

ANALISIS SENTIMEN *TWEET* PENGGUNA *E-COMMERCE* DENGAN MENGGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES*

Anastasya Nadia Puspitasari*¹⁾, Yulian Findawati²⁾, Yunianita Rahmawati³⁾

1. Universitas Muhammadiyah Sidoarjo, Indonesia
2. Universitas Muhammadiyah Sidoarjo, Indonesia
3. Universitas Muhammadiyah Sidoarjo, Indonesia

Article Info

Kata Kunci: Analisis Sentimen; *e-commerce*; Klasifikasi; *Naïve Bayes*

Keywords: *e-commerce*; Classification; *Naïve Bayes*; Sentiment Analysis

Article history:

Received 14 June 2024

Revised 24 July 2024

Accepted 8 August 2024

Available online 1 September 2024

DOI :

<https://doi.org/10.29100/jipi.v9i3.4939>

* Corresponding author.

Corresponding Author

E-mail address:

anastasyanadiap@gmail.com

ABSTRAK

Penelitian ini menganalisis sentimen pengguna yang diungkapkan di jejaring sosial, khususnya Twitter, dalam konteks *e-commerce* menggunakan metode klasifikasi *Naïve Bayes*. Pesatnya pertumbuhan *e-commerce* telah meningkatkan konten buatan pengguna yang mencerminkan opini dan pengalaman berbeda. Dengan menggunakan algoritma *Naive Bayes*, tujuan dari penelitian ini adalah untuk mengklasifikasikan dan memahami opini-opini yang terdapat dalam *tweet*. Emosi ini diklasifikasikan ke dalam kategori *Positive* dan *Negative*. Dalam penelitian ini, pikiran dan perasaan yang diungkapkan pengguna melalui *tweet* dicari, dibaca dan diklasifikasikan menggunakan klasifikasi *Naive Bayes*. Dalam penelitian ini, peneliti menunjukkan cara efektif melakukan segmentasi dan mengklasifikasikan opini *tweet* pengguna *e-commerce* menggunakan metode klasifikasi *Naive Bayes*. Menganalisis konten buatan pengguna memberikan wawasan berharga tentang informasi konsumen. Dari nilai rata-rata tersebut dihasilkan akurasi, presisi dan recall. Dapat disimpulkan bahwa rata-rata akurasi, presisi, dan recall data sebesar 92,00% untuk data akurat, 90,35% untuk data benar, dan 100% untuk data recall.

ABSTRACT

This research analyzes user sentiment expressed on social networks, especially Twitter, in the context of *e-commerce* using the *Naïve Bayes* classification method. The rapid growth of *e-commerce* has given rise to user-generated content that reflects different opinions and experiences. By using the *Naïve Bayes* algorithm, the aim of this research is to classify and understand the opinions contained in tweets. These emotions are classified into positive and negative categories. In this research, thoughts and feelings expressed by users through tweets are searched, read and classified using *Naïve Bayes* classification. In this research, researchers show how to effectively segment and classify *e-commerce* users' tweet opinions using the *Naïve Bayes* classification method. Analyzing user-generated content provides valuable insights into consumer information. From this average value, accuracy, precision and recall are produced. It can be concluded that the average accuracy, precision and recall of data is 92.00% for accurate data, 90.35% for correct data and 100% for recall data.

I. PENDAHULUAN

PERJALANAN *e-commerce* di Indonesia sudah dimulai sejak awal tahun 2000-an, ketika pasar online dan situs periklanan mulai bermunculan. *e-commerce* (perdagangan elektronik) merupakan cabang dari e-bisnis (bisnis elektronik) yang mengaitkan pembelian serta penjualan benda serta jasa lewat Internet. *e-commerce* pula mencakup kegiatan yang menunjang transaksi tersebut, semacam Periklanan, Pemasaran, Layanan Pelanggan, Keamanan, Pengiriman serta Pembayaran. *e-commerce* mengganti seluruh kegiatan pemasaran sekaligus kurangi bayaran operasional kegiatan perdagangan [1]. Namun, sekitar tahun 2020 *e-commerce* benar-benar melesit dengan meningkatnya penetrasi internet, adopsi *smartphone*, dan munculnya platform media sosial. Sekitar 138,09 juta pengguna *e-commerce* aktif pada tahun itu dan terus meningkat sampai pada tahun 2023 mencapai 138,09 juta pengguna menurut website Data Indonesia. Faktor-faktor tersebut menjadi lahan subur bagi tumbuhnya belanja online dan transaksi digital. Pemerintah Indonesia telah memainkan peran penting dalam mempromosikan pengembangan *e-commerce* dengan menerapkan kebijakan dan regulasi yang mendukung seperti

rencana aksi *e-commerce* dan strategi *e-commerce* nasional bertujuan untuk pertumbuhan sektor ekonomi dan menjawab tantangan terkait logistik, sistem pembayaran, dan perlindungan konsumen.

Twitter adalah salah satu platform media sosial yang paling umum digunakan untuk menyampaikan pendapat. *Twitter* adalah sebuah platform media sosial yang memungkinkan pengguna untuk bebas berkomentar atau menuliskan pendapat tentang berbagai hal[2].

Twitter memberikan pengguna kemampuan untuk mengirim pesan secara langsung pada saat itu juga. Melalui *Twitter*, opini pelanggan dapat membantu dalam menentukan situs *e-commerce* online mana yang memberikan layanan terbaik. Banyak komentar di *Twitter* bahkan trending topik di *Twitter* tentang layanan yang diberikan oleh sebuah *e-commerce*. Karena itu, penting untuk menganalisis pandangan dan sentimen pengguna *e-commerce*[3].

