memberikan pendapat terhadap topik, isu, atau peristiwa yang sedang menjadi sorotan saat ini (Ardiani et al., 2020). Data mining merupakan disiplin ilmu yang sangat relevan dan tepat untuk diterapkan dalam kasus seperti ini (Auliadaya et al., 2019). Proses analisis teks menghasilkan output berupa file yang berguna untuk mewakili isi dokumen dengan mencari istilah yang relevan dan memperoleh pemahaman tentang hubungannya (Hasan & Dwijayanti, 2021).

Metode-metode seperti Naïve Bayes Classifier, Random Forest, dan sebagainya merupakan metode yang umum digunakan dalam penerapan analisis sentimen (Arsi et al., 2021). Naïve Bayes Classifier adalah metode Machine Learning yang tidak terikat oleh aturan, yang berarti memiliki kebebasan dalam penggunaannya. Metode Naïve Bayes menggunakan konsep probabilitas untuk menentukan kemungkinan suatu klasifikasi dengan melihat frekuensi kemunculan dalam data latih (Legiawati et al., 2022). Naïve Bayes memiliki fungsi untuk melakukan klasifikasi pada data dan menghitung nilai akurasi dari data tersebut (Arsi et al., 2021). Konsep dasar Naïve Bayes menggunakan Teorema Bayes, yang merupakan sebuah teorema dalam statistik yang digunakan untuk menghitung probabilitas suatu peristiwa berdasarkan informasi yang ada. Algoritma Naïve Bayes digunakan untuk mencari dan menghitung probabilitas tertinggi dalam rangka mengklasifikasikan data uji ke dalam kategori yang sesuai (Tangkelayuk, 2022). Metode ini memanfaatkan klasifikasi Naïve Bayesian yang menggunakan teorema probabilitas, fungsi data mining, dan juga teorema Bayes. Penerapan Multinomial Naïve Bayes adalah untuk menghitung frekuensi kemunculan setiap token atau kata yang terdapat dalam suatu dokumen (Winahyu & Suharjo, 2021). Data akan diuji menggunakan aplikasi *RapidMiner* (Nas, 2021). *RapidMiner* adalah sebuah perangkat lunak yang digunakan untuk melakukan pengolahan data dengan menerapkan algoritma dan prinsip data mining (Irnanda et al., 2022).

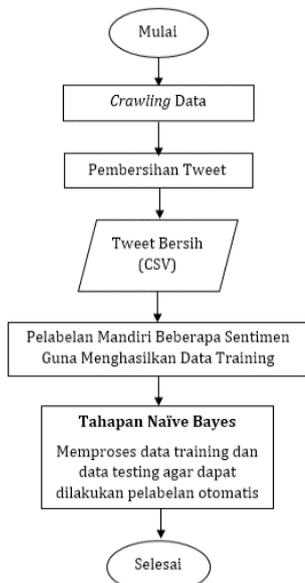
Tujuan dari penelitian ini yaitu untuk mencari tau respon masyarakat yang menggunakan shopee terhadap fitur dan pelayanan yang ada di shopee apakah masyarakat sudah merasa puas atau belum terhadap fitur dan pelayanan yang ditawarkan. Jika respon masyarakat setelah menggunakan shopee merasa kurang puas, maka shopee perlu meningkatkan kualitas dalam fitur dan pelayanan yang diberikan agar shopee bisa meningkatkan minat dan respon masyarakat menjadi terpuaskan. Pada penelitian ini berbeda dengan penelitian terdahulu, peneliti menggunakan aplikasi *RapidMiner* dengan menggunakan metode algoritma *Naïve Bayes* untuk menganalisis dataset *tweets* mengenai respon pengguna shopee yang telah dikumpulkan. Media sosial yaitu twitter digunakan oleh masyarakat untuk membagikan foto, video, dan memberikan pendapat mengenai topik terkini. *RapidMiner* yang digunakan oleh peneliti memiliki berbagai operator machine learning salah satunya yaitu machine learning dengan menggunakan algoritma *Naïve*

Bayes. RapidMiner juga digunakan untuk mengumpulkan dataset dalam bentuk tweet. Penelitian ini menggunakan aplikasi RapidMiner untuk mempermudah, efektif, dan efisien dalam pengerjaannya.

