



Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifing Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes

Sania Wulandari, Firman Noor Hasan*

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

Email: ¹saniawulandari06@gmail.com, ^{2,*}firman.noorhasan@uhamka.ac.id

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

Abstrak—Thrifing merupakan aktivitas belanja barang bekas yang semakin populer di Indonesia terutama dikalangan kaum milenial dan generasi Z sebagai alternatif belanja yang menghemat biaya. Kegiatan thrifing memberikan dampak positif yang jelas bagi masyarakat Indonesia dalam menjaga lingkungan dengan mengurangi pembelian barang baru. Namun aktivitas jual beli thrifing dianggap ilegal dan dapat merugikan industri tekstil dalam negeri. Sehingga analisis sentimen perlu dilakukan untuk mengetahui bagaimana tanggapan masyarakat terhadap aktivitas jual beli thrifing. Penelitian ini bertujuan untuk menghitung jumlah komentar positif dan negatif dari pengguna Twitter, serta mengetahui seberapa akurat algoritma Naïve Bayes digunakan dalam pengklasifikasian. Data yang digunakan diambil dari media sosial Twitter sebanyak 900 tweet, kemudian diolah melalui beberapa tahap lanjutan seperti pre-processing yang terdiri dari cleansing, tokenize, dan filter stopwords. Kemudian pada tahap pelabelan data dibagi menjadi data latih dan data uji dengan rasio 60:40. Setelah diklasifikasikan menggunakan algoritma Naïve Bayes hasil yang didapatkan cenderung positif dengan jumlah komentar positif sebanyak 368 dan sentimen negatif sebanyak 181. Setelah melalui tahap evaluasi, didapatkan nilai accuracy sebesar 95,92%, nilai precision sebesar 95,76%, dan nilai recall sebesar 97,41%. Hasil evaluasi menunjukkan bahwa algoritma Naïve bayes terbukti memiliki keakuratan tingkat tinggi digunakan dalam pengklasifikasian.

Kata Kunci: Thrifing; Anaisis sentimen; Algoritma Naïve Bayes; Twitter; Media sosial

Abstract—Thrifing is an increasingly popular second-hand shopping activity in Indonesia, especially among millennials and generation Z as a cost-saving shopping alternative. Thrifing activities have a clear positive impact on the Indonesian people in protecting the environment by reducing the purchase of new goods. However, thrifing is considered illegal and can harm the domestic textile industry. So sentiment analysis needs to be done to find out how people respond to thrifing activities. This study aims to calculate the number of positive and negative comments from Twitter users, and find out how accurately the Naïve Bayes algorithm is used in the classification. The data used is taken from Twitter social media as many as 900 tweets, then processed through several advanced stages such as pre-processing which consists of cleansing, tokenize, and filter stopwords. Then at the labeling stage the data is divided into training data and test data with a ratio of 60:40. After being classified using the Naïve Bayes algorithm, the results obtained tend to be positive with a total of 368 positive comments and 181 negative sentiments. After going through the evaluation stage, the accuracy value is 95.92%, the precision value is 95.76%, and the recall value is 97.41%. The evaluation results show that the Naïve Bayes algorithm is proven to have a high level of accuracy used in classification.

Keywords: Thrifing; Sentiment Analysis; Naïve Bayes Algorithm; Twitter; Social Media

1. PENDAHULUAN

Salah satu tren berbelanja yang semakin populer di Indonesia terutama dikalangan kaum milenial dan generasi Z disebut dengan thrifing, thrifing adalah kegiatan yang dilakukan para pelanggan dengan mencari barang - barang bekas seperti pakaian, sepatu, aksesoris, dan barang - barang lain yang berkualitas baik dengan harga murah [1]. Thrifing juga mengurangi pemborosan karena barang bekas dalam kondisi bagus masih bisa dipakai kembali daripada di buang ke tempat pembuangan sampah, selain itu membeli barang bekas memungkinkan orang untuk mengekspresikan gaya berpakaian yang unik mereka dapat menemukan pakaian vintage, aksesoris langka atau barang unik lainnya yang mungkin sulit ditemukan di toko konvensional [2].

Dampak positif industri terhadap lingkungan, seperti polusi, air dan limbah tekstil menjadi lebih jelas bagi masyarakat, mereka dapat membantu menjaga lingkungan dengan mengurangi konsumsi barang baru dengan membeli barang bekas [3]. Barang - barang thrifing biasanya bisa didapatkan seperti di toko-toko khusus yang menjual barang bekas seperti bazaar, atau bahkan berbelanja online di platform seperti marketplace atau situs web khusus. Saat ini thrifing sudah dapat dilakukan melalui Media sosial [4].

Twitter adalah media sosial yang memungkinkan penggunanya membuat dan membagikan pesan pendek yang disebut "tweet", yang mencakup hingga 280 karakter yang panjang, teks, video, atau tautan ke situs web lain [5]. Twitter telah berkembang menjadi media sosial penting untuk berbagi informasi secara real-time di seluruh dunia, dan sering digunakan untuk berbagi berita, pemikiran, dan informasi [6]. Media sosial sudah mengubah cara masyarakat bersosialisasi dalam berbagai pengalaman belanja [7].