Berbagai pendapat pelanggan di *Twitter* mengenai situs online marketplace tidak memberikan kesimpulan mengenai situs mana yang memiliki pelayanan terbaik. Ini terjadi karena jumlah komentar yang banyak di *twitter*, bahkan *trending topic* pun hanya menampilkan topik yang sedang banyak diperbincangkan tanpa memberikan kesimpulan apa pun[4].

Analisis sentimen merupakan proses evaluasi atau pengukuran pendapat yang disampaikan tentang suatu objek atau peristiwa dalam bentuk teks, yang dapat diklasifikasikan sebagai *Positive* atau *Negative*. Ada banyak cara untuk mengekspresikan nilai sentimen dalam *tweet*, yang dapat dikenali melalui adanya sentimen yang terpancar di dalamnya. Apabila sebuah *tweet* memiliki sentimen atau kata-kata yang polar atau berlawanan, maka *tweet* tersebut akan dikategorikan sebagai *Positive* atau *Negative*[2]. Fungsi analisis sentimen adalah menggolongkan tiap kutub teks pada berbagai sumber internet maupun media sosial yang berbentuk dokumen atau kalimat kemudian ditentukan apakah katatersebut termasuk kategori *Positive*, netral ataupun *Negative*. Analisis sentiment juga bisa menunjukkan emosi kesedihan, kegembiraan, atau kemarahan[5].

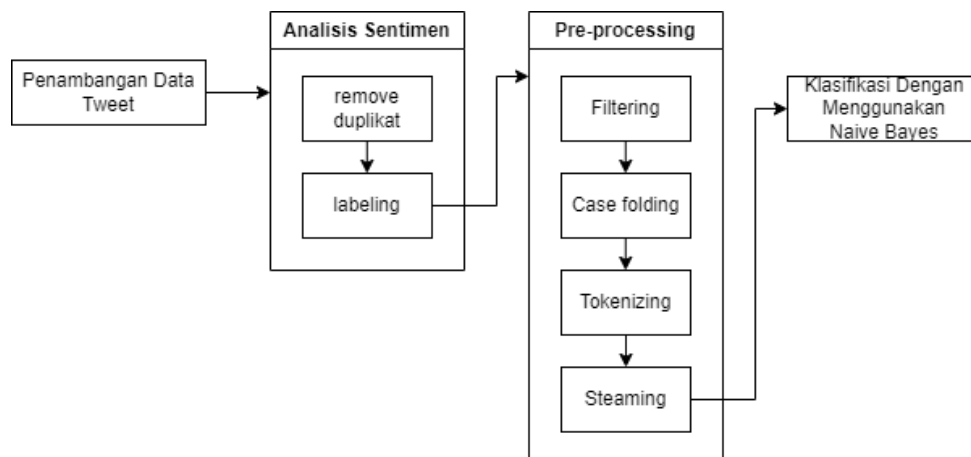
Analisis yang dilakukan pada penelitian ini menggunakan *RapidMiner*. *RapidMiner* adalah platform perangkat lunak penambangan data yang dikembangkan oleh perusahaan. dengan nama yang sama, yang menyediakan lingkungan terintegrasi untuk pembelajaran mesin. Pembelajaran mendalam, penambangan kata, analisis prediktif. Digunakan untuk aplikasi bisnis dan komersial, termasuk penelitian, pendidikan, pelatihan, pembuatan prototipe cepat, dan pengembangan aplikasi, ini mendukung semua aspek proses pembelajaran mesin, termasuk persiapan data, visualisasi hasil, validasi, dan pengoptimalan.[6].

Naïve Bayes adalah teknologi *preprocessing* klasifikasi objek yang menghadirkan skalabilitas, akurasi, dan efisiensi pada klasifikasi teks[7]. Ciri utama dari algoritma *Naïve Bayes* adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian. Kelebihan dari *Naïve Bayes* adalah proses klasifikasi data dapat disesuaikan dengan sifat dan kebutuhan masing-masing[8]. Dengan begitu Metode *Naïve Bayes* memerlukan sedikit data pelatihan untuk membuat perkiraan yang diperlukan untuk proses klasifikasi. Karena variabel independen dipertimbangkan, maka diperlukan distribusi variabel secara acak dalam kelas[9].

II. METODE PENELITIAN

A. Tahapan Penelitian

Tahapan yang akan di lakukan dalam penelitian ini digambarkan dalam diagram alur pada Gambar 1 berikut ini :



Gambar 1 Diagram Alur Tahapan Penelitian

B. Pengumpulan Data

RapidMiner adalah platform perangkat lunak penambangan data yang dikembangkan oleh perusahaan. Seperti namanya, ini menyediakan lingkungan terintegrasi untuk pembelajaran mesin, pembelajaran mendalam, penambangan teks, dan analisis prediktif. *RapidMiner* dikembangkan dengan model open source [10]. Ini dapat digunakan untuk melakukan operasi penambangan data. *RapidMiner* digunakan sebagai alat penambangan. Data dan perpustakaan *Twitter* tersedia di *RapidMiner* dan Proses Pemrosesan Hapus konten duplikat, hapus tautan, dll. Berhenti menggunakan *hashtag*. ada banyak *Library* yang digunakan di *RapidMiner* adalah *Search Twitter*, *Remove duplicate*, *Select attributes*, *Replace* dan *Write csv* [11]. *Library* ini menggunakan API *Twitter* dalam bentuk pencarian, dan dapat mencari *twee* yang sesuai dengan kata kunci yang ditentukan. Peneliti memberikan kata kunci berupa *#Shopee*, *#Tokopedia*, *#Lazada* dan *#Bukalapak Twitter* menyediakan akses ke bagian tertentu dari layanan *Twitter* melalui API kemungkinan pengguna membuat perangkat lunak yang terintegrasi dengan *Twitter*, seperti solusi yang membantu perusahaan membalas dan memberi masukan komentar pelanggan di *Twitter* [6]. Dikarenakan banyak pengguna yang memakai kata kunci tersebut untuk membagikan opini, berjualan serta melakukan promosi.