Penelitian analisis terkait sentimen telah banyak dilakukan sebelumnya. Penelitian terdahulu membahas tentang pembelajaran online dengan menggunakan algoritma Naive Bayes untuk diolah datanya dan pengumpulan dataset yaitu berupa tweet dari twitter, Hasil yang didapat berupa nilai akurasi yaitu sebesar 76,39% (Zusrotun et al., 2022). Berikutnya penelitian terdahulu yang lainnya membahas mengenai analisis sentimen terhadap program kampus merdeka, metode yang digunakan pada penelitian yaitu menggunakan metode algoritma Naive Bayes untuk mengolah dataset dalam bentuk tweet dari twitter, hasil yang didapat yaitu nilai akurasi sebesar 60% (Rahayu et al., 2022).

2. METODE PENELITIAN

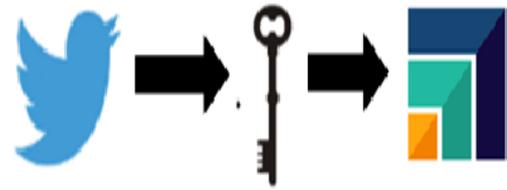
Penelitian yang dilakukan kali ini menggunakan metode algoritma Naive Bayes Classifier dan menerapkan pada aplikasi RapidMiner. Algoritma Naive Bayes Classifier berguna untuk melakukan proses analisis suatu kalimat berupa opini apakah bernilai positif atau negatif. alur proses penelitian Berikut Gambar 1 merupakan alur penelitian yang akan dilakukan oleh peneliti dalam bentuk flowchart.



Gambar 1. Tahapan Proses

2.1. CRAWLING DATA

Pada tahap pertama dilakukannya proses crawling data terlebih dahulu menggunakan RapidMiner. Pada RapidMiner peneliti mengkoneksikan RapidMiner dengan twitter terlebih dahulu dengan request access token untuk mendapatkan API yang berfungsi untuk menghubungkan RapidMiner dengan twitter. Berikut gambar 2 ilustrasi untuk mendapatkan API yang digunakan untuk mengkoneksikan RapidMiner dengan twitter.



Gambar 2. Ilustrasi mendapatkan API untuk mengoneksikan RapidMiner dengan twitter

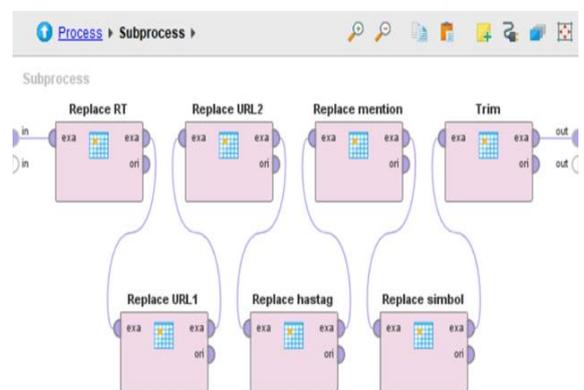
Ketika RapidMiner sudah terkoneksi dengan twitter, peneliti dapat melakukan proses crawling dengan menggunakan operator search untuk mencari data tentang shopee yang ditujukan pada akun resmi twitter shopee yaitu @ShopeeID, dalam bentuk tweet. Data tweet yang telah didapatkan akan disimpan dalam bentuk format csv dengan operator write csv untuk dilakukan ke dalam proses selanjutnya. Berikut Gambar 3 merupakan proses crawling dataset.



Gambar 3. proses crawling dataset.

2.2. PEMBERSIHAN TWEET

Tahap berikutnya adalah melakukan pembersihan tweet yang telah diambil dan disimpan dalam format file csv. Tweet tersebut perlu dibersihkan karena awalnya masih mengandung simbol-simbol seperti [!~@#&*'";:~.,-]. Selain itu, beberapa tweet juga mengandung link dan kata "RT" (Retweet) dalam kalimatnya. Hal ini dapat mengganggu proses analisis sentimen yang dilakukan oleh peneliti. Oleh karena itu, diperlukan pembersihan kalimat pada tweet agar kalimatnya menjadi bersih, tanpa adanya simbol-simbol, kata "RT", link, dan sebagainya. Pembersihan ini dilakukan menggunakan alat yang disebut RapidMiner.