Media sosial seperti Twitter telah menjadi bagian inti dari kehidupan sehari-hari masyarakat di Indonesia dengan banyaknya pengguna yang semakin hari semakin meningkat di Indonesia, dengan meningkatnya pengguna twitter dapat dijadikan sumber data yang melimpah untuk dijadikan informasi mengenai pendapat dan sentimen



masyarakat dengan berbagai macam pengalaman berbelanja thrifting, oleh karena itu dibutuhkan analisis yang sistematis dan mendalam [8].

Naïve Bayes merupakan algoritma klasifikasi yang berdasarkan teori Bayes bahwa setiap fitur dalam kumpulan data tidak bergantung satu sama lain, algoritma ini digunakan untuk memprediksi kelas data dengan menghitung probabilitas kelas berdasarkan fitur-fitur yang ada. Meskipun sederhana, algoritma ini umumnya digunakan dalam klasifikasi teks, dan analisis sentimen [9]. Analisis sentimen adalah proses mengambil dan menilai opini, perasaan, atau sentimen dari teks, tujuannya untuk memahami apakah sebuah teks mengandung sentimen positif dan negatif [10]. Analisis sentimen juga dikenal opinion mining yaitu penggabungan antara text mining dan natural language processing [11].

Text mining mempunyai tujuan untuk mendapatkan data yang bermanfaat untuk sekumpulan data, proses ini dilakukan dengan menggunakan algoritma Naïve Bayes melalui Rapid Miner [12]. Rapid miner digunakan sebagai alat untuk menganalisis data dan mesin data mining. Rapid Miner memiliki beberapa tahap yang harus dijalani untuk memenuhi penelitian ini, proses dimulai dengan pengumpulan data atau crawling data melalui twitter [13]. Setelah proses pengumpulan data selesai, tahapan pembersihan data dilakukan supaya data dapat diolah dengan baik, tahapan cleansing dan filtering tersebut meliputi tokenizing, transform case, stopword removal, setelahnya dilakukan pelabelan dan pengklasifikasian [14]. Dalam menganalisis metode yang dapat digunakan adalah algoritma Naïve Bayes yang terbukti efektif dalam mengklasifikasikan sentimen dalam bentuk teks [15].

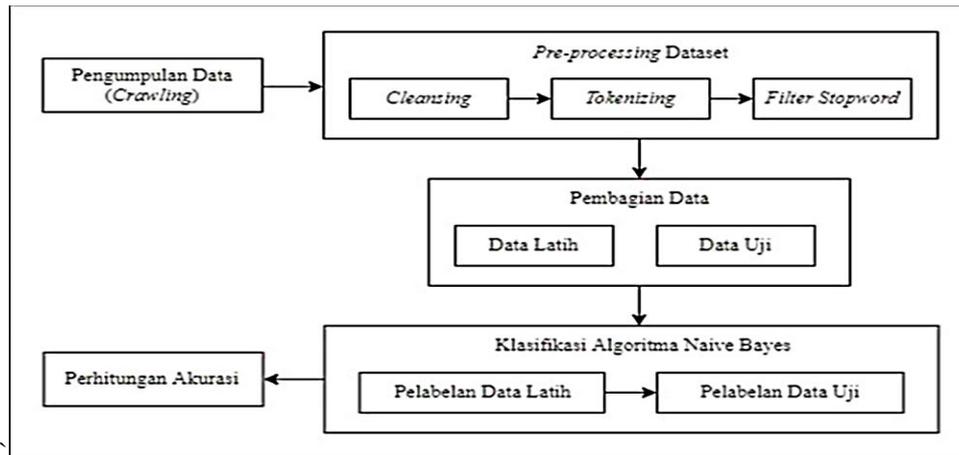
Sebelumnya itu sudah dilakukan penelitian analisis sentimen menggunakan metode naïve bayes yang diteliti oleh Sri Lestari pada Tahun 2021 dengan judul “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naïve Bayes” yang membahas tentang analisis sentimen terhadap jenis vaksin Sinovac untuk Covid-19 di Indonesia yang menjadikan algoritma Naïve Bayes sebagai pengklasifikasinya. Dalam penelitiannya ini bertujuan untuk menilai tanggapan atau pendapat dari masyarakat Indonesia, beserta kelebihan dan kekurangannya. Hasil dari perhitungan sentimen yang dilakukan dalam penelitian ini menunjukkan 86% sentimen positif dan 14% sentimen negatif, masing-masing dengan nilai akurasi 92,96% [16]. Penelitian lainnya yang serupa juga dilakukan oleh Sifa Melina Salsabila dengan judul “Analisis Sentimen Pelanggan Tokopedia Menggunakan Naïve Bayes Classifier” yang membahas tentang analisis sentimen terhadap pelanggan Tokopedia. Penelitian ini bertujuan untuk membuktikan analisis dan pengujian terhadap komentar pelanggan dalam beberapa sentimen positif dan negatif. Hasil dari pengujian menghasilkan nilai akurasi 95,10% dengan class recal 91,46%, precision 100.00%, dan nilai AUC: 0.999 [17]. Penelitian lainnya yang serupa juga dilakukan oleh Vynska Amalia Permadi dengan judul “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura” yang membahas tentang analisis sentimen terhadap review restoran di Singapura. Penelitian ini bertujuan untuk mengevaluasi kemampuan Naïve Bayes dalam mengklasifikasikan sentimen berdasarkan review pelanggan restoran, hasil perhitungan review pelanggan menggunakan metode Naïve Bayes dengan nilai akurasi sebesar 73% [18]. Penelitian yang serupa juga dilakukan oleh Sitti Nurul Jannah Fitriyyah dengan judul “Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes” yang membahas tentang analisis sentimen calon Presiden Indonesia 2019. Penelitian ini bertujuan untuk memahami penerapan metode Naïve Bayes dalam menganalisis sentimen pengguna Twitter dalam 2 kategori (negatif, positif) dan 3 kategori (negatif, positif, netral) [19]. Penelitian yang serupa juga dilakukan oleh Afid Rozaqi dengan judul “Analisis Sentimen Vaksin Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN” yang membahas tentang menilai hasil kinerja klasifikasi sentimen hal akurasi dan presisi, metode yang digunakan untuk menganalisis penelitian ini menggunakan algoritma Naïve Bayes dan K-Nearest Neighbour. Hasil dari percobaan ini menunjukkan akurasi sebesar 78,62% dengan jumlah sentimen positif sebanyak 303 data dan sentimen negatif sebanyak 93 data [20]. Pada penelitian sebelumnya telah dilakukan perbandingan antara algoritma Naïve Bayes dan K-Nearest Neighbour. Sedangkan penelitian ini hanya menggunakan algoritma Naïve Bayes untuk menganalisis sentimen.