Data ini dikumpulkan hingga 42.000 data dari *tweet* selama 7 hari. Dan di ambil hanya 200 data yang akan di analisis dengan melakukan penambahan skor sentimen *Positive* pada data teks dilakukan secara manual. Artinya, tulisan yang mengandung unsur *Positive*, bersifat membangun, dan tidak mengandung unsur kebencian. Dan skor sentimen *Negative*, yaitu kalimat yang mengandung unsur *Negative* dan kritis yang mengarah pada pengabaian. Pengambilan 200 data di ambil sebagai contoh untuk melakukan proses analisis sentimen

C. Pelabelan

Analisis sentimen adalah pendekatan analitis yang digunakan untuk menganalisis sebuah teks. Analisis sentimen ini bertujuan untuk mengetahui subjektivitas opini, hasil review atau *tweet*. Berdasarkan analisis sentimen, opini dari seseorang dapat diklasifikasikan ke dalam berbagai kategori berdasarkan ukuran data dan jenis dokumen [12]. Analisis sentimen di lakukan dengan memberi label pada text dengan memberikan opini dari seseorang dapat diklasifikasikan ke dalam kategori dua kategori yaitu *Positive* dan *Negative*. Data yang di ambil akan dilakukan analisis sentimen sebelum masuk ke tahap preprocessing. Pada titik ini, peneliti melakukan analisis sentimen manual. Pelabelan ini di lakukan oleh Guru Bahasa Indonesia Bu Sri Handayani S.Pd. Di SMP Yayasan Pendidikan Nasional (YAPENAS) dengan membaca setiap kalimat *tweet* satu persatu. Kriteria *tweet* yang di beri label *Positive* adalah *tweet* yang mengandung unsur kalimat kepuasan pelanggan menggunakan aplikasi, sedangkan untuk *tweet Negative* adalah *tweet* yang mengandung kalimat ungkapan kekecewaan pelanggan dalam menggunakan aplikasi, penanganan pihak terkait saat aplikasi mengalami gangguan.

D. Text Preprocessing

Pada tahap *Preprocessing* data, data *tweet* asli akan melakukan proses perubahan menjadi huruf kecil, pembagian kalimat, pemangkasan kata, dan penyaringan. Hasil yang terjadi dari termin ini menghasilkan fitur yg dipergunakan menjadi data pembelajaran mesin sang NBC. Dalam tahap pra-pemrosesan, terdiri dari proses perubahan menjadi huruf kecil, pembagian menjadi token, pemangkasan kata, dan penyaringan [13]. Berikut ini adalah penjelasan dari tahap pra-pemrosesan:

1. Case folding adalah proses *preprocessing* teks yang dilakukan untuk membakukan karakter dalam data. Proses pelipatan huruf besar mengubah semua huruf menjadi huruf kecil. Karakter 'A' sampai 'Z' dalam data diubah menjadi karakter 'a' sampai 'z'.
2. Filtering, menghapus tanda baca berupa *email*, *emoticon* (:@, :* , :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya. Serta menghapus kata yang tidak penting *URL*, *hashtag* (#), *username* (@*username*), kata RT (*retweet*).
3. Steaming, merubah kata dasar menjadi kata yang berimbuhan.
4. Tokenizing, memisah kalimat menjadi kata per kata.

E. Klasifikasi Text Menggunakan Naïve Bayes

Manfaat metode *Naïve Bayes* adalah data pelatihan kecil untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi. Karena dianggap sebagai variabel independen, hanya variasi variabel klasifikasi yang diperlukan untuk menentukan kategori, bukan kategori secara keseluruhan.

Dengan menggunakan data sampel, *Naïve Bayes* menyediakan mekanisme untuk memperkirakan probabilitas posterior $P(y | x)$ dari setiap kelas y yang diberi objek x . Ini dapat digunakan untuk klasifikasi dan fungsi pendukung keputusan lainnya. Dalam istilah yang lebih sederhana, teorema Bayes adalah rumus sederhana untuk menemukan probabilitas bersama dengan probabilitas lain yang spesifik [7]. Klasifikasi adalah proses mengekstraksi data untuk membuat pola guna membuat prediksi. Kelas atau kategori objek dalam database.

Klasifikasi adalah proses dua bagian. Tahapannya adalah tahap pembelajaran dan tahap klasifikasi. Pada tahap pembelajaran, algoritma Pengklasifikasi menganalisis data pelatihan untuk membuat model klasifikasi. Tahap pembelajaran juga dapat dipandang sebagai konstruk fungsi atau peta $Y=f(X)$. Ini Y adalah kelas hasil prediksi dan X adalah *tuple* dengan kelas yang ingin diprediksi. bagian selanjutnya Model yang dihasilkan kemudian digunakan sebagai fungsi pengklasifikasi [14]. Pengklasifikasi adalah fungsi f yang memetakan vektor fitur masukan ke label kelas keluaran di mana X adalah ruang fitur. Biasanya asumsikan. Vektor adalah vektor dari D bilangan real atau D bit biner, tetapi secara umum dapat digabungkan secara diskrit dan berkelanjutan. fungsi khusus bahwa penunjukan kelas tidak terurut (kategorikal) dan saling eksklusif. (jika kita Input milik beberapa kelas, yang disebut masalah multilabel.) Tujuan kita adalah mempelajari f dari a . Set pelatihan menampilkan N pasangan masukan/keluaran (x_n, y_n) , $n = 1 : N$. Data training merupakan data yang digunakan untuk melatih sistem agar mampu mengenali pola yang dicari, sedangkan data testing merupakan data yang digunakan untuk menguji hasil training yang telah dilakukan. [3]

Metode ini di pakai karena tingkat akurasi yang tinggi dengan menggunakan data yang sedikit sehingga meningkatkan efisiensi waktu.