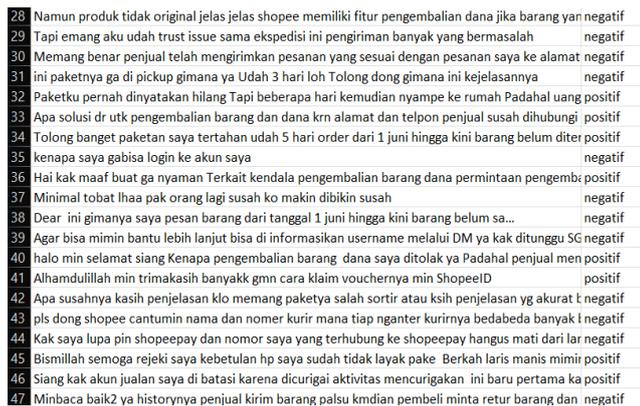
Gambar 4. Tahap Cleansing

Setelah proses pembersihan selesai, tweet yang sudah dibersihkan akan disimpan dalam format file csv. Anda dapat melihat ilustrasi proses pembersihan tweet yang

dilakukan dalam RapidMiner pada Gambar 4. Kemudian dilanjut dengan menghapus data yang kosong menggunakan

2.3. PELABELAN DATASET

Tahapan selanjutnya setelah *tweet* bersih dari kata kata yang tidak diertukan, simbol, dan link telah selesai yaitu peneliti melakukan pelabelan dataset/*tweet* yang telah didapat. Pada proses ini peneliti memberikan label pada *tweet*/sentimen, apakah suatu sentimen bersifat negatif atau positif. Dilakukannya proses ini bertujuan untuk memberikan *training* ke algoritma *Naïve Bayes* agar algoritma *Naïve Bayes* dapat melakukan proses pemberian label kepada sentimen secara otomatis pada tahap selanjutnya. Pada penelitian ini peneliti hanya melabeli 140 dari 200 data yang telah didapat. Data yang tidak diberi label oleh peneliti akan dilakukan proses pemberian label secara otomatis menggunakan algoritma *Naïve Bayes*. Semakin banyak data yang diberi label oleh peneliti, maka akan semakin akurat hasil dari pemberian label secara otomatis oleh algoritma *Naïve Bayes*. Karena pada saat melakukan pemberian label secara otomatis, algoritma *Naïve Bayes* terlebih dahulu melakukan training dengan data yang telah diberi sentimen oleh peneliti. Berikut Gambar 5 yaitu proses pelabelan dataset.



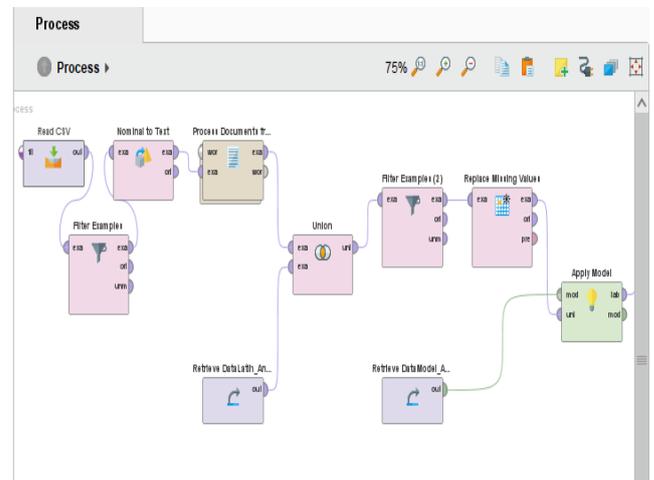
Gambar 5. Proses Pelabelan Dataset

Pada proses ini, peneliti akan memasukan data yang telah diberi label oleh peneliti ke dalam *RapidMiner* karena peneliti menggunakan *RapidMiner* untuk menerapkan algoritma *Naïve Bayes*. *Naïve Bayes* akan mentraining dirinya untuk memahami serta menelusuri semua pola data yang telah peneliti beri label pada sentimen.

3. TAHAPAN NAÏVE BAYES

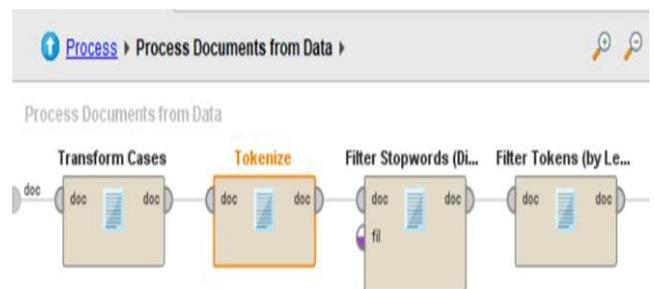
Tahapan selanjutnya yaitu melakukan proses terhadap data yang telah peneliti beri label ke dalam aplikasi *RapidMiner*. Berikut merupakan Gambar 3 mengenai proses pemberian label pada dataset oleh peneliti. Oleh karena itu peneliti telah menjelaskan sebelumnya bahwa banyaknya data yang diberi label oleh peneliti, maka akan lebih akurat *Naïve Bayes* dalam memberikan label secara otomatis. Setelah algoritma *Naïve Bayes* memahami polaritas, maka algoritma *Naïve Bayes* akan menghasilkan model klasifikasi yang nantinya berfungsi untuk melakukan proses pemberian label terhadap sentimen yang bersifat positif atau negatif pada *tweet* yang belum diberi label sentimen oleh peneliti.