Jual beli thrifting merupakan aktivitas ilegal yang dianggap dapat merugikan industri tekstil dalam negeri. Sehingga perlu dilakukan analisis sentimen untuk mengetahui bagaimana tanggapan masyarakat terhadap belanja thrifting, apakah masyarakat memberikan respon positif atau negatif. Hasil analisis sentimen dapat digunakan untuk mengidentifikasi, meningkatkan pemasaran, menentukan tren pasar, dan dapat dijadikan sebagai acuan bagi pemerintahan dalam menentukan kebijakan jual beli thrifting. Naïve Bayes digunakan sebagai metode analisis guna mengevaluasi sentimen tanggapan pengguna internet di Twitter terkait dengan topik belanja thrifting. Penelitian ini bertujuan untuk menghitung jumlah komentar positif dan negatif dari pengguna Twitter dan untuk mengevaluasi sejauh mana metode Naïve Bayes dapat mengukur keakuratan dalam melakukan analisis sentimen.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Naïve Bayes digunakan untuk mengklasifikasi sentimen pada penelitian ini, data yang dikumpulkan berkaitan dengan pengalaman belanja thrifting di Indonesia. Tahap – tahap yang dilalui dalam proses penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menjabarkan mengenai tahap-tahap yang dilakukan dalam penelitian. Tahap penelitian dimulai dengan melakukan pengumpulan data menggunakan aplikasi Rapid Miner yang dihubungkan ke Twitter untuk mendapatkan dataset dengan menggunakan kata kunci “thrifthing”. Tahap selanjutnya adalah pre-processing, data yang telah diperoleh kemudian dibersihkan melalui tahapan pre-processing yang terdiri dari beberapa tahap lagi seperti cleansing, tokenizing, dan filter stopword. Setelah dibersihkan data kemudian dibagi menjadi data latih dan data uji. Proses pelabelan manual dilakukan pada data latih untuk dijadikan sebagai bahan pembelajaran mesin dalam melakukan pengklasifikasian. Sementara pelabelan pada data uji dilakukan dengan menggunakan mesin yaitu dengan mengklasifikasikan data menjadi positif ataupun negatif menggunakan metode Naïve Bayes. Tahap akhir dari penelitian adalah perhitungan akurasi, yang dimana hasilnya berupa accuracy, precision, recall, TNR, dan NPV.

2.1.1 Pengumpulan Data (Crawling)

Langkah awal pengumpulan data atau biasa disebut crawling adalah proses pengambilan data yang diambil dengan keyword thrifthing. Data yang ditarik merupakan tweet yang dibuat oleh para pengguna di Twitter setiap harinya, pada proses pengambilan data dibantu dengan menggunakan Application Performance (API) yang disediakan oleh Twitter [21].

2.1.2 Pre-processing

Saat melakukan analisis sentimen, langkah yang perlu dilakukan adalah preprocessing langkah ini merupakan langkah pertama untuk memaksimalkan hasil dari analisis sentimen [22], kalimat ini akan disesuaikan yang nantinya memudahkan proses pembobotan di tahap selanjutnya [23]. Preprocessing mempunyai beberapa langkah sebagai berikut :

2.1.2.1 Cleansing

Proses membersihkan data berikutnya menghilangkan simbol-simbol seperti username(@), retweet, URL, karakter “!”, angka dan tanda baca yang tidak diperlukan. Agar data memiliki informasi yang diperluka dan relevan untuk penelitian [24].