F. Pengujian *accuracy*, *precision* dan *recall*

Confusion Matrix adalah matriks yang memberikan perkiraan umum nilai yang benar dan digunakan untuk menentukan kinerja pengklasifikasi, yaitu berapa kali kelas X diklasifikasikan dengan benar sebagai suatu kelas padahal hanya ada dua kelas. [15]

Nilai *precision* adalah kecocokan sistem antara informasi yang diberikan oleh sistem untuk menampilkan data kelas *Negative* atau *Positive* dengan benar. Nilai *recall* adalah nilai yang menunjukkan tingkat keberhasilan atau kekhususan dalam menemukan informasi dengan benar tentang data kelas *Negative* atau konten teks *Positive*. Di sisi lain, Nilai Akurasi adalah persentase data *tweet* yang dikenali dengan benar dari data uji. Dengan kata lain, akurasi adalah ukuran seberapa dekat prediksi suatu sistem dengan prediksi manusia [2].

Metode pengukuran untuk information retrieval yang biasa digunakan adalah dengan menggunakan table *Confusion Matrix* seperti pada Tabel I.

TABEL I
SATUAN UNTUK PROPERTI MAGNETIS

	False Positive	False Negative
True Positive	TP (True Positive) Correct result	FP (False Positive) Unexpected result
True Negative	FN (False Negative) Missing result	TN (True Negative) Correct Absence of result

Confusion Matrix dari hasil true Positive (TP), yaitu banyaknya data kelas *Positive* yang diklasifikasi menjadi kelas *Positive* dan True Negative (TN). Dengan kata lain banyaknya data kelas *Negative* diklasifikasikan sebagai kelas *Negative*. Saat ini, False Positive (FP) Banyaknya data kelas *Negative* diklasifikasikan ke dalam kelas *Positive* dan disebut False Negative (FN). Banyaknya data dari kelas *Positive* diklasifikasikan ke dalam kelas *Negative* [4]. Pengujian tersebut secara umum dapat di rumuskan sebagai berikut:

$$precision = \left(\frac{TP}{TP + FP} \right) \times 100\% \quad (1)$$

$$accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100\% \quad (2)$$

$$recall = \left(\frac{TP}{TP + FN} \right) \times 100\% \quad (3)$$

Keterangan :

TP = jumlah data nilai sebenarnya *Positive* dan nilai prediktif *Positive*

FP = jumlah data nilai sebenarnya *Negative* dan nilai prediktif *Positive*

FN = jumlah data nilai sebenarnya *Positive* dan nilai prediktif *Negative*

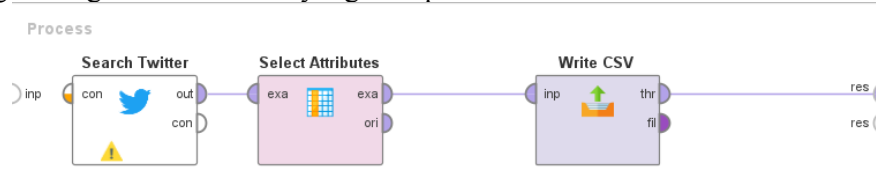
TN = jumlah data nilai sebenarnya *Negative* dan nilai prediktif *Negative*

Ukuran besaran *precision*, *recall*, dan *accuracy* biasanya diberi nilai dalam bentuk presentase antara 1% hingga 100%. Sebuah sistem akan disebut baik Bila taraf *precision*, *recall*, serta *accuracy*-nya tinggi.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data dilakukan pada *RapidMiner* seperti yang tertera pada Gambar 3.1. *Search Twitter* menggunakan API yang disediakan oleh *Twitter* sehingga memungkinkan mengambil data komentar *Twitter*. Lalu *Select attributes* berfungsi sebagai memilih atribut apa saja yang akan di ambil oleh rapid miner. Sedangkan *Write CSV* berfungsi sebagai membuat data yang di dapat di ubah dalam bentuk *CSV*.



Gambar 3.1 Pengumpulan Data

Pengambilan data tersebut hanya mengambil dua kolom attribute yaitu *text* untuk memunculkan kolom komentar yang akan di proses oleh *RapidMiner* dan *id* untuk membedakan komentar tiap pengguna *Twitter* sehingga menghasilkan data seperti pada Tabel II.

TABEL II
SATUAN UNTUK PROPERTI MAGNETIS

text	id
@bukalapak @bukalapak udah lama saldk gua dibekukan padahal gua ga ngapa ngapain udah sering gua buat laporan masih aja ga kasih solusi dan terbilang respond sangat lamban... https://t.co/X0P5eZ215W	1,66E+18
@bukalapak kenapa saya udah ga bisa pasang iklan barang dijual lagi ya sebagai pelapak???	1,66E+18
Shopee, lazada, tokopedia, bukalapak banyak minesnya palingan barang bekas pakai https://t.co/uZJmMFsPK4	1,66E+18
@bukalapak Bukalapak banyak penipuan	1,66E+18
Gak lagi-lagi pake bukalapak, ribet banget	1,66E+18
@bukalapak Komplain melalui live chat hanya disuruh menunggu, dan akhirnya di akhiri secara pihak oleh BL, tanpa dapat solusi	1,66E+18
@bukalapak makan gaji buta aja lo. kendala customer udah 2minggu masih aja ga teratasi. di dm ga bales, tlp gabisa, email malah dioper ke aplikasi, di aplikasi malah ga kehubung sama cs nya. mau nya apaan coba lepas tangan gini?	1,66E+18
Bocor data terbesar w nyata tahun 2017 gara2 Bukalapak. ??? https://t.co/jxoRrSJM	1,66E+18
Proyeksi Direvisi, Saham Bukalapak (BUKA) Berpeluang Uptrend #TrenAsia https://t.co/UDYIYLk1qb	1,66E+18
@DeeHumaira Bukalapak aja lebih cuan	1,66E+18