Kemudian, setelah algoritma *Naïve Bayes* sudah memahami polaritas dari sebuah sentimen, maka peneliti akan menyimpan model *Naïve Baye* yang telah dihasilkan ke dalam operator *store*, seperti pada gambar 6.



Gambar 6. Pemberian Training ke Algoritma Naïve Bayes

Dapat dilihat pada Gambar 3 bahwa terdapat operator data latih dan data model, kedua operator akan digunakan peneliti untuk proses selanjutnya. Pada proses ini peneliti menggunakan operator *Process document* untuk menampung parameter *Transform Cases*, *tokenize*, *Filter Stopword* dan *Filters token by length* di operator *Process Document From Data*. Berikut Gambar 7 merupakan isi didalam *Process Document*.



Gambar 7. Process Document

- A. *Transform Cases* digunakan untuk mengubah semua huruf yang ada di dataset yang telah didapat dari huruf kapital menjadi huruf tidak kapital. Berikut Tabel 1 merupakan contoh data dari tweet sebelum dan sesudah proses *Transform Cases*.

Sebelum		Sesudah	
napasi belanja di shopee tiap pembayaran tu langsung kek loading trus ga selesai pdhl sinyalnya ga knp np anjing aplikasi lu sampah bgt	ShopeeID	napasih belanja di shopee tiap pembayaran tu langsung kek loding trus ga selesai pdhl sinyalnya gak knp np anjing aplikasi lu sampah bgt	shopeeid
Alhamdulillah	min	alhamdulillah	min
terimakasih banyak untuk vouchernya	min ShopeeID	terimakasih banyak untuk vouchernya	min shopeeid
Aku juga terakhir dm cs disuruh isi yang itu sambil coba uninstall reinstall juga		aku juga terakhir dm cs disuruh isi yang itu sambil coba uninstall reinstall juga	

huhuw semoga aja bisa cepet normal Makasih banyak ya kak infonya ShopeeID	huhuw semoga aja bisa cepet normal makasih banyak ya kak infonya shopeeid
Saya mengajukan pengembalian barang dana karena barang yg saya terima itu PALSU Bukan pesanan salah tidak lengkap Tetap kenapa anda menghighlight pesanan saya lengkap dan produk nya dikirim sesuai dengan alamat saya Saya tegaskan lagi saya komplain barang palsu ShopeeID	saya mengajukan pengembalian barang dana karena barang yg saya terima itu palsu bukan pesanan salah tidak lengkap Tetap kenapa anda menghighlight pesanan saya lengkap dan produk nya dikirim sesuai dengan alamat saya Saya tegaskan lagi saya komplain barang palsu shopeeid
Ga peka kita ganti aplikasi aja nah setiap sabtu minggu nahhh Makan shopeefood Kalo diskon ongkirnya di tambah kan bisa tiap hari ShopeeID	Ga peka kita ganti aplikasi aja nah setiap sabtu minggu nahhh Makan shopeefood Kalo diskon ongkirnya di tambah kan bisa tiap hari shopeeid

B. *Tokenize* digunakan peneliti untuk memisahkan kata-kata dalam sebuah *tweet* atau setiap kata dipecah menjadi atribut tersendiri. Berikut Tabel 2 merupakan contoh dari data tweet sebelum dan sesudah proses *Tokenize*.