2.1.2.2 Tokenizing

Tahapan ini merupakan tahap pengolahan teks yang melibatkan pemisahan teks menjadi kata-kata lebih kecil yang bermakna atau “token” [25],

2.1.2.3 Filtering Stopwords

Menghilangkan kata-kata yang sering digunakan namun kurang penting dikenal sebagai filtering stopword, stopword memiliki beberapa contoh dalam bahasa Indonesia seperti “yang”, “dan”, “di”, “dari”, dan lainnya. Stopword bertujuan untuk menghapus kata-kata yang tidak penting menjadi lebih lebih penting [26].

2.1.3 Aloritma Naïve Bayes

Metode Naïve Bayes adalah model algoritma klasifikasi yang sering digunakan untuk memprediksi perhitungan probabilitas di masa depan yang didasarkan pada pengalaman masa lalu. rumus dari metode Naïve Bayes sebagai berikut [27].

$$P(C|A) = \frac{p(A|C).p(C)}{p(A)} \tag{1}$$



Keterangan :

A merupakan data yang kelasnya tidak diketahui, sementara C merupakan hipotesis yang mengacu pada kelas yang spesifik untuk A. $P(C|A)$ mewakili probabilitas posteriori dari C berkaitan dengan A, yang merupakan probabilitas bahwa hipotesis C benar jika kita memiliki data A. $P(C)$ adalah probabilitas hipotesis sebelumnya (prior probability), yaitu probabilitas awal dari kebenaran hipotesis C sebelum melihat data A. $P(A|C)$ adalah probabilitas A berkaitan dengan hipotesis C, yang menunjukkan seberapa mungkin data A akan muncul jika hipotesis C benar. $P(A)$ adalah probabilitas munculnya data A secara keseluruhan.

2.1.4 Perhitungan Akurasi

2.1.4.1 Accuracy

Accuracy mencerminkan sejauh mana model mampu mengklasifikasikan secara tepat. Akurasi merupakan perbandingan antar jumlah prediksi yang benar (positif dan negatif) pada total data. accuracy mengacu pada seberapa dekat nilai prediksi dengan nilai actual. Nilai accuracy dapat dihitung dengan menggunakan persamaan (2). Dengan rumus perhitungan dibawah ini [28].

$$\frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

2.1.4.2 Precision (Posiive Predictive Value)

Tingkat keakuratan antara hasil model dan data yang diminta. Jadi, precision adalah rasio prediksi adalah rasio prediksi benar yang positif. Dari semua kelas positif yang telah di prediksi dengan benar,beberapa banyak data yang benar-benar positif. Nilai precision dapat diperoleh dengan persamaan (3). Dengan rumus perhitungan dibawah ini [28].

$$\frac{TP}{TP+FP} \quad (3)$$

2.1.4.3 Recall (True Posivite Rate)

Menggunakan kesuksesan model dalam menemukan kembali sebuah informasi. sehingga recall adalah rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Nilai recall dapat diperoleh dengan menggunakan persamaan (4) dengan rumus perhitungan dibawah ini [29].

$$\frac{TP}{TP+FN} \quad (4)$$

2.1.4.4 True Negative Rate

True Negative Rate mempunyai nama lain yang disebut nilai specificity yaitu merupakan presentase kelas negatif yang berhasil diidentifikasi sebagai kelas negatif yang disebut tingkat True Negative. Nilai TNR dapat diperoleh dengan menggunakan persamaan (5) dengan rumus perhitungan dibawah ini [30].

$$\frac{TN}{FP+TN} \quad (5)$$

2.1.4.5 Negative Predictive Value

Negative Predictive value adalah presentase data yang terklasifikasi sebagai kelas negatif oleh algoritma pengklasifikasi yaitu data negatif dari keseluruhan data yang diklasifikasikan sebagai kelas negatif. Nilai NVP dapat diperoleh dengan menggunakan persamaan (6) dengan rumus perhitungan dibawah ini [30].

$$\frac{TN}{FN+TN} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data (Crawling)

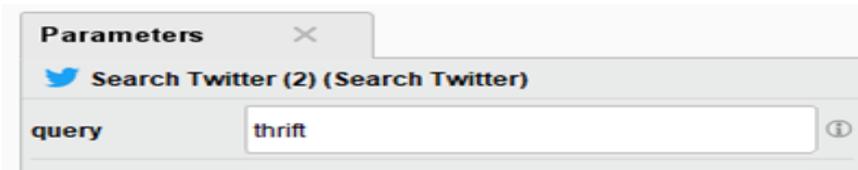
Twitter digunakan sebagai media untuk mengumpulkan data berupa tweet. Data yang dikumpulkan oleh peneliti sebanyak 900 data, dengan jangka waktu pengumpulan data adalah pada tanggal 15 s.d. 18 mei 2023. Gambar 2 adalah ilustrasi crawling data atau proses pengambilan data menggunakan bantuan aplikasi Rapid Miner. Data yang didapatkan bersumber dari Twitter, hasilnya disimpan dalam bentuk format Microsoft Excel lalu di konversi menjadi format CSV (Comma Separated Value).