B. Pelabelan

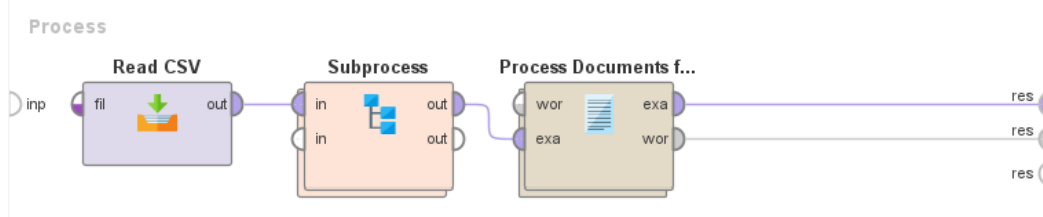
Analisis sentimen dilakukan secara manual. Meliputi 200 data yang di lakukan analisis sentimen dan mendapatkan nilai *Positive* dan *Negative* yang seperti tertera pada contoh twitter dalam Tabel III

TABEL III
SATUAN UNTUK PROPERTI MAGNETIS

tweet	Sentimen
Gak lagilagi pake <i>Bukalapak</i> ribet banget	<i>Negative</i>
Aku suka nya sih <i>Lazada</i> belum ada biaya Admin nya	<i>Positive</i>
<i>Shopee</i> knp skrg nyebelin sih	<i>Negative</i>
<i>Tokopedia</i> hanya pengalihan isu	<i>Negative</i>

C. Text Preprocessing

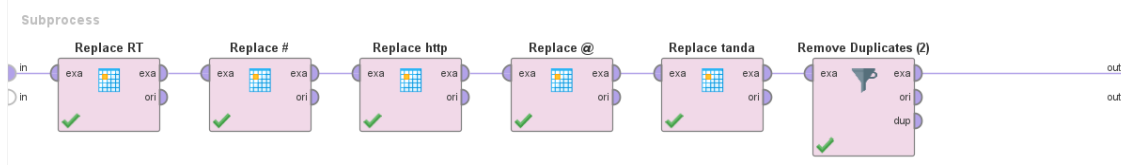
Tahap preprocessing data di bagi menjadi tahapan sub process dan proses dokumen seperti pada Gambar 3.3



Gambar 3.3 Text Preprocessing

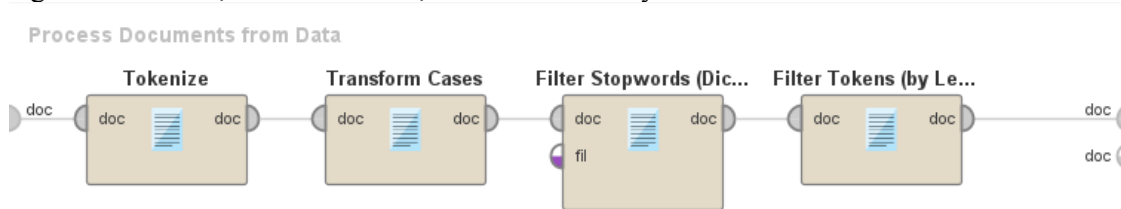
Pada tahap *Preprocessing* data, data *tweet* asli akan melakukan proses perubahan data yang lebih sederhana dimulai dengan sub proses yang terdiri dari *Replace RT* yaitu menghilangkan kata pada RT (*retweet*), *Replace* hastag (#) menghilangkan tanda hastag beserta kata setelah hastag contoh #shoppe, *Replace* http menghilangkan

kata http beserta kata dibelakngnya, *Replace* tanda baca berupa *emoticon* (:@, :* , :D), tanda baca seperti koma (,), titik (.), tanda seru (!). Seperti pada Gambar 3.4



Gambar 3.4 Proses *Text Preprocessing*

Pada tahapan proses dokumen pada Gambar 3.5 terdiri dari tokenize yaitu ,transform cases ,filter stopwords yaitu menghilangkan kata ,dan filter tokens. Tokenize berfungsi sebagai memisah kalimat menjadi kata per kata. Contoh kalimat *Gak lagilagi pake Bukalapak ribet banget*, di bagi menjadi 6 kata yaitu “Gak”, “lagilagi”, “pake”, “Bukalapak”, “ribet”, “banget”. Transform Cases berfungsi sebagaimengubahnya menjadi huruf kecil, huruf besar, dan huruf judul, dan mengembalikan string yang diubah ini. Seperti contoh “Admin” menjadi “admin”. filter stopwords biasanya melibatkan penghapusan kata-kata umum (seperti "dan", "apa", "adalah", dll.) yang sering kali tidak memiliki arti signifikan dalam tugas analisis teks seperti pemrosesan bahasa alami. Dan filter tokens melibatkan pemrosesan sekumpulan token (kata, simbol, atau entitas) berdasarkan kriteria tertentu seperti menghilangkan tanda baca, karakter khusus, atau elemen lainnya.



Gambar 3.5 Proses *Text Preprocessing*

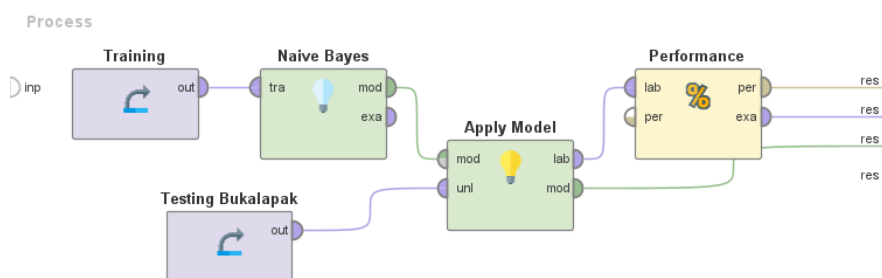
Hasil yang di dapatkan dari melakukan Text Prosesing menghasilkan probabilitas kata dalam tabel meliputi 17 data *Positive* dan 13 data *Negative* dari kata *Bukalapak* menghasilkan total 30 kata. 22 data *Positive* dan 6 data *Negative* dari kata *Lazada* menghasilkan total 28 kata. 16 data *Positive* dan 12 data *Negative* dari kata *Shopee* menghasilkan total 28 kata. 17 data *Positive* dan 11 data *Negative* dari kata *Tokopedia* menghasilkan total 28 kata. Seperti yang di tunjukan pada Tabel IV.