Sebelum	Sesudah
napasi, belanja di shopee tiap pembayaran tu langsung kek loading trus ga selesai pdhl sinyalnya ga knp np anjing aplikasi lu sampah bgt shopeeid	napasih, belanja, di, shopee, tiap, pembayaran, tu, langsung, kek, loding, trus, ga, selesai, pdhl, sinyalnya, gak, knp, np, anjing, aplikasi, lu, sampah, bgt, shopeeid
alhamdulillah min terimakasih banyak untuk vouchernya min shopeeid	alhamdulillah, min, terimakasih, banyak, untuk, vouchernya, min, shopeeid,
aku juga terakhir dm cs disuruh isi yang itu sambil coba uninstall reinstall juga huhuw semoga aja bisa cepet normal Makasih banyak ya kak infonyaa shopeeid	aku, juga, terakhir, dm, cs, disuruh, isi, yang, itu, sambil, coba, uninstall, reinstall, juga, huhuw, semoga, aja, bisa, cepet, normal, makasih, banyak, ya, kak, infonyaa, shopeeid
saya mengajukan pengembalian barang dana karena barang yg saya terima itu palsu bukan pesanan salah tidak lengkap Tetap kenapa anda menghighlight pesanan saya lengkap dan produk nya dikirim sesuai dengan alamat saya Saya tegaskan lagi saya komplain barang palsu shopeeid	saya, mengajukan, pengembalian, barang, dana, karena, barang, yg, saya, terima, itu, palsu, bukan, pesanan, salah, tidak, lengkap, tetap, kenapa, anda, menghighlight, pesanan, saya, lengkap, dan, produk, nya, dikirim, sesuai, dengan, alamat, saya, saya, tegaskan, lagi, saya, komplain, barang, palsu, shopeeid
ga peka kita ganti aplikasi aja nah setiap sabtu minggu nahhh makan shopeefood kalo diskon ongkirnya di tambah kan bisa tiap hari shopeeid	ga, peka, kita, ganti, aplikasi, aja, nah, setiap, sabtu, minggu, nahhh, makan, shopeefood, kalo, diskon, ongkirnya, di, tambah, kan, bisa, tiap, hari, shopeeid

C. *Filter Stopword* digunakan oleh peneliti untuk secara otomatis menghapus kata-kata yang tidak relevan dengan menggunakan *Stopwords* Bahasa Indonesia yang diperoleh dari situs web www.kaggle.com. Berikut Tabel 3 merupakan contoh dari data tweet sebelum dan sesudah proses *Filter Stopwords*.

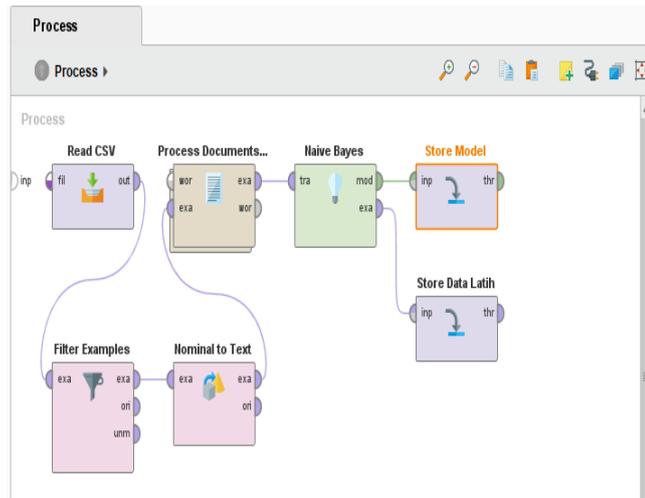
Sebelum	Sesudah
napasih, belanja, di, shopee, tiap, pembayaran, tu, langsung, kek, loding, trus, ga, selesai, pdhl, sinyalnya, gak, knp, np, anjing, aplikasi, lu, sampah, bgt, shopeeid	napasih, belanja, di, tiap, pembayaran, langsung, trus, ga, selesai, sinyalnya, anjing, aplikasi, lu, sampah,
alhamdulillah, min, terimakasih, banyak, untuk, vouchernya, min, shopeeid,	terimakasih, banyak, untuk
aku, juga, terakhir, dm, cs, disuruh, isi, yang, itu, sambil, coba, uninstall, reinstall, juga, huhuw, semoga, aja, bisa, cepet, normal, makasih, banyak, ya, kak, infonya, shopeeid	aku, juga, terakhir, disuruh, isi, yang, itu, sambil, coba, juga, semoga, aja, bisa, cepet, normal, makasih, banyak, ya, infonya
saya, mengajukan, pengembalian, barang, dana, karena, barang, yg, saya, terima, itu, palsu, bukan, pesanan, salah, tidak, lengkap, tetap, kenapa, anda, pesanan, saya, lengkap, dan, produk, nya, dikirim, sesuai, dengan, alamat, saya, saya, tegaskan, lagi, saya, komplain, barang, palsu, shopeeid	saya, mengajukan, pengembalian, barang, dana, karena, barang, saya, terima, itu, palsu, bukan, pesanan, salah, tidak, lengkap, tetap, kenapa, anda, pesanan, saya, lengkap, dan, produk, nya, dikirim, sesuai, dengan, alamat, saya, saya, tegaskan, lagi, saya, komplain, barang, palsu,
ga, peka, kita, ganti, aplikasi, aja, nah, setiap, sabtu, minggu, nahhh, Makan, shopeefood, kalo, diskon, ongkirnya, di, tambah, kan, bisa, tiap, hari, shopeeid	peka, kita, ganti, aplikasi, aja, setiap, sabtu, minggu, makan, kalo, diskon, ongkirnya, di, tambah, kan, bisa, tiap, hari