Gambar 2. Ilustrasi Crawling Data



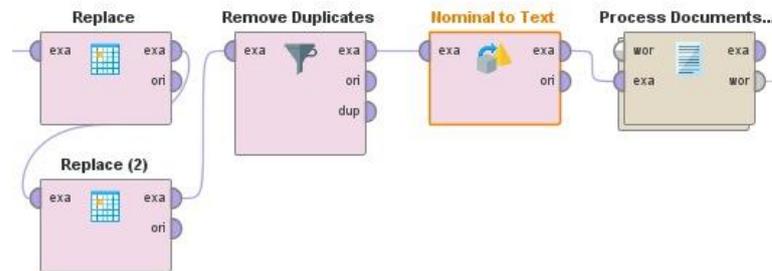
Gambar 3. merupakan kata kunci pada operator search twitter untuk mendapatkan dataset. Dataset yang dikumpulkan berdasarkan parameter search twitter dengan kata kunci “thrift” untuk mendapatkan informasi dari suatu media sosial Twitter..



Gambar 3. Kata Kunci pada Operator Search Twitter

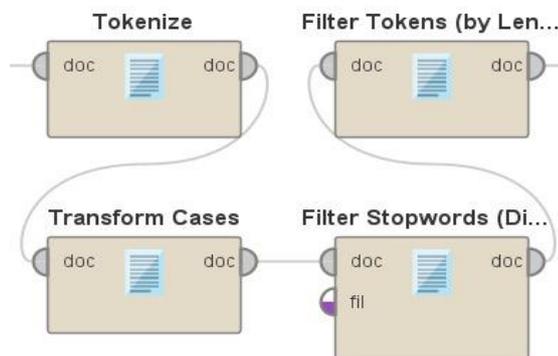
3.2 Pre-processing

Langkah pre-processing yang digunakan untuk penelitian analisis sentimen ini sangat diperlukan, untuk mengubah data yang tidak terstruktur atau sembarang menjadi data yang terstruktur dalam tahap pre-processing terdapat cleansing data, tokenizing, transform case, dan filtering stopwords. Langkah ini digunakan untuk mengubah data yang tidak terstruktur menjadi data yang lebih terstruktur. Gambar 4 menyajikan proses cleansing data yang dimana operator Replace merupakan operator yang berfungsi untuk membersihkan data dan menghilangkan username (@), RT, URL, tanda baca, simbol-simbol, dan angka yang tidak perlu.



Gambar 4. Cleansing Data

Gambar 5 menyajikan operator-operator yang ada didalam operator process document from data. Operator tokenizing merupakan proses memecah teks menjadi bagian-bagian kecil untuk memudahkan analisis sentimen. Operator transform case mengubah huruf dalam teks menjadi huruf besar dan kecil dengan tujuan untuk mengurangi perbedaan anatara huruf besar dan kecil yang dapat mempengaruhi pemrosesan teks yang bertujuan untuk memastikan kalimat dalam data yang memiliki format yang konsisten. Operator filtering stopwords membantu menghilangkan teks seperti tanda baca, karakter, dan angka serta mengoptimalkan data untuk menganalisis.



Gambar 5. Tokenizing, Transform Case, Stopword

Setelah melalui proses pre-processing, dataset berhasil menjadi bersih dari informasi yang tidak berhubungan dengan materi penelitian. Data yang sebelumnya berjumlah 900 data setelah melalui proses cleansing data tersisa menjadi 549 data teks dalam bentuk berupa opini ditujukan pada tabel 1.

Tabel 1. Hasil Pre-processing

Sebelum	Sesudah
RT @kelanapt: gaya amat, biasa juga rebutan jaket di live ig thrift.	gaya amat, biasa juga rebutan jaket di live ig thrift.



Sebelum	Sesudah
orang war tiket coldplay niki gw dan hanin war tiket early bird BAZAAR THRIFT Y2K gara gara murah :VVVVVVVVVV;	orang war tiket coldplay niki dan hanin gw war tiket early bird BAZAAR THRIFT Y2K gara gara murah
Outer hitam kayake relatif gampang dicari. Ke royal/tp aja nder cari di sana banyak, atau kalo low budget ya di thrift shop.;	Outer hitam kayake relatif gampang dicari Ke royaltp aja nder cari di sana banyak atau kalo low budget ya di thrift shop
Kemaren ke senen niatnya mau beli barang buat thrift shop paman, tapi malah puyeng sendiri kebeli cuma 3 biji;	Kemaren ke senen niatnya mau beli barang buat thrift shop paman, tapi malah puyeng sendiri kebeli cuma 3 biji
mak gue skena mana sih buset dapet mulu TNF anjay thrift https://t.co/67ifbvmvKY	mak gue skena mana sih buset dapet mulu TNF anjay thrift