TABEL IV
SATUAN UNTUK PROPERTI MAGNETIS

Data	<i>Positive</i>	<i>Negative</i>	Total
<i>Bukalapak</i>	17	13	30
<i>Lazada</i>	22	6	28
<i>Shopee</i>	16	12	28
<i>Tokopedia</i>	17	11	28

D. Klasifikasi teks menggunakan *Naïve Bayes*

Klasifikasi *Naïve Bayes* dilakukan dengan memasukkan data training pada mesin *Naïve Bayes*. Dilanjutkan dengan membuat apply model untuk menerapkan model data training dan data testing yang akan menghasilkan prediksi data testing. Lalu memasukkan mesin performace untuk menghitung seberapa akurat algoritma *Naïve Bayes* berkerja. Proses klasifikasi dapat terlihat pada gambar 3.6



Gambar 3.6 Proses Klasifikasi *Naïve Bayes*

Dengan menggunakan data sampel, *Naïve Bayes* menyediakan mekanisme untuk memperkirakan probabilitas posterior $P(y | x)$ dari setiap kelas y yang diberi objek x . Probabilitasnya ditunjukkan pada tabel di bawah ini :

Dari tabel di atas dijelaskan bahwa seluruh data yang dijadikan data training menggunakan *RapidMiner* dengan kelas *Positivenya* berjumlah 66 dan kelas *Negativenya* berjumlah 34.

TABEL V
 SATUAN UNTUK PROPERTI MAGNETIS

<i>e-commerce</i>	<i>tweet</i>	Sentimen
<i>Bukalapak</i>	Gak lagilagi pake <i>Bukalapak</i> ribet banget	<i>Negative</i>

Pada Tabel V adalah contoh data yang akan dilakukan proses perhitungan *Naïve Bayes* pada *tweet Bukalapak*

$$P(\text{Gak} | \text{Negative}) = \frac{5+1}{34} = 0,1764705882352941$$

$$P(\text{lagilagi} | \text{Negative}) = \frac{1+1}{34} = 0,0588235294117647$$

$$P(\text{pake} | \text{Negative}) = \frac{2+1}{34} = 0,0882352941176471$$

$$P(\text{Bukalapak} | \text{Negative}) = \frac{13+1}{34} = 0,4117647058823529$$

$$P(\text{ribet} | \text{Negative}) = \frac{3+1}{34} = 0,1176470588235294$$

$$P(\text{banget} | \text{Negative}) = \frac{4+1}{34} = 0,1470588235294118$$

$$P(X | \text{Negative}) = 6,525097867146423$$

Untuk perhitunagn *Positive* sebagai berikut

$$P(\text{Gak} | \text{Positive}) = \frac{7+1}{66} = 0,1212121212121212$$

$$P(\text{lagilagi} | \text{Positive}) = \frac{0+1}{66} = 0,0151515151515152$$

$$P(\text{pake} | \text{Positive}) = \frac{4+1}{66} = 0,0757575757575758$$

$$P(\text{Bukalapak} | \text{Positive}) = \frac{17+1}{66} = 0,2727272727272727$$

$$P(\text{ribet} | \text{Positive}) = \frac{0+1}{66} = 0,0151515151515152$$

$$P(\text{banget} | \text{Positive}) = \frac{8+1}{66} = 0,1363636363636364$$

$$P(X | \text{Positive}) = 7,839915695191408$$

TABEL VI
 SATUAN UNTUK PROPERTI MAGNETIS

<i>e-commerce</i>	<i>tweet</i>	Sentimen
<i>Lazada</i>	Aku suka nya sih <i>Lazada</i> belum ada biaya Admin nya	<i>Positive</i>

Pada Tabel VI adalah contoh data yang akan dilakukan proses perhitungan *Naïve Bayes* pada *tweet Lazada*