D. *Filter Token by Length* digunakan untuk menghilangkan kata yang memiliki jumlah huruf yang telah ditentukan. Peneliti telah menentukan minimal jumlah huruf yaitu sebanyak 4 huruf. Berikut Tabel 4 merupakan contoh dari tweet sebelum dan sesudah proses *Filter Token by Length*.

Sebelum	Sesudah
napasih, belanja, di, tiap, pembayaran, langsung, trus, ga, selesai, sinyalnya, anjing, aplikasi, lu, sampah,	napasih, belanja, di, tiap, pembayaran, langsung, trus, ga, selesai, sinyalnya, anjing, aplikasi, lu, sampah,
terimakasih, banyak, untuk	terimakasih, banyak, untuk
aku, juga, terakhir, disuruh, isi, yang, itu, sambil, coba, juga, semoga, aja, bisa, cepet, normal, makasih, banyak, ya, infonya	aku, juga, terakhir, disuruh, isi, yang, itu, sambil, coba, juga, semoga, aja, bisa, cepet, normal, makasih, banyak, ya, infonya
saya, mengajukan, pengembalian, barang, dana, karena, barang, saya, terima, itu, palsu, bukan, pesanan, salah, tidak, lengkap, tetap, kenapa, anda, pesanan, saya, lengkap, dan, produk, nya, dikirim, sesuai, dengan, alamat, saya, saya, tegaskan, lagi, saya, komplain, barang, palsu,	saya, mengajukan, pengembalian, barang, dana, karena, barang, saya, terima, itu, palsu, bukan, pesanan, salah, tidak, lengkap, tetap, kenapa, anda, pesanan, saya, lengkap, dan, produk, nya, dikirim, sesuai, dengan, alamat, saya, saya, tegaskan, lagi, saya, komplain, barang, palsu,
peka, kita, ganti, aplikasi, aja, setiap, sabtu, minggu, makan, kalo, diskon, ongkirnya, di, tambah, kan, bisa, tiap, hari	peka, kita, ganti, aplikasi, aja, setiap, sabtu, minggu, makan, kalo, diskon, ongkirnya, di, tambah, kan, bisa, tiap, hari

3.1. PELABELAN OTOMATIS DENGAN ALGORITMA NAÏVE BAYES

Pada tahap terakhir akan dilakukan proses pemberian label tweet/sentimen secara otomatis oleh algoritma *Naïve Bayes* yang telah *training* dengan dataset yang telah diberi label sentimen oleh peneliti. Berikut Gambar 8 merupakan proses pemberian label otomatis dengan algoritma *Naïve Bayes* pada aplikasi *RapidMiner*.



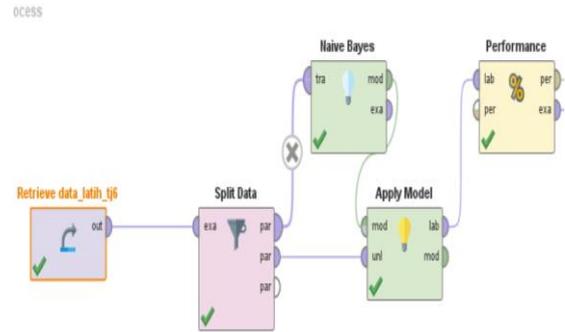
Gambar 8. proses pemberian label sentimen secara otomatis

Proses dari membaca file CSV hingga memproses dokumen dari data (tahap 1 hingga 4) disebut sebagai pengolahan data uji. Pada kotak di bawah ini terdapat operator "Retrieve data latih". Operator Retrieve merupakan langkah yang telah dilakukan oleh peneliti pada proses sebelumnya (terlihat pada Gambar 4). Kemudian, peneliti menggabungkan data uji tersebut dengan operator Retrieve karena isi dari data uji dan data latih berbeda. Oleh karena itu, diperlukan penggabungan isi antara data uji dan data latih menggunakan operator Union. Tabel 1 di bawah ini menjelaskan secara ilustratif bagaimana cara menggabungkan data latih dan data uji agar memiliki isi yang sama. Berikut gambar 9 dibawah ini merupakan hasil analisis sentimen berdasarkan tweet dengan menggunakan *RapidMiner*.