3.3 Pelabelan data/ labeling

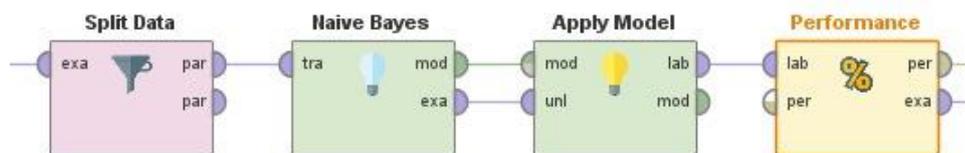
Setelah tahap pre-processing, langkah selanjutnya adalah pelabelan data. Pelabelan data twitter ini dikerjakan dengan manual proses pelabelan ini melibatkan penugasan label pada data latih secara mandiri, Setelah dataset diberi label dengan cara pelabelan ini, dataset tersebut dipakai sebagai referensi untuk analisis data uji menggunakan rapid miner. Kemudian dataset dibagi menjadi dua kelas sentimen yaitu positif dan negatif. Seperti tabel 2 menunjukkan contoh perbandingan hasil analisis sentimen kelas negatif dan positif.

Tabel 2. Pelabelan Data Latih

Label	Teks
Negatif	gaya amat, biasa juga rebutan jaket di live ig thrift.
Positif	orang war tiket coldplay niki dan hanin gw war tiket early bird BAZAAR THRIFT Y2K gara gara murah
Positif	Outer hitam kayake relatif gampang dicari Ke royaltp aja nder cari di sana banyak atau kalo low budget ya di thrift shop
Positif	Kemaren ke senen niatnya mau beli barang buat thrift shop paman tapi malah puyeng sendiri kebeli cuma 3 biji
Positif	mak gue skena mana sih buset dapet mulu TNF anjay thrift

3.4 Klasifikasi Naïve Bayes

Sesudah semua data latih terlabeli, lalu data digunakan untuk acuan pada perkiraan sentimen dan pengujian untuk data testing dengan metode algoritma naïve bayes dalam Rapid Miner.



Gambar 6. Klasifikasi Data (Naive Bayes)

Gambar 6 menyajikan proses klasifikasi menggunakan Naïve Bayes melibatkan pembagian dataset menjadi data latih dan data uji dengan perbandingan rasio 60:40. Data latih digunakan untuk melatih model, sementara data uji untuk menguji kinerja model. Klasifikasi Naïve Bayes digunakan untuk analisis sentimen. Operator Apply Model memprediksi data uji, dan operator performance mengevaluasi kinerja model. Keempat operator ini membentuk siklus pengembangan dan evaluasi model klasifikasi, dimulai dari pembagian dataset, pelatihan model, penerapan model pada data baru, hingga evaluasi kinerja untuk meningkatkan performa model. Proses ini dilakukan menggunakan Rapid Miner. Hasil prediksi sentimen terhadap data uji diperoleh dari proses klasifikasi data sebagai berikut.

Tabel 3. Jumlah Data Sentimen

Jenis Data	Positif	Negatif
Data latih dan Data Uji	368	181

3.5 Evaluasi Hasil Pengujian

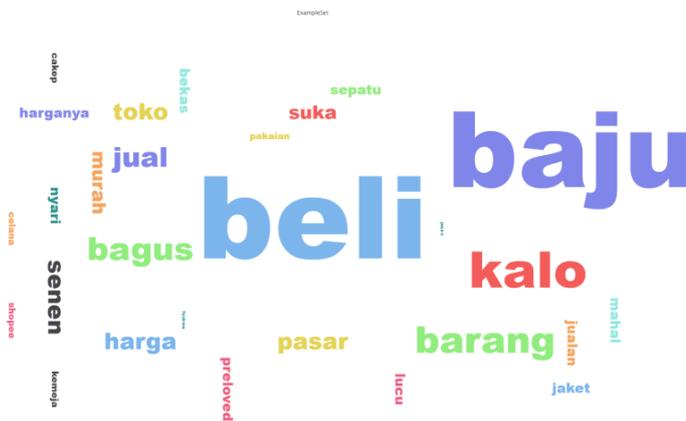
Setelah selesai melakukan proses pengujian dan prediksi sentimen, nilai seperti accuracy, precission, dan recall juga harus dilakukan analisis. Nilai-nilai yang akan di analisis tersebut dapat di perhitungkan terlebih dahulu dengan confusion matrix seperti tabel 4 dibawah ini.



Tabel 4. Confusion Matrix

Prediction Value	Actual Value	
	Positif	Negatif
Positif	TP(113)	FN(3)
Negatif	FP(5)	TN(75)

Confusion matrix mengandung pernyataan yang mengklasifikasikan jumlah data menjadi, benar terprediksi Positif atau True Positive (TP), salah terprediksi Positif atau False Positive (FS), benar terprediksi Negatif atau True Negative (TN), salah terdeteksi Negatif atau False Negative. Selanjutnya confusion matrix tentu akan dipakai dalam menghitung accuracy, recall(True Positive Rate), Precision (Positive Predictive Value), True Negative Rate dan Negative Predictive Value. Serangkaian pengujian telah dilakukan, menurut hasil penelitian ini. accuracy, precision, recall, TNR, dan NPV akan dihitung menggunakan data sentimen yang telah diproses menggunakan algoritma Naive Bayes. Confusion matrix akan digunakan sebagai panduan untuk menghitung kelima persamaan tersebut. Gambar 7 merupakan representasi visual dari kata-kata yang paling sering muncul dalam data yang telah dikumpulkan dalam penelitian ini dalam bentuk wordcloud.