$$P(\text{Aku} | \text{Negative}) = \frac{0+1}{34} = 0,029411765$$

$$P(\text{suka} | \text{Negative}) = \frac{0+1}{34} = 0,029411765$$

$$P(\text{pake} | \text{Negative}) = \frac{2+1}{34} = 0,088235294$$

$$P(\text{nya} | \text{Negative}) = \frac{5+1}{34} = 0,176470588$$

$$P(\text{sih} | \text{Negative}) = \frac{1+1}{34} = 0,058823529$$

$$P(\text{Lazada} | \text{Negative}) = \frac{0+1}{34} = 0,029411765$$

$$P(\text{belum} | \text{Negative}) = \frac{4+1}{34} = 0,147058824$$

$$P(\text{ada} | \text{Negative}) = \frac{0+1}{34} = 0,029411765$$

$$P(\text{biaya} | \text{Negative}) = \frac{0+1}{34} = 0,029411765$$

$$P(\text{Admin} | \text{Negative}) = \frac{0+1}{34} = 0,029411765$$

$$P(\text{nya} | \text{Negative}) = \frac{5+1}{34} = 0,176470588$$

$$P(X | \text{Negative}) = 1,53871$$

Untuk perhitunagn *Positive* sebagai berikut

$$P(\text{Aku} | \text{Positive}) = \frac{0+1}{66} = 0,015151515$$

$$P(\text{suka} | \text{Positive}) = \frac{4+1}{66} = 0,075757576$$

$$P(\text{pake} | \text{Positive}) = \frac{4+1}{66} = 0,075757576$$

$$P(\text{nya} | \text{Positive}) = \frac{9+1}{66} = 0,151515152$$

$$P(\text{sih} | \text{Positive}) = \frac{6+1}{66} = 0,106060606$$

$$P(\text{Lazada} | \text{Positive}) = \frac{0+1}{66} = 0,348484848$$

$$P(\text{belum} | \text{Positive}) = \frac{4+1}{66} = 0,075757576$$

$$P(\text{ada} | \text{Positive}) = \frac{0+1}{66} = 0,015151515$$

$$P(\text{biaya} | \text{Positive}) = \frac{17+1}{66} = 0,272727273$$

$$P(\text{Admin} | \text{Positive}) = \frac{0+1}{66} = 0,015151515$$

$$P(\text{nya} | \text{Positive}) = \frac{9+1}{66} = 0,151515152$$

$$P(X | \text{Positive}) = 1,52159$$

TABEL VII
 SATUAN UNTUK PROPERTI MAGNETIS

<i>e-commerce</i>	<i>tweet</i>	Sentimen
<i>Shopee</i>	<i>Shopee knp skrg nyebelin sih</i>	<i>Negative</i>

Pada Tabel VII adalah contoh data yang akan dilakukan proses perhitungan *Naïve Bayes* pada *tweet Shopee*

$$P(\text{Shopee} | \text{Negative}) = \frac{12+1}{34} = 0,1764705882352941$$

$$P(\text{knj} | \text{Negative}) = \frac{1+1}{34} = 0,0588235294117647$$

$$P(\text{skrng} | \text{Negative}) = \frac{1+1}{34} = 0,0882352941176471$$

$$P(\text{nyebelin} | \text{Negative}) = \frac{2+1}{34} = 0,4117647058823529$$

$$P(\text{sih} | \text{Negative}) = \frac{1+1}{34} = 0,058823529$$

$$P(X | \text{Negative}) = 2,49135$$

Untuk perhitunagn *Positive* sebagai berikut

$$P(\text{Shopee} | \text{Positive}) = \frac{15+1}{66} = 0,1764705882352941$$

$$P(\text{knj} | \text{Positive}) = \frac{0+1}{66} = 0,0588235294117647$$

$$P(\text{skrng} | \text{Positive}) = \frac{0+1}{66} = 0,0882352941176471$$

$$P(\text{nyebelin} | \text{Positive}) = \frac{1+1}{66} = 0,4117647058823529$$

$$P(\text{sih} | \text{Positive}) = \frac{6+1}{66} = 0,106060606$$

TABEL VIII
 SATUAN UNTUK PROPERTI MAGNETIS

<i>e-commerce</i>	<i>tweet</i>	sentimen
<i>Tokopedia</i>	<i>Tokopedia hanya pengalihan isu</i>	<i>Negative</i>

Pada Tabel VIII adalah contoh data yang akan dilakukan proses perhitungan *Naïve Bayes* pada *tweet Tokopedia*

$$P(X | \text{Positive}) = 8,94331$$

$$P(\text{Tokopedia} | \text{Negative}) = \frac{5+1}{34} = 0,1764705882352941$$

$$P(\text{hanya} | \text{Negative}) = \frac{1+1}{34} = 0,0588235294117647$$

$$P(\text{pengalihan} | \text{Negative}) = \frac{2+1}{34} = 0,0882352941176471$$

$$P(\text{isu} | \text{Negative}) = \frac{13+1}{34} = 0,4117647058823529$$

$$P(X | \text{Negative}) = 0,000377151$$

Untuk perhitunagn *Positive* sebagai berikut

$$P(\text{Tokopedia} | \text{Positive}) = \frac{5+1}{66} = 0,1764705882352941$$

$$P(\text{hanya} | \text{Positive}) = \frac{1+1}{66} = 0,0588235294117647$$

$$P(\text{pengalihan} | \text{Positive}) = \frac{2+1}{66} = 0,0882352941176471$$

$$P(\text{isu} | \text{Positive}) = \frac{13+1}{66} = 0,4117647058823529$$

$$P(X | \text{Positive}) = 2,46644$$

E. Pengujian accuracy, precision dan recall

Proses *Naïve Bayes* yang telah dilakukan dengan data pelatihan 200 data dan data pengujian 25 di setiap masing masing data *Bukalapak, Lazada, Shopee* dan *Tokopedia*.

$$\text{precision} = \left(\frac{15}{15 + 2} \right) \times 100\% = 88,24\%$$

$$\text{recall} = \left(\frac{15}{15 + 0} \right) \times 100\% = 100\%$$

$$\text{accuracy} = \left(\frac{15 + 8}{15 + 8 + 2 + 0} \right) \times 100\% = 92,00\%$$

accuracy: 92.00%

	true Positive	true Negative	class precision
pred. Positive	15	2	88.24%
pred. Negative	0	8	100.00%
class recall	100.00%	80.00%	

Gambar 3.7 Hasil *Confusion Matrix Bukalapak*

Untuk pengujian *confusion matrix* pada data *Bukalapak* menghasilkan nilai *precision* 88,24%, nilai *recall* 100% dan nilai *accuracy* 92% sesuai pada Gambar 3.7.

accuracy: 88.00%

	true Positive	true Negative	class precision
pred. Positive	20	3	86.96%
pred. Negative	0	2	100.00%
class recall	100.00%	40.00%	

Gambar 3.8 Hasil *Confusion Matrix Lazada*

Untuk pengujian *confusion matrix* pada data *Lazada* menghasilkan nilai *precision* 86,96%, nilai *recall* 100% dan nilai *accuracy* 88% sesuai pada Gambar 3.8.