Sentimen	text
negatif	paket dinyatakan hilang ditahan lokasi sortir gas tolong bantu banget
negatif	napsai belanja shopee pembayaran langsung loading trus selesai pdhl akhirnya anjing aplikasi sampah shopeed
negatif	pake anteraja bermasalah terus kalo liang kebobol kapok ekspedisi
negatif	aduh kemarin kemas catusu mulu paket hilang nartikan paket
negatif	sampe kurirnya main
negatif	hajar pengalaman dibidang ekspedisi bisnis disana bhuheuehue
negatif	kemaspa paketlalu picup gimana udah hari tolong gimana kejelasannya
positif	athamdulillah untung penerima paket menghugungi sosial mediau ambil paketya
negatif	ngukatin hilangin ekspedisi toko ngirim
negatif	dibatalin toko kma barang stuck dipihak ekspedisi
negatif	samaspe minggu paket stuck dsatu daerahpa komplan dinyatakan hilang paket
negatif	paket nyampe kemas besok dibutuhkan pantas nari parah banget shopee spress nyampe balikin
negatif	gatau kemarin sempat admin shopee care datadatu akun sampe kajelasan
negatif	hahaha penyakit sicepat kurirnya sengkaja ngpasasin deadline paketan reguler barang posisi barang elektronik barang line haul udara barang line haul darat...
negatif	pesan paket trus status pengiriman udah diterima neima mint tolong shopeeud
negatif	pake kurir udah kapok
negatif	lyaaa donglot banget pake anteraja

Gambar 9. Hasil Analisis sentiment

3.2. PENGUJIAN EVALUASI



Gambar 10. Proses Evaluasi dengan Confussion Matrix

Ketika Proses pelabelan secara otomatis oleh algoritma *Naïve Bayes* telah selesai, maka akan dilanjutkan dengan melakukan proses pengujian terhadap algoritma *Naïve Bayes* dengan menggunakan *confusion matrix* pada aplikasi *RapidMiner*. Pada saat proses pengujian dengan *confusion matrix* peneliti menggunakan data testing untuk dilakukan pengujian pada *confusion matrix*. Pada operator split data peneliti menggunakan perbandingan 70:30 untuk data *training* dan data *testing*. Hasil yang diperoleh dari dilakukannya pengujian dengan *confusion matrix* yaitu nilai *recall*, *precision*, dan *accuracy*. Berikut tabel 5 dibawah ini adalah hasil dari pengujian *confusion matrix*.

Tabel 5. Confussion Matrix

Accuracy: 77.50%

	True Positif	True Negatif	Class Precision
Pred. Positif	20	3	86.96%
Pred. Negatif	6	11	64.71%
Class recall	76.92%	78.57%	

Dari hasil yang didapat dari pengujian evaluasi dari algoritma *Naïve bayes* dengan menggunakan *confusion matrix* didapat hasil yaitu nilai *accuracy* yang diperoleh sebesar 77,50%. Setelah itu proses evaluasi yang diperoleh sentimen positif memperoleh *precision* sebesar 86,96% dan *recall* sebesar 76,92%. Sedangkan untuk sentimen yang bersifat negatif memperoleh *precision* sebesar 64,71% dan *recall* sebesar 78,57%.

4. KESIMPULAN

Dapat diambil kesimpulan pada penelitian ini bahwa algoritma *Naïve Bayes* bisa melakukan proses prediksi terhadap sentimen secara otomatis tentang ulasan penggunaan shopee oleh masyarakat, baik nilai sentimen bernilai positif atau negatif. Dataset yang melalui proses algoritma *Naïve Bayes*, maka akan mendapatkan 156 sentien negatif dan 46 sentimen positif dari 200 dataset yang telah melalui proses analisis sentimen. Banyaknya jumlah sentimen negatif yang didapat, maka dapat disimpulkan bahwa shopee masih memiliki banyak kekurangan pada pelayanannya sehingga respon para pengguna shopee banyak yang berpendapat negatif terhadap shopee. Akurasi dari algoritma *Naïve Bayes* pada penelitian ini yaitu sebesar 77,50% dan pada saat pengujian evaluasi pada sentimen negatif memperoleh nilai *precision* sebesar 64,71% dan

recall sebesar 78,57%. Sedangkan untuk sentimen positif memperoleh *precision* sebesar 86,96% dan *recall* sebesar 76,92%. Kekurangan pada penelitian ini ialah jumlah dari dataset yang digunakan. Peneliti hanya menggunakan 200 dataset untuk melakukan proses algoritma *Naïve Bayes*, sehingga hasil yang diperoleh masih bisa ditingkatkan. Saran penelitian selanjutnya, saat proses crawling data diharapkan memperoleh data sebanyak banyaknya karena semakin banyak data yang dipakai maka hasil yang didapat akan lebih baik dan sesuai dengan yang diharapkan.