Gambar 7. Wodrcload Kata yang Sering Muncul

Wordcloud diatas, menunjukan kedua kata yang sering digunakan adalah kata “beli” teridentifikasi sebanyak 115 kali sedangkan kata “baju” yang teridentifikasi sebanyak 106 kali muncul. Perhitungan dilakukan dalam confusion matrix, untuk menghitung nilai Accuracy, Recall (True Positive Rate – TPR), Precision (True Negative Rate – TNR), dan Negative Predictive Value NVP). Tujuan dari nilai accuracy adalah untuk mengevaluasi sejauh mana algoritma klasifikasi Naive Bayes dapat memprediksi secara tepat data uji. Selanjutnya perhitungan Recall (true positive rate (TPR) untuk menilai sejauh mana algoritma klasifikasi Naive Bayes dapat mengenali sentimen positif dengan tepat. Kemudian kita melakukan perhitungan precision (Positive Prediction Value (PPV) untuk mengukur akurasi prediksi positif algoritma Naive Bayes. Selain itu, kita juga menghitung Rate True Negative (TNR) untuk menilai kemampuan algoritma klasifikasi Naive Bayes dalam mendeteksi sentimen negative dengan akurat. Terakhir, ada perhitungan Negative Predictive Value (NPV) yang digunakan oleh algoritma klasifikasi Naive Bayes untuk mengevaluasi akurasi prediksi negatif. Hasil perhitungan dari nilai-nilai ini ditunjukkan di bawah ini :

Accuracy

$$\text{Accuracy} = \frac{113+75}{113+3+5+75}$$

$$\text{Accuracy} = 0.9592$$

$$\text{Accuracy} = 95.92\%$$

Recall (True Predictive Value)

$$\text{Recall} = \frac{113}{113+3}$$

$$\text{Recall} = 0.9741$$

$$\text{Recall} = 97,41\%$$

Precision (Positive Prediction Value)

$$\text{Precision (PPV)} = \frac{113}{113+5}$$

$$\text{Precision (PPV)} = 0.9576 = 95,76\%$$

True Negative Rate (TNR)

$$\text{True Negative Rate} = \frac{75}{5+113}$$

$$\text{True Negative Rate} = 0.6335 = 63,35\%$$

Negative Predictive Value (NPV)



$$\text{Negative Predictive Value} = \frac{75}{5+75}$$

$$\text{Negative Predictive Value} = 0.9615 = 96,15\%$$

Tabel 5. merupakan hasil perhitungan pada Rapid Miner. Perhitungan ini menunjukkan seberapa baik algoritma Naive Bayes menganalisis sentimen. Dengan accuracy 95,92%, model klasifikasi ini sangat tepat dalam prediksi sentimen positif dan negatif. Selain itu, perhitungan recall (TPR dan TNR) yang mencapai 97,41% dan 63,35% menunjukkan kemampuan model untuk mendeteksi dengan akurat sentimen positif dan negatif. Selain itu, perhitungan precision (PPV) dan NPV masing-masing menunjukkan hasil 95,76% dan 96,15%. Dari hasil masing-masing perhitungan ini, dapat disimpulkan bahwa model klasifikasi ini berhasil memprediksi sentimen positif dan negatif.

Tabel 5. Hasil Perhitungan pada Rapid Miner

Accuracy: 95.92%			
	True Positif	True Negatif	Class Precision
Pred. Positif	113	3	97.41%
Pred. Negatif	5	75	93.75%
Class Recall	95.76%	96.15%	

4. KESIMPULAN

Menganalisis sentimen masyarakat Indonesia mengenai belanja thrifting dari tweet di Twitter menggunakan algoritma Naïve Bayes untuk penelitian ini. Data yang digunakan sebanyak 900 tweet dan diolah menjadi lebih terstruktur melalui tahap preprocessing yang terdiri dari cleansing, tokenizing dan filtering stopwords. Penelitian ini menggunakan pembagian data dengan ratio 60:40. Setelah melalui rangkaian proses dalam melakukan analisis sentimen, didapatkan hasil sentimen positif lebih banyak dibandingkan dengan sentimen negatif dengan jumlah sentimen positif sebanyak 368 tweet dan jumlah sentimen negatif sebanyak 181 tweet. Dari hasil analisis sentimen dapat disimpulkan bahwa opini masyarakat Indonesia mengenai belanja thrifting cenderung positif karena dengan thrifting masyarakat dapat membeli barang-barang branded yang masih layak pakai atau membeli barang-barang unik dengan harga murah sehingga mengurangi terjadinya pemborosan. Hal tersebut membuat masyarakat merasa puas belanja thrifting sehingga hasil analisis mendapatkan banyak sentimen positif. Penggunaan algoritma Naïve Bayes dalam penelitian ini terbukti akurat dengan hasil accuracy mencapai 95,92%, recall sebesar 97,41%, precision sebesar 95,76. Algoritma Naïve Bayes juga mampu memprediksi sentimen negatif cukup akurat terbukti dengan didapaknya nilai TNR (True Negative Rate) mencapai 63,35% dan mampu mengevaluasi akurasi prediksi negatif dengan sangat baik, terbukti dari NPV (Negative Predictive Value) mencapai 96,15%. Dengan demikian, hasil penelitian ini menunjukkan bahwa algoritma Naïve Bayes sangat efektif dalam menganalisis sentimen dari data yang telah dikumpulkan.