$$\text{precision} = \left(\frac{10}{10 + 0} \right) \times 100\% = 100\%$$

$$\text{recall} = \left(\frac{10}{10 + 0} \right) \times 100\% = 100\%$$

$$\text{accuracy} = \left(\frac{10 + 15}{10 + 15 + 0 + 0} \right) \times 100\% = 100\%$$

accuracy: 100.00%

	true Negative	true Positive	class precision
pred. Negative	10	0	100.00%
pred. Positive	0	15	100.00%
class recall	100.00%	100.00%	

Gambar 3.9 Hasil *Confusion Matrix Shopee*

Untuk pengujian *confusion matrix* pada data *Shopee* menghasilkan nilai *precision* 100%, nilai *recall* 100% dan nilai *accuracy* 100% sesuai pada Gambar 3.9.

$$precision = \left(\frac{16}{16 + 3} \right) \times 100\% = 84,21\%$$

$$recall = \left(\frac{16}{16 + 0} \right) \times 100\% = 100\%$$

$$accuracy = \left(\frac{16 + 6}{16 + 6 + 0 + 3} \right) \times 100\% = 88,00\%$$

accuracy: 88.00%

	true Positive	true Negative	class precision
pred. Positive	16	3	84.21%
pred. Negative	0	6	100.00%
class recall	100.00%	66.67%	

Gambar 3.10 Hasil *Confusion Matrix* Tokopedia

Untuk pengujian *confusion matrix* pada data *Tokopedia* menghasilkan nilai *precision* 84,21%, nilai *recall* 100% dan nilai *accuracy* 88% sesuai pada Gambar 3.10.

IV. KESIMPULAN

Studi ini menunjukkan bagaimana analisis sentimen *tweet* pengguna *e-commerce*, menggunakan metode klasifikasi *Naïve Bayes*, secara efektif membedakan dan mengkategorikan sentimen. Dengan menganalisis konten buatan pengguna, ini memberikan wawasan berharga tentang perspektif pelanggan. Dari 200 data yang di uji menghasilkan 92,00% untuk total data *accuracy*. Dilihat dari data yang diuji dapat disimpulkan bahwa *Shopee* mendapatkan hasil tertinggi yaitu *recall true positive* 100%, *precision prediction positive* 100% dan nilai *accuracy* 100%.

DAFTAR PUSTAKA

- [1] Ayu, S., & Lahmi, A. (2020). Peran e-commerce terhadap perekonomian Indonesia selama pandemi Covid-19. *Jurnal Kajian Manajemen Bisnis*, 9(2), 114. <https://doi.org/10.24036/jkmb.10994100>
- [2] Wandani, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, Issue 2)
- [3] Bayhaqy, A., Sfenianto, S., Nainggolan, K., & Kaburuan, E. R. (2018, July 2). Sentiment Analysis about e-commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. 2018 International Conference on Orange Technologies, ICOT 2018.
- [4] Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes Muljono, A., Putri Artanti, D., Syukur, A., Prihandono, A., & Rosal Moses Setiadi, D. I. (2018). Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang. <http://twitter.com>
- [5] Minfo Polgan, J., Melina Salsabila, S., Alim Murtopo, A., Fadhilah, N., Sentimen Pelanggan Toko Online Tokopedia Menggunakan Metode Naive Bayes Classifier, A., Informatika, T., & YMI Tegal, S. (n.d.). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naive Bayes Classifier. www.tokopedia.com
- [6] Nofitri, R., & Irawati, N. (2019). ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE Rapid Miner. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteksi.v5i2.365>
- [7] https://doi.org/10.1007/978-1-4899-7502-7_581-1
- [8] Gunawan, B., Sasty, H., #2, P., Esyudha, E., & #3, P. (2018). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. 4(2), 17–29. www.femaledaily.com
- [9] Widya Ningsih, E. (2019). Penerapan Algoritma Naive Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus. <https://doi.org/10.31294/jtk.v4i2>
- [10] Komara, D. A., & Hadiapurwa, A. (2022). PUBLIS JOURNAL AUTOMATING Twitter DATA COLLECTION: A Rapid Miner-BASED CRAWLING SOLUTION. 6. <http://journ.umpo.ac.id/index.php/PUBLIS>
- [11] Yan, K., Arisandi, D., & Tony, J. (n.d.). *Jurnal Ilmu Komputer dan Sistem Informasi ANALISIS SENTIMEN KOMENTAR NETIZEN TWITTER TERHADAP KESEHATAN MENTAL MASYARAKAT INDONESIA*.
- [12] Wongkar, M., & Angdresy, A. (2019, October 1). Sentiment Analysis Using Naïve Bayes Algorithm Of The Data Crawler: Twitter. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*. <https://doi.org/10.1109/ICIC47613.2019.8985884>
- [13] Garbian Nugroho, D., Herry Chrisnanto, Y., Wahana Jurusan Informatika, A., & Matematika dan Ilmu Pengetahuan Alam Universitas Jenderal Achmad Yani Jalan Terusan Jenderal Sudirman, F. (n.d.). ANALISIS SENTIMEN PADA JASA OJEK ONLINE MENGGUNAKAN METODE NAIVE BAYES.
- [14] Susana, H., & Suarna, N. (2022). PENERAPAN MODEL KLASIFIKASI METODE NAIVE BAYES TERHADAP PENGGUNAAN AKSES INTERNET Program Studi Teknik Informatika STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 3) Program Studi Rekayasa Perangkat Lunak STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 4) Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon. *Jurnal Sistem Informasi Dan Teknologi Informasi*, 4(1), 1–8.
- [15] Aditya, D., Mubarak, A., Kom, M., & Susanti, S. (2019). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier (Studi Kasus: Komentar Publik Kepada Tri Indonesia). *JURNAL INFORMATIKA*, 6(1), 1–8. <http://ejournal.bsi.ac.id/ejournal/index.php/ji/article/view/-https://doi.org/10.1109/ICOT.2018.8705796>