DAFTAR PUSTAKA

- Afandi, I. R., Hasan, F. N., Rizki, A. A., Pratiwi, N., & Halim, Z. (2022). Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes. *Jurnal Linguistik Komputasional*, 5(2), 63–70. <https://doi.org/https://doi.org/10.26418/jlk.v5i2.107>
- Alfandi Safira, & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONASI: Jurnal Sistem Informasi*, 5(1), 59–70. <https://doi.org/10.31849/zn.v5i1.12856>
- Alghifari, D. R., Edi, M., & Firmansyah, L. (2022). Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. *Jurnal Manajemen Informatika (JAMIKA)*, 12(2), 89–99. <https://doi.org/10.34010/jamika.v12i2.7764>
- Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(2), 183. <https://doi.org/10.26418/justin.v8i2.36776>
- Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 231–237. <https://doi.org/10.29207/resti.v5i2.2698>
- Auliadaya, A. S., Rizki, M., & Azhary, M. A. F. (2019). Analisa Pola Pembelian Produk Pada Toko Cimahi-Apriori. *Manajemen Informatika*, 9(1), 58–69.
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Hamid, A. I. (2020). Analisis Semiotik Meme Anies Baswedan Banjir Jakarta. *DIALEKTIKA KOMUNIKA: Jurnal Kajian Komunikasi Dan Pembangunan Daerah*, 8(1), 36–45. <https://doi.org/10.33592/dk.v8i1.554>
- Hasan, F. N., & Dwijayanti, M. (2021). Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier. *Jurnal Linguistik Komputasional*, 4(2), 52–58. <https://doi.org/https://doi.org/10.26418/jlk.v4i2.61>
- Hasugian, A. H., Fakhriza, M., & Zukhoiriyah, D. (2023). Volume 6; Nomor 1. *Januari*, 6, 98–107. <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- Irnanda, K. F., Windarto, A. P., & Damanik, I. S. (2022). Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner. *JURIKOM (Jurnal Riset Komputer)*, 9(1), 122. <https://doi.org/10.30865/jurikom.v9i1.3836>
- Kurnia, W. (2023). *Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent*. 4(1), 138–143.
- Laurensz, B., & Eko Sedyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Legiawati, N., Hermanto, T. I., & Ramadhan, Y. R. (2022). Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 930. <https://doi.org/10.30865/jurikom.v9i4.4629>
- Nas, C. (2021). Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5. *Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 131–145. <https://doi.org/10.34010/jamika.v11i2.5506>
- Rahayu, I. P., Fauzi, A., & Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. *Jurnal Sistem Komputer Dan Informatika (JSON) Hal: 296–, 301(2)*, 25–38. <https://doi.org/http://dx.doi.org/10.30865/json.v4i2.5381>
- Rahma, A., & Delliana, S. (2022). Pengaruh Konten Youtube Ria Ricis Terhadap Agresi Verbal Netizen. *Translitera: Jurnal Kajian ...*, 11(1), 106–119. <http://ejournal.unisablitar.ac.id/index.php/translitera/article/view/1668%0Ahttps://ejournal.unisablitar.ac.id/index.php/translitera/article/download/1668/1244>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686. <https://doi.org/https://doi.org/10.24176/simet.v10i2.3487>
- Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>
- Winahyu, J., & Suharjo, I. (2021). Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, 10(2), 206. <https://doi.org/10.23887/karmapati.v10i2.36609>
- Yanti, S. D., Astuti, S., & Safitri, C. (2023). Pengaruh Pengalaman Belanja Online Dan Kepercayaan Terhadap Minat Beli Ulang Di Tiktok Shop (Studi Kasus Mahasiswa Fkip Uhamka 2018). *Jurnal EMT*

KITA, 7(1), 47–61.
<https://doi.org/10.35870/emt.v7i1.728>
Zusrotun, O. P., Murti, A. C., & Fiati, R. (2022). *Sentimen Analisis Belajar Online di Twitter Menggunakan Naïve Bayes* *Jurnal Nasional Pendidikan Teknik*

Informatika : JANAPATI | 311. 11, 310–320.
<https://doi.org/https://doi.org/10.23887/janapati.v11i3.49160>