UCAPAN TERIMAKASIH

Terima kasih kepada pihak-pihak yang telah berkontribusi pada penelitian jurnal ini, semoga hasil penelitian ini bermanfaat.

REFERENCES

- [1] S. Nadhila, M. Muzhirah, H. Sajali, and M. Andinata, "Eksistensi Diri Remaja Dalam Penggunaan Pakaian Bekas (Studi Kasus Pada Konsumen Thrifting Pajak Melati Medan)," *Innov. J. Soc. Sci. Res.*, vol. 3, no. 3, pp. 2436–2446, 2023, doi: 10.31004/innovative.v3i3.2403.
- [2] M. Hayati and N. Susilawati, "Thrifting sebagai presentasi diri Mahasiswa di Pasar Putih Bukittinggi," *J. Perspekt.*, vol. 4, no. 3, p. 359, Sep. 2021, doi: 10.24036/perspektif.v4i3.460.
- [3] D. Rahmayanti, "Pengaruh Fashion Involvement, Promosi, Religiusitas terhadap Impulse Buying dengan Shopping Emotion sebagai Variabel Intervening pada Pakaian Thrift di Salatiga," vol. 3, no. 1, pp. 40–51, 2023, doi: 10.19105/mabny.v3i01.8988.
- [4] N. L. Adji and D. Claretta, "Fenomena thrift shop dikalangan remaja: studi fenomenologi tentang thrift shop di kalangan remaja Surabaya," *Dawatuna J. Commun. ion Islam. Broadcast.*, vol. 3, no. 1, pp. 36–44, 2023, doi: 10.47476/dawatuna.v3i1.2201.
- [5] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 542, 2022, doi: 10.30865/mib.v6i1.3530.
- [6] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 296, 2022, doi: 10.30865/json.v4i2.5381.
- [7] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020, doi: 10.30865/mib.v4i3.2140.
- [8] M. Oktaviana, Z. A. Achmad, H. Arviani, and K. Kusnarto, "Budaya komunikasi virtual di Twitter dan Tiktok: Perluasan makna kata estetik," *Satwika Kaji. Ilmu Budaya dan Perubahan Sos.*, vol. 5, no. 2, pp. 173–186, 2021, doi:



- 10.22219/satwika.v5i2.17560.
- [9] Y. A. V. Gunawan, N. A. S. ER, I. B. M. Mahendra, I. M. Widiartha, I. G. N. A. C. Putra, and I. G. A. G. A. Kadyanan, "Analisis Sentimen Ulasan Aplikasi Transportasi Online Menggunakan Multinomial Naïve Bayes dan Query Expansion Ranking," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 1, p. 121, 2022, doi: 10.24843/jlk.2022.v11i01.p13.
- [10] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [11] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [12] Y. A. Singgalen, "Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2486.
- [13] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: 10.26418/jlk.v5i1.79.
- [14] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [15] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [16] S. Lestari and S. Saepudin, "Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, p. 163–170, 2021.
- [17] S. M. Salsabila, A. Alim Murtopo, and N. Fadhillah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," *J. Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.
- [18] Vynska Amalia Permadi, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," *J. Buana Inform.*, vol. 11, pp. 141–151, 2020, doi: 10.24002/jbi.v11i2.3769.
- [19] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.
- [20] A. Rozaqi, A. Triayudi, and R. T. Aldisa, "Analisis Sentimen Vaksinasi Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 184, 2022, doi: 10.30865/json.v4i1.4907.
- [21] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier," *J. Asimetrik J. Ilm. Rekayasa Inov.*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asiimetrik.v4i1.3577.
- [22] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [23] 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," *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [24] 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," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 27–29, 2020, doi: 10.30864/jsi.v15i1.332.
- [25] U. Gunadarma, U. Gunadarma, K. Positif, K. Negatif, and N. Bayes, "Analisis Sentimen Terhadap Universitas Gunadarma Berdasarkan Opini Pengguna Twitter Menggunakan Metode Naive Bayes Classifier," *J. Ilm. Komputasi*, vol. 19, no. 4, pp. 507–521, 2020, doi: 10.32409/jikstik.19.4.354.
- [26] A. Saputra and F. Noor Hasan, "Analisis Sentimen Terhadap Aplikasi Coffee Meets Bagel Dengan Algoritma Naïve Bayes Classifier," *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 2, no. 2, pp. 465–474, 2023, doi: 10.54443/sibatik.v2i2.579.
- [27] A. Kusuma and A. Nugroho, "Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naive Bayes," *J. Ilm. Teknol. Inf. Asia*, vol. 15, no. 2, p. 137, 2021, doi: 10.32815/jitika.v15i2.557.
- [28] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: 10.33557/jurnalatrik.v22i1.843.
- [29] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [30] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.